Genetic Investigations of Canine Idiopathic Epilepsy and Movement Disorders

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Biology, Medicine and Health

2023

Christopher A Jenkins

School of Health Sciences

Contents

Word count: 65321

List of figures

List of tables

List of appendices

Abbreviations

µg	Microgram
µL	Microlitre
µm	Micrometre
A	Adenine
ABI	Applied Biosystems
ACCPN	Agenesis of the Corpus Callosum with Peripheral Neuronopathy
AD	Autosomal Dominant
ADHD	Attention Deficit Hyperactivity Disorder
ADPEAF	Autosomal-Dominant Partial Epilepsy with Auditory Features
AEM	Antiepileptic Medication
AHT	Animal Health Trust
ALS	Amyotrophic Lateral Sclerosis
AMP	Adenosine Monophosphate
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
AR	Autosomal Recessive
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
BAC	Bacterial Artificial Chromosome
BBBCS	Big Brainy Border Collie Study
BE	Beagle
BFJE	Benign Familial Juvenile Epilepsy
BOAS	Brachycephalic Obstructive Airway Syndrome
bp	Base Pair
BVA	British Veterinary Association
BWA	Burrows-Wheeler Aligner
C	Cytosine
ca/co	Case/Control
CaTS	Power calculator for two stage association studies
CCD	Cerebellar Cortical Degeneration
cDNA	Complementary DNA
CECS	Canine Epileptoid Cramping Syndrome

CFA	*Canis Familiaris* Chromosome
Chr	Chromosome
CI	Confidence Interval
CME	Clathrin-Mediated Endocytosis
CMSD	Canine Multiple System Degeneration
CNS	Central Nervous System
CNV	Copy Number Variant
contig	Contiguous Sequence
CSF	Cerebrospinal Fluid
CT	Computed Tomography
CT	Cycle Threshold
DBVDC	Dog Biomedical Variant Database Consortium
DM	Degenerative Myelopathy
DMD	Duchenne Muscular Dystrophy
DNA	Deoxyribonucleic Acid
DWLM	Dandy-Walker-Like Malformation
EAST	Epilepsy, Ataxia, Sensorineural Deafness, Tubulopathy
EBV	Estimated Breeding Values
EDTA	Ethylenediaminetetraacetic Acid
EEG	Electroencephalogram
ER	Endoplasmic Reticulum
EU	European Union
FAM	Carboxyfluorescein
FFPE	Formalin-Fixed Paraffin-Embedded
FWM	Foliary White Matter
G	Guanine
GABA	γ-aminobutyric acid
GAS	Online genetic association study power calculator
GATK	Genome Analysis Toolkit
Gb	Gigabase
GD	Great Dane
GL	Granular Layer
GPI	Glycosylphosphatidylinositol

GR	Golden Retriever
GTP	Guanosine Triphosphate
GWAS	Genome-Wide Association Study
H-bond	Hydrogen Bond
IBS	Identity-By-State
ID	Identification
IDEAS	Idiopathic Epilepsy and Anxiety Study
IE	Idiopathic Epilepsy
IGV	Integrative Genomics Viewer
ILAE	International League Against Epilepsy
Info	IMPUTE2's Imputation Certainty Metric
ISCGB	Italian Spinone Club of Great Britain
ISDS	International Sheep Dog Society
IVETF	International Veterinary Epilepsy Task Force
KASP	Competitive Allele-Specific PCR
kb	Kilobase
kcal	Kilocalorie
KCGC	Kennel Club Genetics Centre
KCNIP	Voltage-Gated Potassium Channel-Interacting Protein
kDa	Kilodalton
Kv	Voltage-Gated Potassium Channel
LD	Linkage Disequilibrium
LINE-1	Long Interspersed Element-1
LOWESS	Locally Weighted Scatterplot Smoothing
LR-X	Labrador Retriever Cross Breed
MA	Mechanically Activated
MAF	Minor Allele Frequency
Mb	Megabases
MDS	Multidimensional Scaling
MHC	Major Histocompatibility Complex
ML	Molecular Layer
mol	Mole
MRC	Medical Research Council

MRI	Magnetic Resonance Imaging
mRNA	Messenger RNA
mRNA-seq	mRNA Sequencing
mtDNA	Mitochondrial DNA
n	Sample Size / Number
NB	Norwegian Buhund
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NGS	Next Generation Sequencing
NMDA	N-Methyl-D-Aspartate
NTC	No Template Control
OMIA	Online Mendelian Inheritance in Animals
OMIM	Online Mendelian Inheritance in Man
PC	Purkinje Cell
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PED	Paroxysmal Exercise-Induced Dyskinesia
PGSD	Paroxysmal Gluten-Sensitive Dyskinesia
PKD	Paroxysmal Kinesigenic Dyskinesia
PNKD	Paroxysmal Nonkinesigenic Dyskinesia
POAG	Primary Open Angle Glaucoma
PRA	Progressive Retinal Atrophy
PRCD	Progressive Rod-Cone Degeneration
PRS	Polygenic Risk Score
PRT	Parson Russell Terrier
PxD	Paroxysmal Dyskinesia
QQ	Quantile-Quantile
R^2	Squared Pearson Correlation Coefficient
RIPA	Radioimmunoprecipitation Assay
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RT-qPCR	Reverse Transcription Quantitative Polymerase Chain Reaction

RVC	Royal Veterinary College
SAMS	Spinocerebellar Ataxia with Myokymia, Seizures, or Both
SCA	Spinocerebellar Ataxia (Autosomal Dominant)
SCAR	Spinocerebellar Ataxia (Autosomal Recessive)
SCAX	Spinocerebellar Ataxia (X-Linked)
SDCA	Spongy Degeneration with Cerebellar Ataxia
SDCA1	Spongy Degeneration with Cerebellar Ataxia, Subtype 1
SDCA2	Spongy Degeneration with Cerebellar Ataxia, Subtype 2
SDS	Sodium Dodecyl Sulfate
SeSAME	Seizures, Sensorineural Deafness, Ataxia, Mental Retardation, and Electrolyte Imbalance
SINE	Short Interspersed Nuclear Element
SINEC	Canine Short Interspersed Element
SNP	Single Nucleotide Polymorphism
SPG76	Spastic Paraplegia 76
STR	Short Tandem Repeat
SV	Structural Variant
T	Thymine
TAG	Triacylglycerol
UCSC	University Of California Santa Cruz Genomics Institute Genome Browser
UCSF	University Of California San Francisco
UH	University Of Helsinki
UK	United Kingdom
UM	University of Manchester
USA	United States of America
UTR	Untranslated Region
UU	University of Utrecht
V	Hungarian Vizsla
WGS	Whole Genome Sequence
WMLS	Weill-Marchesani-Like Syndrome
XB	Siberian Husky Cross Breed
χ^2	Chi-Squared

Abstract

Dogs represent a fascinating species with dramatic variation in traits across hundreds of breeds. Genetic bottlenecks, inbreeding, and selection for extreme phenotypes have led to the accumulation of deleterious genetic variants and inherited diseases within breeds. Inherited canine neurological diseases such as movement disorders (e.g. ataxia and paroxysmal dyskinesia) and idiopathic epilepsy (IE) often have an impact on quality of life and life expectancy and can have a high prevalence in some breeds. The welfare impact means that elucidating the underlying genetics is of veterinary importance, and many canine inherited diseases are naturally occurring disease models of human disease.

Whole-genome sequencing of two Norwegian Buhund siblings with progressive cerebellar ataxia led to the identification of a single nucleotide polymorphism (SNP) in *KCNIP4*. This gene had not been implicated previously in hereditary ataxia in any species. The findings suggest an important role for Kv4 channel complex KCNIP accessory subunits in the cerebellum.

The feasibility and accuracy of genotype imputation of a study set using a reference panel comprising breed-specific array data and multi-breed variant data derived from whole genomes was assessed. Canine genotype imputation can be effective and accurate and was utilised for the IE and paroxysmal dyskinesia studies.

A preliminary genome-wide association study (GWAS) of paroxysmal dyskinesia was conducted in the Norwich Terrier dog breed, and analysis of 44 SNPs in an independent dataset identified five genomic regions, and genes, for potential future investigation.

Separate GWAS were performed for IE in the Italian Spinone and Border Collie breeds, and independent validation sets were utilised for replication analyses. Variants with evidence of association were incorporated into weighted risk scores, and their ability to predict disease status was assessed. The GWAS findings were not reproduced in the validation sets of the Italian Spinone study. However, the findings in the Border Collie IE study demonstrated evidence of replicable association and a three-SNP genetic risk score showed potential for predicting disease status. The findings suggest that mode of inheritance for IE is not monogenic in either breed and implicate three genes in the Border Collie.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given the University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, the University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in the University's policy on Presentation of Theses.

Acknowledgments

I would like to thank my supervisors Janine Lamb, Sally Ricketts, Lorna Kennedy, Bill Ollier, and Andrea Short for all their support and advice throughout my PhD. The past six and a half years have been extremely challenging at times and their encouragement has been invaluable. Sally Ricketts deserves a special mention, a great friend and mentor as well as a supervisor, without whom none of this would have been possible.

I would also like to thank Cathryn Mellersh, head of the Kennel Club Genetics Centre, for supporting my PhD research and allowing me to do this work as part of, and alongside, my role as a research assistant within the group. The closure of the Animal Health Trust in 2020 made the future of the group and my PhD research uncertain and I am extremely grateful to Cathryn for everything she did to keep myself and the group going during this difficult period. I would like to thank everyone at the University of Cambridge Department of Veterinary Medicine, especially David Sargan, who made our group feel welcome and helped us get up and running again after our move to the department. My friends and colleagues in the Kennel Club Genetics Centre also deserve my thanks for their support and for making this PhD a rewarding and enjoyable experience.

I am thankful for the Animal Health Trust for giving me the opportunity to do this PhD, for funding the first four years of tuition fees, and for all my former colleagues who made the Animal Health Trust a fantastic place to work. It was a unique organisation, and its closure a huge loss. I would particularly like to thank Luisa De Risio, whose expertise as a neurologist was crucial to many of the projects described in this thesis. I must also thank Wisdom Health for funding the final two years of tuition fees, allowing me to continue my PhD. I am grateful to my collaborators and funders, acknowledged in each manuscript, and to all the dogs and dog owners that contributed DNA samples and completed questionnaires, who made the projects described in this thesis possible.

Finally, I owe a huge thank you to my friends and family for getting me through the last six and a half years. Thank you to my parents Andy and Sally Jenkins, to my brother Will Jenkins, and to my grandparents, for all their support and for encouraging me to ask questions and to be interested in finding answers, and for always believing in me. Last but by no means least, I must mention my Border Collie Ed, a source of comfort, entertainment, and companionship, and most likely the only individual mentioned here to have never asked me how much longer I had left until submission!

The author

I graduated from the University of East Anglia in 2010 with my undergraduate degree, BSc Biological Sciences. I was keen to work in research, particularly genetics, and went on to study at Cranfield University to obtain a master's degree (MSc Molecular Medicine). The research project that was the subject of my MSc thesis was to develop an *in vitro* axonal growth model, using a neuroblastoma cell line. Working on this project gave me my first insight into the fascinating field of neuroscience, and later led to me seizing the opportunity to undertake a PhD with a focus on neurological disease. After graduating I worked at the University of Sheffield as a research technician in the Department of Molecular Biology and Biotechnology on a short grant, before joining the Kennel Club Genetics Centre at the Animal Health Trust as a research assistant in 2014. I worked on a range of projects investigating the genetics of canine inherited diseases, gaining experience and an understanding of the canine genetics field. In 2016 I took the exciting opportunity to start this part-time PhD investigating the genetics of movement disorders and idiopathic epilepsy, registered at the University of Manchester while continuing to work as a research assistant within the Kennel Club Genetics Centre.

The last six and a half years have been a challenging time to work in the Kennel Club Genetics Centre, with the closure of the Animal Health Trust, the impacts of COVID-19, and reduced funding. However, I have enjoyed my time working on this PhD project and I have found the research and working in this field extremely rewarding. Working with breed clubs allows you to see the direct impact of your research, and the canine genetics field is small, collaborative, and supportive with some excellent researchers. This PhD has allowed me to apply for, and receive, my first grant funding as a principal investigator (PetPlan Charitable Trust project number S18-688-726), and to be a secondary investigator on another (Dogs Trust). I have published two peer reviewed manuscripts as a first named author as a direct result of this PhD, and written three more which are planned for publication. The PhD has also led to me presenting results at four international conferences, at two of which I was invited to give a full talk and at a third a short talk in support of my poster. I have gained experience in a wide range of methodologies, and through this project I have discovered an affinity and real satisfaction in the use of computational approaches to analyse genetic data.

Papers and conference presentations arising during this PhD:

Papers:

Jenkins CA, Dog Biomedical Variant Database Consortium, Schofield EC, Mellersh CS, De Risio L, Ricketts SL (2021) Improving the resolution of canine genome-wide association studies using genotype imputation: a study of two breeds. Animal Genetics.

Jenkins CA, Kalmar L, Matiasek K, Mari L, Kyöstilä K, Lohi H, Schofield EC, Mellersh CS, De Risio L, Ricketts SL (2020) Characterisation of canine KCNIP4: A novel gene for cerebellar ataxia identified by whole-genome sequencing two affected Norwegian Buhund dogs. PLoS Genetics. 16(1):e1008527. doi: 10.1371/journal.pgen.1008527.

Mari L, Matiasek K, **Jenkins CA**, De Stefani A, Ricketts SL, Forman O, De Risio L (2018) Hereditary ataxia in four related Norwegian Buhunds. Journal of the American Veterinary Medical Association. 15;253(6):774-780. doi: 10.2460/javma.253.6.774. PMID: 30179085.

Conference presentations:

Poster: "Idiopathic epilepsy in the Border Collie: using genome-wide association study and whole genome sequencing approaches to identify genetic risk factors". 11th International Conference on Canine and Feline Genetics and Genomics. Huntsville, Alabama, USA, 2022.

Short talk and poster: "Validation and application of genome-wide imputation as a strategy to facilitate the study of canine complex inherited disease". 10th International Conference on Canine and Feline Genetics and Genomics. Bern, Switzerland, 2019.

Talk: "Identification and characterisation of a mutation associated with cerebellar ataxia in the Norwegian Buhund dog breed". Companion Animal Genetic Health Conference. Edinburgh, UK, 2018. **Published abstract:** Selected canine abstracts from the Companion Animal Genetic Health conference 2018 (CAGH 2018): Canine Genetics and Epidemiology. Canine Genet Epidemiol 5 (Suppl 2), 7 (2018). https://doi.org/10.1186/s40575-018-0062-z

Talk: "The Use of Whole Genome Sequencing to Investigate the Genetic Cause of Hereditary Ataxia in Two Norwegian Buhund Siblings". 9th International Conference on Canine and Feline Genetics and Genomics. Saint Paul, Minnesota (USA), 2017.

Rationale for submission in journal format

In the process of this PhD I have investigated four neurological diseases, two of which were movement disorders and two idiopathic epilepsy, each in a different dog breed. The approaches for each were distinct and tailored to the breed and disease investigated. Each study was therefore suitable for writing as separate chapters and manuscripts, and for publishing independently. As part of the PhD the use of genome-wide genotype imputation was investigated, looking at accuracy and best practices for using this tool for canine genetic research. All five chapters were therefore suitable for publication, and the results are important and relevant to the canine genetics and wider genetic research communities, and it was therefore important to ensure that the findings were disseminated. For these reasons, discussions with my PhD supervisors determined that it would be appropriate to submit my thesis in journal format. Journal format in this instance is comprised of an introduction, aims and objectives, five results chapters that are each a separate manuscript, and a general discussion.

Contributions

The results chapters within this PhD thesis include two published manuscripts and three manuscripts which are planned for publication. In all cases I am the first named author and have written the manuscript and performed the majority of the work presented including contributing to study design and performing data collection, analysis, and interpretation of results. The contributions of co-authors are stated at the start of each chapter, and the contributions of any individuals not included as co-authors are given in the acknowledgements section towards the end of each chapter.

1 Introduction

Canis lupus familiaris, the domestic dog, is a fascinating species that consists of approximately 400 unique breeds with dramatic and extensive variation in morphological and behavioural traits [1]. Dogs have been companions for humans and have played a role in human society for many thousands of years [2]. Dogs are scientifically interesting in their own right, and are also a unique model of human diseases [3, 4]. In the following chapter an introductory overview of canine genetics will be provided, followed by a focus on the canine neurological disorders that are the subject of this thesis: idiopathic epilepsy and movement disorders.

1.1 Origin of domestic dogs and the generation of specific breeds

1.1.1 Domestication of canids and genetic bottlenecks

Domestic dogs, *Canis lupus familiaris*, are highly selected variants of a subspecies of the wolf *Canis lupus*; it is possible for wolves and domestic dogs to mate and produce fertile offspring. Pedigree dogs are a product of restricted inbreeding and artificial selection for desired traits. Modern domestic dogs have therefore undergone two major genetic bottlenecks in their history [5]: the divergence from wolves followed by the relatively recent creation of pedigree breeds. A genetic bottleneck is a dramatic drop in the number of breeding individuals in a population, which results in a loss of genetic diversity that is then followed by population expansion.

Dogs are thought to be the earliest domesticated animal [2, 6]. Domestication of the dog from wolves is understood to have occurred between 11,000 and 35,000 years ago, although the timing and location of the domestication event or multiple domestication events remain controversial [2, 6]. Out of the currently extant species of wild canids, such as coyotes and jackals, domestic dogs are most closely related to the grey wolf and analysis of mitochondrial DNA (mtDNA) supports wolf ancestry for modern dogs [2, 7]. Dogs may not have modern wolves as a direct common ancestor; it could be a now extinct subspecies of wolf [8].

The second major genetic bottleneck for the dog was the formation of breeds [5]. Most modern dog breeds came into existence in the last 200 to 300 years [9]. The Kennel Club in the United Kingdom (UK) and the American Kennel Club in the United States of America (USA), and other registering organisations, were established in the 1800s and required

both parents of registered dogs to be registered members of the same breed. This resulted in the isolation of each breed as a breeding population. The establishment of breed standards for behaviour and physical appearance has caused strong selective pressure for specific traits within breeds.

1.1.2 Breeds, selection, and inbreeding

Dogs are an excellent example of how physical and behavioural phenotypes can be diversified through domestication and selection. The >400 different breeds and breed varieties recognised around the world demonstrate a huge range of morphology and behavioural characteristics which make the dog extremely diverse phenotypically [1] (**Figure 1.1**). Many dog breeds were originally bred for a specific purpose, such as guarding, companionship, herding, hunting; or were bred for their novelty. The Kennel Club in the UK categorises dog breeds into seven groups: hound, working, terrier, gundog, pastoral, utility, and toy. Dog breeds can also be divided into ancient breeds and modern breeds. The ancient breeds, as their name suggests, have historic origins of typically from 500 to thousands of years ago, and are genetically divergent to modern breeds [10]. Most modern breeds were produced by the controlled breeding practices used in the Victorian era.

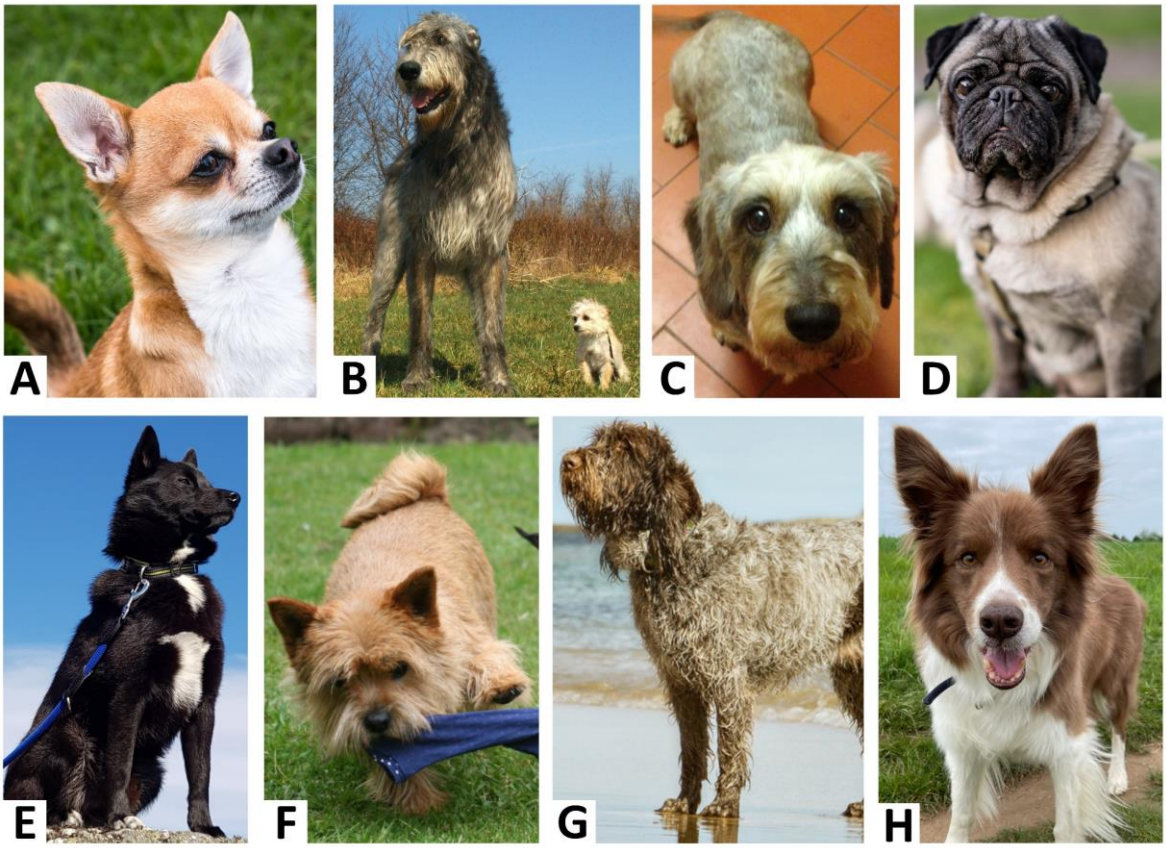


Figure 1.1. Photographs demonstrating some of the morphological diversity of dog breeds in general and of the four breeds that are the focus of this PhD thesis. A) Chihuahua (smooth coat), toy group. B) Irish Wolfhound (left), hound group, depicted with a small mixed breed dog (right). C) Dachshund (wire haired), hound group. D) Pug, toy group. E) Norwegian Buhund, pastoral group. F) Norwich Terrier, terrier group. G) Italian Spinone, gundog group. H) Border Collie, pastoral group. E-H) Breeds which were the focus of studies as part of this PhD project. Breed groups are as defined by the UK Kennel Club.

The restriction of breeding to within the breed registries, and other breeding practices, has resulted in pedigree dogs becoming highly inbred. The UK dog population became increasingly inbred between 1980 and 2000, but inbreeding has since levelled off, coinciding with the relaxation of quarantine laws which allowed increased numbers of imported dogs into the breeding population [11]. One breeding practice which increases inbreeding has been the use of popular sires: more female dogs are used for breeding than males, and some males can sire many litters and account for a large proportion of the puppies born for a breed each year. Inbreeding is a key consideration when implementing new genetic tests for disease-causing mutations, or when attempting to breed away from an inherited condition for which the aetiology is unknown. Selection away from a disease can result in the reduction of the breeding population and for

numerically small breeds this can be a particular problem, possibly necessitating the introduction of dogs from a different breed.

1.2 Clinical consequences of breeding purebred dogs

1.2.1 Rigorous selection for extreme phenotypes

Breeders of purebred dogs aim for their dogs to fit the breed standard: guidelines for what is considered the ideal example of the breed, including appearance and temperament. This can lead to selection for extreme phenotypes such as large or small size, or short muzzles. The selection for extremes of phenotype can result in unintended harmful consequences to the health of dogs, leading to welfare concerns.

The short-muzzled, or brachycephalic, breeds demonstrate how extreme phenotypes can lead to health problems. French Bulldogs, Bulldogs, Pugs, and other breeds with shortened skulls and muzzles, are prone to the brachycephalic obstructive airway syndrome (BOAS) respiratory disorder caused by obstruction of the airways by soft tissue not reduced in proportion with the skull [12-14]. BOAS is a considerable welfare concern; it causes difficulty breathing and dogs to struggle with exercise and heat, and results in a shortened lifespan. To reduce the frequency of BOAS there is a need to breed away from the short skull and muzzle phenotype, along with managing other factors which are associated with an increase in a brachycephalic dog‘s risk of BOAS. These include stenotic (closed) nostrils, and obesity [13, 14].

Another way in which selection for desirable breed phenotypes can have a negative impact on health is by increasing the frequency of genetic variants which produce the desired trait, but which also cause a disease phenotype. The insertion of a retro-gene copy of the fibroblast growth factor 4 (*FGF4*) gene on canine chromosome 12 (CFA12) is associated with both short limbs and Hansen’s type I intervertebral disc disease across multiple breeds [15, 16]. This gene insertion variant is semi-dominant for height, i.e. heterozygotes for this insertion are shorter than those without a single copy and homozygotes are shorter still; breeders selecting for the desirable short phenotype could therefore be unintentionally increasing the frequency of the disease allele. Genetic variants within *ADAMTS17* are associated with primary open angle glaucoma (POAG) in the Petit Basset Griffon Vendeen and Shar Pei breeds [17, 18], and have also been linked with short stature [19]. Mutations within *ADAMTS17* in humans cause Weill-Marchesani-

like syndrome (WMLS), a connective tissue disorder with phenotypes that include glaucoma and short stature [20]. This could therefore represent another example of selection for height having the unintended consequence of increasing the frequency of a disease-causing variant.

1.2.2 The effect of inbreeding on haplotype structure

If two genomic locations on the same chromosome are inherited together more or less frequently than would be expected if they were randomly associated with each other (as calculated from their respective allele frequencies in the population), they are considered to be in Linkage Disequilibrium (LD); i.e. an unequal level of expected linkage frequency. The dog genome exhibits extensive LD, with long haplotype blocks, within breeds [5]. This is due to the relatively recent origin of dog breeds from small founder populations; there have not been enough generations for randomisation of linked genes through cross-over events and non-random mating. By contrast, across breeds the dog population has short haplotype blocks, reflecting the short-range LD of the ancestral population from which modern breeds are derived. When breeds were formed from a small subset of the entire population, the diverse combinations of short haplotype blocks present in the ancestral population became long haplotypes specific to breeds due to the relatively low diversity of the chromosomes in the founding population.

The widespread inbreeding, relatively recent bottlenecks, and extensive LD have resulted in long runs of genome homozygosity, covering a large proportion of the genome [5]. The long runs of homozygosity represent regions of the genome with identical haplotypes on both chromosomes. Long regions of the genome around a variant causing a trait being selected for can become fixed or have reduced heterozygosity [21]. Through these selective sweeps, and extensive LD and long runs of homozygosity, variants conferring risk of disease can be more likely to be linked to the region under selection. The population bottlenecks and reduction of breeding population size caused by the selective breeding involved in domestication and breed formation have resulted in the accumulation of deleterious genetic variants throughout the canine genome [22]. The number of deleterious variants is highest in regions that contain variants causing desirable traits that have been positively selected for by dog breeders [22].

The extensive LD and homozygosity has both advantages and disadvantages for the experimental approaches used to investigate canine inherited disease. As a result of

widespread LD smaller sample sets and genotyping arrays with fewer genetic markers are necessary for genome-wide association studies (See chapter 1.4.2.2 for a description of this approach) to identify regions of the genome associated with disease in dog breeds than for comparable human studies [5]. However, the long haplotypes can also make it harder to narrow down the associated region to the variant causing disease because of the many linked variants over a broad, homozygous, region.

1.3 Breed health monitoring techniques and breeding tools

Inherited diseases in pedigree dog breeds are a major welfare concern. Responsible dog breeders and breed clubs are keen to eradicate these diseases from their breeds while still maintaining the breed standard. Health schemes that monitor inherited diseases through clinical testing of disease phenotypes [23-25] (https://www.bva.co.uk/Canine-Health-Schemes/Eye-scheme/, accessed 24/02/2023), and commercially available DNA tests for known disease-associated variants [26], have been widely implemented with the aim of reducing the incidence of canine inherited diseases and improving canine health.

1.3.1 Health schemes

National kennel clubs have an important role in the implementation of breeding tools and strategies for the promotion of health and welfare [27]. Organisations such as the UK Kennel Club, International Sheep Dog Society (ISDS), and British Veterinary Association (BVA) have implemented schemes that aim to help responsible breeders avoid producing dogs affected by inherited disease and improve the health of their breed.

The BVA/UK Kennel Club hip and elbow dysplasia schemes, which use the scoring of factors seen in radiographs to describe the condition of a dog's hips or elbows, have been running since 1984 [23]. The schemes have been implemented extensively for the numerically large Labrador Retriever breed. The scores for each dog are recorded on a database and can be accessed through the Kennel Club website. The voluntary scheme is designed to allow dog breeders to make informed breeding decisions to reduce disease prevalence. The scheme has been broadly successful, and many breeds have shown a genetic trend towards improved hip and elbow condition [28]. The UK Kennel Club has introduced "estimated breeding values" (EBV) which are calculated from pedigree data, combining a dogs phenotype data with that of its relatives weighted using pedigree information [28, 29]. This requires a large number of dogs within a breed to have been

included in the hip and elbow dysplasia schemes [28]. EBVs are more accurate than phenotype measurement for predicting genetic risk of disease, and therefore are expected to have a higher efficacy for efforts to improve the health and welfare of the affected breeds.

The BVA/UK Kennel Club/ISDS eye scheme aims to allow breeders to avoid breeding dogs affected by inherited eye conditions [24, 25] (https://www.bva.co.uk/Canine-Health-Schemes/Eye-scheme/, accessed 24/02/2023). A certificate is given for all dogs tested through the scheme with details of their eye test results for each inherited condition included. The scheme aims to continually evaluate the evidence for eye conditions which may be inherited so that only conditions known to be inherited are included, and emerging conditions with strong evidence of inheritance can be added. Eye diseases screened for by the BVA/UK Kennel Club/ISDS eye scheme include, for example, hereditary cataract and progressive retinal atrophy (PRA) (https://www.bva.co.uk/Canine-Health-Schemes/Eye-scheme/, accessed 24/02/2023).

Health schemes such as these in the UK run by the BVA, Kennel Club, and ISDS, provide vast databases which when combined with pedigree information can become powerful research tools [23, 28, 30]. The eye schemes implemented by various organisations around the world have contributed to the considerable success in investigating the genetic cause of inherited eye conditions [31].

1.3.2 DNA testing

According to Online Mendelian Inheritance in Animals (OMIA), at least 338 putative causal variants have been found for monogenic canine inherited diseases and traits [32] (http://omia.org/, accessed 24/03/2023). There are DNA tests commercially available for over 152 variants associated with an inherited disease either in a single dog breed or across multiple breeds [26]. Important examples of canine DNA tests include those for eye diseases such as PRA and neurological diseases including ataxia [26]. DNA tests, which are typically for autosomal recessive inherited diseases, allow dog breeders to make informed breeding decisions based on if a dog is heterozygous, or homozygous for the disease-causing variant or the non-disease-causing allele. The subsequent disease status for a fully penetrant autosomal recessive disease is highly predictable in progeny when the genotypes of the parents are known. In this way dog breeders can avoid producing dogs affected by a disease, and work to reduce the frequency of the disease-causing

variant within the breed, a strategy that has been successfully implemented in several breeds [33].

1.4 Canine genetic research

1.4.1 Resources for canine genetic research

The investigation of canine inherited disease and traits has been facilitated by the development of resources and technologies dedicated to the canine genome. Early resources included microsatellite linkage maps and bacterial artificial chromosome (BAC) libraries [34, 35]. A key development was the use of fluorescent *in situ* hybridisation to identify the 38 canine autosomes, a challenge due to the small size and acrocentric morphology (centromeres located towards one end) of many of the canine chromosomes [36]. Sequencing of the canine genome (approximately 2.41 gigabases (Gb) long) later provided a powerful tool for canine genetic research [5]. It took six years from the first canine microsatellite linkage map to sequencing the first canine genome, and the field has continued to progress rapidly alongside the developments made in human genetic research.

1.4.1.1 Microsatellite linkage map

A linkage map contains the relative locations of markers on a chromosome as determined by the frequency of recombination. Microsatellite repeats are useful as markers for linkage maps because they are easy to genotype, are polymorphic (variable), and in mammalian genomes are randomly, and widely, distributed [34]. Early linkage maps for the canine genome included comparatively low numbers of markers. One of the first linkage maps, published in 1997, included 150 microsatellite markers: 139 in 30 different linkage groups and 11 which were informative but not linked to any of the other markers [34]. The largest linkage group included nine markers. Soon after (1999), an expanded linkage map was developed which included 276 markers, 268 assigned to 40 linkage groups; 29 linkage groups had ordered markers, and three were assigned to specific chromosomes [37]. Although none of the markers included were located on the male Y chromosome, five were X-linked. It was estimated that over 55% of the genome was flanked and covered by markers in this linkage map. Construction of a linkage map with approximately 3,000 microsatellite markers (published in 2010) was facilitated by the availability of a canine reference genome [38]. A comprehensive fine-scale genetic map

was later produced using 3.5 million autosomal single nucleotide polymorphisms (SNPs), and 198,000 SNPs on the X chromosome, identified by whole-genome sequencing of 51 free-ranging, non-breed, dogs [39].

1.4.1.2 BAC library

A BAC is a modified bacterial F factor plasmid into which fragments of eukaryotic genomes can be inserted [40]. BAC vectors contain sequences for restriction enzyme recognition sites that enable the incorporation of DNA to form a recombinant molecule which can then be inserted into a bacterial host. BAC vectors carry F Factor genes for replication, and at least one antibiotic resistance marker allowing easy isolation of the bacteria into which the BAC has been successfully inserted. The bacteria can be cultured to generate colonies containing many identical copies of the original DNA fragment. A library of BAC clones can contain a copy or copies of every part of an organism's genome: a genomic BAC library. BACs are useful for the generation of genomic libraries because they can carry fragments up to approximately 300 kilobase (kb), are easy to purify, and are stable for many generations of growth [35, 40]. A genomic BAC library for the canine genome (Dobermann Pinscher) was constructed in 1999 and made available to the academic community [35]. The library included around 166,000 clones, with a mean insert size of 155kb, and was predicted to give a redundancy (coverage) of 8.1, or in other words the entire genome was represented over eight times.

1.4.1.3 The dog genome sequence

Arguably the most powerful resource for canine genetic research is the availability of a high-quality genome sequence for the dog. The first draft dog genome sequence, from a male black standard poodle called Shadow, was published in 2003 [41]. This draft sequence had 1.5-fold sequence coverage, with many gaps, and was estimated to cover approximately 78% of the genome. This incomplete sequence was enough, however, to identify 18,473 putative canine orthologues for human genes. A higher quality canine genome sequence was published in 2005 for a female Boxer called Tasha [5]. This Boxer was chosen because of its low levels of heterozygosity, the lowest among dogs tested for a set of loci, making assembly of the genome simpler due to the high similarity of the two copies of each chromosome [5, 31]. However, later analysis showed the heterozygosity in other breeds is not greatly different to that seen in Tasha. The use of a female dog meant that the X chromosomes and autosomes had equal coverage. The Boxer genome had

approximately 7.5-fold sequence coverage, and had an initial assembly, named CanFam 1.0, which was closely followed by the updated CanFam 2.0 (BROADD2) assembly [5]. The contigs (contiguous sequences, i.e. sets of overlapping sequence reads) were around 50 times larger than those in the poodle sequence, with half of the bases located in contigs over 180kb in size. This meant that for most genes the sequence would be within a single contig, without gaps. The quality of the base calls was also high; quality scores were above 40 for 98% of bases. The quality scores used were Phred quality scores; a score of 40 means that there is 1 in 10,000 probability of an incorrect base call [42, 43]. In 2011 the CanFam3.1 canine genome assembly was released (GenBank Assembly ID GCA_000002285.2). The CanFam 3.1 assembly had substantially larger contigs; half of bases were in contigs over 267kb in size and was more complete with fewer gaps and smaller total gap size. The availability of a high-quality canine genome sequence facilitated the use of techniques such as genome-wide genotyping arrays and next generation sequencing, allowing rapid advancement in canine genetic research.

For a decade CanFam 3.1 was the reference genome used by most researchers in canine genetics. Recently, however, improved canine genome assemblies have become available, generated using long-read sequencing technologies [44-46]. One of the assemblies, Dog10K_Boxer_Tasha (GenBank GCA_005444595.1), is an improved assembly for the Boxer, Tasha [46]. Assemblies have also been developed for other breeds, including the German Shepherd (UU_CFam_GSD_1.0, GenBank GCA_011100685.1), and Great Dane (UMICH_Zoey_3.1, GenBank GCA_005444595.1) [44, 45]. The new assemblies have increased contiguity, and many of the remaining gaps have been filled in. Genome annotation has also been improved; the new Boxer assembly includes over 1200 protein-coding transcripts not previously annotated on CanFam 3.1 [46]. The availability of genome assemblies from different breeds will allow an improved understanding of the genetic variation between breeds. For example, much of the variation observed between the Great Dane genome UMICH_Zoey_3.1 and CanFam 3.1 (Boxer) was structural, accounting for > 13.2 megabases (Mb) of sequence difference, and these structural variants were primarily canine short interspersed element (SINEC) and long interspersed element-1 (LINE-1) sequences [45]. The new reference genomes present a challenge for the canine genetics field, however, in that researchers will need to choose to which reference they align resequencing data. Their choice may depend on which trait or disease they are investigating, or in which breed. Researchers will want to ensure

compatibility between their own data, and with that of collaborators, but they are unlikely to wish to align all their data to all available reference genomes. To do so would be expensive in time and resources. To ensure the canine genetics field remains highly collaborative it is likely to be advantageous for a primary reference to be chosen. Alternatively, a pan-genome that incorporates variation in genome sequences across multiple breeds could have great utility [47].

1.4.2 Approaches for canine genetic research

1.4.2.1 Linkage analysis

Genetic linkage studies use extensive pedigrees to identify regions of the genome inherited with the disease. To do this effectively, inbred colonies of dogs can be necessary. An illustrative example of a successful genetic linkage study in the dog is the identification of a locus for canine progressive rod-cone degeneration (PRCD), a type of PRA, using a colony of Miniature Poodles, Beagles, and Beagle-crossbred dogs [48]. Informative pedigrees were generated through the breeding of known affected dogs with unrelated dogs to generate carriers, followed by backcrossing carriers with affected dogs (**Figure 1.2**). An informative pedigree allows the carrier or affected status of dogs to be determined, permitting investigation of marker segregation with disease, and therefore marker linkage with the disease locus. Genotyping of 100 microsatellite markers identified linkage between a marker and the PRCD locus [48]. Genotyping of additional microsatellite markers, including gene-specific microsatellites, allowed the relative mapping location of the PRCD locus, close to the centromere on chromosome 9, to be determined.

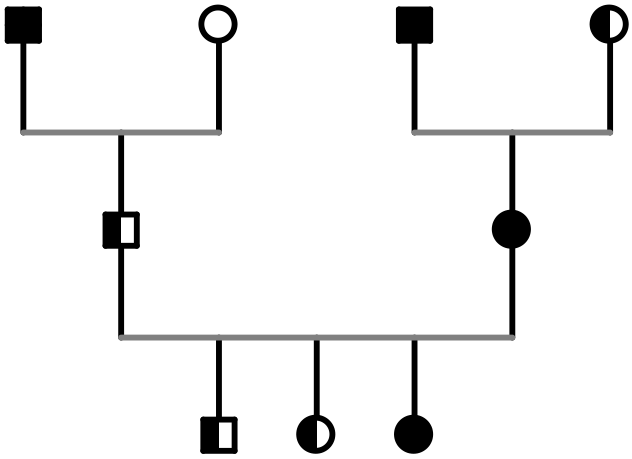


Figure 1.2.Example of an informative pedigree used for linkage analysis. The pedigree diagram was recreated from Acland et al (1998) [48]. Squares represent male dogs, circles

female dogs. Solid black shapes are homozygous PRCD cases, hollow are unaffected dogs homozygous 'normal' at the PRCD locus. Half-filled shapes represent heterozygous 'carriers'.

1.4.2.2 Genome-wide association studies

Genome-wide association studies (GWAS) utilise SNP arrays to genotype markers distributed across the genome in a set of cases (individuals with the phenotype of interest) and a set of controls (individuals without the phenotype of interest) [49, 50]. GWAS can also be used to investigate quantitative traits, such as height, across a population-based cohort [49, 50]. Statistical analysis is carried out, testing differences in SNP allele frequency, to investigate if any SNP markers are associated with the phenotype being investigated. Regions of the genome associated with the trait can be identified and then subsequently investigated to identify genetic variants contributing to a phenotype, which could be a behavioural or morphological trait, or an inherited disease. A key advantage of GWAS is that it is hypothesis-free; i.e. it is not reliant on any knowledge of the underlying molecular biology of the disease or trait being investigated, such as is required for candidate gene studies, and can therefore identify novel regions associated with disease [51].

The development of SNP arrays for the dog was facilitated by the availability of a canine genome sequence. The Illumina Canine HD beadchip array was designed by the LUPA Consortium, which was a collaboration of canine geneticists and veterinarians in Europe [3]. The array was released prior to the 2011 LUPA publication, and until recently this was the largest genotyping array available for the dog, with 173,662 SNP markers. In 2017 the Axiom Canine HD Array was released, which genotypes up to 729,642 markers. The marker density available for a GWAS can be increased using genotype imputation (see chapter 1.4.2.5).

The thousands of SNP markers used in a GWAS analysis means that a huge number of independent tests are carried out, and this results in a very high probability of Type I error (i.e. false positive results) [52]. To account for this, it is necessary to correct for multiple testing, such as by setting a stringent threshold for significance utilising the Bonferroni correction (dividing 0.05 by the number of variants included in the GWAS). However, the Bonferroni corrected threshold can be too conservative because LD between markers

means that the number of true independent tests are lower than the number of SNPs, and this can increase the chances of Type II error (i.e. false negative results).

Population structure in GWAS, for example if cases and controls are not evenly distributed across sub-populations, can result in false-positive statistical associations of SNP markers due to differing allele frequencies between populations, potentially obscuring meaningful associations with disease [49, 53]. Statistical mixed models can be used to compensate for population stratification [53, 54]. A mixed model generates variables for each individual using a relatedness matrix and uses variables modelled using both fixed and random effects to account for population structure in an association analysis.

Two main diagnostic plots are used to visualise GWAS data and detect potential population structure or systematic bias: quantile-quantile (QQ) plots, and multidimensional scaling (MDS) plots. QQ plots compare the results obtained to those that would be expected in hypothetical GWAS data to determine if more statistically significant associations were observed than would be expected by chance (**Figure 1.3**), which would be seen as a wide-ranging deviation from a null line that represents the expected P-values and represents an inflation of test statistics [49]. If this is due to population stratification, adjustment using a mixed model as described above will mitigate against these effects. An MDS plot uses values extracted from the genotyping data by comparing the genetic distance between the individuals in a relationship matrix [55]. Two values representing genetic distance are plotted on a scatter plot to give a two-dimensional representation of the relatedness of the individuals included in the GWAS (**Figure 1.4**). Population stratification would be observed if cases and controls formed distinct or diverging clusters.

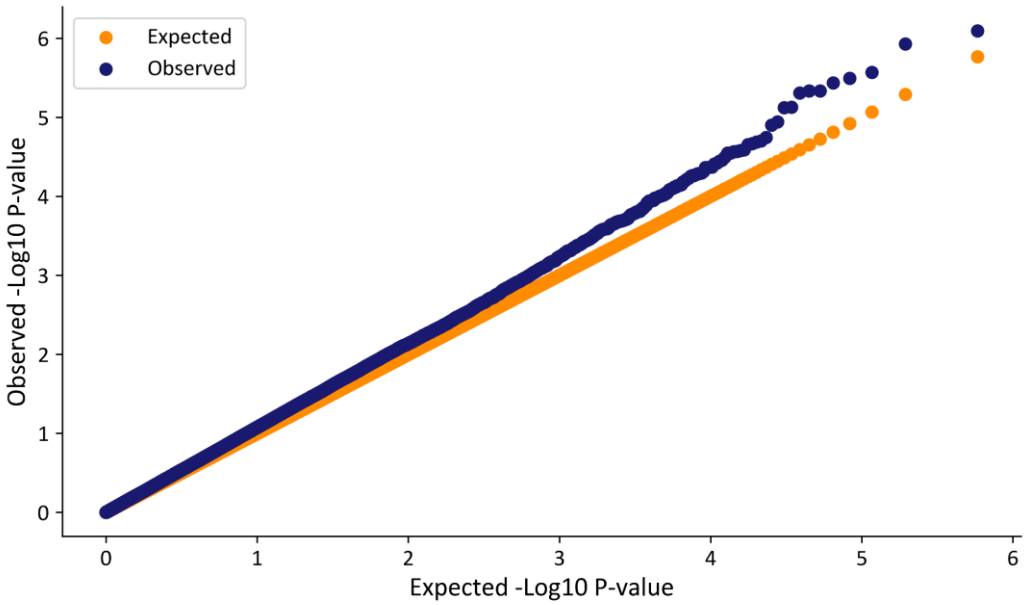


Figure 1.3. Example QQ plot showing mixed model adjusted -Log10 P-values from a GWAS. The GWAS of Border Collie dogs included 291,450 SNPs (including imputed data up to Axiom Canine HD array SNP density), and 104 idiopathic epilepsy cases and 167 controls (see chapter 7).

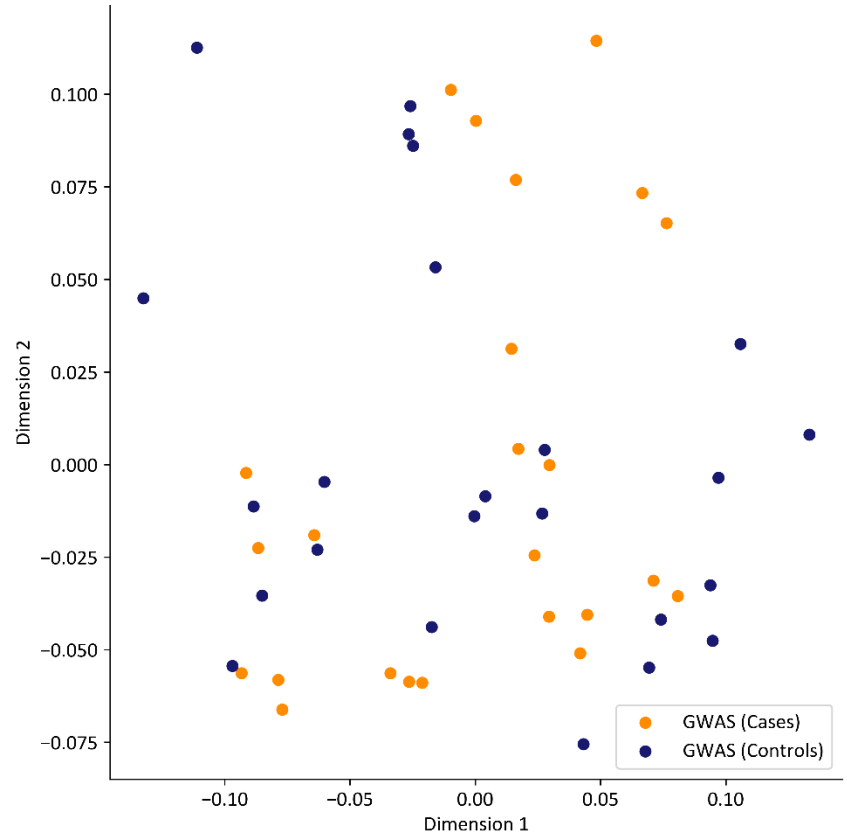


Figure 1.4. Example MDS plot showing cases and controls from a GWAS. MDS plot is from chapter 5 and was generated for Norwich Terrier dogs (24 paroxysmal dyskinesia cases and 24 controls) genotyped using the Axiom Canine HD array.

1.4.2.3 Candidate gene analysis

A candidate gene approach is a way of investigating the genetic basis of a disease or trait if there is a strong biological rationale implicating a specific gene with the disease [56]. A gene or genes of interest are sequenced and interrogated for potential causal variants [57]. Alternatively, the candidate genes can be investigated by genotyping tagging microsatellite (or any other type, for example SNP) markers [58]. The choice of gene(s) is based on prior knowledge, which may be incomplete or inaccurate, for example if the gene is known to be associated with a phenotype in humans, a different breed, or in a knock-out mouse model. The candidate gene approach is a rapid and cost-effective way of identifying potentially causal variants. Due to the prior knowledge required, discoveries made in this way will not be novel, but can allow the development of a DNA test for the dog breed or breeds under investigation. However, candidate gene studies can be subject to bias, and both type I and type II error [59].

An example of a canine candidate gene study utilising microsatellite markers was the identification of gene mutations in *HSF4* that are associated with hereditary cataract in multiple breeds [58]. This study genotyped markers flanking 20 candidate genes for cataracts. The genes were considered candidates for canine cataracts because they had previously been implicated in cataracts in humans or mice. The microsatellites were genotyped in affected dogs, carriers, and unaffected dogs of the Staffordshire Bull Terrier, American Cocker Spaniel, Golden Retriever, and Miniature Schnauzer breeds. An association with disease was found for *HSF4* in the Staffordshire Bull Terrier. Sequencing the gene identified mutations which were present in affected Staffordshire Bull Terriers, Boston Terriers and Australian Shepherds.

A candidate gene approach using next generation DNA sequencing methodology to re-sequence the protein-coding exons of the canine Duchenne Muscular Dystrophy (*DMD*) gene identified the single base pair deletion that has been suggested to cause dystrophin-deficient muscular dystrophy in a Norfolk terrier dog [57].

1.4.2.4 Next generation DNA sequencing

DNA sequencing technologies underwent rapid development during the time of the human genome project (1995 - 2003) and in the years after [60]. The sequencing of the first genomes relied on versions of Sanger sequencing which involved the fragmentation of the genome, subcloning of the fragments, and semi-automated sequencing of each

fragment in individual reactions using capillary electrophoresis to determine the sequence [5, 40, 41, 61]. NGS, or massively parallel sequencing, technologies make possible a dramatically higher throughput. The Illumina next-generation sequencing technology sequences all of the fragments of an entire genome in parallel on a single flow-cell (chip) [61].

NGS can be used to identify candidate causal variants for canine inherited disease in multiple ways. Genome-wide mRNA sequencing (mRNA-seq) can be used to sequence all transcripts expressed in a tissue, and then candidate genes interrogated for potentially causal variants [62]. Targeted sequencing of a region of the genome identified through GWAS or linkage analysis can be used to interrogate the region to find the likely-causal variant [63, 64]. It has become feasible, as the cost of NGS has decreased, to sequence the entire genome of affected dogs and compare the variants identified to genomes of control dogs to identify mutations that segregate with disease [65, 66].

To ensure that the most use is made of the huge amount of data being produced through sequencing the whole genomes of dogs, international consortia have been formed to share the data between research groups. An example is the Dog Biomedical Variant Database Consortium (DBVDC), which shares variant data from over 600 canine whole genome sequences (WGS) [67]. The data shared by the DBVDC includes information about the breed of the dog sequenced and the DBVDC member who submitted the data, but no information about the disease status of the dog. In this way data can be shared without compromising the research interests of the consortium members, but the information can be obtained from the individual members at their discretion if necessary. The database allows the members to screen variants of interest within a large panel of genomes of many different breeds, something that would not be possible for smaller research groups which may not have the resources or facilities necessary to generate such a database on their own. The consortium ensures that data generated for the specific task of identifying a disease-causing variant by a research group can continue to be a useful resource to multiple research groups once the initial research objectives has been accomplished. The DBVDC variant data has contributed to the research of over 50 different inherited canine diseases [67]; including diseases with an autosomal recessive, autosomal dominant, X-linked, and presumed complex mode of inheritance.

Another international collaborative effort, Dog10k, aims to generate high quality WGS data for 10,000 canids [68, 69]. The data will facilitate investigations into the genetics of traits and disease, and an improved understanding of genetic diversity across breeds and the effect of domestication. Variant data from the Dog10k WGS will be shared as the project progresses, and added to the genome-wide variant data of over 528 purebred dogs, 36 mixed breed dogs, 104 village and feral dogs, and 54 canids from five wild species, that has already been generated by the canine genetics community and is publicly available [69, 70].

1.4.2.5 Genome-wide genotype imputation

Genotype imputation is a computational approach that allows unobserved genotypes for individuals in a dataset (the study dataset) to be predicted using variant data for a set of different individuals with observed genotype data at a higher density (the reference panel) [71, 72]. Statistical models are utilised to extrapolate from haplotype patterns observed in the reference panel and predict unknown genotypes in the study dataset [72, 73]. Genome-wide genotype imputation has typically been used to enable data from different arrays to be combined in meta-analysis and to increase the resolution of GWAS data generated using a genotyping array, although it has also been applied to fill in the gaps in low-coverage WGS [74, 75]. The use of genotype imputation is well established in human genetics, and the approach is increasingly being used in canine genetics [76-79]. The accuracy of imputation can be dependent on the reference panel used, with large datasets including a combination of population-specific and more divergent and cosmopolitan individuals (for example, other dog breeds) providing best results [73, 76]. The increasing availability of large datasets of canine genetic variation, through sequencing collaborations such as the DBVDC and Dog10k, is invaluable for generating multi-breed reference panels for canine genotype imputation [67, 69].

It has been reported that higher density marker sets could facilitate the identification of most moderate- to large-effect loci underlying canine complex inherited diseases [80]. Imputation of older datasets to increase SNP density is therefore likely to be important for canine complex disease research.

1.5 Scientific interest in the dog as a natural disease model

In addition to the veterinary importance of identifying the variants causing canine inherited disease, and the welfare impact this can have on breeds, canine genetic research is also of general scientific interest. The dog can be an interesting model for disease because it is a large mammal that shares a similar living environment with humans; dogs live in the same homes as humans, in towns and cities, and are exposed to the same environmental factors that can have an impact on health. Many inherited canine diseases, including ataxia and epilepsy, are close analogues of human disease, and can have a similar aetiology [4, 63, 81-83]. OMIA includes over 550 canine traits or diseases cross referenced to corresponding phenotypes recorded on online Mendelian inheritance in man (OMIM), and have therefore been identified as potential models of human traits [32, 84] (http://omia.org/, accessed 24/03/2023). The LUPA consortium aimed to use canine inherited disease as a model to identify genes of relevance to human disease [3]. As mentioned in chapter 1.2.2 the haplotype structure of the canine genome can be advantageous for genetic studies. Extensive LD means that smaller sample sets and genotyping arrays with fewer genetic markers can be required for GWAS [5]. Other advantages of the dog as a disease model are the availability of genealogical data from kennel clubs and other organisations, and phenotypic data accrued through health schemes (see chapter 1.3.1 for more information) [23, 28, 30, 31].

Genetic studies into canine inherited diseases similar to human conditions have identified variants in genes previously associated with human disease, and also in novel genes not previously implicated [81]. Novel inherited neurological disease-associated genes identified in the dog have later been demonstrated to have a role in human inherited disease, for example *CAPN1* for cerebellar ataxia in Russell Terrier Group dogs and autosomal recessive hereditary spastic paraplegia or spastic ataxia in humans [63, 82, 83].

Canine epilepsy is highly similar clinically to human epilepsy and represents an important naturally and spontaneously occurring disease model [4, 85]. Experimentally induced rodent models of epilepsy have been extensively used for comparative research, but often have differences to the naturally occurring epilepsy in humans [4]. Non-human primates have similarly been used to study experimentally induced seizures, as have cats [85]. Epilepsy in cats can occur naturally, but idiopathic epilepsy (IE) is relatively uncommon in comparison to dogs, with some studies reporting incidences from 0% to

59% in cats with seizures [86], whereas IE has been found to be the most common cause of seizures dogs [87]. Other advantages of the dog as a model organism are that the clinical surveillance of disease by specialist veterinary neurologists can be extensive, and anti-epileptic medications used in humans often work well in dogs [4].

1.6 Movement disorders

1.6.1 Definitions and characteristics in humans

Movement disorders are a heterogeneous group of diseases of the nervous system that cause excessive movement (hyperkinesia), abnormal movement (dyskinesia), reduction of voluntary and reflexive movement (hypokinesia), or repetitive involuntary movements (dystonia) [88, 89]. Abnormalities in the basal ganglia are associated with most movement disorders, but there are some exceptions [89, 90]. Ataxia (incoordination) is a movement disorder often caused by abnormalities of the cerebellum. Some movement disorders can lack structural pathology observable through computed tomography (CT) or magnetic resonance imaging (MRI) (see chapter 1.6.3), making diagnosis challenging [89].

1.6.1.1 Ataxia

The term 'ataxia' is used to refer to both the clinical phenotype of incoordination and to a group of degenerative neurological diseases which are characterised by progressive ataxia [91]. Ataxia can be caused by dysfunction in the cerebral cortex, cerebellum, or basal ganglia [92]. This group of disorders can have an acquired, genetic, or unknown, cause. Causes of acquired ataxia include trauma, tumours, and demyelinating diseases such as multiple sclerosis. Ataxia with a genetic cause can be autosomal recessive or autosomal dominant. In some cases it is caused by mutations in the mitochondrial DNA, or ataxia can be X-linked such as in the case of Fragile X tremor-ataxia syndrome [92].

Spinocerebellar ataxias are clinically and genetically heterogeneous [91, 93, 94]. Although the name indicates clinical changes in both the cerebellum and spinal cord, the spinal cord is not affected in many diseases within this group and there are often other regions of the nervous system involved [91]. Age of onset for spinocerebellar ataxia varies widely, from childhood to late in life [91, 93, 94]. Spinocerebellar ataxia, SCA, refers to the inherited diseases with an autosomal dominant mode of inheritance [91]. The acronym SCAR is used to denote autosomal recessive ataxias, and SCAX those which are caused by variants on the X chromosome [93, 94]. Each distinct spinocerebellar ataxia is named SCA, SCAR, or SCAX,

followed by a number representing the order in which the genetic cause was determined. There are exceptions to this naming convention, mostly for ataxias defined less recently [95]. SCAs and SCARs are relatively rare genetic diseases, with pooled average prevalence estimates of 2.7 SCA cases per 100,000 and 3.3 SCAR cases per 100,000 in a systematic review of population-based prevalence studies [96]. There are over 40 distinct SCAs, which can be split into those caused by repeat expansion mutations (where regions of repetitive nucleotide sequences are increased in length, often causing long stretches of glutamine), and the rare SCAs caused by non-repeat mutations [91]. Variants can be pathogenic through a broad range of mechanisms, including; proteotoxicity, RNA toxicity, ion channel dysfunction, impaired mitochondrial function, and disruption of nuclear functions. Over 100 genes have been associated with SCARs, also covering a wide array of pathologies, including; impaired DNA damage repair, disruption of mitochondrial homeostasis, impaired phospholipid or sphingolipid metabolism, loss of autophagy-lysosomal activity, and damaged cilia function [93]. For most SCAs and SCARs there is currently no treatment [91, 93].

The most common ataxia is Friedreich Ataxia, which is a SCAR (although it isn't included within the SCAR nomenclature) caused by mutations in *FXN*, the gene encoding Frataxin, a mitochondrial protein that has a role in intracellular iron homeostasis and which is strongly expressed in the heart, dorsal root ganglia, and in the cerebellum [93, 95, 97, 98]. In most affected individuals Friedreich Ataxia is caused by homozygous GAA repeat expansions in the first exon of *FXN*, the remainder are compound heterozygotes for a repeat expansion and a second variant within the same gene [98]. This disorder begins in childhood, with an average age of onset of 10 years, and presents with ataxia, decreased reflexes, sensory neuropathy (loss of sensation), scoliosis (curved spine), and abnormal eye movement [98, 99]. Friedreich Ataxia eventually progresses to an inability to walk, and affected individuals have a reduced lifespan. The progressive ataxia is caused by degeneration of the cerebellum in combination with sensory neuropathy and involvement of the vestibular nerve [99].

1.6.1.2 Paroxysmal dyskinesia

Paroxysmal (meaning occurring periodically with a well-defined onset and cessation of clinical signs) movement disorders include the episodic ataxias and paroxysmal dyskinesias [100]. Paroxysmal dyskinesias (PxDs) are a group of relatively rare movement

disorders characterised by short-lived recurrent episodes of involuntary abnormal movement without loss of consciousness [101]. PxDs can be genetic, they can be caused by trauma, or they can be secondary to vascular, metabolic, or immune-mediated syndromes [102]. There are considerable similarities between PxDs and epilepsies, and often overlap in disease phenotypes and aetiologies [103].

PxDs have been divided into three groups based on the cause or trigger that precedes the attacks: kinesigenic (PKD), nonkinesigenic (PNKD), and exercise-induced (PED) [104]. PKDs are triggered by an abrupt movement such as standing up, and the most common clinical manifestation is brief uncontrollable contraction of the muscles (dystonia) of limbs either on one side of the body, or (less commonly) both [101, 104, 105]. PNKD attacks are not caused by movement, but typically by alcohol, coffee, or strong emotions such as stress, and can also occur unprovoked [101, 104, 105]. PNKD attacks are similar to those of PKD, but longer in duration and typically do not occur as often. PED episodes are typically caused by extended periods of exertion, and are usually characterised by dystonia [101, 104, 105].

Despite the above classification of the PxDs into three distinct groups, these syndromes are heterogeneous and affected individuals can have triggers for their episodes that fit within more than one of these definitions [106, 107]. Where genetic causes have been identified it can be more appropriate to refer to gene-specific phenotypes. At least 34 genes have been implicated in PxDs, including genes encoding proteins with roles in synaptic function, transmembrane transporter proteins, proteins involved in the second messenger intracellular signalling systems, ion channels, and mitochondrial proteins [107]. For this brief overview of the PxDs in humans the focus will be the three original classifications and the genes predominantly, or first, associated with them.

Variants in *PRRT2*, which encodes proline-rich transmembrane protein 2, have been associated with a high proportion of PKD, the most common PxD phenotype [106, 107]. PRRT2 is thought to have a role in normal transmission at synapses in the cerebellum [108]. Variants in *PRRT2* act in an autosomal dominant manner with incomplete penetrance and cause a spectrum of phenotypes which can be exclusively PKD or benign familial infantile seizures, or benign familial infantile seizures in combination with PKD [107]. More rarely, *PRRT2* mutations are associated with either of the other PxD classifications (PNKD and PED) or with any of a range of neurological symptoms including

episodic ataxia, childhood-absence epilepsy, and febrile seizures [106, 107]. PKD is treated effectively with anticonvulsants, and, although quality of life can be poor in undiagnosed cases, patients tend to have a normal life expectancy [101].

Variants in the *PNKD* gene (previously called *MR-1*) cause up to 71% of PNKD, the second most common PxD phenotype [107]. The PNKD protein has a role in synaptic neurotransmitter release [109]. Alternate splicing produces at least three different PNKD proteins, with differing expression patterns, including a long (PNKD-L) (which is expressed only in the CNS), medium (PNKD-M), and short (PNKD-S) splice variant [107]. Mutations in each of these have been associated with neurological disorders, although only those affecting the long and short proteins have been associated with PNKD, and one of these variants presents with PNKD and episodic ataxia. Other variants affecting the long and medium proteins cause various phenotypes including PKD and seizures, and episodic ataxia [107]. Variants in *PNKD* tend to cause PNKD in an autosomal dominant manner [106]. The knowledge of factors (including alcohol, caffeine, and stress) that trigger PNKD attacks allows patients to be taught to reduce their exposure, and PNKD attack frequency has been reported to reduce with age [101].

The first gene to be associated with PED was the glucose transporter 1 (GLUT1) gene *SLC2A1*, although mutations within this gene only explain a small proportion of PED cases and the phenotype is heterogeneous [106, 110]. GLUT1 has an important role in transporting glucose across the blood-brain barrier [111]. Mode of inheritance is typically autosomal dominant for PED, although other *SLC2A1* associated (GLUT1 deficiency) syndromes can be autosomal recessive [106]. GLUT1 deficiency can cause PED, in some rarer cases PNKD, and other phenotypes include epilepsy and ataxia. Modifying the patient's diet can be used to treat GLUT1 deficiency PED, typically through a ketogenic diet [112]. Treatment with the synthetic, medium-chain triglyceride 'triheptanoin' has also been shown to be effective [113].

1.6.2 Canine movement disorders

There are well characterised examples of hereditary ataxia and paroxysmal dyskinesia in dogs, some of which have considerable similarity to the movement disorders found in humans [114-116]. There has been success in investigating the genetic causes for several canine movement disorders, and some of the genes implicated in dogs had not previously been associated with a disease phenotype in humans. Canine movement disorders could

therefore be an important model of human disease and may help improve understanding of the underlying molecular mechanisms involved.

There is a scarcity of publications reporting the incidence and prevalence of movement disorders in the general canine population, or in specific breeds. The hereditary ataxias as a group are a key cause of movement disorders in dogs, though the specific diseases can be rare [114]. Studies of cerebellar degeneration in American Staffordshire Terriers and Scottish Terriers estimated a prevalence of 1 in 400 and 1 in 1,335 respectively, by comparing the number of dogs registered with a kennel club within a given range of years to the number of dogs known to have been diagnosed with the disease [117, 118]. This approach will at best give an estimate of the minimum prevalence because the studies are unlikely to identify all affected dogs. Exercise-induced ataxia was reported in 0.4% of Australian Greyhounds following races in a study that observed starters for 536 races [119]. An estimated prevalence of PxD in the Norwich Terrier of 13% has been reported through a breed-wide UK survey [120]. However, in this questionnaire-based study there was a high likelihood of bias caused by owners of affected dogs being more likely to participate. In another breed-wide survey, this one for Dutch Border Terriers born between 1998 and 2001, or between 2003 and 2006, episodes with abnormal movements or posture were reported for 7.6% and 4.8% of dogs respectively [121].

1.6.3 Canine ataxia

1.6.3.1 Clinical characteristics and heterogeneity

In this section the clinical characteristics and heterogeneity of canine ataxia will be outlined. The two key categories of canine ataxias which have been described are cerebellar cortical degeneration and spinocerebellar ataxia, but there are canine ataxias that do not fit within either of these groups.

1.6.3.1.1 Cerebellar cortical degeneration

Cerebellar cortical degeneration is characterised by degeneration of the cerebellar cortex, and manifests clinically with signs reflecting cerebellar dysfunction [114]. These signs can include a wide-based stance, dysmetria (over or under extending limbs when moving), swaying of the trunk, and a broad, low frequency, tremor (intention tremor).

The neuropathology of cerebellar cortical degeneration includes the loss of Purkinje neurons, and atrophy of the molecular layer, granular layer, and white matter [114]. The

neuronal degeneration varies between breeds. Purkinje cell loss and granular cell layer depletion has been reported in the majority of cerebellar cortical degeneration cases, including in the American Staffordshire Terrier, Beagle, Finnish Hound, Old English Sheepdog, Gordon Setter, Hungarian Vizsla, and Scottish Terrier breeds [62, 65, 117, 118, 122-124]. Thinning of the cerebellar molecular layer was also described in American Staffordshire Terriers, Scottish terriers, Old English Sheepdogs, Hungarian Vizslas [117, 118] [65, 123]. The subsequent shrinkage of the cerebellum can be detectable through MRI [118, 122].

Age of onset for cerebellar cortical degeneration is variable between breeds, with some having a neonatal or juvenile onset, and others presenting with the disease when adult. For example, in Beagles the ataxia is noticeable from when they start walking [62]. In the Finnish Hound the clinical signs have an onset ranging from four to 12 weeks [122]. Onset in Hungarian Vizslas has been reported at about three months of age [65]. Old English Sheepdogs and Gordon Setters develop signs of progressive ataxia between six months and four years of age [123]. Onset in the American Staffordshire Terrier is typically in adult dogs and most are diagnosed with cerebellar cortical degeneration between 4 and 6 years of age [118].

Disease progression can also be variable between breeds [114]. In some breeds progression can be rapid and result in early euthanasia, in others slow progression and stabilisation results in a relatively mild phenotype for the dog's entire life. For example, progression is typically slow in Gordon Setters and Scottish Terriers; whereas ataxia is severe and progressive in the Beagle and rapidly progressive in the Finnish Hound [62, 117, 122, 125].

1.6.3.1.2 Spinocerebellar ataxia

Spinocerebellar degeneration affects the medulla, spinal cord, or both, in addition to (or without) affecting the cerebellum [114]. There is variation between breeds, and multiple forms have been found within breeds.

There are at least two forms of spinocerebellar ataxia in Russell group terriers (Jack Russell Terriers, Parson Russell Terriers, and Russell Terriers) [63, 66]. One form is referred to as spinocerebellar ataxia with myokymia, seizures, or both (SAMS) [66]. This form of ataxia has an age of onset between two and ten months and is characterised by

cerebellar ataxia in combination with myokymia (involuntary quivering muscle contractions) and/or seizures [66]. The second form of spinocerebellar ataxia in Russell group terriers has a later onset, usually between six months to one year of age, and has not been reported to exhibit myokymia or seizures [63, 66]. An inherited ataxia with similar clinical signs and histopathology to that seen in Russell group terriers has been described in Smooth-Haired Fox Terriers. In both breeds the ataxia, which is progressive, is characterised by hypermetria (over-extended gait) and a gait that is described as dancing or bouncing [126, 127]. The histopathology findings show degenerative changes most pronounced in the spinal cord, without any reported cerebellar degeneration [126-128].

Multiple forms of cerebellar ataxia have also been described in the Malinois variety of Belgian Shepherd Dogs. Spongy degeneration with cerebellar ataxia (SDCA) in the Malinois has an age of onset ranging from 4 to 7 weeks [129-131]. Clinical signs have some heterogeneity, with moderate to severe ataxia observed in all cases, but additional signs including seizures and central blindness are seen in a subset of dogs that also show rapid disease progression. SDCA histopathology demonstrates spongy degeneration of the cerebellar nuclei and the presence of vacuoles within the granular cell layer and foliate white matter extensively, and some scattered lesions within the Purkinje cell layer and elsewhere in the brain [129]. Another spinocerebellar ataxia in the Malinois has some similarity to the SAMS reported in Russell group terriers, and presents with severe and progressive ataxia, moderate paraparesis (partial inability to move legs), and muscle contractions that resemble myokymia [132]. The histopathology shows swelling of axons and the presence of vacuoles, which is most pronounced throughout the spinal cord but also present within the dorsal and ventral nerve roots, brain stem, and cerebellum. Progressive ataxia, muscle spasm, and short episodic spastic fits in dogs from 12-14 days old has been described in yet another form of inherited ataxia in the Malinois variety of Belgian Shepherd Dogs [133]. This spinocerebellar ataxia, referred to as central nervous system (CNS) atrophy with cerebellar ataxia, shows degeneration within both the spinal cord and cerebellum, in addition to lesions elsewhere in the brain.

In the Italian Spinone, spinocerebellar ataxia is characterised by wide-based stance, hypermetria, compromised balance, and truncal swaying [64]. Age of onset is typically four months; the ataxia is progressive, and deterioration of balance eventually reaches a point that the affected dogs cannot stand or walk at approximately one year of age. There

is no substantial Purkinje cell loss or granular cell layer depletion observed in histopathological analysis [64].

Spinocerebellar ataxia in two litters of Alpine Dachsbracke dogs exhibited severe early onset ataxia and signs of impaired vision [134]. The neuropathology of this ataxia has similarities to SDCA, with spongy degeneration characterised by vacuoles and astrogliosis (increase in the number, and changes in morphology, of astrocytes) throughout the white and grey matter of the entire brain. The Purkinje and granular cell layers of the cerebellum are relatively spared, but the molecular layer shows mild astrogliosis and vacuolisation.

Cerebellar ataxia in the black Norwegian Elkhound is reported to have an onset of approximately four weeks of age [135]. Signs of degeneration within the cerebellum and brain stem are observed in histopathological analyses in the form of axonal swelling and spheroids within the Purkinje cell axons in the granular cell layer and vacuoles in the brain stem.

1.6.3.1.3 Other canine ataxias

The ataxias are phenotypically diverse, and some will not fit neatly within either of the above categories. Some ataxias do not show degeneration of the cerebellum or of the medulla or spinal cord. Other disorders for which ataxia is the primary sign show malformation, rather than degeneration, of the cerebellum.

Coton de Tulear dogs affected by neonatal cerebellar ataxia are unable to walk, and this ataxia is apparent from the age that their siblings begin to show coordinated movement [136]. The ataxia, which is severe from the start, does not worsen but also does not improve over time. The cerebellum and spinal cord do not show any lesions or substantial degeneration. However, signs of synaptic abnormalities are observed in Purkinje cells and the molecular layer.

An example of a canine ataxia characterised by a cerebellar malformation is Dandy-Walker-like malformation (DWLM) in the Eurasier breed [137]. Dandy-Walker malformation is a condition described in humans characterised by a specific set of malformations of the brain, including agenesis of the cerebellum [137, 138]. Cerebella in DWLM affected dogs lack the caudal sections of the vermis and cerebellar hemispheres [137]. DWLM cases show early onset, non-progressive, mild to moderate ataxia from the

age that dogs begin to walk; and follow-up in older dogs has demonstrated that the cerebellar ataxia can improve over time.

Ataxia can be the predominant sign for diseases that affect tissues in addition to those within the central nervous system, such as that seen in a neurodegenerative vacuolar storage disease in the Lagotto Romagnolo breed [139]. Progressive cerebellar ataxia in Lagotto Romagnolo dogs has variable age of onset (four months to four years) and a varying rate of progression between dogs. Neuropathological findings show degeneration of the cerebellum, including loss of Purkinje cells and depletion of the granular cell layer, and cytoplasmic vacuolisation throughout the central and peripheral nervous system [139]. Vacuolisation is also observed within the secretory epithelium and mesenchymal cells.

1.6.3.2 Diagnosis

Diagnosis of hereditary ataxia typically involves the exclusion of other possible causes [114]. Blood and cerebrospinal fluid (CSF) testing, carried out to investigate possible metabolic, inflammatory, or infectious diseases, are typically normal for hereditary ataxias. MRI is used to identify inflammation or structural irregularities. MRI can be useful in diagnosis of dogs with ataxia caused by cerebellar malformation [137]. Atrophy of the cerebellum can sometimes be seen on MRI images for dogs with hereditary ataxia, but this is not always the case. For example, MRI images of the brain in Italian Spinoni affected by spinocerebellar ataxia, Russell group terrier SAMS cases, and Belgian Shepherd Dogs with SDCA, do not show any significant differences to those of unaffected dogs [64, 66, 129]. Post-mortem histopathology can be necessary to determine the presence or absence of cerebellar or spinocerebellar degeneration.

1.6.3.3 Genetics

Genetic testing for putative ataxia-causing mutations can be extremely useful for diagnosis.

There has been considerable success in investigating the genetic causes of canine inherited ataxias, and a summary of what is known about the genes involved is given in **Table 1.1**. At least 15 genes have been implicated in canine ataxias, and the putative causal variants have been identified for over 15 breeds of dog; some genes have been associated with ataxia in multiple breeds and some individual breeds have multiple forms

of ataxia each putatively caused by a variant in a different gene (**Table 1.1**). Eight of the genes have been associated with inherited ataxia in humans. As demonstrated above, the inherited canine ataxias are heterogeneous and can be difficult to categorise, particularly prior to post-mortem histopathology. As causal genetic variants are identified, it could become more appropriate to categorise the canine ataxias by the associated gene, or by the gene's function, or even by the molecular pathophysiological mechanisms underlying the disease where this can be elucidated. There are at least five genes associated with canine ataxia that have roles in ion transport (ion channels, ion pumps, or ion transporters) (section 1.6.3.3.1); two (or potentially three) genes with roles in autophagy (section 1.6.3.3.3); two involved in the signalling pathway of the neurotransmitter glutamate (section 1.6.3.3.2); two with functions relating to the degradation of proteins (section 1.6.3.3.4); and two genes that have been linked to normal development of the cerebellum (section 1.6.3.3.5) (**Table 1.1**). Genes with roles in lipid metabolism, mitochondrial function, neuroprotective signalling pathways, and selenium transport, have also been identified (section 1.6.3.3.6). The genes associated with canine inherited ataxia are discussed in the following sections, grouped by their suggested biological roles.

To the author's knowledge, at the time of writing, all canine inherited ataxias for which a putative disease-causing variant has been elucidated are inherited in an autosomal recessive manner (**Table 1.1**). Linkage analysis, genome-wide association analysis, and homozygosity mapping for cerebellar ataxia in Scottish Terriers identified an associated region on the X chromosome, but a candidate causal variant has not been found and pedigree analysis was unable to confirm an X-linked mode of inheritance [140]. Some of the genes associated with autosomal recessive ataxia in dogs are known to cause autosomal dominant conditions in humans (**Table 1.1**).

Table 1.1. Genes implicated in canine inherited ataxia.

Gene	Gene function / role	Dog breed	Canine ataxia (inheritance[+])	Discovery methodology in the dog	Human ataxia* (inheritance[+])
GRM1	Glutamate signalling pathway	Coton de Tulear	Neonatal cerebellar ataxia (AR)	GWAS (12 cases, 12 controls) Candidate gene exon sequencing [141]	SCA44 (AD) [142] SCAR13 (AR) [143]
SPTBN2	Glutamate signalling pathway	Beagle	Cerebellar cortical degeneration (AR)	mRNA-seq candidate gene study (one case) [62]	SCA5 (AD) [144] SCAR14 (AR) [144]
SEL1L	ER-associated protein degradation pathway	Finnish Hound	Cerebellar cortical degeneration (AR)	GWAS (13 cases, 18 controls) Exon sequencing in region [122]	N/A
CAPN1	Protease, neuroprotective signalling pathways	Parson Russell Terrier	Spinocerebellar ataxia (AR)	GWAS (16 cases, 16 controls) Targeted NGS [63]	Spastic ataxia / spastic paraplegia 76 (SPG76) (AR) [82, 83]
RAB24	Autophagy	Old English Sheepdog and Gordon Setter	Cerebellar cortical degeneration (AR)	GWAS (14 cases, 40 controls) Targeted NGS Genotyping additional breeds [123]	N/A
KCNJ10	Potassium channel	Russell group terrier	Spinocerebellar ataxia / spinocerebellar ataxia with myokymia, seizures, or both (SAMS) (AR)	WGS (one case, 81 controls) [66]	Epilepsy, ataxia, sensorineural deafness, tubulopathy (EAST) / seizures, sensorineural deafness, ataxia, mental retardation, and electrolyte imbalance (SeSAME) (AR) [145, 146]
		Smooth-Haired and Toy Fox Terrier	Hereditary ataxia (AR)	Genotyping previously identified variant [147]	
		Belgian Shepherd Dog (Malinois)	Spongy degeneration with cerebellar ataxia, subtype 1 (SDCA1) (AR)	Linkage analysis (six cases, 18 controls) Homozygosity mapping (six cases) WGS (three cases, 140 controls) [131]	

Gene	Gene function / role	Dog breed	Canine ataxia (inheritance[+])	Discovery methodology in the dog	Human ataxia* (inheritance[+])
ATG4D	Autophagy	Lagotto Romagnolo	Cerebellar ataxia / Neurodegenerative Vacuolar Storage Disease (AR)	Linkage analysis (two cases, four controls) Homozygosity mapping (three cases) WGS (One case, 118 controls) [139]	N/A
ITPR1	Calcium channel	Italian Spinone	Spinocerebellar ataxia (AR)	Microsatellite homozygosity mapping (six cases, six controls) Microsatellite linkage analysis (13 cases and 47 controls) Targeted NGS (two cases, three controls) [64]	SCA15/16 (AD) [148, 149] SCA29 (AD) [150] Gillespie syndrome (AD/AR) [151]
VLDLR	Cerebellum development	Eurasier	Cerebellar hypoplasia / Dandy-Walker-like malformation (DWLM) (AR)	GWAS (nine cases, 11 controls) Homozygosity mapping (nine cases) WGS (one case, 47 controls) [152]	Cerebellar hypoplasia (AR) [153]
SNX14	ER-associated lipid metabolism, intracellular mitochondrial trafficking, (possible) autophagy	Hungarian Vizsla	Cerebellar cortical degeneration (AR)	WGS (one case, 13 controls) [65]	SCAR20 (AR) [154]
ATP1B2	Sodium - potassium pump	Belgian Shepherd Dog (Malinois)	Spongy degeneration with cerebellar ataxia, subtype 2 (SDCA2) (AR)	Linkage analysis (four cases, nine controls) Homozygosity mapping (four cases) WGS (one case) [130]	N/A
SCN8A	Voltage gated sodium channel	Alpine Dachsbracke	Spinocerebellar ataxia (AR)	Linkage analysis (two cases, five controls) Homozygosity mapping (four cases, eight controls) WGS (one case, 600 controls) [134]	N/A

Gene	Gene function / role	Dog breed	Canine ataxia (inheritance[+])	Discovery methodology in the dog	Human ataxia* (inheritance[+])
SLC12A6	Potassium and chloride ion cotransporter	Belgian Shepherd Dog (Malinois)	Spinocerebellar ataxia (AR)	Whole exome sequencing (two cases, two unaffected parents) [132]	N/A
SELENOP	Selenium storage and transportation, (possible) synaptic signalling	Belgian Shepherd Dog (Malinois)	CNS atrophy with cerebellar ataxia (AR)	Linkage analysis (four cases, six controls) Homozygosity mapping (four cases) WGS (one case, 735 controls) [133]	N/A
HACE1	Protein degradation / cerebellum development	Norwegian elkhound (Black)	Cerebellar ataxia (AR)	WGS (three cases, nine related controls) [135]	Intellectual disability, spasticity and abnormal gait (AR) [155]

The biological function of the encoded proteins, and the canine and human ataxias they are reported to cause, and the methodology used in the genetic study for the dog. * Neurological diseases where ataxia is a predominant sign. Other inherited conditions may have been associated with the gene listed. + Modes of inheritance: autosomal recessive (AR), autosomal dominant (AD). GWAS: Genome-wide association study. WGS: whole genome sequencing. ER: Endoplasmic reticulum. mRNA-seq: mRNA sequencing. NGS: Next generation sequencing. SCA: Spinocerebellar ataxia (autosomal dominant). SCAR: Spinocerebellar ataxia (autosomal recessive).

1.6.3.3.1 Ion transport

KCNJ10 encodes the inwardly rectifying potassium channel Kir4.1; which is expressed in the central nervous system, particularly in astrocytes which have an important function in the clearance of potassium ions from the extracellular space; and has been implicated in EAST (Epilepsy, Ataxia, Sensorineural deafness, Tubulopathy) or SeSAME (Seizures, Sensorineural deafness, Ataxia, Mental retardation, and Electrolyte imbalance) syndrome in humans [145, 146, 156]. A missense variant within the *KCNJ10* gene is associated with SAMs in Russell group terriers and both Smooth-Haired and Toy Fox Terriers [66, 147]. Tenterfield Terriers heterozygous for the variant and unaffected by ataxia have also been identified, potentially indicating that this variant could be involved in ataxia in yet another breed [147].

ITPR1 encodes inositol 1,4,5-trisphosphate receptor type 1, which is a calcium channel that regulates intracellular calcium signalling [157, 158]. A GAA repeat expansion within intron 35 of *ITPR1* is associated with spinocerebellar ataxia in the Italian Spinone [64]. The size of the expansion in affected dogs is 300 to 650 GAA repeats, whereas the wild-type alleles have a range of seven to 22 repeats. A similar GAA intronic repeat expansion, in Frataxin, causes Friedreich Ataxia in humans; and intronic repeats cause at least four SCAs [91, 98]. *ITPR1* has been associated with SCA15 (also called SCA16), SCA29, and Gillespie syndrome in humans [148-151].

There are at least four genetically distinct, autosomal recessive, forms of inherited ataxia in the Belgian Shepherd Dog (Malinois) breed [130-133]. Three of the variants putatively found to cause ataxia in the Belgian Shepherd are in genes involved in ion transport [130-132]. A variant within the second exon of potassium channel gene *KCNJ10* is associated with spongy degeneration with cerebellar ataxia, subtype 1 (SDCA1) [131]. This represents a second, distinct, variant within *KCNJ10* that is associated with canine ataxia in addition to that seen in Russell group terriers and Smooth-Haired and Toy Fox Terriers. A 227 bp short interspersed nuclear element (SINE) insertion within the second exon of the *ATP1B2* gene is associated with spongy degeneration with cerebellar ataxia, subtype 2 (SDCA2) [130]. *ATP1B2* encodes the β_2 subunit of Na^+/K^+-ATPase [130, 159, 160]. The Na^+/K^+-ATPase, and the β_2 subunit, have an important role in maintaining neuronal transmembrane ionic gradients through transport of two potassium ions across the cellular membrane, into the cell, in exchange for three sodium ions [161]. An insertion/deletion variant in a third gene,

SLC12A6, has been putatively found to cause another form of spinocerebellar ataxia in the Belgian Shepherd Dog breed [132]. The *SLC12A6* gene encodes K^+-CL^- cotransporter 3 (KCC3) which is highly expressed in the brain (including the cerebellum), and variants within the gene are implicated in agenesis of the corpus callosum with peripheral neuronopathy (ACCPN) [162, 163]. ACCPN is a neurodegenerative disease, but has a different phenotype to canine spinocerebellar ataxia [132, 163].

SCN8A encodes sodium voltage-gated channel α subunit 8, which is the alpha subunit of $Na_v1.6$, a voltage-gated sodium channel with a role in the regulation of action potential formation, and nerve conduction velocity, in the central nervous system [164, 165]. A missense variant in the final exon of *SCN8A* is associated with spinocerebellar ataxia in Alpine Dachsbracke dogs [134]. Mutations within *SCN8A* have been implicated in a range of human neurological conditions including paroxysmal dyskinesia, a spectrum of epilepsy phenotypes, and cognitive impairment with or without cerebellar ataxia [107, 166, 167].

1.6.3.3.2 The signalling pathway of the neurotransmitter glutamate

Neonatal cerebellar ataxia in Coton de Tulear dogs is putatively caused by a 62-bp insertion in exon 8 of *GRM1* [141]. *GRM1* encodes the metabotropic glutamate receptor mGluR1 which is highly expressed in Purkinje cells, and has been implicated in the human spinocerebellar ataxias SCAR13 and SCA44 [142, 143]. The G-protein coupled mGluR1 is key to the postsynaptic response to glutamate and is important for cerebellar function [168].

A form of cerebellar cortical degeneration in Beagles is putatively caused by an 8 bp deletion in *SPTBN2*, which encodes β-III spectrin [62]. β-III spectrin stabilises the plasma membrane localisation and modulates the activity of the neuronal glutamate transporter EAAT4 and the metabotropic glutamate receptor mGluR1 [169, 170]. *SPTBN2* has been associated with the spinocerebellar ataxia SCA5/SCAR14 in humans [144].

1.6.3.3.3 Autophagy

The protein encoded by *RAB24* has a role in autophagy, which is the process through which long-lived proteins and organelles are transported to lysosomes and degraded [171]. A SNP in *RAB24* has been identified as the likely genetic cause of cerebellar cortical degeneration in Old English Sheepdogs and Gordon Setters [123]. The role of *RAB24* in autophagy includes clearance of autophagic vacuoles [171]. Defects in autophagy in mice

have been demonstrated to cause neurodegeneration which includes loss of Purkinje cells, and impaired autophagy has been seen in human SCAs [172-174].

The ATG4D (autophagy related 4D, cysteine peptidase) protein is also involved in autophagy, and loss of this gene in mice and zebrafish results in cerebellar neurodegeneration [139, 175]. A non-synonymous variant within *ATG4D* is associated with cerebellar ataxia in the Lagotto Romagnolo breed [139]. Immunohistochemical analysis of affected canine cerebellum has demonstrated that the autophagy pathway was disrupted [139].

1.6.3.3.4 Protein degradation

SEL1L has a role in the pathway for transportation of misfolded protein from the endoplasmic reticulum (ER) to the cytosol (a process called dislocation) for degradation by the proteosome [176]. A non-synonymous SNP in exon 19 of *SEL1L*, resulting in a serine to proline substitution in the encoded protein, putatively causes cerebellar cortical degeneration in Finnish Hounds [122].

Genetic investigations of ataxia in the black Norwegian Elkhound found that a single base deletion in exon 11 of *HACE1* causes the disease [135]. *HACE1* (HECT Domain And Ankyrin Repeat Containing E3 Ubiquitin Protein Ligase 1) encodes an E3 ubiquitin ligase strongly expressed in the brain that targets specific proteins for degradation by the 26S proteasome [177]. HACE1 has a role in the regulation of RAC1, a GTPase involved in cerebellar development, and has been implicated in an autosomal recessive human neurodevelopmental disease with clinical signs that include cognitive impairment and hypotonic or ataxic movement [155, 178, 179].

1.6.3.3.5 Cerebellum development

A single base deletion in the *VLDLR* gene putatively causes Dandy-Walker-like malformation in the Eurasier breed [152]. *VLDLR* encodes a receptor for Reelin, which has important roles in the development of the brain, and notably the cerebellum; mutations within *VLDLR* have been implicated in cerebellar hypoplasia in humans [153, 180].

1.6.3.3.6 Genes with other roles and functions

Spinocerebellar ataxia in Russell terrier group dogs can be associated with a missense mutation in *CAPN1*, which encodes a subunit of calcium dependent cysteine protease

calpain-1 [63]. Calpain-1 is activated by calcium influx into the cell, and its protease activity triggers various signalling pathways including those involved in neuronal survival and synaptic plasticity [181]. Mutations in *CAPN1* had not previously been associated with ataxia in any species, however, since the *CAPN1* variant for cerebellar ataxia in the Russell terrier group dogs was identified at least 43 causative variants within this gene have been associated with a form of spastic ataxia in humans [82, 83].

A SNP at the splice donor site of *SNX14* exon 26, that alters splicing and extends the exon by 275 bp, putatively causes cerebellar cortical degeneration in the Hungarian Vizsla breed [65]. The protein encoded by *SNX14*, sorting nexin 14, is associated with a form of cerebellar ataxia in humans (SCAR20) [154]. *SNX14* is highly expressed in the mouse brain, particularly during neuron development, and knockdown of its expression impairs neuronal excitability and synaptic function in mouse cortical neurons [182]. *SNX14* has a role in lipid homeostasis between ER, lysosomes, and ER-derived lipid droplets [183]. *SNX14* has been implicated in the regulation of mitochondrial transport in axons through management of microtubule organization via an interaction with spastin [184]. Mutations within *SNX14* have been associated with lysosome-autophagosome dysfunction and slower autophagosome clearance, but other work has shown normal autophagosome-lysosome function in SNX14-deficient cells [183, 185].

The fourth genetically distinct form of ataxia in the Belgian Shepherd, CNS atrophy with cerebellar ataxia, is putatively caused by the deletion of the entire protein-coding sequence of the *SELENOP* gene [133]. *SELENOP* encodes selenoprotein P, which has a key role in the storage and transportation of selenium. Selenium is important for normal brain function, and selenoprotein P itself may have a role in synaptic signalling [186].

1.6.4 Canine paroxysmal dyskinesia

1.6.4.1 Clinical characteristics and heterogeneity

The classification of canine PxDs can be difficult [187, 188]. They are phenotypically heterogeneous but also often share characteristics, and most canine PxDs have similarities to human PNKD, limiting the usefulness of the human PxD classification system in dogs [187, 188]. The canine PxDs could be classified by aetiology; there are at least four canine PxDs with a putative genetic cause, and there are examples of PxDs influenced by diet or that are reported to be drug-induced [189-196]. There are also canine PxDs which are considered likely to be inherited, but for which genetic causes have

not yet been identified [120, 121]. In this section canine PxDs for which a putative genetic cause has been identified, or for which a genetic study has been reported, are discussed.

1.6.4.1.1 Episodic falling – Cavalier King Charles Spaniel

Episodic falling in the Cavalier King Charles Spaniel is characterised by attacks that are prompted by exercise, excitement, or stress [189, 197]. The disorder has been categorised as a PNKD [187]. Age of onset for episodic falling in the Cavalier King Charles Spaniel is variable, ranging from 3 months to 4 years [197]. Episodic falling attacks are brief (lasting up to 5 minutes) but variable in length and are typified by dystonia [189, 197]. Progressive hypertonicity (stiffening) of muscles generally causes the dog to assume an unusual posture with a lowered head, arched spine, and stiff limbs. The episodes typically result in the dog collapsing, but consciousness is maintained throughout the episode.

1.6.4.1.2 Canine epileptoid cramping syndrome (CECS) – Border Terrier

CECS, or Spikes disease, is a heterogeneous paroxysmal dyskinesia. In one study, episodes were reported to not be evoked by sudden movements and mostly have no obvious trigger, whereas another observed that episodes are triggered by sudden movements in most cases [195, 198]. Both studies reported that episodes can be induced by stress or anxiety. CECS has been classified as a PNKD, although based on some of the above-mentioned precipitating events it is possible it could also be labelled as a PKD in some individuals by certain criteria [104, 187, 195]. There is evidence that CECS is a gluten-sensitive disorder caused by gluten in the diet, and should therefore be referred to as paroxysmal gluten-sensitive dyskinesia (PGSD) [194, 195]. Border terriers affected by CECS or PGSD have attacks which have a broad range of durations, lasting from 30 seconds to 2.5 hours [195, 198]. Episodes can occur in clusters of up to three in a day, but with long periods free from attacks in between (multiple weeks or months). Age of onset is also highly variable, with the first attack occurring between two months and eight years [195, 198]. Attacks typically affect all limbs, head, and neck; and are typified by walking difficulties, tremors, and dystonia. Affected dogs can also show signs of gastrointestinal disease [195, 198].

1.6.4.1.3 Paroxysmal dyskinesia - Soft-coated Wheaten Terriers

Episodes of PxD in the Soft-coated Wheaten Terrier breed can be precipitated by stress, excitement, or anxiety, but in most cases episodes have no clear trigger [191, 199]. The PxD in this breed is similar to human PNKD. Episode frequency is variable, from one a week to over 10 a day, and duration ranges from several minutes to over four hours. Median age of onset has been reported to be 2.25 or 2.5 years, but can range from 10 months to 12 years [191, 199]. Episodes are characterised by involuntary hyperkinetic irregular movements: rapid flexion and extension of pelvic limbs or dystonia, or both [199]. Dogs remain conscious during episodes and able to walk, suggesting that movement can be semi-purposeful.

1.6.4.1.4 Paroxysmal exercise-induced dyskinesia - Shetland Sheepdogs

A form of paroxysmal dyskinesia, categorised as PED, in the Shetland Sheepdog breed has been described in four cases and is reported to have episodes triggered primarily by exercise or stress but which can occur while at rest or during activity [192]. The episodes, which are progressive and can last minutes or hours, affect all limbs and are typified by generalised ataxia, hypermetria and muscular hypertonia, and dystonia. Transient hypoglycemia (low blood glucose level), increased lactate excretion, and metabolic acidosis have been reported in affected dogs. The most severe episodes can render the dog unable to walk. Episode frequency can range from years free from episodes, one to two per year, or an episode every 2 to 3 days [192].

1.6.4.1.5 Paroxysmal dyskinesia - Markiesje

The Markiesje breed is numerically small and currently recognised only by the Dutch Kennel Club. Juvenile paroxysmal dyskinesia in this breed is severe and rapidly progressive and the dyskinesia is present whenever dogs are walked or stressed, with the result that it is almost constant while the dogs are moving but absent when dogs are at rest [193]. It is arguable that this PxD most closely fits within the PKD classification. The clinical signs include tetraparesis (weakness of all four the limbs), dystonia, and cramping; affected dogs fall over when they try to walk [193]. Age of onset in this breed is early, approximately 10 weeks of age. The severity of this PxD leads to euthanasia in most cases, and histopathology shows mild degeneration of the brain stem and spinal cord and in the skeletal muscle some mild denervation atrophy.

1.6.4.2 Diagnosis and treatment

Dogs affected by paroxysmal dyskinesia often do not have any substantial abnormalities that are clinically or neurologically observable between episodes [115]. MRI and CSF are used to determine if a PxD is primary (i.e. no evidence of pathogenic abnormalities, or inherited) or secondary (symptomatic with evidence of pathogenic abnormalities) [187]. Videos of episodes in combination with medical history can be used for diagnosis [187, 194]. Diagnosis is typically based on the overall phenomenology of the attacks, but differentiation from focal seizures can be challenging [120, 188, 198]. Intraictal (between seizure) electroencephalogram (EEG) monitoring would ideally be used to differentiate PxD from epileptic seizures, but this is not usually feasible. Serological testing for gluten-specific antibodies can be used to assist diagnosis of PGSD in Border Terriers [194, 195].

Episodic falling in Cavalier King Charles Spaniels can be effectively treated with clonazepam (a benzodiazepine) [115]. Clonazepam is a potent anticonvulsant which supresses neuronal activity by hyperpolarising neurons, binding to the GABA receptor and increasing its activity [200]. The treatment of CECS in Border Terriers with anticonvulsants such as phenobarbital and diazepam is typically ineffective [198]. It has been reported that CECS / PGSD can be responsive to a gluten-free diet [194, 195]. Treatment with the carbonic anhydrase inhibitor acetazolamide has been shown to be an effective treatment for some cases of PxD in the Soft-coated Wheaten Terrier breed; benzodiazepines and other antiepileptic drugs can be beneficial but to a lesser extent [199]. Treatment of PED in Shetland Sheepdogs using the anticonvulsants phenobarbital, levetiracetam, or diazepam has been reported to be ineffective; however dietary changes, stress-reduction, and treatment with acetazolamide or zonisamide (reported for a single case) may be beneficial. PxD in the Markiesje does not respond to treatment [193].

1.6.4.3 Genetics

At the time of writing, four different genes, with diverse biological functions, have been associated with canine PxD (**Table 1.2**) [189-193]. The studies that implicated these genes in canine PxD have used either a GWAS or WGS approach. Although some of these genes have been identified as causing a range of phenotypes in humans or other species, some of which are neurological, they have not yet been reported to be associated with human conditions for which paroxysmal dyskinesia is the predominant phenotype. Also discussed in this section is the GWAS of PxD in the Border Terrier that was unable to identify any

regions of the genome associated with the disease which is postulated to have a complex mode of inheritance [121].

Table 1.2. Genes associated with canine paroxysmal dyskinesia.

Gene	Gene function / role	Dog breed	Canine PxD (inheritance*)	Discovery methodology
BCAN	Perineuronal nets: cell adhesion, cell motility, axonal guidance, synapse plasticity	Cavalier King Charles Spaniel	Episodic falling (AR)	GWAS and targeted resequencing [189, 190]
PIGN	GPI biosynthesis	Soft-Coated Wheaten Terriers	Paroxysmal Dyskinesia (AR)	WGS [191]
PCK2	Gluconeogenesis, regulation of mitochondrial respiration	Shetland Sheepdogs	Paroxysmal exercise-induced dyskinesia (AD)	WGS [192]
SOD1	Copper / zinc superoxide dismutase, regulation of superoxide radical levels	Markiesje	Juvenile paroxysmal dyskinesia (AR)	GWAS and exon sequencing [193]

The biological function of the encoded proteins, and the methodology used in the genetic studies that implicated them. *Modes of inheritance: autosomal recessive (AR), autosomal dominant (AD). GWAS: Genome-wide association study. WGS: whole genome sequencing. Glycosylphosphatidylinositol (GPI).

Episodic falling in the Cavalier King Charles Spaniel is inherited in an autosomal-recessive manner [189]. A GWAS consisting of 31 affected dogs and 38 unaffected controls genotyped using the Illumina Canine HD bead chip, and targeted resequencing of the region identified on chromosome 7, found a 15,724 bp deletion that included the first three exons of *BCAN* [189]. A different group also independently identified the same deletion as associated with episodic falling, also using a GWAS approach [190]. The *BCAN* gene had not previously been associated with paroxysmal dyskinesia in any species. These studies demonstrate the utility of GWAS for detecting candidate disease-causing variants in genes which have not previously been associated with a disease phenotype and would not be identified through a candidate gene sequencing approach. *BCAN* encodes brevican which is a proteoglycan highly expressed only in the brain and which localises at perineuronal nets, extracellular matrix structures that have a role in cell adhesion and motility, axonal guidance, and the plasticity of synapses [201, 202].

The mode of inheritance of PxD in the Border Terrier has not yet been established, although an increased prevalence in some specific lines has been observed; most cases have unaffected parents; and approximately half have at least one affected sibling [121]. PxD in the Border Terrier is most likely to have a complex mode of inheritance [121]. A GWAS including 110 Border Terriers affected by PxD, and 120 unaffected by the disease, was unable to identify any regions of the genome significantly associated with PxD [121]. The dogs included in the GWAS were from Finland, The Netherlands, and Germany. For most dogs, the only phenotypic information available was owner-reported, and the affected dogs showed a wide range of clinical features. Affected dogs were not tested for gluten-sensitivity. PxD in this breed is heterogeneous, and epilepsy, which can be difficult to differentiate from PxD in some cases, is known to be prevalent within the breed [194, 195, 203]. Multiple genetically distinct disorders, each of which may be polygenic or have contributing environmental factors, may therefore have been included within the case definition.

PxD is inherited in an autosomal recessive manner in Soft-Coated Wheaten Terriers [191]. Sequencing whole genomes of two affected Soft-Coated Wheaten Terriers and filtering the identified variants for those predicted to have a high effect and that segregated with disease using 100 WGS of unaffected dogs, identified a nonsynonymous SNP in the *PIGN* gene. Genotyping additional dogs demonstrated that the mutation segregated with the disease. The enzyme encoded by *PIGN* has a role in glycosylphosphatidylinositol (GPI)

biosynthesis [204]. GPI is a complex glycophospholipid that is attached to proteins and anchors them to the cell surface. *PIGN* is one of over 30 genes involved in GPI biosynthesis, and mutations within these genes cause a wide range of phenotypes [205]. Mutations in *PIGN* have been implicated in autosomal recessive disorders with a wide spectrum of epilepsy phenotypes and movement disorders [206].

A whole genome sequencing study of two Shetland Sheepdog PxD cases identified 10 heterozygous variants by filtering for protein-changing variants heterozygous or homozygous in both cases but absent from the genomes of control individuals (648 dogs of genetically diverse breeds and eight wolves) [192]. Genotyping the variant within the putative candidate gene *PCK2* in all four affected dogs and unaffected dogs of the same breed and a range of genetically diverse breeds found that the variant segregated with disease in an autosomal dominant manner. *PCK2* encodes the mitochondrial isoform of phosphoenolpyruvate carboxykinase, which has an important function in gluconeogenesis (the generation of glucose from noncarbohydrate substrates), and PCK2 has a role in the regulation of mitochondrial respiration [207]. In humans phosphoenolpyruvate carboxykinase deficiency has been suggested to cause hypoglycemia and liver impairment, and a mutation within *PCK2* has been associated with a form of glaucoma [208, 209]. There is some limited evidence of low glucose in PxD affected dogs with the *PCK2* variant but affected dogs were not reported to have been assessed for glaucoma [192].

Juvenile paroxysmal dyskinesia in Markiesje dogs has an autosomal recessive mode of inheritance [193]. A GWAS of five cases and 18 controls, followed by the sequencing of the exons of genes within the identified region of chromosome 31, identified a frameshift-causing insertion/deletion in the first exon of the *SOD1* gene that segregated with disease [193]. *SOD1* encodes copper / zinc superoxide dismutase which has a role in regulating the levels of superoxide radicals, a by-product of cellular respiration, by converting superoxide radicals to oxygen and hydrogen peroxide; and can also have functions in metabolic signalling and transcriptional regulation [210]. Missense mutations within the *SOD1* gene have been associated with canine degenerative myelopathy (DM), and it is thought that the mutations cause the disease through a gain of function that leads to aggregate formation [211-213]. Mutations within *SOD1* cause human amyotrophic lateral sclerosis (ALS), for which canine degenerative myelopathy is a model, also by causing protein misfolding and the formation of aggregates in a way thought to be

prion-like [214]. Juvenile paroxysmal dyskinesia in the Markiesje breed has a very different phenotype to DM, most likely because it is caused by a loss of function mutation that is effectively a complete knock-out of the *SOD1* gene in comparison to the gain of function missense mutations that are thought to cause DM and ALS [193].

1.7 Epilepsy

1.7.1 Epilepsy definition and characteristics in humans

Epilepsy is a disease of the brain typically defined as two or more unprovoked seizures which have occurred over 24 hours apart [215]. The International League Against Epilepsy (ILAE) definition of epilepsy also includes cases where only a single seizure has occurred but the probability of a second occurring within the next 10 years is similar to cases where two have occurred (60%).

A seizure is defined as the temporary manifestation of abnormal neuronal activity in the brain which is synchronous or excessive, and the process that leads to seizure generation is called ictogenesis [216, 217]. A seizure occurs when the normal balance between neuronal excitation and inhibition in the brain is skewed [217, 218]. Seizures are categorised by their onset, which can be focal, generalised, or unknown [216]. Focal seizures originate in neural networks located in one or more localised brain regions or in a single hemisphere, either in a distinct location or with a wide distribution [217, 219]. Generalised seizures originate and rapidly spread within neural networks in both hemispheres of the brain.

The physiological signs of seizures are extremely diverse [219]. Focal seizures can be categorised into those that impair awareness and those that do not [216]. Focal, generalised, and unknown onset seizures can all have motor or non-motor forms [216]. Generalised tonic-clonic motor seizures present as an initial stiffening (tonic) followed by jerking movements (clonic) of all limbs bilaterally and symmetrically and result in a loss of consciousness [218]. A seizure can have a focal onset and then spread to both hemispheres and cause a tonic-clonic seizure, previously called a secondarily generalised seizure but now referred to as focal to bilateral tonic-clonic [216]. Tonic or clonic signs can also occur separately as signs of generalised or focal motor seizures. Motor seizures with both onsets can also cause sudden and extremely brief movements of muscles, called myoclonic seizures; focal myoclonic seizures typically affect only individual muscles

or muscle groups [216, 218]. Generalised onset atonic motor seizures result in a loss of muscle tone, often characterised by the head dropping [218, 219]. Atonic seizures can also have a focal onset [216]. Non-motor generalised seizures are referred to as absence seizures and can present as gazing with no response to external auditory stimuli [216, 218]. Non-motor focal seizures can have autonomic, behavioural, cognitive, emotional, or sensory signs [216].

After the type of seizure has been established, diagnosis can determine the type of epilepsy that an individual has: focal, generalised, combined generalised and focal (where an individual has both focal and generalised seizures), and unknown (when the available information is insufficient) [220]. Once epilepsy type is established, and if sufficient diagnostic information is available, an epilepsy syndrome can be determined based on features that tend to occur together such as seizure-types, age of onset, triggers, and comorbidities.

1.7.2 Epilepsy genetics in humans

The aetiology of epilepsy can be genetic, immune, infectious, metabolic, or structural (abnormalities visible on neuroimaging) [220]. The structural and metabolic aetiologies can both have underlying genetic causes. A genetic epilepsy is one in which a causal genetic mutation is either known or presumed (based on a family history or on clinical research into a particular epilepsy syndrome) [220]. There has been a lot of progress in recent years; however, for many cases of epilepsy with a presumed genetic cause the underlying variant or gene has not been identified [220-222]. The genetic epilepsies include numerous rare monogenic diseases, and there are also more common epilepsies that are thought to have an oligogenic (a relatively low number of variants conferring disease risk with a modest-to-high effect size) or polygenic (hundreds or even thousands of low effect risk variants) modes of complex inheritance with possible environmental and epigenetic factors also contributing to disease risk [217, 221, 222].

Most of the epilepsy-associated genes have been identified through discovery of rare variants for monogenic disorders [217, 222]. NGS approaches, including whole-exome sequencing and WGS, have been effective in the identification of candidate epilepsy-causing variants, although these have mostly been limited to variants within the protein-coding regions [217, 222]. WGS is used in diagnostic genetic testing of epilepsy, and can lead to the discovery of novel genes [217, 221, 222]. There are 227 epilepsy-associated

gene and loci entries on OMIM as of 2023 [84] (https://omim.org/, accessed 24/03/2023). The Genomics England PanelApp (a curated collection of publicly available gene panels used for prioritising variants in genetic testing) genetic epilepsy syndromes (version: 3.0) gene panel includes 489 genes, short tandem repeats (STRs), or copy number variants (CNVs) that have a diagnostic level of supporting evidence for disease association [223, 224] (https://panelapp.genomicsengland.co.uk, accessed 23/01/2023).

Considerable success has been had with the identification of genes associated with the epileptic encephalopathies, which are (individually) rare severe early onset progressive conditions characterised by cognitive and behavioural impairments caused by the epileptic activity [222, 225, 226]. A notable example is the sodium channel α1 subunit gene *SCN1A*, mutations within which cause the majority of the cases of Dravet syndrome, a severe early-childhood-onset epileptic encephalopathy [227, 228]. The variants within *SCN1A* that cause Dravet syndrome are mostly *de novo* (new variants present in the offspring but not in the parents), and many of the epileptic encephalopathies for which a genetic cause has been identified can be caused by *de novo* mutations [222, 228-230]. Along with *SCN1A*, most of the genes initially implicated in epileptic encephalopathies encoded ion channels, leading to them being referred to as channelopathies [225]. However, genome-wide screening has led to the discovery of epileptic encephalopathy associated genes that encode proteins with a diverse range of functions **Figure 1.5**.

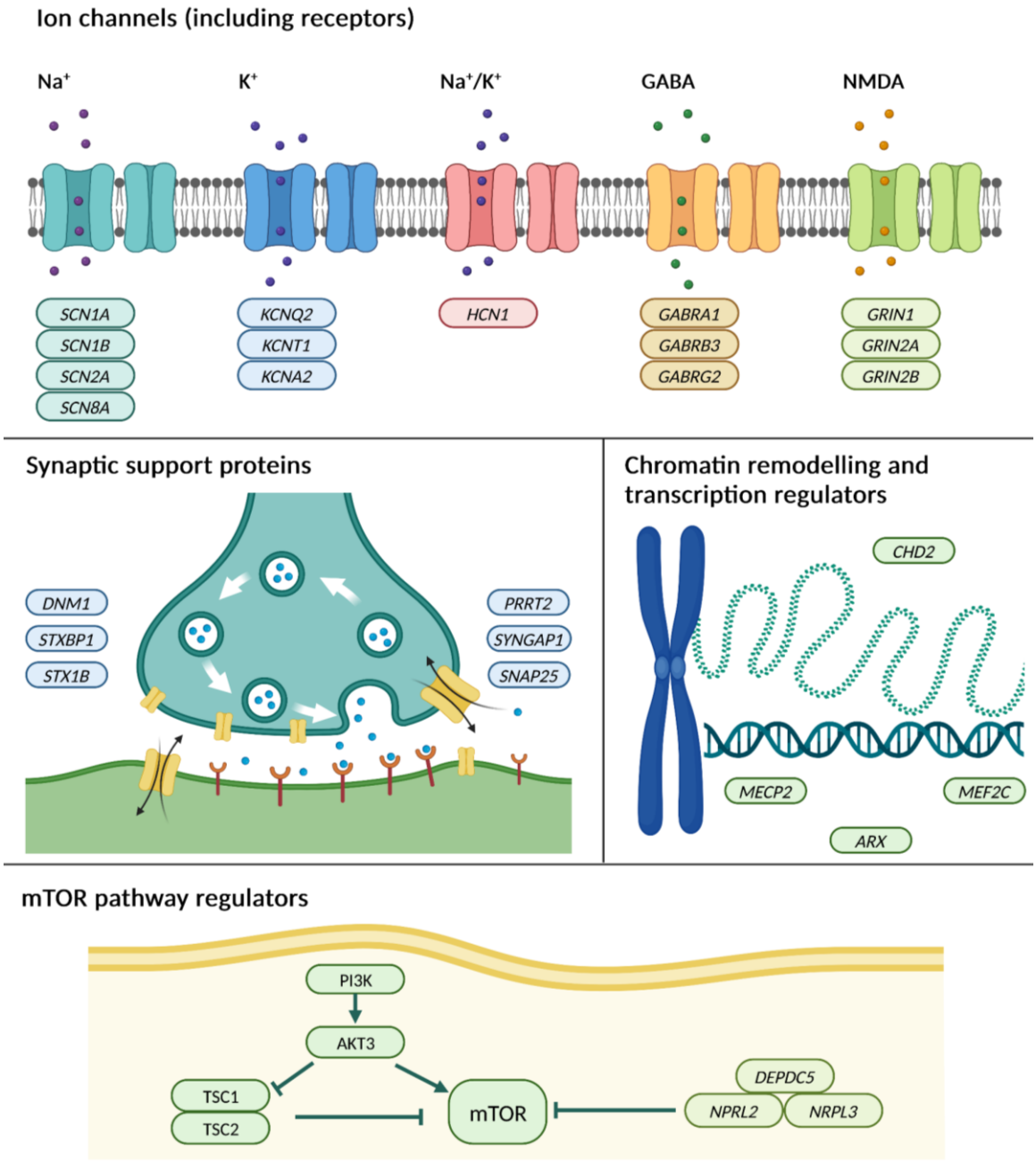


Figure 1.5. Simplified illustrations demonstrating some examples of cellular functions, pathways, and genes implicated in epilepsy. Figure adapted from Ellis et al (2020) [221]. Created with BioRender.com.

Success has been more modest in the elucidation of the underlying genetic causes of the common epilepsies, which comprise the genetic generalised epilepsies (previously called idiopathic generalised epilepsies) and focal epilepsies [220, 222, 231]. Although rare monogenic causes have been identified for some, most of the common epilepsies are thought to be caused by an oligogenic or polygenic combination of rare and common variants [222, 232]. Ultra-rare variants in genes known to cause rare dominant epilepsy disorders have been shown to contribute to both genetic generalised epilepsy and focal epilepsy [233, 234]. The investigation of the common variants involved in epilepsy has been hindered by limited study power, the clinical heterogeneity of epilepsy makes generating the necessary large case sets challenging. However, GWAS meta-analyses combining epilepsy cohorts have produced some positive results; a more recent genome-wide multi-ethnic mega-analysis (analysis of pooled raw data from multiple studies) that comprised 15,212 epilepsy cases and 29,677 controls identified 16 loci associated with common epilepsies (11 of which were not previously known as epilepsy genes) [231, 235].

1.7.3 Canine idiopathic epilepsy

Canine epilepsy is a neurological condition of great concern to many dog breeders and owners, and it is also relatively common in many pedigree dog breeds. One study has estimated the prevalence of canine epilepsy to be 0.62% across all UK breeds and mixed-breeds; some sub-populations of individual breeds have an estimated prevalence of up to 18.3% [203, 236]. Epilepsy can have a negative impact on quality of life, both for the affected dog and for the owner [237, 238]. Epilepsy is the most common diagnosis for dogs with new-onset seizures [239].

IE in dogs is defined by the International Veterinary Epilepsy Task Force (IVETF) as epilepsy with a known or suspected genetic cause or as epilepsy for which the cause is unknown; and for which a structural cerebral pathology has not been identified and is not suspected [240]. IE has been found to be the most common diagnosis for dogs with epilepsy, ahead of structural epilepsy [87, 239, 241, 242]. It has been estimated that just over half of dogs undergoing MRI for epileptic seizures have IE, and IE is the most common diagnosis when the dogs undergoing MRI are aged six months to six years [242].

Not all epilepsy with a genetic cause is idiopathic; structural epilepsy, which encompasses disorders characterised by a cerebral or intracranial pathology (including developmental, degenerative, and neoplastic diseases) that causes epileptic seizures, can have a genetic

cause [240]. Canine Lafora disease is a progressive neurodegenerative disease that causes myoclonic epilepsy and other clinical signs such as ataxia, vision loss, and cognitive decline [243, 244]. In canine Lafora disease a repeat expansion within the NHL repeat containing E3 ubiquitin protein ligase 1 (*NHLRC1*) gene results in the accumulation of Lafora bodies (insoluble glycogen) in neurons and causes neurodegeneration [243, 245]. Lafora disease is therefore categorised by the IVETF as a structural epilepsy because of these structural changes in the brain [240]. Lafora disease putatively caused by repeat expansions within *NHLRC1* has been identified in multiple dog breeds [243, 244, 246, 247]. Another genetic canine epilepsy that fits within the IVETF definition of structural epilepsy is the recently reported severe early onset neurodegenerative disease with mitochondrial respiratory deficiency and epileptic encephalopathy in the Parson Russell Terrier breed [248]. Affected dogs present with severe seizures that rapidly worsen until the dog dies as a result of the disease or euthanasia. The histopathological findings, which show severe neuronal degeneration and necrosis throughout the brain, accumulation of rounded and swollen mitochondria in neurons, and a build-up of amyloid-β, conform to the IVETF definition of structural epilepsy [248]. An in-frame deletion in the nuclear-encoded pitrilysin metallopeptidase 1 (*PITRM1*), a mitochondrial protease, has been associated with this disorder [248].

In the following sections the focus will be on the clinical characteristics, epidemiology, diagnosis, treatment, and genetics of canine IE.

1.7.3.1 Clinical description, characteristics, and heterogeneity

There is considerable variation between the clinical characteristics of IE in different breeds of dog [249]. The predominant type of seizure observed can be different between breeds; seizure type can be focal, generalised, or focal onset progressing to generalised [250]. The manifestation of the focal or generalised seizures can also vary between dogs, and the most commonly reported seizure phenotype can depend on the breed of dog [249]. The average ages of onset, and survival time, have also been reported to vary between breeds. For some breeds different preictal (before seizure) and postictal (after seizure) signs have been observed. Neuro-behavioural comorbidities which develop at disease onset can be common in some breeds [251]. Below are summaries of the clinical characteristics of epilepsy with a focus on the Belgian Shepherd, Border Collie, Irish Wolfhound, Italian Spinone, and Petit Basset Griffon Vendeen breeds; benign familial

juvenile epilepsy (BFJE) in the Lagotto Romagnolo; and Rhodesian Ridgeback generalised myoclonic epilepsy with photosensitivity.

1.7.3.1.1 Seizure type

The Goenendael and Tervueren varieties of the Belgian Shepherd breed most commonly suffer from focal seizures and over half of affected dogs are reported to have focal seizures that progress to generalised seizures [252, 253]. It has been reported that one quarter of affected Belgian Shepherd dogs have focal seizures that do not become generalised, and some experience seizures that are generalised from onset. Border Collies diagnosed with IE primarily suffer from generalised seizures, typically tonic-clonic; seizures can have a focal onset, and focal seizures which do not progress to secondary generalisation have also been reported [254, 255]. A retrospective study of Irish Wolfhounds in 120 related litters reported that most dogs had grand mal (now referred to as generalised tonic-clonic) seizures [236]. However, the authors did not state the proportion of dogs that had this type of seizure, or if any had seizures with a focal onset. Italian Spinoni with IE have also been reported to typically present with generalised tonic-clonic seizures, with focal onset and secondary generalisation reported in approximately half of affected dogs [251]. In the Petit Basset Griffon Vendeen breed most epilepsy-affected dogs suffer from focal seizures; dogs that experience focal seizures that progress to become generalised are slightly more common than those that only have focal seizures [256].

1.7.3.1.2 Seizure duration

Seizure duration can vary, and seizures are reported to last between 30 seconds and 2.5 minutes in Belgian Shepherd dogs and between one to three minutes in the Petit Basset Griffon Vendeen breed [252, 256]. The frequency of seizures is also highly variable, some Border Collies can have multiple seizure days each week while others are reported to have a single seizure day per year, and seizure frequency in the Italian Spinone breed varies from as many as 11 seizure days each month down to fewer than one per month [251, 254].

1.7.3.1.3 Cluster seizures and status epilepticus

Cluster seizures (multiple seizures occurring unusually close together, more than one seizure in a day, with full recovery between) and status epilepticus (an unusually

prolonged seizure, or multiple seizures without full recovery in between) are characteristic of epilepsy which is more severe; survival time is reportedly reduced in Border Collies that have had status epilepticus and Italian Spinone dogs that have suffered from cluster seizures, and euthanasia is a more likely outcome when cluster seizures are more frequent [240, 251, 254, 257]. In a study of 407 dogs of various breeds affected by idiopathic epilepsy, cluster seizures occurred in 41% of dogs [257]. Cluster seizures have been reported to occur in a third of affected Belgian Shepherd dogs [258]. A high proportion (59 - 94%) of Border Collies have been reported to have had cluster seizures, and this can be associated with a poorer quality of life [254, 255]. Border Collies also have an increased probability of cluster seizures when compared to the Labrador Retriever breed as a baseline, giving further evidence for the more severe phenotype in this breed [259]. In the breed-wide Italian Spinone survey cluster seizures had occurred in 73% of affected dogs [251]. Status epilepticus appears to be rare in IE-affected Belgian Shepherd dogs, whereas it has been reported to occur in 33-53% of affected Border Collies and 21% of affected dogs in the Italian Spinone breed [251, 254, 255, 258].

1.7.3.1.4 Seizure triggers

Seizure-precipitating events have been reported in some breeds, and studies which are not breed specific have found that owners frequently claim to be able to identify seizure triggers and predict when a seizure is likely to occur, although these studies can be affected by recall bias [260-264]. Owners of 22% of Belgian Shepherd dogs report that seizures can be provoked by stress or hyperactivity, but seizures also occur when the dog is at rest [252]. In Border Collies most seizures occur whilst sleeping or at rest and seizures can have no apparent precipitating event, but owners also report stress as a potential trigger [254, 255].

1.7.3.1.5 Preictal signs

Preictal signs, which occur before a seizure and can last multiple days, are relatively rare (4%) in Belgian Shepherd dogs but when they occur they are characterised by dogs behaving in an unusually dispirited way [252]. Over half of Border Collie cases have been observed to have preictal signs, the most commonly reported of which were the seeking of the owners attention and restlessness [254, 255]. Less commonly reported preictal symptoms in the Border Collie include salivation, aggression and vomiting. The preictal

signs in this breed are typically observed less than 30 minutes before seizure onset. In some Border Collie cases, signs, such as reduced responsiveness to commands, or a single forelimb becoming lame, are observed between 24 and 48 hours before a seizure [254]. Preictal signs, including anxiety, seeking owner's attention, barking or whining, and aggressive behaviour, have been reported in 57% of Italian Spinoni [251].

1.7.3.1.6 Postictal signs

Many affected Border Collies take from six hours to multiple days to recover completely from seizures [254]. This postictal phase commonly manifests as restlessness. Other common postictal signs reported for the Border Collie breed include polyphagia (increased appetite), polydipsia (excessive thirst), temporary blindness, and over half of dogs become lethargic [254, 255]. A postictal phase, of variable length, has been reported to occur in all affected dogs in the Italian Spinone breed and signs include disorientation, ataxia, blindness, and sleeping deeply [251].

1.7.3.1.7 Age of onset

Age of IE onset varies widely between dogs and breeds. The mean age of onset in a study of Belgian Shepherd dogs was 3.3 years, and a similar mean age of onset was observed in the Italian Spinone breed (3.2 years) [251, 252]. In the Belgian Shepherd a range of six months to eight years was reported for age of onset, and a range of 11 months to six years was seen in the Italian Spinone [251, 252]. In the Border Collie age of onset has a median of 2.37 - 2.79 years, although age of onset is variable [254, 255]. Age of onset in a study of epilepsy in the Petit Basset Griffon Vendeen breed ranged from two months to 6.5 years (median two years) [256]. For the majority of Irish Wolfhounds the first seizure has been reported to occur by three years of age [236].

1.7.3.1.8 Survival time and epilepsy as a cause of death

Epilepsy can be associated with an earlier age of death [236, 254, 258]. In the Border Collie breed it has been reported that dogs with an age of onset of less than two years, or for which at least one episode of status epilepticus has occurred, have a significantly decreased survival time after diagnosis [254]. In the Italian Spinone breed the mortality rate due to causes related to epilepsy has been estimated to be 32% [251]. Epilepsy was the most common cause of death in a breed wide survey of Danish Petit Basset Griffon Vendeen dogs registered with the Danish Kennel Club between 1999 and 2008, most of

the dogs that died were euthanised due to poor seizure control [256]. Some breeds have been reported to have a tendency towards a more severe epilepsy and shorter survival time: a study of 136 dogs of various breeds with epilepsy onset at less than one year of age, 13 of which were Border Collies, found that Border Collies had a shorter survival time than other IE-affected dogs [265].

1.7.3.1.9 Lagotto Romagnolo benign familial juvenile epilepsy (BFJE)

BFJE occurs in young Lagotto Romagnolo dogs, has a mean age of onset of 6.3 weeks, and completely remits by 13 weeks (although some adult dogs have been observed with the disease) [266]. Affected dogs present with simple (where consciousness is not affected) or complex (where consciousness is impaired) focal seizures with tremors, stiffness, and ataxia. Some dogs can become ataxic between seizures, but this is no longer observed once the seizures remit [266]. It has been reported that BFJE-affected dogs exhibit behaviour resembling human attention deficit hyperactivity disorder (ADHD) in the form of increased excitability and inattention [267]. This ADHD-like behaviour continues after the seizures have remitted.

1.7.3.1.10 Rhodesian Ridgeback generalised myoclonic epilepsy

Generalised myoclonic epilepsy with photosensitivity has been reported in Rhodesian Ridgeback dogs [268]. In 35% of affected dogs the seizures are triggered by visual stimuli such as flashing light, and seizures are characterised by myoclonic jerks, progressing to generalised tonic-clonic seizures in 38% of dogs. In some dogs absence seizures (staring episodes) have been observed [269]. Age of onset is between 6 weeks and 1.5 years of age (mean six months), and the dogs that progress to generalized seizures do so within six months of disease onset [268]. Most dogs are reported to have myoclonic jerks daily, some dogs have up to 150 a day.

1.7.3.2 Epidemiology

Although some epidemiological studies have been carried out for canine idiopathic epilepsy, they have mostly been limited to retrospective, cross-sectional, and observational studies, and as such they are unable to determine cause and effect in relation to epilepsy and its risk factors. Currently ongoing longitudinal studies hope to overcome this limitation [270].

1.7.3.2.1 The prevalence of epilepsy

There is considerable variation in prevalence of epilepsy between breeds. In a population of dogs attending 92 primary veterinary clinics in the UK, part of the VetCompass project, a prevalence of 0.62% was observed, with 539 dogs affected by epilepsy with no known cause [203]. **Figure 1.6** summarises the estimated prevalence of epilepsy of unknown cause in the most common dog breeds in this study. It was found that Border Terriers and German Shepherd Dogs had the highest odds ratios when compared to crossbreed dogs, with 2.7 and 1.9 times the odds, respectively, of developing epilepsy. However, the Border Terrier breed is known to have a high prevalence of PxD which can be challenging to differentiate from focal seizures [121, 198]. Another study analysing data from VetCompass, that included 455,553 dogs and looked at seizure prevalence during 2013, found a 0.82% prevalence of seizures [271]. This study included seizures of all causes, which most likely accounts for the increased prevalence in comparison to the previous study which looked only at epilepsy, although IE has been found to be the most common cause of seizures in the VetCompass data [87].

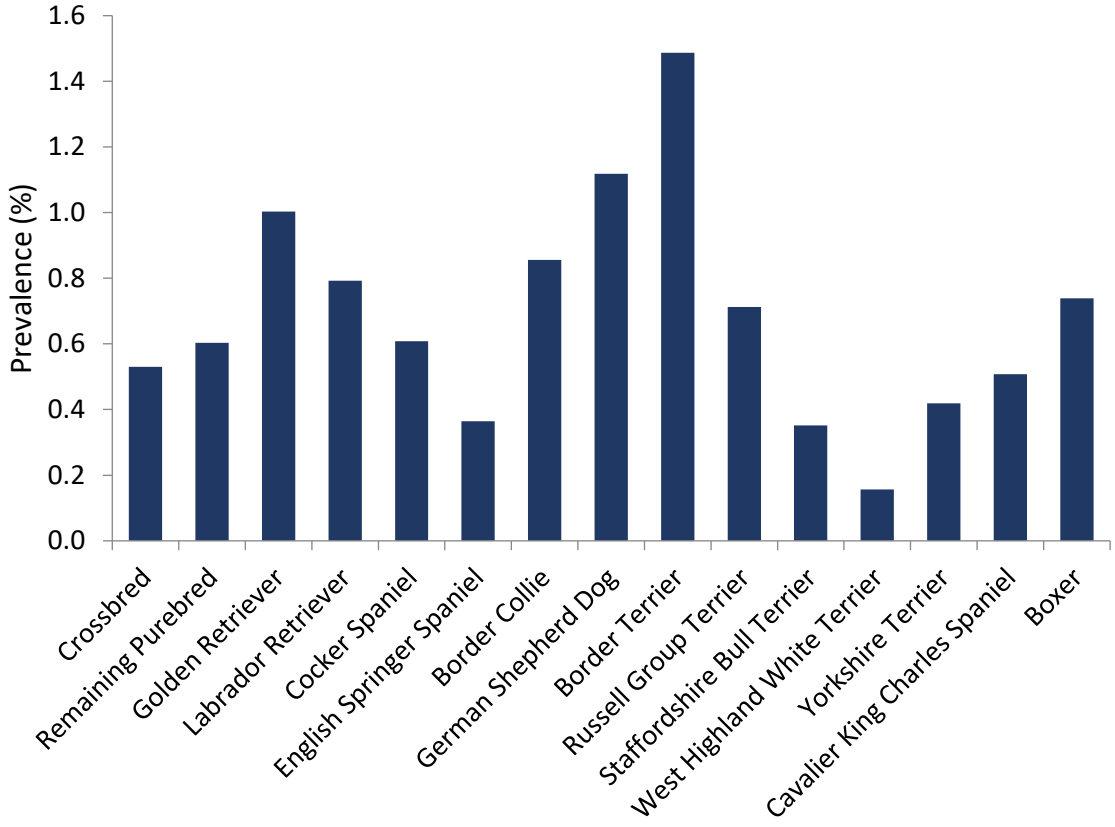


Figure 1.6. Estimated prevalence of epilepsy of unknown cause in some of the most common breeds of dog. Graph generated using data from Kearsley-Fleet et al (2013) [203]

The estimated prevalence found in breed specific studies for some breeds for which genetic studies have now been carried out are shown in **Table 1.3**. The studies were all questionnaire-based, and are therefore subject to bias; owners of affected dogs could be more likely to complete a questionnaire. For one study of epilepsy in the Belgian Shepherd breed, if all dogs for which a questionnaire was not returned were unaffected the prevalence would be 3.9% in this breed, as opposed to the estimated 9.5% [252]. In the same scenario for the Petit Basset Griffon Vendeen study the prevalence within the entire population would be 5.1% instead of 8.9%. The Belgian Shepherd dog study's authors suggest that there may have been owners of additional affected dogs who did not return questionnaires and young dogs within the population which had not yet had their first seizure, and so the prevalence may have been underestimated [252]. The authors of the study of Petit Basset Griffon Vendeen dogs also argued that some owners of dogs with IE could be dissuaded from responding to the questionnaire due to not wanting to be associated with owning dogs with epilepsy [256]. The increased prevalence observed in an extended family of Belgian Shepherd dogs provided additional evidence of an underlying genetic factor for the disease [253]. An increased occurrence of epilepsy was observed in some Petit Basset Griffon Vendeen litters, again supporting a genetic cause [256]. A study of related litters of Irish Wolfhounds found a high prevalence of IE, however, the study only included litters containing affected dogs or litters for which both parents, in different litters, had produced affected offspring [236].

Table 1.3. Prevalence estimates for epilepsy in breed specific studies.

Breed	**Study Population**	**Prevalence**
Belgian Shepherd dog	Danish Kennel Club registered 1995 - 2004 (Goenendael and Tervueren varieties)	9.5% [252]
	Extended family of 199 dogs	33% [253]
Irish Wolfhound	120 related litters (796 dogs)	18.3% [236]
Italian Spinone	UK Kennel Club registered 2000 - 2011	5.3% [251]
Petit Basset Griffon Vendeen	Danish Kennel Club registered 1999 - 2008	8.9% [256]

1.7.3.2.2 Sex bias and sex as a risk factor for canine epilepsy

It has been suggested that the risk of being diagnosed with IE varies between sexes [272]. **Figure 1.7** summarises the representation of male and female dogs in three large, across-

breed, country-specific studies investigating the epidemiology of epilepsy. All three studies demonstrate an over representation of male dogs. In a UK study of 1,260 epileptic dogs an overrepresentation of male dogs (63%) was observed [273]. Kearsley-Fleet et al (2013) found that male dogs were more than 1.5 times more likely to have epilepsy [203]. In a study of 665,000 insured dogs in Sweden, male dogs were overrepresented in the group of 5,013 dogs with at least one record of an insurance claim for epilepsy (1.4:1) [274]. Male dogs have also been observed to have a higher risk of seizures of any cause [271].

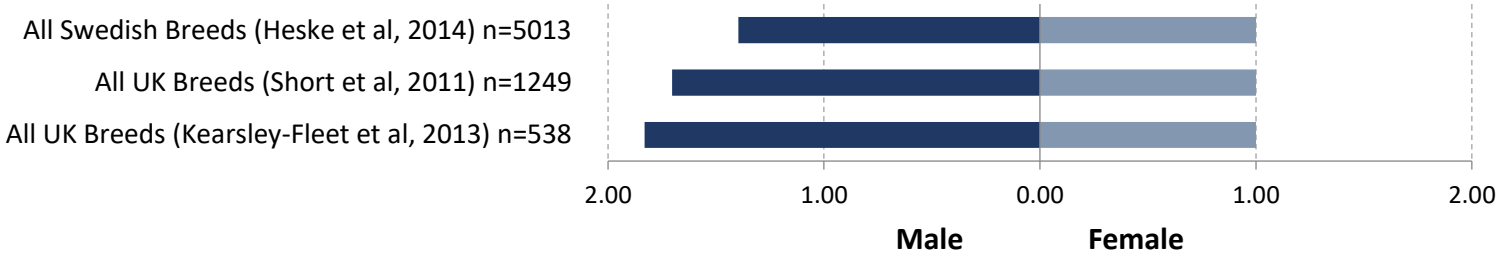


Figure 1.7. The ratio of male and female dogs of all breeds with epilepsy in three country-specific studies. The total number of affected dogs included is indicated for each study. The sex least represented was used as the reference (1) for each study.

It is possible that there is breed-specific variation in the sex distributions of dogs with epilepsy. **Figure 1.8** summarises some examples of the sex distributions in breed-specific epilepsy sample sets. Some studies demonstrate an overrepresentation of male dogs with IE in concordance with the across-breed studies outlined above, whereas some show no bias or a bias towards female dogs with IE. Examples of studies which have evidence of a sex bias towards male dogs include the studies for the Irish Wolfhound, Italian Spinone, and Petit Basset Griffon Vendeen breeds [236, 251, 256]. In studies for the Belgian Shepherd breed the distribution of affected dogs between the sexes varies by study, suggesting sampling variation; although the majority of the studies suggest that a dog's sex does not have an effect on its likelihood of developing IE in this breed it is possible there is some country-specific variation [258, 275, 276]. For example, a Danish study had an over-representation of female dogs affected by IE [252]. There is also variation in the proportion of affected dogs that are male or female between Border Collie studies [254, 255]. The definition and diagnosis of epilepsy is variable between studies, and study design also varies and can be subject to sample collection bias. Some of these studies have a very small sample size and are unlikely to be representative of the general population.

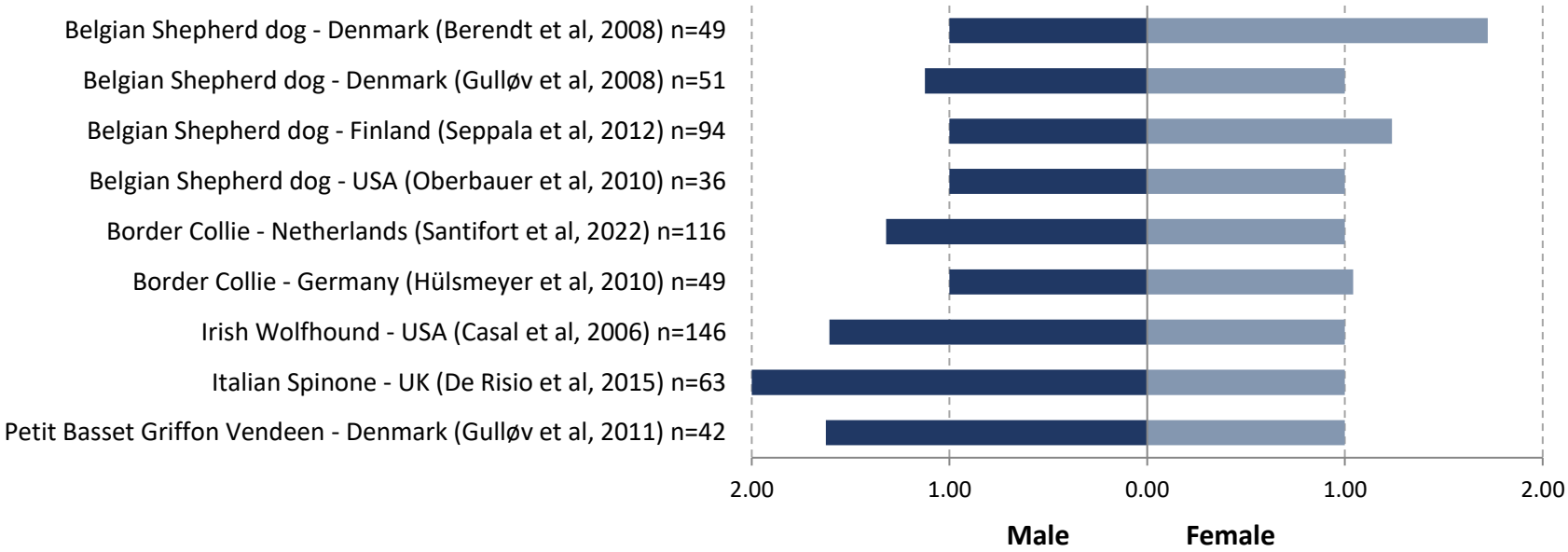


Figure 1.8. The ratio of male and female dogs with epilepsy in breed-specific studies. The main country in which dogs were located, and the total number of affected dogs included, is indicated for each study. The sex least represented was used as the reference (1) for each study.

Sex or reproductive status may also influence a dog's risk of being diagnosed with IE, or disease severity. A study that reviewed the medical records of 407 dogs with IE found that cluster seizures were more likely for unneutered dogs [257]. The study also found that intact females had cluster seizures more frequently than neutered females. However, a study of 384 dogs of various breeds and mixed breeds treated at a UK specialist canine epilepsy clinic found no association between sex and neuter status and cluster seizures [259]. A disproportionate number of neutered dogs with epilepsy has been described in comparison to non-epileptic dogs, but this is possibly a result of owners having their dogs neutered after epilepsy diagnosis in the belief that it will reduce seizure frequency [273]. A large cohort study with data including the age of disease onset and the date of neutering would be needed to investigate this possibility. The apparent association between neuter status and IE was not replicated in a later study [203].

The effect of sex or neuter status on disease progression or outcome may vary depending on breed. A study of Danish Belgian Shepherd dogs found that intact animals were more likely to be euthanised due to epilepsy [252]. Two studies of Border Collies affected by IE did not find any association between sex or reproductive status and seizure frequency, age of onset, or survival time [254, 255]. A study that investigated if there was an association between neutering and IE onset and severity in Labrador Retrievers and Border Collies diagnosed with IE, using primary-care clinical data including timing of neutering, observed that most cases were neutered before epilepsy onset and found that age of onset and occurrence of cluster seizures were unaffected by neuter status [277]. Another finding from this study was an increased survival time in dogs that were not neutered prior to epilepsy onset. In Irish Wolfhounds epilepsy was reported to have a later onset in male dogs [236]. Survival time was longer for female dogs and for neutered animals in a study in the Italian Spinone breed, however, this was confounded by more female dogs being neutered in comparison to males [251]. Sex and neuter status were not associated with survival in the Italian Spinone breed when analysed using a multivariable Cox regression model.

1.7.3.3 Diagnosis and treatment

Epilepsy across breeds is a collection of neurological conditions, with differing aetiologies and seizures as a key phenotype. This can make diagnosis challenging, requiring collection

of extensive clinical phenotype data [250]. The wide variety of possible aetiologies makes diagnosis of idiopathic epilepsy a process of elimination, and diagnosis is by exclusion of the alternatives. Diagnosis of epilepsy requires differentiation from phenotypically similar conditions such as PxD or exercise-induced collapse [250, 278]. Treatment options for canine idiopathic epilepsy have broadened recently, although phenobarbital and potassium bromide are still the most commonly used [200, 279]. The following section describes the diagnosis and treatment of canine idiopathic epilepsy in more detail.

1.7.3.4 The diagnosis of canine idiopathic epilepsy

The IVETF has recommended a three-tier system for diagnosis of canine IE [250]. The lowest level of confidence is tier I, and tier III is the highest. The requirement of each tier includes those of the tier below it with additional stringencies. The key criteria for each tier are summarised in **Table 1.4**. It has been estimated that approximately half of UK seizure-affected dogs are evaluated diagnostically to the level of IVETF Tier I or higher [87].

Table 1.4. The criteria for the three-tier system for the diagnosis of canine IE.

Tier	Criteria *
I	- At least two seizures have occurred 24 or more hours apart - Clinically unremarkable inter-ictal (between seizure) physical and neurological examinations - No abnormalities observed in blood and urine tests - Dog is aged between six months and six years at onset
II	- Unremarkable: o Fasting and post-prandial (after a meal) bile acids o MRI o CSF
III	- EEG results typical of IE

* Each tier's requirements include those of the previous, lower diagnostic confidence, tier. MRI: Magnetic resonance imaging. CSF: Cerebrospinal fluid. EEG: Electroencephalogram. IE: Idiopathic epilepsy.

Diagnosis can be made challenging by the reluctance of owners to proceed with tests, possibly due to procedures such as MRI being both expensive and invasive. One questionnaire-based study found that only 59.5% of Petit Basset Griffon Vendeen dogs with IE were diagnosed with the disease by a veterinarian before taking part in the study [256]. Over a quarter of owners had not informed their veterinarian about their dog's seizures, and for the remainder their veterinarian had been unable to diagnose IE.

1.7.3.5 The treatment of canine idiopathic epilepsy

1.7.3.5.1 Phenobarbital

The most commonly used antiepileptic medication (AEM) in small animals is the barbiturate phenobarbital [200]. Barbiturates are CNS depressants, some of which are commonly used as anaesthetic agents. Phenobarbital, in contrast to other barbiturates, works as an anticonvulsant at lower doses without causing anaesthesia [200, 280]. Barbiturates such as phenobarbital increase the binding of the inhibitory neurotransmitter GABA (γ-aminobutyric acid) to its receptor. GABA receptor activation inhibits the postsynaptic neuron, increasing the potential needed for depolarisation. Barbiturates also inhibit the activity of the excitatory neurotransmitter glutamate. In combination these effects of phenobarbital raise the seizure threshold [200]. Phenobarbital can have a high efficacy in dogs, decreasing seizure frequency in 60-93%, and is relatively safe. Common adverse effects of this AEM include drowsiness, ataxia, excessive hunger, thirst, and urination [200]. Many of these adverse effects occur at the start of treatment, and decrease as the dog develops tolerance to the AEM [200, 279]. There are also some less commonly reported adverse effects, including liver damage (hepatotoxicity).

1.7.3.5.2 Bromide

Bromide, typically administered as potassium bromide, is used in combination with phenobarbital to treat dogs with seizures which are refractory to phenobarbital monotherapy [200, 279]. Phenobarbital and bromide work synergistically to improve seizure control in epileptic dogs, and bromide treatment on its own is also possible but is not approved in most EU countries. When used in combination with bromide the dose of phenobarbital can often be reduced [200]. Common adverse effects of bromide

treatment include those of phenobarbital, with the addition of pelvic limb weakness, vomiting, and diarrhoea [279]. The combination of phenobarbital with bromide can amplify these side effects. The exact mechanism for bromide's effect has not yet been determined, although it may function through the stabilisation of neuronal cell membranes [200].

1.7.3.5.3 Imepitoin

Phenobarbital and bromide have both been in use for a long time, whereas imepitoin was developed more recently [200, 279, 281]. Imepitoin was initially intended for use in humans but was eventually developed for the treatment of canine epilepsy due to a better pharmacokinetic profile (the pharmacokinetics was found to be too variable in humans and affected by factors such as smoking status) [279, 281]. Imepitoin acts as a low affinity partial agonist of the benzodiazepine recognition site of the $GABA_A$ receptor, and potentiates its inhibitory effects [200, 281]. The lower affinity of imepitoin at the recognition site overcomes some of the adverse effects of benzodiazepines which are full agonists of this site, such as diazepam which is used to treat status epilepticus in dogs [200, 279, 281]. Imepitoin has been in use for a relatively short amount of time, so its side effects and possible reactions with other drugs are perhaps not so well characterised as phenobarbital or bromide, but it has been suggested that this AEM has less harmful adverse effects [279, 281]. The most commonly reported adverse effects of imepitoin include ataxia and polyphagia (excessive eating or hunger) [279, 282].

1.7.3.5.4 Other AEMs

Phenobarbital and imepitoin are the only AEMs approved for first-line treatment of canine IE [279]. The remaining AEMs are only used if phenobarbital or imepitoin treatment, as a monotherapy or in combination together or with potassium bromide, fails to reduce seizure frequency.

Levetiracetam's exact mechanism of action has not been determined, but is thought to regulate neurotransmitter release from vesicles [200]. Zonisamide's mechanism of action is also not well characterised, although it reduces sodium and T-type calcium channel activity [283]. Levetiracetam and zonisamide are both used for the treatment of epilepsy

in humans [200]. Felbamate is thought to inhibit N-methyl-D-aspartate (NMDA) excitatory signalling activity [200]. Topiramate's anticonvulsant effect is due to it acting on multiple pathways; it has been suggested that it inhibits sodium channels and AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) signalling, and increases inhibitory GABA activity [284]. The mechanism of action of gabapentin and pregabalin as anticonvulsants is uncertain [200]. Despite being structurally analogous to GABA, they do not function as GABA agonists [285]. They are thought to act on a subunit of voltage-gated calcium channels, the alpha2-delta protein, and reduce the release of neurotransmitters. Gabapentin is also used to treat neuropathic pain [200].

1.7.3.5.5 The use of diet to treat epilepsy

There is some evidence that changing the diet of IE affected dogs may help treat the disease [286-289]. A medium-chain TAG ketogenic diet, which is high in fat (using medium-chain fatty acids as an alternative fat source) and low in proteins and carbohydrates, has been suggested to be beneficial for some dogs affected by epilepsy [286]. The diet is considered a potential way to improve quality of life when AEM are ineffective, although the exact mechanism of action remains unknown. The ketogenic medium-chain triglyceride diet has also been suggested to help reduce ADHD-like behaviour in dogs with epilepsy [287]. Supplementation of a dog's diet with medium-chain triglycerides may have potential for the reduction of seizure frequency and has been suggested to improve cognition in dogs diagnosed with epilepsy [288, 289].

1.7.3.5.6 Treatment and treatment outcome in specific breeds

Belgian Shepherd dogs diagnosed by general practitioners are administered AEMs in 37% of cases, all of which are prescribed either phenobarbital on its own or, less commonly, in combination with potassium bromide [252, 258]. In 16% of cases inability to control seizures results in the dog being euthanised [252]. Remission has been reported to occur in 13.7% of dogs for which AEMs have not been administered [258].

Most Border Collies diagnosed with idiopathic epilepsy are administered AEM treatment continually [254]. A high proportion of these dogs are treated with more than one AEM [254, 255], typically phenobarbital and potassium bromide in combination. Some Border

Collies are given additional treatments along with phenobarbital and potassium bromide, namely one of carbamazepine, levetiracetam, or zonisamide [254]. Of the Border Collies treated with two or more AEMs, 71% are resistant to treatment and have one or more seizures each month [254]. Remission is achieved in some dogs (18%), mostly through treatment but in some cases remission is spontaneous. Adverse effects of treatment are reported by 67% of owners and cause a loss of working ability in over half of dogs [254].

A survey of Italian Spinone dogs born between 2000 and 2011 found that the majority of Italian Spinoni diagnosed with idiopathic epilepsy are treated with AEM [251]. Most dogs are treated using one or two AEMs, but some can be treated with as many as five. Treatment with more than one AEM is more common than monotherapy. The most administered AEMs are phenobarbital and potassium bromide [251]. Adverse effects of treatment are reported in 82% of Italian Spinone dogs; over half of owners consider the adverse effects to be acceptable. It has been reported that remission through treatment is achieved for very few Italian Spinone dogs (6%) [251].

A study of Petit Basset Griffon Vendeen dogs registered with the Danish Kennel Club between 1999 and 2008, the only published report of epilepsy prevalence and clinical characteristics in this breed, did not report any details of epilepsy treatment or treatment effectiveness in the dogs included [256]. This was also the case for the study of epilepsy in Irish Wolfhounds of 120 related litters [236].

1.7.3.6 Genetics

A genetic basis for IE is highly likely given the increased risk and high prevalence of the disease in some pedigree dog breeds, and the evidence for an underlying genetic cause in many human epilepsies [203, 217, 221, 222, 236, 251-253, 256]. Pedigree analyses for the Border Collie, Belgian Shepherd dog, and Irish Wolfhound have provided further evidence that IE has an underlying genetic cause within these breeds [236, 253, 254]. Some of these pedigree studies have suggested that an autosomal recessive mode of inheritance is possible; however, the Irish Wolfhound pedigree study and the GWAS findings for various breeds suggest a complex mode of inheritance is likely [80, 236, 253, 254, 290-293].

Although the probable genetic causes have been identified for some rare, monogenic, canine epilepsies, the underlying genetics remains unclear for most canine IE; it is likely that a combination of rare and more common variants, with smaller effects on conferring risk of disease, are involved, as is hypothesised for the common epilepsies in humans [222, 232, 268, 294]. The extent of the contribution of environmental causes and triggers is unknown; longitudinal studies may help elucidate this [270].

The following section summarises the studies which have so far investigated the genetics of IE in the dog. Some earlier studies looked at linkage using microsatellite markers, but GWAS have had the most success at identifying genomic regions associated with epilepsy: three genes and one haplotype containing a number of potential candidate genes have been identified as associated with canine epilepsy using this method to date [80, 268, 276, 290, 291, 294].

1.7.3.6.1 A SNP in LGI2 is associated with benign familial juvenile epilepsy (BFJE) in the Lagotto Romagnolo

An autosomal recessive mode of inheritance was suggested for BFJE based on the analysis of a large multinational pedigree [294]. A GWAS using DNA from 11 affected Lagotto Romagnolo dogs and 11 unaffected littermates identified a region associated with BFJE on chromosome 3 [294]. The region contains the epilepsy candidate gene *LGI2* (leucine-rich gene, glioma-inactivated, 2), sequencing of which identified an A to T substitution in affected dogs that introduces a stop codon. The SNP has a strong association with BFJE, but shows incomplete penetrance [294]. Some affected dogs are heterozygous and, rarely, unaffected dogs can be homozygous for the mutation. LGI2 belongs to a group of neuronally secreted proteins, with roles in the development and function of the nervous system, that includes LGI1, mutations in which putatively cause autosomal-dominant partial epilepsy with auditory features (ADPEAF) in humans [295-298]. The protein-truncation causing SNP identified in *LGI2* prevents secretion of the protein and interaction with the synaptic membrane proteins ADAM22 and ADAM23 [294].

1.7.3.6.2 A common risk haplotype in ADAM23

A linkage study using 410 microsatellite markers across the genome was initially used to investigate the genetics of IE in the Belgian Shepherd dog breed [275]. The study included DNA from 74 affected, 281 unaffected, and 11 dogs with unknown epilepsy status, and constructed pedigrees using an additional 344 dogs. Six tentative loci associated with IE were identified: three on chromosome 2, and one on each of chromosome 6, chromosome 12, and chromosome 37 [275].

A GWAS of 40 Belgian Shepherd dog IE cases and 44 controls identified a locus on chromosome 37 associated with the disease, which overlapped with the region found previously on this chromosome [275, 276]. The association was subsequently replicated by a GWAS using a higher density SNP array, a larger set of Belgian Shepherd dogs (93 cases, 162 controls, with a 16% overlap with the previous study), and a cohort of dogs from three additional breeds (Beagle, Finnish Spitz, and Schipperke, 157 cases and 179 controls in total) [290]. The most significantly associated SNP was in *ADAM23* (a disintegrin and metalloproteinase 23). Resequencing *ADAM23* in the Belgian Shepherd, and genotyping the identified variants in the four breeds, found a six-variant risk haplotype with low penetrance which covers *ADAM23* exons 5 to 11, although a causal mutation has not been identified [290].

The *ADAM23* risk haplotype was later genotyped in eight more breeds: Australian Shepherd (59 cases, 71 controls), Finnish Lapphund (35 cases, 50 controls), Irish Setter (19 cases, 36 controls), Kromfohrländer (46 cases, 29 controls), Labrador Retriever (29 cases, 50 controls), Miniature Pinscher (16 cases, 15 controls), Pyrenean Shepherd (17 cases, 23 controls), and Whippets (17 cases, 25 controls) [291]. A combined analysis of all eight breeds showed an association between the one of the variants (CanFam3.1, chromosome 37, 15,108,593 bp) and IE with strong statistical significance ($P= 4.6 \times 10^{-6}$). The risk haplotype was associated with IE ($P < 0.05$) in four of the breeds: Australian Shepherd, Kromfohrländer, Labrador Retriever, and Whippet [291]. The Finnish Lapphund did not show any association between IE and the haplotype; the other three breeds showed a trend towards association although not statistically significant. The *ADAM23* risk haplotype alleles were major (common) alleles in all eight breeds.

ADAM23 is one of three ADAM proteins (ADAM11, ADAM22 and ADAM23) that lack metalloprotease catalytic activity, are expressed in the brain, and act as receptors for LGI proteins [299]. ADAM23 is important for synapse maturation and function, has a role in the regulation of voltage-gated potassium currents, and it is known to interact with the epilepsy-associated proteins LGI1 and LGI2 [294, 299-301].

1.7.3.6.3 A potential second risk haplotype for IE in the Belgian Shepherd dog

A recent GWAS of IE in the Belgian Shepherd breed, including 20 cases and 45 controls, replicated the association on chromosome 37 by finding a region with suggestive association with IE that is in LD with *ADAM23* [292]. The strongest association in this study, however, was on chromosome 14, upstream of the *RAPGEF5* (Rap guanine nucleotide exchange factor) gene [292]. Downregulation of *RAPGEF5* has been suggested to have a potential role in epileptogenesis and seizure onset in a rat model of temporal lobe epilepsy [302]. Logistic regression of the chromosome 14 and 37 regions found that a two-loci model was highly predictive of IE status in the GWAS [292]. The model had high sensitivity but low specificity, i.e. many unaffected dogs have the risk alleles, and was not validated in an independent set of cases and controls. The study also found that the two risk loci do not account for all epilepsy risk, suggesting that there are additional risk loci or environmental risk factors yet to be identified [292]. Reanalysis of the data from the 2012 Belgian Shepherd dog IE study, this time imputed to allow analysis of 2,012,439 markers, replicated the chromosome 37 locus; however, the analysis was unable to replicate the association on chromosome 14, with or without the inclusion of the strongest associated marker on chromosome 37 as a fixed effect [303]. Another study, including an additional 47 cases and 74 controls, was also unable to replicate the association on chromosome 14, although an interaction between the two loci was reported to elevate IE risk [304]. It is possible that different risk loci are present within different Belgian Shepherd populations, or that the chromosome 14 association is a false positive. Genotyping large, independent, study sets within and across the different populations is necessary to validate the identified risk haplotype.

1.7.3.6.4 Irish Wolfhound

Hayward et al (2016) [80] used an approach that included 4,224 dogs from 150 breeds in addition to mixed breed and village dogs and performed GWAS for 12 different phenotypes. Within this were 34 Irish Wolfhounds affected by IE and 168 unaffected controls. This within-breed GWAS identified an epilepsy-associated haplotype, present in 38% of Irish Wolfhound cases and 11% of controls, which covers 13.5 Mb on chromosome 4 and contains a number of epilepsy candidate genes. However, the study did not attempt to narrow the haplotype or identify the gene or genes involved.

1.7.3.6.5 A deletion in DIRAS1 is associated with generalised myoclonic epilepsy in Rhodesian Ridgeback dogs

A combined GWAS, whole exome sequencing, and whole genome sequencing approach was used to identify a 4 bp deletion in *DIRAS1* as the cause of generalised myoclonic epilepsy with photosensitivity in Rhodesian Ridgeback dogs [268]. Whole exome sequencing of two unrelated affected dogs and 169 unaffected dogs of other breeds identified a 4 bp deletion in the second exon of *DIRAS1* that causes a frameshift and the loss of the stop codon, extending the encoded protein by 104 amino acids. A GWAS of 10 affected and 18 unaffected dogs confirmed that this region was associated with disease, and the deletion was found in the WGS of an affected dog and not in 99 unaffected dogs of other breeds. Genotyping additional Rhodesian Ridgebacks demonstrated that the deletion fully segregates with the disease. *DIRAS1* (DIRAS family GTPase 1) is a GTPase that may have a role in synaptic function by regulating the release of the neurotransmitter acetylcholine and could be involved in neuronal development [305, 306].

1.7.3.6.6 Petit Basset Griffon Vendeen

A GWAS of 23 Petit Basset Griffon Vendeen IE cases and 30 unaffected controls identified three regions with suggestive associations with IE on chromosomes 13, 24, and 35 [293]. The study identified a candidate gene near to each locus but sequencing the exons of the genes failed to identify any variants that could confer disease risk. The candidate genes investigated were *DOK5* (docking protein 5), *NRN1* (neuritin 1), and *FAM135b* (family with

sequence similarity 135 member B) [293]. This study hasn't ruled out the possibility that variation within regulatory elements within the regions could underlie IE risk, and studies with larger case-control sets would be necessary to determine the validity of the identified regions.

1.7.3.6.7 Studies with negative results

The publication of results that may be inconclusive is helpful for the research community to improve our knowledge of the inheritance of canine epilepsy and to allow researchers to build on previous studies.

Microsatellite markers located near to 52 candidate genes were used for linkage analysis in 31 IE affected and 60 unaffected dogs of the Vizsla breed, and association analysis in the Beagle, English springer spaniel, and Greater Swiss Mountain Dog breeds (24 cases and 24 controls of each breed) [307]. The study did not identify any genes that reached Bonferroni-corrected statistical significance for association with IE; this suggests that either these candidate genes are not involved in IE in these breeds, or the study did not have sufficient power to find an association and therefore provides evidence that canine IE is not monogenic in these breeds.

For a number of breeds the GWAS approach has so far been unsuccessful for investigating the genetics of IE [291]. GWAS including 50,000 SNP markers did not identify any associations which reached genome-wide significance in the Kromfohrländer (21 cases and 21 controls) or Miniature Pinscher (15 cases and 15 controls). A GWAS of 173,662 SNP markers was used for the Finnish Lapphund (40 cases and 97 controls) and Pyrenean Shepherd (20 cases and 27 controls) breeds and also did not identify any significantly associated regions. The inability of these GWAS to find an IE-associated region, particularly for the studies using the less SNP dense genotyping array, could be due to disease associated regions not being tagged by markers but it is most likely due to the fact that three of these GWAS used very small sample sets and therefore lacked statistical power to find common loci of modest effect on disease risk. These negative results give further evidence that in many breeds canine IE may not be monogenic and is likely to have a complex mode of inheritance.

1.7.3.6.8 The genetics of refractory IE

A survey of owners of IE affected dogs, general practice veterinarians, and veterinary neurologists found the research area of highest priority to be the management of antiepileptic medication and the development of new anti-seizure drugs [308]. It has been suggested that responsiveness to antiepileptic medication is affected by genetic factors, and this is supported by the differences in IE severity and the variation in prevalence of refractory IE between breeds; observable, for example, in the reports that most Belgian Shepherd dogs are treated with a phenobarbital monotherapy whereas treatment with more than one AEM is more common than monotherapy in Border Collies and Italian Spinone dogs [251, 252, 254, 258].

The *ABCB1* (ATP binding cassette subfamily B member 1) gene, also known as *MDR1* (multidrug resistance protein 1), has been investigated as a candidate gene for refractory IE [309-312]. *ABCB1* encodes an efflux transporter that has a role in the excretion of drugs over the blood-brain barrier, and variants within the gene can lead to ivermectin sensitivity in dogs [313, 314]. A 4 bp deletion that putatively causes ivermectin sensitivity in dogs was suggested to be associated with better seizure outcome in homozygous dogs in a study of 29 dogs with epilepsy [310]. A study that included 25 Border Collies with idiopathic epilepsy, of which 13 were refractory to phenobarbital treatment, found that a SNP (c.-6-180T > G) in the first intron of *ABCB1* was significantly more common in dogs resistant to treatment than the IE affected dogs that were responsive [309]. This variant was found to have a high frequency in 472 Border Collies in Japan, but phenobarbital-resistance was not investigated [311]. In a subsequent study of 45 refractory and 50 responsive dogs of various breeds including mixed breed dogs and Border Collies there was no association between the c.-6-180T > G SNP and refractoriness in an across breed analysis; instead, the study found a within-breed association between the homozygous T/T genotype with either refractoriness or responsiveness depending on the breed [312]. This suggests that any effect of variants within *ABCB1* on response to treatment may be dependent on other risk factors, genetic or environmental. The study included four or fewer dogs of each breed with refractory epilepsy, but 21 mixed-breed dogs, which limits its power to find associations within specific breeds. The three studies investigating *ABCB1* and refractory epilepsy all have limited sample set sizes, and the inconclusive

findings suggest that larger study sets are necessary to enable any conclusions to be reached on the role of *ABCB1* in the responsiveness to antiepileptic medication [309, 310, 312].

1.8 Summary

The dog is an interesting model for inherited neurological disease. Inbreeding and genetic bottlenecks have led to an increased prevalence of specific inherited disease within certain breeds and has had effects on the haplotype structure of the genome that can be both beneficial and challenging for genetic research. Inherited movement disorders and idiopathic epilepsy have a high prevalence in some breeds and can have an impact on length of life or quality of life.

Ataxia in the dog is predominantly inherited in an autosomal recessive manner, and WGS of small numbers of cases has been effective in the identification of putative causal variants. The canine paroxysmal dyskinesias for which putative causal variants have been identified have had autosomal recessive or autosomal dominant modes of inheritance, and GWAS or WGS have been used successfully in these cases, but there is evidence that inheritance is likely polygenic or complex in some breeds. Idiopathic epilepsy is highly likely to have a complex mode of inheritance in most breeds, and the GWAS approach has been the most successful in elucidating the underlying genetics. Higher density genotyping arrays are valuable when investigating complex disease, and genotype imputation would facilitate the generation of high-density GWAS datasets.

The identification of genetic factors contributing to the risk of developing idiopathic epilepsy and movement disorders in the dog could have major impact on the way these conditions are diagnosed, treated, and selected against. Such information could also provide invaluable insights into equivalent human conditions.

2 Aims and objectives

2.1 Aim

The aim of this PhD project is to identify genetic factors contributing to the risk of developing idiopathic epilepsy and movement disorders in the dog. These neurological disorders are distinctive but can overlap phenotypically, and previous research has identified examples of these diseases with both monogenic and suspected oligogenic or multigenic modes of inheritance. This PhD project will explore the genetic and computational approaches available for investigating canine neurological diseases, developing, and utilising, tools to best elucidate the underlying genetics.

2.2 Objectives

1. To whole genome sequence two Norwegian Buhund siblings affected by cerebellar ataxia and perform analyses to identify and characterise potentially causal variants. This incurable progressive cerebellar ataxia is likely rare [315], although it could potentially be an emerging disease in the breed, and previous research has demonstrated that WGS of small numbers of cases can be an effective methodology for investigating the genetics of canine inherited ataxia [65, 66].

2. To validate and implement genome-wide genotype imputation to impute Illumina CanineHD datasets for the Border Collie and Italian Spinone up to the genotype density possible through the Axiom Canine HD array, assessing the effect of breed and reference panel size on imputation accuracy. Such an approach could be a cost-effective use of new and existing datasets and be a useful tool for complex disease research in this PhD and for future studies.

3. To carry out a GWAS of PxD in the Norwich Terrier and to investigate any disease-associated loci in a large population-based dataset. Although a pedigree study supplied evidence that PxD in this breed is inherited [120], it could not establish mode of inheritance and there are no previous publications describing the underlying genetics. Mode of inheritance for canine PxD can be monogenic or potentially multigenic [121, 189-193]. This study will therefore be the first to

explore the underlying genetics; the GWAS approach is expected to identify disease-associated loci if PxD in the Norwich Terrier is monogenic or to give a preliminary indication of mode of inheritance if PxD is complex.

4. To carry out a GWAS of IE in the Italian Spinone and to investigate any disease-associated loci in independent sample sets. IE has a high prevalence in the Italian Spinone and a severe clinical course [251], making it a welfare concern for the breed. As for Norwich Terrier PxD, there have not been any previous published studies describing the mode of inheritance for IE in the Italian Spinone. Although monogenic epilepsies have been reported [268, 294], for other breeds a complex mode of inheritance for IE is suspected [80, 236, 253, 254, 290-293]. This study will be the first GWAS of IE in the Italian Spinone to be described and will supply evidence of mode of inheritance, identify disease-associated loci if IE is monogenic in this breed, and potentially indicate risk-associated loci for a multigenic disorder.

5. To carry out a GWAS and meta-analysis of IE in the Border Collie and to investigate any disease-associated loci in a large independent case-control set. Although the prevalence of IE in the Border Collie has not been investigated in a breed-wide survey, an increased prevalence has been reported [203], and the numerically large size of this breed and severe clinical course suggest that even with a relatively low prevalence a large number of dogs could be severely affected making the disease a major welfare concern [254, 255, 316] (https://www.thekennelclub.org.uk/media-centre/breed-registration-statistics/, accessed 16/01/2023). The lack of significant findings in an unpublished GWAS of IE in the Border Collie that was conducted as part of the LUPA consortium [3] suggested that the disease is likely oligogenic or multigenic. Power calculations suggest that a study including a minimum of 100 cases and 150 controls would have sufficient power to identify variants of moderate or intermediate effect (**Appendix i.i**).

3 Manuscript 1 - Characterisation of canine *KCNIP4*: A novel gene for cerebellar ataxia identified by whole-genome sequencing two affected Norwegian Buhund dogs

This manuscript has been published in PLOS Genetics

Full citation:

Jenkins CA, Kalmar L, Matiasek K, Mari L, Kyöstilä K, Lohi H, Schofield EC, Mellersh CS, De Risio L, Ricketts SL (2020) Characterisation of canine KCNIP4: A novel gene for cerebellar ataxia identified by whole-genome sequencing two affected Norwegian Buhund dogs. PLoS Genetics. 16(1):e1008527. doi: 10.1371/journal.pgen.1008527.

3.1 Authors and affiliations

Christopher A Jenkins

Kennel Club Genetics Centre, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Lajos Kalmar

Department of Veterinary Medicine, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

Kaspar Matiasek

Section of Clinical & Comparative Neuropathology, Centre for Clinical Veterinary Medicine, Ludwig-Maximilians-Universität Munich, München, Germany

Lorenzo Mari

Neurology/ Neurosurgery Service, Centre for Small Animal Studies, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Kaisa Kyöstilä

Department of Veterinary Biosciences, and Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland

Folkhälsan Research Center, Helsinki, Finland

Hannes Lohi

Department of Veterinary Biosciences, and Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland

Folkhälsan Research Center, Helsinki, Finland

Ellen C Schofield

Kennel Club Genetics Centre, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Cathryn S Mellersh

Kennel Club Genetics Centre, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Luisa De Risio

Neurology/ Neurosurgery Service, Centre for Small Animal Studies, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Sally L Ricketts

Kennel Club Genetics Centre, Animal Health Trust, Newmarket, Suffolk, United Kingdom

3.2 Authors contributions

CAJ contributed to study design, performed experimental work and analysis and interpretation of data, and wrote the manuscript. SLR conceived of, designed, and supervised the study, and provided critical review of the manuscript. LDR contributed to study conception and provided clinical expertise. LK provided expertise and contributed to study design relating to *in silico* protein analysis. KM performed the immunohistochemical analysis. LM contributed case definition and clinical expertise. CSM contributed to study conception and funding. ECS oversaw genome alignment and data curation and provided bioinformatics support including scripts for WGS analysis. KK and HL contributed to sample collection.

3.3 Abstract

A form of hereditary cerebellar ataxia has recently been described in the Norwegian Buhund dog breed. This study aimed to identify the genetic cause of the disease. Whole-

genome sequencing of two Norwegian Buhund siblings diagnosed with progressive cerebellar ataxia was carried out, and sequences compared with 405 whole genome sequences of dogs of other breeds to filter benign common variants. Nine variants predicted to be deleterious segregated among the genomes in concordance with an autosomal recessive mode of inheritance, only one of which segregated within the breed when genotyped in additional Norwegian Buhunds. In total this variant was assessed in 802 whole genome sequences, and genotyped in an additional 505 unaffected dogs (including 146 Buhunds), and only four affected Norwegian Buhunds were homozygous for the variant. The variant identified, a T to C single nucleotide polymorphism (SNP) (NC_006585.3:g.88890674T>C), is predicted to cause a tryptophan to arginine substitution in a highly conserved region of the potassium voltage-gated channel interacting protein KCNIP4. This gene has not been implicated previously in hereditary ataxia in any species. Evaluation of KCNIP4 protein expression through western blot and immunohistochemical analysis using cerebellum tissue of affected and control dogs demonstrated that the mutation causes a dramatic reduction of KCNIP4 protein expression. The expression of alternative *KCNIP4* transcripts within the canine cerebellum, and regional differences in KCNIP4 protein expression, were characterised through RT-PCR and immunohistochemistry respectively. The voltage-gated potassium channel protein KCND3 has previously been implicated in spinocerebellar ataxia, and our findings suggest that the Kv4 channel complex KCNIP accessory subunits also have an essential role in voltage-gated potassium channel function in the cerebellum and should be investigated as potential candidate genes for cerebellar ataxia in future studies in other species.

3.4 Author summary

Hereditary ataxias, which are a group of disorders characterised by incoordination of movement, are typically incurable and there are often no disease-modifying treatments available. Canine hereditary ataxias are a notable group of movement disorders in dogs, and represent well characterised naturally occurring disease models of ataxia that can help improve our understanding of the underlying biology of the disorder in both dogs and humans. We used the whole genome sequences of two affected siblings to investigate the genetic cause of a slowly progressive form of hereditary ataxia in the

Norwegian Buhund dog breed, and identified a single base change within the *KCNIP4* gene. We have characterised the expression of *KCNIP4* in the dog, and investigated the effect of the identified mutation. This gene has not previously been implicated in inherited ataxia in any species, and our findings suggest that this and related genes represent potential candidates for ataxia in future studies in other species. Our findings will allow dog breeders to avoid producing affected dogs, reduce the disease allele frequency, and eventually eliminate the disease from the breed, through the use of a DNA test.

3.5 Introduction

Hereditary ataxias are a group of movement disorders, typified by incoordination of gait, limbs, or eyes, primarily caused by inherited dysfunction of the cerebellum and/or its afferent or efferent pathways [92]. In humans autosomal recessive and dominant forms of hereditary ataxia have been reported, in addition to mitochondrial, and, in the case of fragile X tremor-ataxia, X-linked forms. Inherited ataxias are typically not curable and there are often not any disease-modifying treatments available [92, 317].

Multiple examples of hereditary ataxia have been described in purebred dogs [32, 64, 65, 114, 131, 132, 134]. Although the specific diseases are often rare, and the genetic mutations can be breed-specific, hereditary ataxia is a key cause of movement disorders in dogs. Canine hereditary ataxia is typically inherited in an autosomal recessive manner, and putatively disease-causing variants have been identified for some breeds [32, 64, 65, 114, 130-132, 134, 147, 318]. Some of the genes implicated in canine ataxia had not previously been associated with disease in humans [63, 122, 123, 130], whereas other forms of canine ataxia are associated with variants within the same genes that are associated with well characterised forms of human ataxia [62, 64, 65, 134]. Canine hereditary ataxia is a naturally occurring disease model and research into the genetics of ataxia in purebred dogs can help improve the understanding of the underlying molecular mechanisms of human disease.

Hereditary ataxias in humans are classified by mode of inheritance, whereas in dogs, where the underlying genetic basis is often less well defined, have recently been classified based on the clinical signs and the neuronal structures affected [92, 114]. The five classes

of canine hereditary ataxia are cerebellar cortical degeneration (CCD), spinocerebellar degeneration, canine multiple system degeneration (CMSD), cerebellar ataxias without significant neurodegeneration, and episodic ataxias [114].

Different genetic approaches have successfully been used to investigate the causal mutations for canine hereditary ataxia. Homozygosity mapping, linkage analysis, and targeted resequencing was used to investigate spinocerebellar ataxia in the Italian Spinone, and identified a GAA repeat expansion in *ITPR1*, a calcium channel that regulates intracellular calcium levels [64]. mRNA sequencing and a candidate gene approach was used to identify an 8 bp deletion in *SPTBN2* (β-III spectrin, involved in the development of Purkinje cells) associated with CCD in Beagles [62]. Genome-wide association studies (GWAS) have been used to identify likely causal mutations for CCD in Finnish Hounds (*SEL1L*, targeted degradation of misfolded or unassembled peptides), and Old English Sheepdogs and Gordon Setters (*RAB24*, which has a role in autophagy); and spinocerebellar ataxia in Russell Terrier Group dogs (*CAPN1*, which encodes μ-calpain, a subunit of calcium dependent cysteine protease) [63, 122, 123]. Whole genome sequencing of single cases of ataxia has been used more recently to successfully identify likely causal variants for CCD in the Hungarian Vizsla (*SNX14*, which has a role in the maintenance of neuronal excitability and synaptic transmission, and is associated with cerebellar ataxia in humans) [65] and spinocerebellar ataxia in Russell Group Terriers (*KCNJ10*, a potassium channel gene) [66]. Both of these studies used control genomes, and predicted effect on the protein sequence, to filter common benign variants, and then identified variants within candidate genes for genotyping in additional dogs.

The clinical and histopathological characteristics of hereditary cerebellar ataxia in the Norwegian Buhund have been described previously in a study of four cases presenting with mild and slowly progressive cerebellar ataxia, characterised by a broad base stance and hypermetria in all four limbs, truncal ataxia, and fine head tremors [315]. At referral two dogs (siblings) were aged 12 weeks, and two other affected dogs were aged 16 and 20 weeks. Histopathological analysis of the cerebellum showed minor signs of degeneration and reduced expression of Purkinje cell differentiation markers calbindin-D-28K and ITPR1 in some cerebellar regions [315]. Pedigree analysis suggested an autosomal recessive mode of inheritance. Genome-wide mRNA-sequencing of two

affected siblings and subsequent investigation of 20 candidate genes in this dataset did not identify any potentially causal variants for the disease [315].

The present study aimed to identify the underlying genetic cause of ataxia in the Norwegian Buhund breed using whole genome sequencing of the two affected siblings. A mutation within *KCNIP4*, a novel gene for cerebellar ataxia, was identified.

3.6 Results

3.6.1 Whole genome variant filtering

Whole genome sequences of two ataxia-affected Norwegian Buhund siblings were initially compared with the Boxer reference sequence CanFam3.1 and whole genome sequences of 44 dogs unrelated to the two cases and of 29 different breeds (control genomes). Variants identified were filtered to leave only those which were homozygous in both of the affected dogs, but which were homozygous for the reference or an alternative allele in the control dogs. Our hypothesis was that the causal variant is rare and private to the Norwegian Buhund breed. From this, a total of 26,073 segregating variants were identified (**Appendix ii.i**). Additional filtering by variant effect, leaving only high-effect variants that were predicted to directly affect a protein-coding sequence, or disrupt a transcript, reduced the number to 121.

The 121 variants were then further filtered using whole genome sequence variant data from the Dog Biomedical Variant Database Consortium (DBVDC) [67]. The consortium dataset included 361 additional whole genome sequences comprising 96 different pure breeds, three wolves, and seven types of mixed-breed dog. The variants were filtered to leave only those which were absent in all of the consortium genomes. This left 16 high-effect variants that were homozygous in the two affected dogs and were not present in any of the other whole genome sequences. Seven of the 16 variants were predicted to be tolerated and benign by two variant effect prediction tools; SIFT and PolyPhen-2. The nine remaining variants were taken forward for genotyping in additional Norwegian Buhunds.

3.6.2 Variant segregation in additional dogs

Each of the nine variants was initially genotyped in 14 additional, unaffected, Buhunds. These included two full siblings of the affected dogs and two obligate carriers which were

identified through the pedigree (**Appendix ii.ii**). Only one of the nine variants segregated as would be expected for an autosomal recessive mode of inheritance: a nonsynonymous SNP in the *KCNIP4* gene (NC_006585.3:g.88890674T>C, XP_003434448.1:p.(Trp142Arg)) (**Appendix ii.iii**). This variant was genotyped in a further 56 Buhunds (archived DNA samples collected between 2008 and 2011) not reported to have ataxia (a total of 70 tested "UK Buhund set 1"). Of these 70 dogs 24 were heterozygous for the *KCNIP4* variant (T/C), and the other 46 were homozygous for the CanFam3.1 reference allele (T/T) (**Table 3.1**). None of the unaffected dogs were homozygous for the *KCNIP4* variant. Genotyping unaffected Norwegian Buhunds sampled in 2017 revealed that this contemporary set of 36 UK dogs ("UK Buhund Set 2") included three heterozygotes, and within 40 dogs sampled in Finland ("Finnish Buhund Set") one heterozygote was identified (**Table 3.1**). Neither of these sample sets included any dogs homozygous for the variant. Two additional Norwegian Buhunds were previously diagnosed with cerebellar ataxia in 1998 and 2002. Genotyping these additional cases, in addition to the two siblings used for whole genome sequencing, confirmed that all four affected dogs were homozygous for the *KCNIP4* variant (**Table 3.1**).

Table 3.1. Genotypes of Norwegian Buhunds and a multi-breed panel for the *KCNIP4* variant (NC_006585.3:g.88890674T>C)

	T/T	T/C	C/C	Total
Buhund Cases	0	0	4	4
UK Buhund Set 1	46	24	0	70
UK Buhund Set 2	33	3	0	36
Finnish Buhund Set	39	1	0	40
Multi-breed Panel	359	0	0	359
Total	477	28	4	509

For further validation and to investigate if the variant is confined to the Norwegian Buhund breed a panel of 359 dogs of 122 other breeds was genotyped. The *KCNIP4* variant was not present in this multi-breed panel (**Table 3.1**) (**Appendix ii.iv**).

Variant data for additional whole genome sequences subsequently became available after the initial analysis. These additional genomes included 140 in-house genomes and 255

genomes in the DBVDC consortium [67], all of which were homozygous for the reference allele.

In total the *KCNIP4* variant was assessed in 802 whole genome sequences, including dogs of 158 breeds, 13 mixed breed dogs, and eight wolves (**Appendix ii.v**), and genotyped in an additional 505 unaffected dogs (including 146 Norwegian Buhunds), and only the four affected Buhunds (two whole genome sequences and two genotyped) were homozygous for the variant. These results demonstrate that the variant segregates with disease and is confined to the Norwegian Buhund breed.

3.6.3 Bioinformatics tools predict the *KCNIP4* variant to be deleterious

The *KCNIP4* variant is a nonsynonymous T/C SNP causing a tryptophan to arginine amino acid change. The nucleotide, and codon within which the *KCNIP4* variant is located, is conserved across 99 species of mammal (UCSC) (**Appendix ii.vi**). Tryptophan is highly conserved at this location, and so is the flanking amino acid sequence in 19 species of mammal (16 primates) (UCSC) (**Appendix ii.vii**). SIFT predicted the variant to be "deleterious" (SIFT value: 0), and Polyphen-2 "probably damaging" (Polyphen-2 value: 0.99) [319, 320], both based on Ensembl transcript ENSCAFT00000026195.4. The effect of the variant was also assessed using a third tool, Mutation Taster, which predicted the variant to be "disease causing" (transcript Genbank ID: NM_025221.6, probability: 0.99) [321].

3.6.4 At least five *KCNIP4* transcripts with alternative first exons are expressed in canine cerebellar tissue

There are four NCBI RefSeq *KCNIP4* transcripts, and seven Ensembl *KCNIP4* transcripts, annotated for the canine genome (**Appendix ii.viii**). Two of the canine Ensembl transcripts (ENSCAFT00000060142.1, ENSCAFT00000083618.1) match canine RefSeq transcripts (XM_014112663.2, XM_003434400.4), making a total of nine unique *KCNIP4* transcripts annotated for the canine genome. Three of the NCBI RefSeq canine *KCNIP4* transcripts correspond to human Ensembl transcripts and transcripts reported previously to be expressed in human and mouse cerebellum [322] (**Appendix ii.viii**). The corresponding protein RefSeq for canine transcripts XM_014112663.2, XM_003434400.4, and XM_005618660.3 align to the protein sequences for human transcripts KCNIP4-1bΔ2

(ENST00000382148.7), KCNIP4-1dΔ2 (ENST00000382150.8), and KCNIP4-1eΔ2 (ENST00000382149.9) respectively (**Appendix ii.viii**) (**Figure 3.1A**). These three transcripts were confirmed using RNA sequencing data for canine cerebellum from dogs of multiple breeds, including the two affected Norwegian Buhund siblings.

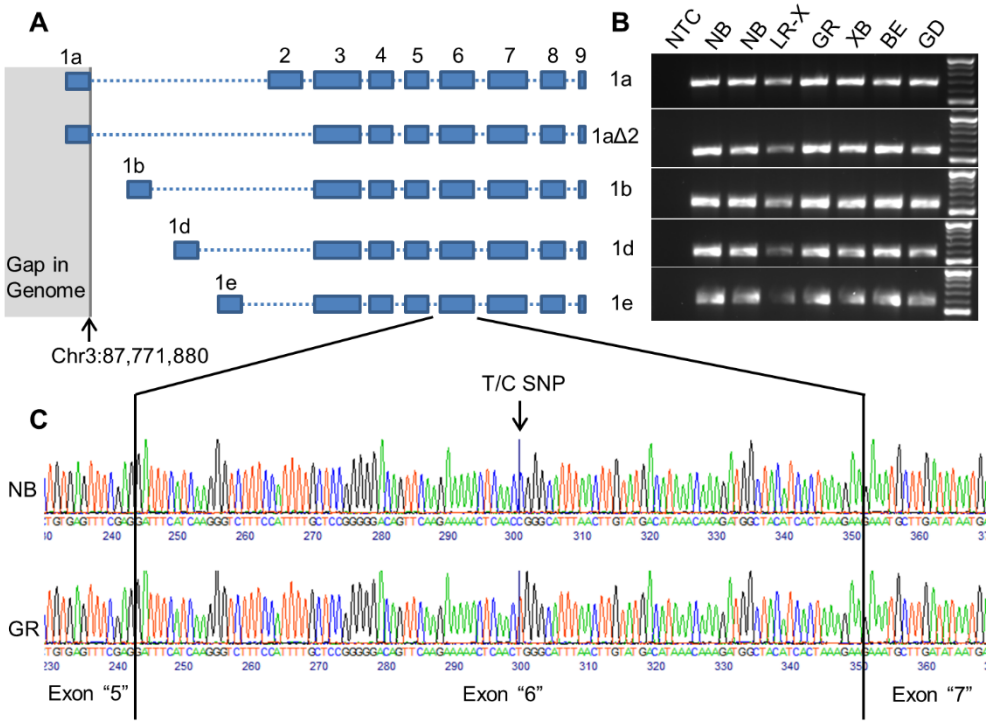


Figure 3.1. Canine *KCNIP4* transcript exon composition and confirmation of the expression of the transcripts, including the exon containing the identified variant, in canine cerebellum. A) An illustration of the exon composition of the five canine KCNIP4 transcripts identified using mRNA and whole genome sequencing data and alignment with known human transcripts. KCNIP4-1a and KCNIP4-1aΔ2 share exon 1a as their first exon, whereas all other transcripts have unique first exons. In these five canine transcripts, exon 2 is only present in KCNIP4-1a. Exons 3 to 9 are shared across all five transcripts. Coordinates for the first and second exons of each transcript are in Appendix ii.viii. The genomic region containing exons 3 to 9 as labelled in the Figure is Chr3: 88,780,584-88,894,638. The genomic region containing all of the transcripts is Chr3: 87,771,818-88,894,638. B) RT-PCR of five KCNIP4 transcripts in canine cerebellum samples for seven dogs. Primers and expected product sizes are given in Appendix ii.xii. From left to right: no template water control (NTC); two ataxia-affected Norwegian Buhunds (NB); Labrador Retriever cross breed (LR-X); Golden Retriever (GR); Siberian Husky cross breed (XB); Beagle (BE); Great Dane (GD). C) Sanger sequencing chromatogram demonstrating that exon "6", containing the mutation, is present in the transcripts expressed in the canine cerebellum. The sequences shown are for KCNIP4-1bΔ2. NB: Norwegian Buhund, GR: Golden Retriever.

Two other transcripts (KCNIP4-1a and KCNIP4-1aΔ2) which have been reported to be expressed in human and mouse cerebellum do not have corresponding canine RefSeq transcripts [322] (**Appendix ii.viii**). An assessment of the syntenic region for the first exon for these transcripts in CanFam3.1 revealed that there is a gap in the canine genome at this location (CanFam3.1 Chr3:87,771,558-87,771,880) (**Figure 3.1A**). When translated into the amino acid sequence, unmapped reads from the RNA sequencing data which were paired to the mapped reads of the flanking exons (exon 2 and exon 3), aligned with the N-terminal protein sequence corresponding to the first exon of human transcripts KCNIP4-1a (ENST00000382152.7), which contains "exon 2", and KCNIP4-1aΔ2 (ENST00000447367.6), in which "exon 3" is the second exon (**Figure 3.1A**) (**Appendix ii.viii**). Reads extending upstream into the gap in the genome from the gap's 3' end (chr3:87,771,880) aligned with part of the missing sequence, confirming that this gap in the genome is the location of the unaligned exon. A comparison of the identified transcripts, and the sequence and genomic position of canine Ensembl transcripts ENSCAFT00000090603.1 and ENSCAFT00000090861.1, revealed that they are partial canine transcript annotations for KCNIP4-1a and KCNIP4-1aΔ2, but are lacking most of the first exon as a result of the gap in the genome (**Appendix ii.viii**).

There are also four annotated canine transcripts, one RefSeq and three Ensembl, which do not match any of the known human *KCNIP4* Ensembl transcripts (**Appendix ii.viii**). Two of these transcripts, ENSCAFT00000079461.1 and ENSCAFT00000061238.1, are indicated in the annotation to have an additional exon between exons 3 and 4. This additional exon is not seen in any of the human transcripts or any of the other annotated canine transcripts, and was not observed in canine cerebellum mRNA sequencing data. Ensembl transcript ENSCAFT00000026195.4 appears to be an incorrect amalgamation of the first exon of KCNIP4-1eΔ2 (XM_005618660.3), the second exon of KCNIP4-1a, and exons 3 to 9 which are present in all of the *KCNIP4* transcripts (**Figure 3.1A**). This transcript was not seen in the mRNA sequencing data. Canine RefSeq transcript XM_536275.6 is stated to start with the second exon of KCNIP4-1a, followed by exons 3 to 9. This combination of exons is not seen in any of the annotated human Ensembl transcripts.

RT-PCR of RNA extracted from cerebellum tissue samples from two ataxia-affected Buhunds and five unaffected dogs of other breeds confirmed that at least five transcripts

for *KCNIP4*, with alternative first exons, are expressed in the canine cerebellum (**Figure 3.1B**). The variant identified in this study is located in exon 6, which Sanger sequencing of the RT-PCR products for the Buhund siblings and a Golden Retriever confirmed is expressed in all five of these transcripts (**Figure 3.1C**). All of the confirmed transcripts contain exons 3 to 9.

3.6.5 RT-qPCR of *KCNIP4* expression

The relative expression of *KCNIP4* in cerebellar tissue samples from the two sibling cases and five unaffected dogs, in comparison to the ubiquitously expressed TATA box binding protein (*TBP*) gene, was assessed using RT-qPCR. The assay for *KCNIP4* was designed with primers in exons 3 and 4, both of which are present in all five of the transcripts shown to be expressed in the canine cerebellum, allowing the assay to quantify total *KCNIP4* transcript expression.

Relative quantification of *KCNIP4* was suggestive of reduced expression in cerebellar tissue from the two affected dogs ($\Delta\Delta Cq = 0.497$) (**Figure 3.2A**), although the change in expression was not statistically significant (Student's T-test, $P = 0.07$).

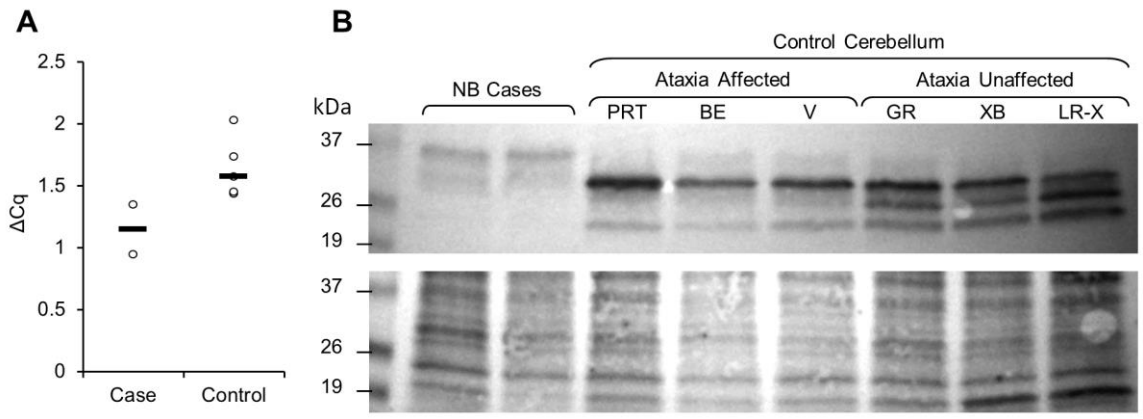


Figure 3.2. RT-qPCR and western blot analysis of KCNIP4. A) Relative quantification (ΔCq) of *KCNIP4* in comparison to the ubiquitously expressed control gene TBP in cerebellar tissue samples of two ataxia affected Buhund siblings (Case) and five ataxia unaffected dogs (Labrador Retriever cross-breed, Siberian Husky cross-breed, Beagle, Golden Retriever, and a Great Dane) (Control). Black lines show group median. B) Western blot comparing KCNIP4 protein expression in the cerebellum tissue lysate of two Norwegian Buhund cases and six control dogs. Top panel: Anti-KCNIP4 western blot. Bottom panel: Ponceau S total protein loading control. From left to right: two ataxia affected Norwegian Buhunds (NB); Three ataxia-affected control dogs of other breeds (Parson Russell Terrier (PRT); Beagle (BE); Hungarian Vizsla (V)); three ataxia-unaffected control dogs of other breeds (Golden Retriever (GR); Siberian Husky cross breed (XB); Labrador retriever cross breed (LR-X)). Approximate sizes of protein ladder bands are indicated on the left (kDa).

3.6.6 Western blot shows loss of KCNIP4 in cerebellum of cases

Western blot analysis was carried out for KCNIP4 using cerebellum tissue lysate from two cases and six control dogs. Three of the controls were affected by cerebellar ataxia but were of other breeds for which the putative causal mutation is known [62, 63, 65], and the remaining three were unaffected by ataxia. The western blot showed a clear change in KCNIP4 expression in the cerebellum tissue from the Buhund case (**Figure 3.2B**). Three bands are visible for the control tissues, at approximately 28 kDa, 26 kDa, and 22 kDa. The band with the highest saturation in all controls is at 28 kDa, with the lower molecular weight bands observed to be fainter in the three controls which were affected by genetically distinct forms of cerebellar ataxia. The two lower bands are absent, or too faint to observe, in the two Buhund cases. The 28 kDa band appears to still be present in

the Buhund cases, but at a much lower saturation. A fourth band is observable for all eight tissue lysates at approximately 37 kDa, and has a higher saturation in the two Buhund cases. The different bands may represent the differently sized isoforms of KCNIP4, although the band at 37 kDa does not fit within the size ranges of any of the known KCNIP4 isoforms (**Appendix ii.viii**). The western blot analysis indicates a dramatic reduction of KCNIP4 expression in dogs homozygous for the *KCNIP4* variant, and could suggest a complete loss of some, if not all, KCNIP4 isoforms.

3.6.7 Immunohistochemistry shows a reduction of KCNIP4 expression in the cerebellum of an affected Buhund

Expression of KCNIP4 protein was identified immunohistochemically throughout synaptic glomeruli of the cerebellar granular cell layer and basket cells of the molecular layer in all dogs, but staining intensity was considerably lower in the Buhund with two copies of the mutation (**Figure 3.3**). As functional differentiation of Purkinje cells appears to underlie regional differences [315], sagittal and transverse sections of all cerebellar lobuli were examined. Reduction of KCNIP4 expression was seen throughout the entire cerebellar cortex but more extensive in areas where Purkinje cells showed least expression of calbindin and ITPR-1 in the histopathological analysis carried out previously for the clinical characterisation of the disease [315].

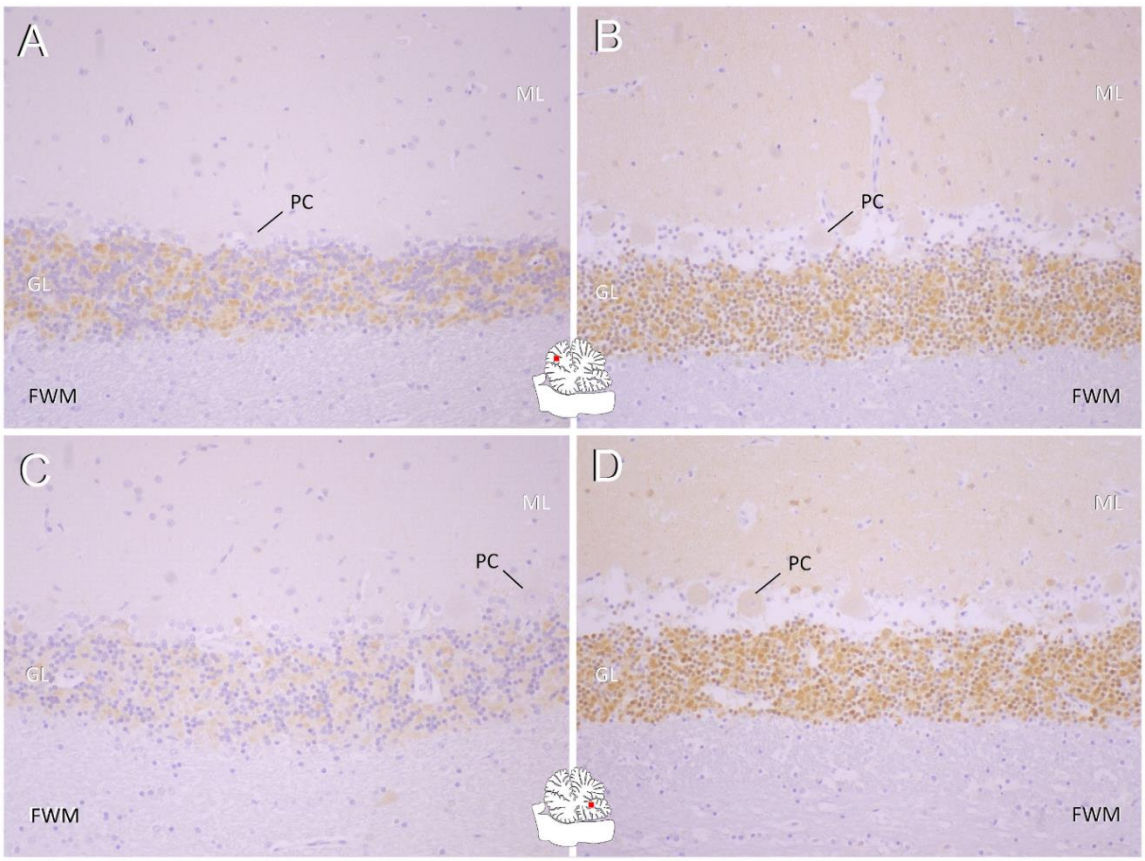


Figure 3.3. Immunohistochemical analysis of KCNIP4. Immunohistochemical expression patterns of KCNIP4 in the cerebellum of an affected Buhund (A, C) compared to that of two control dogs (B,D).Top panels: Rostral lobe. Bottom panels: Caudoventral vermis. The Buhund shows a diffusely decreased immunopositivity within the molecular layer (ML) and granular layer (GL) (brown staining). Thereby, rostral lobe areas (A) present with slightly stronger signal intensity than the caudoventral vermis (C). This difference matches to regional differences in Purkinje cell marker expression demonstrated previously [315]. PC: Purkinje cell; FWM: foliary white matter.

3.6.8 *In silico* protein analysis of the 3D structure of KCNIP4 suggests that the mutation affects protein stability and function

The structure of one KCNIP4 isoform, KChIP4a (KChIP4 was a previous abbreviation for KCNIP4), has been determined through X-Ray diffraction (PDB ID 3DD4). KChIP4a aligns perfectly with XP_003434448.1 and KCNIP4-204 (KCNIP4-1dΔ2). This allowed the use of online tools for predicting the effect of the variant on protein stability. The Eris server, which uses discrete molecular dynamics, was used to predict the effect of the mutation

on the free energy of the protein's structure [323]. The tryptophan to arginine substitution caused a predicted ΔΔG of 7.31 kcal/mol, which indicates a dramatic decrease in stability.

The 3D protein structure was used to investigate the physical location of the amino acid substitution within the protein. The tryptophan residue affected is within the hydrophobic core of KCNIP4, and is predicted to interact with the N-terminal helix which sits within the groove of the protein (**Figure 3.4A**). The substitution replaces the non-charged, non-polar, hydrophobic tryptophan with the positively charged, polar, arginine. This is likely to have an impact on the hydrophobicity of the protein's core. The arginine residue is predicted to interact with different helical structures within the protein compared to tryptophan, no longer interacting with the N-terminal helix (**Figure 3.4B**). The arginine residue is also predicted to overlap, and clash, with neighbouring residues (**Figure 3.4C**).

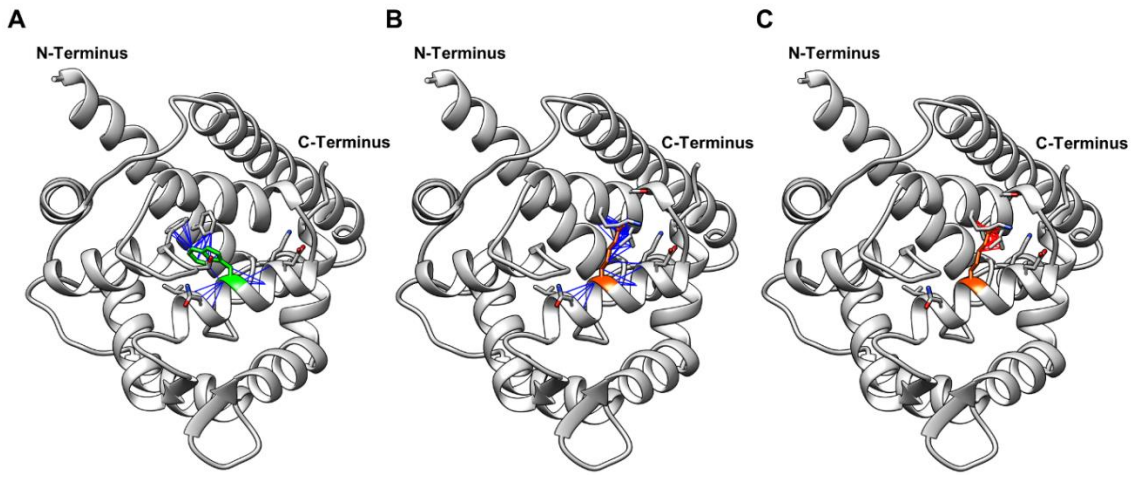


Figure 3.4. 3D models of KCNIP4 with and without the amino acid substitution caused by the mutation identified. (A) KCNIP4 with tryptophan (green) at position 142. (B) KCNIP4 with arginine (orange) at position 142. Blue lines in A) and B) indicate predicted interactions between atoms (including polar and nonpolar interactions). (C) KCNIP4 with arginine (orange) at position 142, with red lines indicating clashes (interactions where atoms are too close together).

3.7 Discussion

In the present study we used whole genome sequencing to identify the likely causal mutation for a recently characterised cerebellar ataxia in the Norwegian Buhund dog breed [315]. Whole genome sequencing of single cases of ataxia has recently been used to successfully identify mutations putatively causing CCD in the Hungarian Vizsla and spinocerebellar ataxia in Russell Group Terriers [65, 66]. Taken together this suggests that cerebellar ataxias in the dog are particularly amenable to this approach for the identification of candidate causal mutations. The approaches used in these two previous studies, however, used the candidacy of the genes in which variants were found to filter for potentially causal variants [65, 66]. The present study used a much larger number of in-house and consortium control sequences to filter variants to a level manageable for follow-up, allowing the discovery of a likely-pathogenic variant in *KCNIP4*, a novel gene for cerebellar ataxia. A previous study that used GWAS and targeted resequencing to investigate a different form of canine cerebellar ataxia in Russell Terrier Group dogs identified a variant in CAPN1, which had also not previously been implicated in ataxia in humans [63]. Mutations in this gene were later demonstrated to cause autosomal-recessive hereditary spastic paraplegia in humans, which is a condition that can present with cerebellar signs and which is associated with ataxia in some cases [82]. Canine hereditary cerebellar ataxia, as a naturally occurring disease model, therefore represents a resource for the identification of novel genes that should be considered as potential candidates when investigating similar conditions in humans. The discovery of novel genes for hereditary ataxia (and other rare, autosomal recessive, or autosomal dominant diseases) in the dog, when only very small numbers of cases are available, is therefore becoming increasingly possible with the rapid expansion of publicly available whole genome sequence datasets [67]. In the future this could make clinical diagnostic sequencing affordable and efficient for emerging Mendelian canine conditions.

The mutation identified is a nonsynonymous SNP causing a tryptophan to arginine amino acid substitution within a highly conserved region of KCNIP4. Voltage-gated potassium channel-interacting proteins (KCNIP, previously called KChIP) are four calcium binding proteins which interact with voltage-gated potassium (Kv) channels and modulate A-type potassium currents [324, 325] (**Appendix ii.ix**). Mutations in *KCNIP4* have not been

associated with cerebellar ataxia previously in any species to the authors' knowledge. However, mutations within KCND3, which encodes the voltage-gated potassium channel α-subunit Kv4.3, have been found to cause spinocerebellar ataxia type 19/22 in humans [326-330]. Mutations in KCND3 reduce trafficking and cell surface expression of KV4.3 and suppress the amplitude of the potassium current or affect channel gating [331]. KChIP4a, an isoform of KCNIP4, has been shown to interact with Kv4.3 and to modulate its inactivation [332, 333].

RT-qPCR analysis did not demonstrate a statistically significant reduction in the expression of the *KCNIP4* transcript in the affected Buhunds, although it was suggestive of reduced expression; whereas a dramatic reduction in KCNIP4 protein expression was observed in the cerebellum of cerebellar ataxia-affected Norwegian Buhunds, through both western blot and immunohistochemistry. It has been demonstrated previously that changes in the protein expression of the Kv4 accessory subunits (KCNIP1, KCNIP2, KCNIP3, and KCNIP4) do not necessarily reflect transcriptional expression [334]. It has been suggested that the formation of complexes between the accessory subunits and the α-subunits (Kv4.2 or Kv4.3) stabilises the proteins, leading to increased levels of each protein [334]. The tryptophan to arginine substitution identified in the current study is predicted to be destabilising and damaging to protein function. If the predicted effect on protein function prevents the formation of complexes, and this precludes the stabilisation of the protein, combined with the mutation's destabilising effect this could potentially explain the dramatic reduction in protein expression observed without a significant drop in transcript expression. Only RNA from two cases was available for the RT-qPCR analysis; inclusion of a larger number of cases would be necessary to confirm this finding.

Although the western blot and immunohistochemical analysis indicate a considerable reduction in KCNIP4 protein expression in affected dogs, the exact epitope of the antibody used is unknown. The immunogen used by the manufacturer consisted of the majority of the protein's amino acid sequence, and the epitope has not been mapped. This means that the epitope could be in the region of the mutation, and the mutation could potentially have prevented antibody binding and thus mimicked an apparent fall in protein expression. However, the immunohistochemistry demonstrated that there was still some expression in a case, and that the expression pattern matched that seen

previously for KCNIP4 expression in humans and mice [322], and the variability in expression in the case match that of the Purkinje cell markers demonstrated in the histopathological analysis carried out previously [315]. This indicates that the antibody used is still capable of binding the protein containing the variant.

An additional band, at a greater atomic mass than the known KCNIP4 isoforms, was observed on the western blot. This additional band with a mass of ~37 kDa, which had a higher intensity in the two cases, could potentially represent an increase in expression of a different accessory Kv4 subunit, observable as a result of cross-reactivity of the antibody used. Loss of one of the KCNIPs has previously been observed to result in an increase in protein expression of the other, non-disrupted KCNIPs [334]. This compensatory mechanism is thought to occur through competition between the KCNIPs for binding to α-subunits [334, 335]. The stabilisation that occurs through binding to the other subunits results in an increase in protein levels.

The three main bands are observed for the control tissues in the western blot, at approximately 28 kDa, 26 kDa, and 22 kDa, are expected to represent the various isoforms of KCNIP4, which fit within this approximate size range (**Appendix ii.viii**). The band with the highest saturation in all controls, at 28 kDa, has a size closest to that expected of KCNIP4-1a. The lower two bands were observed to be fainter in the three controls which were affected by other forms of cerebellar ataxia, in comparison to those unaffected by ataxia. We propose that this difference could be a result of cerebellar degeneration, particularly for the Beagle and Hungarian Vizsla CCD cases, which are forms of ataxia characterised by the loss of Purkinje cells and depletion of the granular cell layer [62, 65]. The degradation of the cerebellum and cell types known to express the KCNIP4 protein could feasibly have resulted in the reduction in KCNIP4 expression observed in comparison to the unaffected controls. In comparison, the Norwegian Buhund cerebellar ataxia cases, which had two copies of the KCNIP4 variant but negligible cerebellar degeneration observed in the histopathological analysis [315], demonstrated almost complete loss of KCNIP4 protein expression in the western blot analysis, and dramatically reduced expression in the immunohistochemistry, indicating that in these individuals the loss of protein is caused by the mutation, not the cerebellar degeneration.

Human KCNIP genes have seven 3' exons (exons 3 to 9, **Figure 3.1A**) which are identical in length and have highly similar sequences between the four genes [322]. The 5' exons for the four genes are extremely dissimilar, and each KCNIP gene has multiple 5' exons which are unique. The high homology of the genes indicates shared protein function, and the conservation of the amino acid sequence of exons 3 to 9 suggests that variants in these regions are likely to be damaging to function. The variant identified in this study is located within exon 6, one of the highly conserved 3'exons, in a region of the protein which has identical amino acid sequences in the four canine proteins. The bioinformatics tools we have used show that the mutation in this conserved region affects the hydrophobic core of the protein, and predict it to be destabilising and damaging to function.

In addition to demonstrating that a variant within *KCNIP4* is associated with cerebellar ataxia in the Norwegian Buhund dog breed, we have characterised the gene's expression in the canine cerebellum. RT-PCR analysis and in situ hybridisation studies have demonstrated that *KCNIP4* is expressed in the cerebellum in humans and mice [322]. We have demonstrated that at least five *KCNIP4* transcripts are expressed in canine cerebellum, and that they share high sequence similarity with the human transcripts, all sharing the seven highly conserved 3' exons but with alternative 5' exons aligning with those seen in humans. A previous in situ hybridisation study in mice demonstrated that *KCNIP4* is expressed in the Purkinje cell layer and granular layer neurons of the cerebellum [322], and an immunohistochemical study in the rat showed protein expression throughout the cerebellum but particularly in the granular cell layer [336]. We demonstrate that KCNIP4 protein is also expressed strongly in the granular layer of the canine cerebellum. One transcript was only found in the kidney in humans [322]. Consistent with this, this transcript, KCNIP4-1cΔ2, was not identified in the canine cerebellum. However, the first exon of this transcript is a shortened version of the first of the 3' exons shared by all transcripts, which makes it impossible to assay using the methods used here; primers designed for this first exon would amplify all KCNIP4 isoforms. We were unable to locate the 5' untranslated region (UTR) for the KCNIP4-1cΔ2 transcript in canine cerebellum mRNA sequencing data.

In the immunohistochemical analysis we observed that expression of KCNIP4 in the canine cerebellum was predominantly within the granular cell layer, and KCNIP4

expression was strong throughout the synaptic glomeruli. Previous research has suggested a role for A-type potassium channels in the regulation of postsynaptic excitability of granule cell dendrites at the synapse between granule cells and mossy-fibre cells [336]. We suggest that a theoretical effect of the mutation described in the present study could therefore be hyperexcitability of postsynaptic granule cell membranes and/or receptor potentials being lost in the dendritic tree instead of travelling to granule cell soma and via parallel fibres to Purkinje cells. We theorise that the features observed in Purkinje cells [315] could be secondary to uncoordinated signalling from granule cells. Interestingly, an in situ hybridisation study in mice demonstrated that the KCNIPs that are expressed in the cerebellum, including KCNIP4, are expressed at a higher level in the rostral lobe when compared to the caudal vermis [322]. Our immunohistochemical findings showed the expression of KCNIP4 protein was lowest in the caudoventral vermis of cases when compared to the rostral lobe, and this is also the region where the previous histopathological findings showed the least expression of Purkinje cell differentiation marker proteins (calbindin and ITPR-1) [315].

A potential limitation of the present study is the small number of affected dogs included in the research; increasing the risk of a false positive finding. This is a result of the numerically small size of the Buhund breed, and the limited availability of samples from affected dogs. However, the availability of obligate carriers and other individuals from the extended pedigree moderates this as a limitation. Genotyping large numbers of unaffected breed matched controls, and multiple representatives of many different breeds, also minimised the effect of this limitation. We were also unable to breed-match when performing RT-qPCR, western blot and immunohistochemistry due to the lack of availability of relevant tissue, but this limitation was minimised by the use of control samples from dogs of 3-6 diverse breeds. Only a single control gene (TATA box binding protein (*TBP*)) was used for RT-qPCR normalisation, the use of multiple control genes would be more robust. The antibody used for western blot and immunohistochemistry was not validated in-house; testing it using a recombinant protein would confirm that it performs as expected. A potential limitation of the WGS analysis pipeline used here is that it would not have detected structural variants or insertions and deletions larger than approximately 75 bp. Given that we identified a compelling candidate causal variant in

our current pipeline, we did not further investigate larger structural or non-coding variants. However, both would require the use of gene candidacy to generate a realistic number of variants for follow-up. Future refinement of structural variant callers will enable their incorporation into pipelines for efficient analysis of WGS, with necessary harmonisation amongst datasets.

In summary, by whole genome sequencing two Norwegian Buhund siblings, and filtering against a bank of genomes of dogs of other breeds, we have identified a mutation associated with cerebellar ataxia in this breed. The mutation is in a gene, *KCNIP4*, not previously implicated in this disease in any species, and these findings could therefore inform research into inherited ataxia of unknown aetiology in humans. This research has led to the development of a DNA test for cerebellar ataxia in the Norwegian Buhund breed, allowing dog breeders to avoid producing affected dogs, reduce the allele frequency, and eventually eliminate the disease from the breed.

3.8 Methods

3.8.1 Ethics statement

Samples were collected from privately owned pet dogs from the general population, and samples from the UK were in the form of buccal swabs or residual blood samples taken as part of a veterinary procedure. This study was approved by the Animal Health Trust ethics committee (AHT06-09). EDTA blood samples (3 ml) from 40 Norwegian Buhunds donated to research were collected under the permission of animal ethical committee of County Administrative Board of Southern Finland (ESAVI/343/04.10.07/2016) and all experiments were performed in accordance with relevant guidelines and regulations and with owners' written consent.

3.8.2 Sample collection and DNA extraction

Samples from four Norwegian Buhunds diagnosed with cerebellar ataxia by veterinary neurologists at the Animal Health Trust Centre for Small Animal Studies, Newmarket, were included in this research. Buccal swabs and post mortem tissue samples (stored in RNAlater) were collected from two affected siblings diagnosed in 2008 (aged 12 weeks (0.2 years)). For the dogs diagnosed in 1998 and 2002 (aged 16 and 20 weeks,

respectively (0.3 and 0.4 years)) residual formalin-fixed paraffin-embedded (FFPE) tissue samples, collected for histopathology at the time of diagnosis, were used. Clinical descriptions and details of diagnosis have been described previously [315].

For validation of potential causal variants, samples from three sets of unaffected Norwegian Buhunds were utilised. The first set of DNA samples from 70 UK dogs were archived samples which had previously been collected for an unrelated study of inherited cataracts (unpublished) between 2008 and 2011, with the exception of one dog sampled in 2015 ("UK Set 1"). The dogs in this set were aged between 0.3 and 15.3 years (mean age 5.5 years) at sample collection, and none had been reported to have ataxia by their owners. Two sets of DNA samples were collected in 2017 from Norwegian Buhunds reported to have no signs of ataxia; one set from 36 dogs in the UK ("UK Set 2") and another from 40 dogs living in Finland ("Finnish Buhund Set"). Dogs in "UK Set 2" were aged 0.9 to 8.8 years (mean age 4.2 years), and in the Finnish Buhund Set were aged between 0.6 and 9.5 years (mean age 3.2 years).

DNA was extracted from buccal swabs using the QIAamp Midi Kit (Qiagen), or whole blood using a standard chloroform protocol (Cytiva Nucleon). DNA extraction from FFPE samples was carried out using Recoverall total nucleic acid isolation kit (Ambion). In the Finnish cohort, genomic DNA was extracted from the white blood cells using a semi-automated Chemagen extraction robot (PerkinElmer Chemagen Technologie GmbH, Baeswieler, Germany) according to the manufacturer's instructions. DNA concentrations were measured using Qubit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Nanodrop ND-1000 UV/Vis Spectrophotometer (Nanodrop technologies, Wilmington, Delaware, USA) and samples were stored at –20 °C.

3.8.3 RNA sequencing

Genome-wide RNA sequence analysis was carried out, using extant data generated as described previously [315]. Visual inspection of gene transcripts was carried out using the Integrative Genomics Viewer (IGV) [337].

3.8.4 Whole genome sequencing

Whole genome sequencing, including library preparation, of the two affected Norwegian Buhund siblings was carried out by Edinburgh Genomics laboratories, University of Edinburgh, using Illumina 150 bp paired-end sequencing (approximately 40X coverage). The library preparation method used was TruSeq DNA Nano (Illumina). Sequence reads were aligned to the canine reference genome CanFam3.1 using the Burrows-Wheeler Aligner (BWA-MEM), and SNP and in-del variants were called using the Genome Analysis Toolkit (GATK) Haplotypecaller (v3.6) using GATK best practices [338, 339]. Consequence predictions were designated for each variant using the Variant Effect Predictor (Ensembl), and variant calls for genomes of 44 unrelated dogs of 29 other breeds, were used to filter variants [340] (**Appendix ii.i**). The initial 44 control genomes, and the additional 140 that later became available for assessment of candidate variants, had been accrued over time for other research and as a resource.

3.8.5 Genotyping

The initial genotyping for candidate variants identified as potentially causal was performed using Sanger sequencing for eight SNPs, and fragment length analysis for a deletion, on an ABI 3130XL Genetic Analyzer. Primer sequences are available in **Appendix ii.x**. Hotstartaq plus (Qiagen) was used for the initial PCR prior to sequencing, using the following cycling conditions: 95°C for 5 minutes, followed by 35 cycles of 30 seconds each at 95°C and 57°C, and 72°C. Sanger sequencing used Bigdye terminator v3.1 ready reaction mix (Applied Biosystems), using standard cycling conditions: 96°C for 30 seconds, followed by 44 cycles of 4 seconds at 92°C, 4 seconds at 55°C and 1 minute 30 seconds at 60°C. Fragment length analysis was carried out using a three primer system, utilising a forward primer designed with a “tail” at the 5’ end and a fluorescently labelled (FAM) primer complementary to the “tail” on the forward primer [341]. Hotstartaq plus (Qiagen) was used for PCR, and cycling conditions were as follows: 94°C for 4 minutes, followed by 30 cycles of 30 seconds each at 94°C and 57°C, and 1 minute at 72°C. This was followed by eight cycles of 30 seconds each at 94°C and 50°C, and 1 minute at 72°C. The cycling was concluded with a 30 minute extension step at 72°C.

Further genotyping for the *KCNIP4* variant utilised an allelic discrimination method, using an ABI StepOne real-time thermal cycler. The assay primer and probe sequences are in **Appendix ii.xi**. The Kapa probe fast master mix (Kapa Biosystems) was used for genotyping all dogs, except those for which samples were collected as FFPE tissue. To optimise genotyping of the latter by limiting the levels of PCR inhibitors, DNA isolated from FFPE samples was further purified by ethanol precipitation before genotyping. FFPE derived DNA samples were genotyped using the Taqpath Proamp master mix (Applied Biosystems). For allelic discrimination assays using KAPA Probe Fast and DNA from buccal swabs or blood samples, a fast ramping speed was used and the following cycling conditions were used: 30 seconds Pre-PCR read at 25°C, 95°C holding stage for 3 minutes, 40 cycles of 95°C for 3 seconds and 60°C for 10 seconds, followed by a 30 second post-PCR read step at 25°C. Allelic discrimination assays using Taqpath Proamp used the standard ramping speed and the following cycling conditions: 60°C for 30 seconds pre-PCR read, 5 minute 95°C holding stage, 40 cycles of 15 seconds at 95°C and 1 minute at 60°C, and 60°C for 30 seconds post-PCR read step.

3.8.6 RT-qPCR

RNA was extracted from cerebellum samples collected post-mortem from the two affected Norwegian Buhund siblings and five other dogs unaffected by ataxia of different breeds (Labrador Retriever cross-breed, Siberian Husky cross-breed, Beagle, Golden Retriever, and a Great Dane) Elimination of genomic DNA and reverse transcription was carried out using the Quantitect cDNA synthesis kit (Qiagen). Assay primers and probe were designed for qPCR (using the PrimeTime qPCR Assay Design Tool, IDT) with the fluorescently labelled probe overlapping the boundaries of exons present in all alternative *KCNIP4* transcripts (Forward: GCCCAGAGCAAATTTACCAAG, probe: AAGAATGAGTGTCCCAGCGGTGT, reverse: CGGAAAGAACTGCGAGTAAATC). The qPCR was carried out using Luna Universal qPCR Master Mix (NEB) and an ABI StepOnePlus real-time PCR system, and comparative CT analysis used for relative quantification compared to an assay for the ubiquitously expressed *TBP* gene [342]. Reaction efficiency was calculated by performing a seven point, doubling dilution, standard curve. The reaction efficiencies of both the *KCNIP4* and *TBP* assays were 99.2%.

3.8.7 PCR and Sanger sequencing of transcripts

PCR primers were designed to be specific to each transcript, with a unique forward primer designed to overlap the boundary of the first and second exon and a reverse primer in the final exon which all transcripts share: GCATGGAGCGCATTATGTTT (**Appendix ii.xii**). Hotstartaq plus (Qiagen) was used for PCR of the cDNA produced by reverse transcription for the two Buhunds and five control dogs (see above), and cycling conditions were as follows: 95°C for 5 minutes, followed by 35 cycles of 30 seconds each at 95°C and 57°C, and 72°C. This was followed by final step at 72°C for 5 minutes. Agarose gel electrophoresis (1.5% agarose, 100v for one hour) was carried out using 3 µL of each product and images were taken. Sanger sequencing was carried out on the products for the two Norwegian Buhunds and Golden Retriever, as described above in section 3.8.5.

3.8.8 Western blot

Cerebellum tissue from the two cerebellar ataxia-affected Norwegian Buhund siblings and six control dogs were homogenised in radioimmunoprecipitation assay (RIPA) buffer (Cell Signalling Technology). The six control dogs included three dogs affected by cerebellar ataxia, but clear of the *KCNIP4* mutation and of breeds for which the putative causal mutation is known (one of each of Parson Russell Terrier, Beagle, and Hungarian Vizsla). The Parson Russell Terrier was diagnosed with late-onset spinocerebellar ataxia, and both the Beagle and Hungarian Vizsla with CCD, and all three were homozygous for the applicable variant for their form of ataxia [62, 63, 65]. The remaining three control dogs were unaffected by ataxia and clear of the mutation: one of each of Golder Retriever, Siberian Husky cross breed, Labrador Retriever cross-breed. Total protein lysate (15 µl, 180 - 255 µg) was diluted in Laemmli sample buffer (BIO-RAD) and separated by SDS-polyacrylamide gel electrophoresis using a 4 - 20% Mini-PROTEAN TGX Stain-Free precast gel (BIO-RAD). Wet transfer was used to transfer proteins onto a nitrocellulose membrane (0.45 µm). Immunoblotting was carried out using a rabbit monoclonal primary antibody targeting KCNIP4 (abcam, ab203831). The immunogen for the primary antibody used was a recombinant protein covering all of KCNIP4 (KCNIP4-1a, ENST00000382152.7), from the first amino acid to the C-terminus. Protein band detection utilised the WesternBreeze anti-rabbit chromogenic kit (Invitrogen) which uses a goat anti-rabbit antibody and

conjugated alkaline phosphatase for detection. Ponceau S stain was used as a total protein loading control.

3.8.9 Immunohistochemistry

Tissue expression of KCNIP4 was evaluated immunohistochemically in transverse and sagittal sections of cerebella of one affected Buhund sibling and three neurologically healthy control dogs (Labrador Retriever, Australian Shepherd, Jack Russell Terrier). The staining employed the same antibody as used for western blot (as described above), polymer technology and a diaminobenzidine tetrahydrochloride detection kit. Slides were counterstained with haematoxylin and routinely coverslipped using xylene-based mounting medium.

3.8.10 *In silico* protein analysis

In silico protein analysis was carried out using a 3D model (PDB ID 3DD4) obtained from the PDB-REDO Databank (pdb-redo.eu). The model had been refined and rebuilt from the original PDB model which had been determined through X-ray diffraction [333]. The model was visualised using UCSF Chimera software [343], and the two missing loops were modelled using Modeller [344]. Heteroatoms were removed from the PDB file, and a model of the protein containing the variant was created using Chimera's Rotamers tool, choosing the arginine rotamer with the highest probability in the Dunbrack library [345]. Clashes and contacts were predicted using Chimera's "Find Clashes/Contacts" tool which uses van der Waals radii to find interatomic clashes and contacts. The software's default clash and contact criteria were used. Clashes: van der Waals overlap >=0.6 angstroms, subtracting 0.4 for H-bonding pairs. Contacts: van der Waals overlap >=-0.4 angstroms.

3.9 Acknowledgements

We would like to thank all owners who provided samples from their dogs for this study.

We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G090074791070) for the generation of the sequencing data for the genome-wide mRNA of two Norwegian Buhunds, and the whole genomes of 24 dogs. The sequencing of the whole genomes of four dogs were carried out by the University of Cambridge. The whole

genomes of the two Norwegian Buhund cases and 155 other dogs were sequenced by Edinburgh Genomics laboratories, University of Edinburgh. The whole genome sequencing of 82 of the dogs used as controls was part of the Give a Dog a Genome project. We thank the Dog Biomedical Variant Database Consortium (DBVDC; Gus Aguirre, Catherine André, Danika Bannasch, Doreen Becker, Brian Davis, Cord Drögemüller, Kari Ekenstedt, Kiterie Faller, Oliver Forman, Steve Friedenberg, Eva Furrow, Urs Giger, Christophe Hitte, Marjo Hytönen, Vidhya Jagannathan, Tosso Leeb, Hannes Lohi, Cathryn Mellersh, Jim Mickelson, Leonardo Murgiano, Anita Oberbauer, Sheila Schmutz, Jeffrey Schoenebeck, Kim Summers, Frank van Steenbeek, Claire Wade) for sharing whole genome sequencing data from control dogs. We also acknowledge all canine researchers who deposited dog whole genome sequencing data into public databases.

3.10 References

See chapter 9

3.11 Supplementary tables and figures

See chapter 10

4 Manuscript 2 - Improving the resolution of canine genome-wide association studies using genotype imputation: a study of two breeds

This manuscript has been published in Animal Genetics

Full citation:

Jenkins CA, Dog Biomedical Variant Database Consortium, Schofield EC, Mellersh CS, De Risio L, Ricketts SL (2021) Improving the resolution of canine genome-wide association studies using genotype imputation: a study of two breeds. Animal Genetics.

4.1 Authors and affiliations

Christopher A Jenkins

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Dog Biomedical Variant Database Consortium

Dog Biomedical Variant Database Consortium (DBVDC): Gustavo Aguirre, Catherine André, Danika Bannasch, Doreen Becker, Brian Davis, Cord Drögemüller, Kari Ekenstedt, Kiterie Faller, Oliver Forman, Steven Friedenberg, Eva Furrow, Urs Giger, Christophe Hitte, Marjo Hytönen, Vidhya Jagannathan, Tosso Leeb, Hannes Lohi, Cathryn Mellersh, James R. Mickelson, Leonardo Murgiano, Anita Oberbauer, Sheila Schmutz, Jeffrey J. Schoenebeck, Kim M. Summers, Frank G. van Steenbeek, Claire Wade. The affiliations of the DBVDC members are listed in **Appendix iii.i**.

Ellen C Schofield

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Cathryn S Mellersh

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Luisa De Risio

Neurology / Neurosurgery Service, Centre for Small Animal Studies, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Current address: Linnaeus Veterinary Ltd, United Kingdom

Sally L Ricketts

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

4.2 Authors contributions

CAJ contributed to study design and funding, performed analysis and interpretation of data, and wrote the manuscript. SLR conceived of, designed, and supervised the study, and provided critical review of the manuscript. LDR contributed to sample collection and funding. CSM contributed to funding and provided critical review of the manuscript. ECS oversaw genome alignment and data curation and provided bioinformatics support. Members of the Dog Biomedical Variant Database Consortium provided WGS variant data and provided critical review of the manuscript.

4.3 Summary

Genotype imputation using a reference panel that combines high-density array data and publicly available whole genome sequence consortium variant data is potentially a cost-effective method to increase the density of extant lower-density array datasets. In this study three datasets (two Border Collie; one Italian Spinone) generated using a legacy

array (Illumina CanineHD, 173,662 SNPs) were utilised to assess the feasibility and accuracy of this approach and to gather additional evidence for the efficacy of canine genotype imputation. The cosmopolitan reference panels used to impute genotypes comprised dogs of 158 breeds, mixed breed dogs, wolves, and Chinese indigenous dogs as well as breed-specific individuals genotyped using the Axiom Canine HD array. The two Border Collie reference panels comprised 808 individuals including 79 Border Collies and 426,326 or 426,332 SNPs; and the Italian Spinone reference panel comprised 807 individuals including 38 Italian Spinoni and 476,313 SNPs. A high accuracy for imputation was observed, with the lowest accuracy observed for one of the Border Collie datasets (mean $R^2 = 0.94$) and the highest for the Italian Spinone dataset (mean $R^2 = 0.97$). This study's findings demonstrate that imputation of a legacy array study set using a reference panel comprising both breed-specific array data and multi-breed variant data derived from whole genomes is effective and accurate. The process of canine genotype imputation, using the valuable growing resource of publicly available canine genome variant datasets alongside breed-specific data, is described in detail to facilitate and encourage use of this technique in canine genetics.

4.4 Introduction

Genotype imputation is a computational method that predicts missing genotypes in a dataset of genotyped individuals, using a reference panel of individuals genotyped at a higher density [71, 72]. Imputation can enable meta-analyses of data generated using different arrays that include differing sets of SNP markers, and can increase the resolution of GWAS datasets by increasing SNP density and allowing inclusion of SNPs not genotyped on that array [346]. Genotype imputation is a well-established tool in human genetics, facilitated by the availability of large datasets of human genetic variation, such as the HapMap [347]; 1,000 Genomes Project [348]; and Haplotype Reference Consortium [349], that can be used as reference panels for imputation of GWAS array data [73]. The reference panels used can include a mixture of both population-specific panels and more divergent and cosmopolitan panels. An inclusive approach, using a reference panel with a composite of individuals closely related to the study population and individuals from other populations, can improve imputation accuracy [73].

Genotype imputation has also been established in other mammalian species, such as horse [350-353], cattle [354-356], pig [357, 358], and sheep [359, 360]. The feasibility of using genotype imputation in the domestic dog has also been demonstrated; examples include imputation from a theoretical very low-density array up to the commonly used Illumina CanineHD BeadChip array [77], and imputation from an array up to whole genome level (resulting in 4.9 million to 24 million variants) [76, 78]. Furthermore, genotype imputation has been shown to facilitate the identification of potentially novel loci for complex traits in dogs and the refining of intervals for known associated loci [78].

It has been demonstrated that to impute genotypes accurately in the dog a number of reference panel individuals specific to the breed of the dogs in the study set are required in combination with individuals of multiple other breeds [76]. Genome sequence consortia could be invaluable resources for this approach, particularly for the generation of a multi-breed reference panel [67, 69]. Such consortia have produced large variant datasets that are, or will become, publicly available. However, genome consortia datasets may include relatively few dogs of each breed, and many of the less common breeds may not be represented at all. Despite the decreasing cost of whole genome sequencing, generating a breed-specific component of a whole genome reference panel may be unfeasible for smaller studies. In recent years, however, a new higher density genotyping array for the canine genome has become available: the Axiom Canine HD array, which genotypes over 710,000 markers. Genotyping a set of breed-specific individuals using this array for use in a reference panel for imputation is comparatively cost-effective. Before the development of the Axiom Canine HD array, the 173,662 SNP Illumina CanineHD array had been used extensively for research since 2011 [3], meaning long running and ongoing studies often have extant datasets generated using this array. Applying genotype imputation to bring existing datasets up to marker densities comparable with the newer Axiom array could be an attractive way to utilise the wealth of data already available and increase the resolution and concomitant power of datasets.

There is still a need to build evidence for the optimum size of the breed-specific component of canine reference panels, and to examine how this may vary by breed. To date there is also a scarcity of literature outlining in detail the process of imputation in the dog, and to the authors' knowledge no publications describing in detail the

imputation of canine genotypes from the commonly used high density Illumina array up to the newer and increasingly utilised higher density Axiom array. This knowledge would be highly valuable to many researchers without the resources to generate large WGS datasets, with current WGS consortia containing only limited numbers of individuals of most breeds. This study intended to address these points and to provide further evidence towards a best practice method for accurate imputation in the dog.

The aim of the present study was to validate the use of genome-wide genotype imputation to impute extant Illumina CanineHD datasets up to the genotype density possible through the Axiom Canine HD array. Three Illumina datasets of two different breeds (two Border Collie and one Italian Spinone dataset) were imputed, assessing the effect of breed and reference panel size on imputation accuracy.

4.5 Materials and methods

The steps involved in preparing datasets for imputation, and the datasets used, which are described in detail below, are summarised in **Figure 4.1**. For this study, each reference panel was assembled using data from three datasets: a breed-specific dataset (either Border Collie or Italian Spinone) genotyped using the Axiom Canine HD array; and two sets of array marker data extracted from WGS datasets (one in-house WGS dataset including 186 dogs of multiple breeds, and a consortium (Dog Biomedical Variant Database Consortium, DBVDC) WGS dataset comprising 577 dogs of multiple breeds; 28 Chinese indigenous dogs, and eight wolves). These reference panels were used to impute Axiom genotypes in three study sets that had been genotyped using the Illumina CanineHD array ('Border Collie Set 1', 'Border Collie Set 2', and 'Italian Spinone') (**Figure 4.1**).

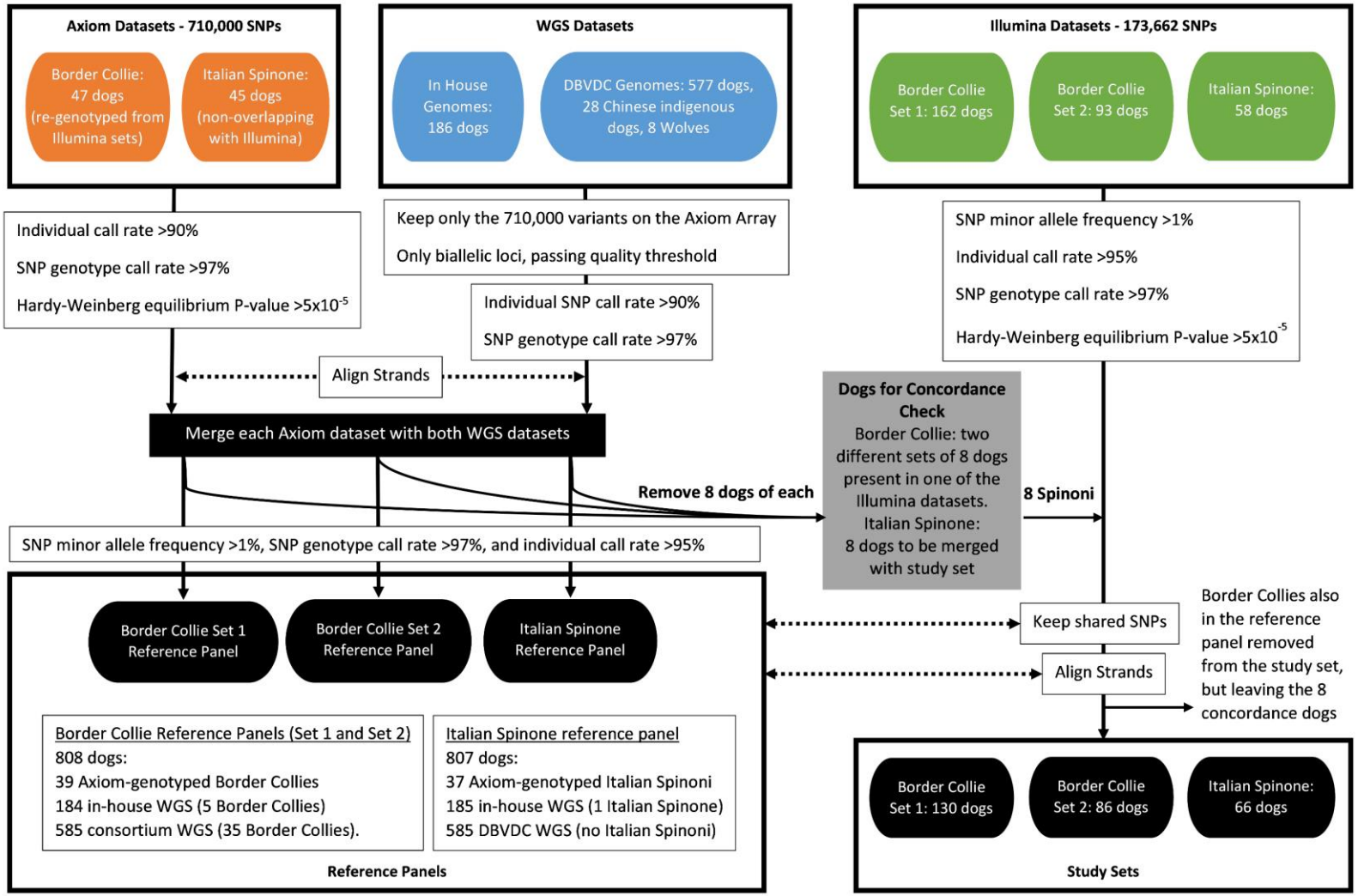


Figure 4.1. Flowchart to illustrate dataset processing for imputation study sets and reference panels.

4.5.1 Array-genotyped datasets for breed-specific reference panels

The Axiom Canine HD array genotype datasets, one each for the Italian Spinone and Border Collie breeds, were processed for quality control using the Axiom Analysis Suite and the Best Practices Workflow. Genotype data were available for 47 dogs (579,158 SNPs) in the Border Collie dataset, and 45 dogs (593,264 SNPs) in the Italian Spinone dataset (**Figure 4.1**).

4.5.2 WGS for the multi-breed reference panel

As stated above, two sets of WGS were used to make up the multi-breed component of the reference panel (**Figure 4.1**). The first set consisted of 186 in-house WGS of dogs, representing 93 breeds and five mixed breed dogs, accrued over time for other research and as a resource (average coverage >30X, lowest coverage 11X) (**Appendix iii.ii**). The second set was an international consortium (DBVDC) dataset that included sequence variant data for an additional 577 dogs (117 breeds, in addition to mixed breed dogs), eight wolves, and 28 Chinese indigenous dogs [67]. The genomes included in the DBVDC had an average of approximately 24X coverage, and a minimum of 10X coverage.

The Axiom Canine HD array SNPs were extracted from the two sets of WGS variant data using VCFtools (v0.1.15) [361] to allow the data to eventually be merged with the breed-specific Axiom array genotype data. A minimum quality score (minQ) was set to 20 to exclude genotypes with quality scores (Phred) below this threshold, and only biallelic loci were extracted. The output files produced by VCFtools were in PLINK ped and map format [55].

4.5.3 Aligning variant datasets from the Axiom array and WGS

Genotype data from both WGS datasets (in-house WGS dataset and DBVDC WGS dataset), and each Axiom dataset, were filtered using PLINK (v1.07) to exclude individuals genotyped for <90% of the SNPs, and to exclude SNPs that were called in <97% of individuals. Axiom datasets were also filtered to exclude SNPs with a Hardy-Weinberg equilibrium P-value $<5x10^{-5}$ (**Figure 4.1**). None of the datasets were filtered by MAF at this stage to retain as many SNPs as possible prior to merging.

The statistical software package Stata (Stata 15. College Station, TX, USA) was used to identify genotypes with strands that did not match between the datasets, and SNPs or variants which were insertions, deletions, or not biallelic across the datasets. Although only strand-flipped SNPs that were not between complementary bases (i.e. T/C, A/G) could be identified using this method, the small number that were found (Border Collie: n=78; Italian Spinone: n=91) indicated that the number of missed flipped SNPs is likely to be negligible. The identified insertions, deletions, and SNPs which were not biallelic were excluded, and the strands of the strand-flipped SNPs were aligned using PLINK (v1.07).

4.5.4 Merging the datasets to make a reference panel

For each of the two breeds, a combined reference panel was created using the Axiom array marker variants extracted from the two sets of WGS and the appropriate breed-specific Axiom canine HD array genotype dataset. To facilitate this, these three datasets were processed to keep only unique SNPs (i.e. removing SNPs within the same dataset that had different array IDs, but the same genomic position) that were present in all three (**Figure 4.1**).

4.5.5 Study sets

Two Border Collie GWAS sets ('Border Collie Set 1' included 162 dogs, 'Border Collie Set 2' comprised of 93 dogs) and one Italian Spinone set (58 dogs), all previously genotyped using the Illumina CanineHD array, were used in this research (**Figure 4.1**). The Border Collie Illumina GWAS sets were genotyped at different times and therefore retained as separate study sets to preserve data quality and account for any between-run variability, as good practice for downstream use of the data in GWAS meta-analysis [362]. Datasets were filtered to remove individuals with genotype call rates <95%, SNP call rates <97%, MAF <1%, and Hardy-Weinberg equilibrium P $<5x10^{-5}$. A more stringent individual genotype call rate was used, in comparison to the initial filtering of the reference panel datasets, for consistency across chromosomes, to prevent individuals from later being removed by the filtering carried out for each chromosome prior to haplotype phasing. Only SNPs present in the corresponding reference panel were retained (**Figure 4.1**).

4.5.6 Dogs for analysing genotype concordance and imputation accuracy

The two Border Collie Illumina study sets contained dogs (33 in Set 1, 14 in Set 2) that were re-genotyped on the Axiom array (47 total) and which would therefore be part of the reference panel. All except eight of these re-genotyped dogs (selected at random to be kept in for use in calculating imputation accuracy and genotype concordance) were removed from each study set (**Figure 4.1**). The two different sets of eight dogs for concordance calculations, one set for each Border Collie study set, were independently removed from the Border Collie reference panel. Each set of eight dogs was therefore present in one of the two original Illumina datasets, but there were no overlaps between each study set and its respective reference panel. This resulted in a different reference panel for each of the two Border Collie Illumina datasets (**Figure 4.1**). For the Italian Spinone there were no individuals that were present in both Illumina and Axiom datasets to use for assessing imputed genotype concordance. Instead, eight dogs genotyped using the Axiom array were selected at random to be excluded from the reference panel, filtered to leave only the SNPs present in the Illumina study set, and merged with this dataset (**Figure 4.1**).

4.5.7 Summary of the final reference panels

The pooled reference panels were filtered for SNP MAF <1%, SNP call rate <97%, and individual call rate (95%).

The final Border Collie reference panels were each comprised of 808 dogs: 39 Axiom-genotyped Border Collies, 184 in-house WGS (5 Border Collies), and 585 DBVDC WGS (35 Border Collies) (**Figure 4.1**). The Italian Spinone reference panel included 807 dogs: 37 Axiom-genotyped Italian Spinoni, 185 in-house WGS (1 Italian Spinone), and 585 DBVDC WGS (no Italian Spinoni). Each reference panel included dogs of 158 breeds, 12 mixed breed dogs, six wolves, and 28 Chinese indigenous dogs.

To investigate the relationship between the number of breed-specific reference individuals and accuracy, two additional reference panels were produced for Border Collie Set 1, one without the 35 DBVDC Border Collies (‘44 Border Collie Reference Panel’) and a second with half of the in-house WGS and genotyped Border Collies removed at random

('22 Border Collie Reference Panel'). The dogs were removed from the reference panel before filtering SNPs again as above.

4.5.8 Multidimensional scaling (MDS) plot of Set 1 Border Collies

To assess for the presence of any population stratification between the Axiom genotyped, in-house WGS, and DBVDC Border Collies; and the Illumina genotyped Border Collies; an MDS plot of Border Collies included in the Border Collie Set 1 reference panel and study set was generated using PLINK (v1.90). The data for only the Border Collies was extracted from the Border Collie Set 1 reference panel and filtered to keep only the 100,535 SNPs also present in the Border Collie Set 1 study set. The resulting dataset was merged with the study set. The MDS plot included 39 Axiom-genotyped Border Collies, five in-house WGS Border Collies, 35 DBVDC WGS Border Collies, and 130 Border Collie Set 1 study set dogs.

4.5.9 Aligning study set variant datasets with reference panel variant datasets

The strands of the Illumina study set genotype data needed to be aligned with that of the reference panel before imputation could be carried out (**Figure 4.1**). A considerable number of discrepancies were identified when comparing the Illumina strand annotations to those of the Axiom/WGS data. This could have been due to the Illumina CanineHD BeadChip probes being originally designed using the previous canine reference genome build BROADD2 whereas the Axiom Canine HD array and WGS were CanFam3.1. To identify all of the SNPs that needed to be strand flipped, flanking DNA information provided in the annotation documents for each of the two genotyping arrays was used. Ten bases of the upstream and downstream sequence for each of the SNPs were extracted from the annotation file and were compared between arrays. The strands of the study set SNPs that were not on the same strand between datasets were aligned (**Figure 4.1**).

4.5.10 Haplotype phasing and imputation

The reference panel and study sets were split by chromosome for haplotype phasing and imputation. Only the autosomes were used for imputation. Each individual in the reference panel and study set needed to pass a genotype rate threshold of 90% for each

chromosome. Three individuals (originally part of the DBVDC WGS set) were excluded for the chromosome 9 (CFA 9) reference panel because they failed to pass this threshold.

The Border Collie Set 1 reference panel included 426,326 SNPs; and the Border Collie Set 2 reference panel included 426,332 SNPs. The Italian Spinone reference panel contained 476,313 SNPs. In the reduced Border Collie Set 1 reference panels, the number of SNPs were: 44 Border Collie Reference Panel, 426,235 SNPs; 22 Border Collie Reference Panel, 426,154 SNPs.

Haplotype phasing of reference panels and study sets was carried out using SHAPEIT (v2, r904) [363]. Genotype imputation was carried out using IMPUTE2 (IMPUTE v2.3.2) [72, 73]. A publicly available canine genetic map was used for haplotype phasing and imputation [38]. A window size of 2 Mb was used for haplotype phasing, and the effective population size was set at 200 for both phasing and imputation [76].

4.5.11 Analysis of imputed genotypes

To assess accuracy of imputed genotypes, the predicted allele 'dosage' produced by imputation was compared to the 'known' genotypes in the array data for eight different dogs from each study set. After exclusion of the observed Illumina array genotypes, the squared Pearson correlation coefficient (R^2) was calculated for each individual to give an indication of accuracy for each chromosome. Genotype concordance (%) was also calculated after converting the allele dosages provided by IMPUTE2 to binary genotypes using PLINK (v1.90) (calls with uncertainty >0.1 were called as missing).

IMPUTE2 produces a metric, called Info, for each SNP that describes the reliability of the imputed genotypes. An Info score is a value typically between 0 and 1, with scores closer to 1 indicating greater certainty. The Info scores were split into 10 groups to allow visualisation of the data and comparison with previous studies, and the concordance of the SNPs with known heterozygous or homozygous genotypes in the eight dogs were analysed.

4.6 Results and discussion

4.6.1 Imputation accuracy and concordance, and comparison with previous studies

After filtering the SNPs as would typically be carried out for a GWAS (Hardy-Weinberg $P<5x10^{-5}$, call rate <97%, MAF <5%) the number available for analysis was on average (mean) three times higher than that of the study set (**Table 4.1**). This increase in SNP number and therefore density would be expected to reduce the gaps between genotyped SNPs, increasing the likelihood of a SNP tagging a risk-conferring variant in a GWAS (dependent on local LD structure). This also allows meta-analysis with data genotyped on the higher density Axiom array, without sacrificing a large proportion of the available data. However, the number of imputed SNPs is limited by the number within the reference panel, which is dependent on the allele frequencies within the breed. This can be seen clearly when comparing the relative sizes of the Border Collie and Italian Spinone reference panels and the number of SNPs in the resulting imputed datasets (**Table 4.1**).

Table 4.1. SNPs in each dataset before and after imputation.

Dataset	Study Set Dogs (n)	Study Set SNPs (n)	Total SNPs After Imputation (n)	SNPs Passing Quality Control* (n)
Border Collie Set 1	130	100,535	426,326	310,617
Border Collie Set 2	86	105,443	426,332	310,300
Italian Spinone Set	66	104,432	476,313	341,854

* Hardy-Weinberg $P>5x10^{-5}$, call rate >97%, MAF >5%

Across the three imputed datasets, genotype dosages produced were highly correlated (>0.94) with the known genotypes provided by the array (**Table 4.2**). After conversion of the predicted dosages to binary genotype format, the percentage of genotypes concordant between the imputed data and array data was high (≥96.9%), demonstrating that genotype imputation was very accurate for all three datasets (**Table 4.2**). The concordances observed for the three sets imputed in this study are higher than that observed in a previous study also using IMPUTE2 but a smaller multi-breed reference panel to impute genotypes in Standard Poodles up to whole genome level (94.1%), and comparable to the same study's results for the Boxer when using different software for imputation (Beagle 4.0, [364])(97.8%) [76]. This previous study used a reference panel

with a multi-breed component of 63 dogs representing 14 different breeds, and 19 breed-specific dogs (Standard Poodles or Boxers depending on the study set). When the breed-specific dogs were excluded from the study's reference panel, or only dogs of other breeds were included, accuracy dropped. The present study utilised reference panels of over 800 dogs from 158 breeds (including breed-specific dogs), and accuracy was high for both Border Collies and Italian Spinoni. The inclusion of individuals in reference panels from other populations not matched to the study set (in addition to population-matched individuals) has also been shown to be effective for achieving optimum accuracy in the imputation of genotypes in human studies, by improving imputation of alleles less common in the study population which may be poorly represented in population-matched individuals [73].

Table 4.2. Imputation accuracy across the three study datasets.

Dataset	Mean R^2	Genotype Concordance (%)	Worst Chr (Mean R^2)	Best Chr (Mean R^2)	Individual Lowest R^2*	Individual Highest R^2*
Border Collie Set 1	0.94	96.9	CFA 6 (0.92)	CFA 5 (0.96)	0.89	0.98
Border Collie Set 2	0.96	97.7	CFA 21 (0.94)	CFA 23 (0.97)	0.93	0.97
Italian Spinone Set	0.97	98.2	CFA 36 (0.95)	CFA 7 (0.98)	0.94	0.99

*Lowest or highest R^2 observed in an individual dog

The concordance for the three sets in the present study was also higher than the highest concordance observed (92.7%) in another study that imputed genotypes of multiple dog breeds up to whole genome level using a multibreed reference panel of 365 WGS that included minimal (between 10 and 16) breed-specific dogs [78]. This highlights again the importance of breed-specific individuals in reference panels for canine genotype imputation accuracy. Including population-matched individuals has been demonstrated to be important for the accuracy of imputation of genotypes in human studies. Similarly, increasing the number of breed-matched individuals in reference panels can improve imputation accuracy in cattle [355].

Both of the two aforementioned canine studies [76, 78] imputed from the Illumina CanineHD array or a comparable array up to whole genome level, whereas the present study imputed up to the Axiom array, a comparatively lower proportion of SNPs. It is possible that imputing a greater proportion of SNPs increases error rate. However, previous work has suggested that it is the density of the known SNPs (the number of existing genotypes) in the study set that has the greatest impact on accuracy, not the number of missing SNPs that need to be imputed to bring the study set up to the size of the reference panel [77, 365]. It could be that studies imputing to whole genome level impute a greater proportion of SNPs with low MAF. Alleles with the lowest frequencies are well established as having a reduced accuracy when imputed, particularly for heterozygous loci [73, 77, 78].

4.6.2 Variation in imputation accuracy across chromosomes and study individuals

Accuracy was moderately consistent across autosomes, although some variation was observed (**Table 4.2**, **Figure 4.2**). There was no correlation between chromosome size and imputation accuracy in this or a previous study [76]. However, a correlation between accuracy and chromosome size was seen in the other study that imputed up to genome level [78]. Imputation accuracy was also variable across individuals (**Table 4.2**, **Figure 4.3**). Border Collie Set 1 showed the biggest difference in mean R^2 values between the individuals (and, to a lesser extent, chromosomes) with the highest and lowest accuracies.

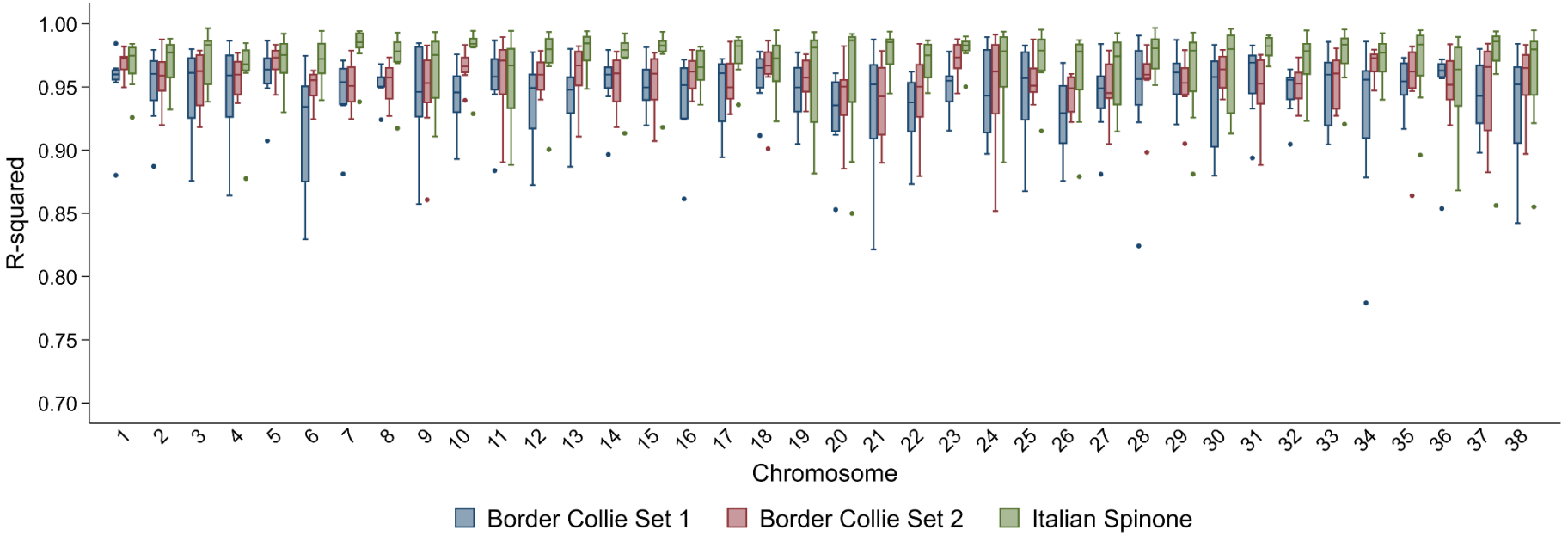


Figure 4.2. Accuracy of imputation for each chromosome in Italian Spinone and Border Collie datasets. The graph shows the R^2 of imputed calls and known genotypes. Boxes are 25th to 75th percentiles, with lines for the median. Whiskers indicate upper and lower adjacent values; outliers are shown using dots. Truncated y-axis starts at 0.7.

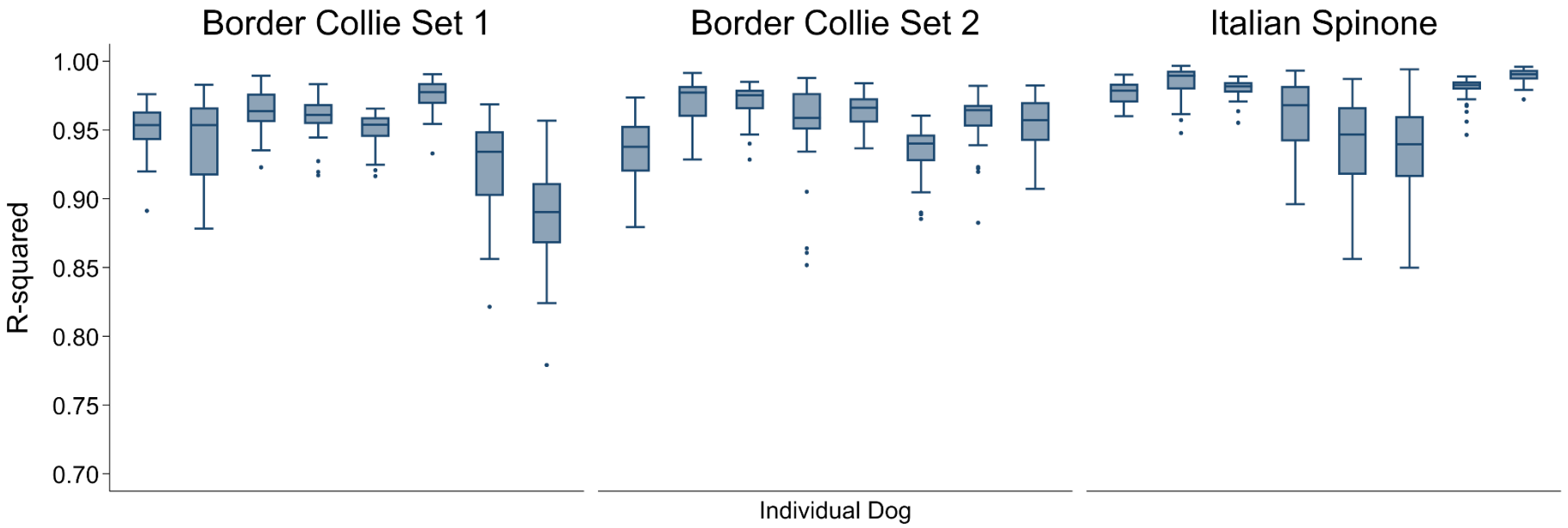


Figure 4.3. Accuracy of imputation for each concordance-tested individual (n = 8 for each set) in Italian Spinone and Border Collie datasets. The graph shows the R^2 of imputed calls and known genotypes. Boxes are 25th to 75th percentiles, with lines for the median. Whiskers indicate upper and lower adjacent values; outliers are shown using dots. Truncated y-axis starts at 0.7.

4.6.3 Study-specific differences and the effect of reducing the number of breed-specific reference panel individuals on imputation accuracy

Border Collie Set 1 had the lowest imputation accuracy, and the highest accuracy was observed for the Italian Spinone dataset (**Table 4.2**, **Figure 4.2**), despite the Italian Spinone reference panel including only 38 breed-specific dogs, whereas the Border Collie reference panels contained more than double the number (79 Border Collies). This indicates that the relationship between accuracy and the size of the breed-specific component of the reference panel reaches a plateau, and that other factors also have a role. To test this hypothesis, Border Collie Set 1 was imputed using two other reference panels: one without any of the DBVDC Border Collies ('44 Border Collie Reference Panel'), and one with half of the remaining Border Collies ('22 Border Collie Reference Panel')(**Figure 4.4**). The 44 Border Collie Reference Panel did not materially reduce imputation accuracy (R^2 = 0.94) (**Figure 4.4**). Using the 22 Border Collie Reference Panel had a greater effect, bringing the accuracy down to R^2 = 0.92 (**Figure 4.4**). This suggests that above 44 breed-specific dogs in the reference panel, imputation accuracy plateaued for the Border Collie, and that other factors caused this dataset to be imputed at a lower accuracy than the Italian Spinone set. The multi-breed reference panel used in this study included more dogs from more breeds than those described for previous studies [76, 78], therefore it is possible that the large number and diversity of haplotypes present limited the effect of reducing the number of breed-specific dogs on accuracy. Since differences between the levels of inbreeding and LD in the Border Collie and Italian Spinone breeds could also be contributing to some of the variation in accuracy observed; future work could compare imputation accuracy across many different breeds when using the same sized reference panel. The reduced accuracy in Border Collie Set 1 when compared to Border Collie Set 2 suggests differences in the sample populations or potentially lower DNA quality and therefore reduced genotype reliability in Set 1.

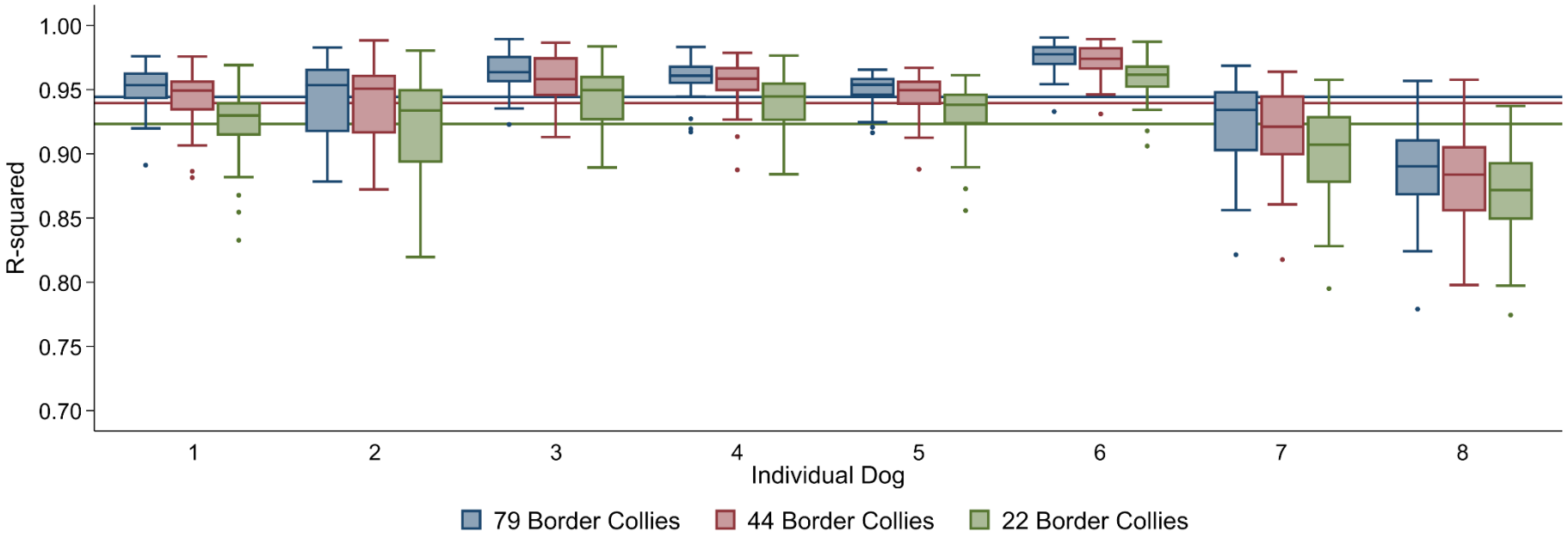


Figure 4.4. Accuracy of imputation for each concordance tested dog from Border Collie Set 1 and each of three reference panels containing decreasing numbers of Border Collies. The graph shows the R^2 of imputed calls and known genotypes. Boxes are 25th to 75th percentiles, with lines for the median. Whiskers indicate upper and lower adjacent values; outliers are shown using dots. Lines show mean R^2 for each reference panel. Truncated y-axis starts at 0.7.

A study of imputation in sheep showed that including more closely related individuals in the reference panel can improve imputation accuracy [359] and previous research has indicated that including related individuals can also increase accuracy in the dog [77], although the effect seen was minimal. The Border Collie breed is numerically much larger than the Italian Spinone, and the dogs included in the reference panel are therefore likely to be less closely related to those in the study set. The DBVDC is an international consortium, and the consortium Border Collies could therefore be expected to originate from populations less closely related to the study set, which were predominantly UK dogs, compared to the dogs used for array genotyping or WGS in the UK, which were also predominantly UK dogs. This could also partially explain why removing these dogs had only minimal effect on accuracy. To examine this, an MDS plot of Set 1 Border Collies (reference panel and study set) was generated using SNP data common to all four datasets (Axiom-generated Border Collies; in-house and DBVDC WGS-derived Border Collies; Illumina-genotyped Border Collie Set 1) (**Appendix iii.iii**). This demonstrated that the reference panel captures the study-set individuals effectively, and in particular that the combination of the Axiom and WGS-derived reference panels appears to give the greatest coverage of individuals. However, as the majority of the DBVDC individuals cluster with a close group (**Appendix iii.iii**), it may be that the limited number of haplotypes in this group means that removing the DBVDC Border Collies had a smaller effect than removing a further 22 dogs which may have been more distributed. Future research that examines imputation accuracy in breeds with known differences between geographical populations, such as the Retriever breeds [366, 367], would help to elucidate this.

Differences in the approaches used to calculate accuracy between the two breeds could also explain some of the differences observed. The dogs used to calculate concordance in the Italian Spinone dataset had been genotyped on the Axiom array before being filtered to keep only Illumina array SNPs before imputation. This created an artificial low-density dataset. By contrast, the Border Collies used to calculate concordance had been genotyped on both arrays, and the Illumina dataset imputed. Differences between accuracy of arrays, and errors in genotype calls when retesting, introduced discrepancies between the Border Collie datasets, whereas the Italian Spinone concordance dogs had

identical genotypes between the reference panel and artificially created Illumina study set dogs. This means that accuracies are not directly comparable, although it does give an indication of the real differences.

4.6.4 Imputation accuracy stratified by IMPUTE2's imputation certainty ('Info') metric

The accuracy of imputation across the range of the 'Info' statistic, split into 10 'Info groups', was assessed. The concordance of homozygous SNPs was consistently high across the Info groups, but heterozygous genotypes had a low concordance in the lower Info groups (**Figure 4.5**), consistent with earlier canine research [76]. Most SNPs fell within either the very lowest Info group or the higher Info groups, which is also similar to previously published findings [76]. When the grouped Info scores were compared to the expected allele frequency provided by the IMPUTE2 software, a positive trend was observed (**Appendix iii.iv**), however, this was skewed by the lowest and highest Info score groups containing the majority of the SNPs with low frequency alleles (**Appendix iii.v**). The Info metric produced by IMPUTE2 can be used to filter the imputed SNPs to remove those for which there is a lower imputation certainty. The results from this study indicate that the optimum threshold to use for filtering by Info will vary depending on the breed of dog in the dataset imputed. A higher threshold might be necessary for the Border Collie, compared to the Italian Spinone, to ensure highest accuracy without excluding too many useful SNPs (**Figure 4.5**). However, the majority of the SNPs with lower imputation certainty will be filtered out of downstream GWAS analyses by MAF (**Appendix iii.iv** and **Appendix iii.v**).

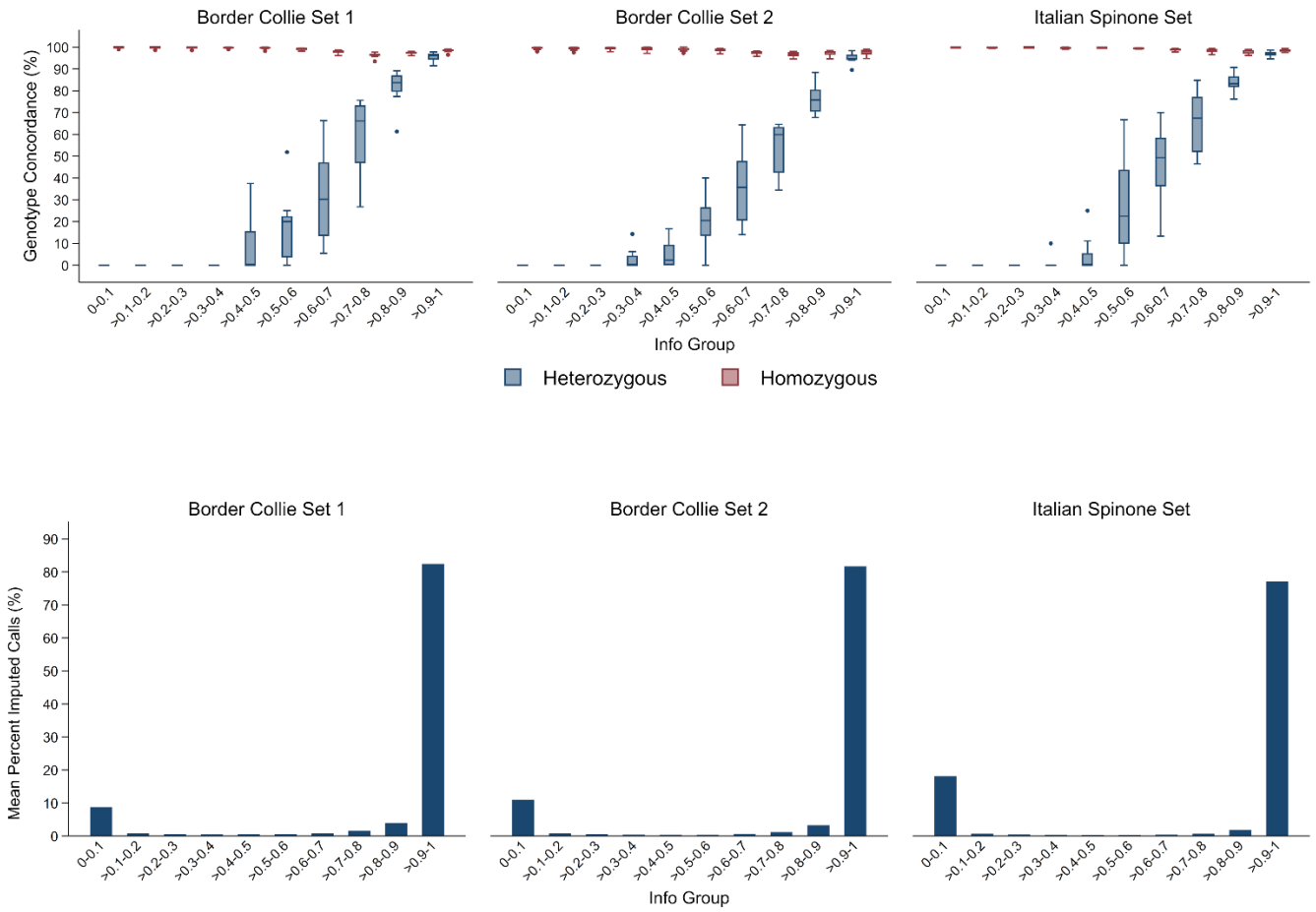


Figure 4.5. A comparison of imputation accuracy and predicted certainty. Top: percent of concordant genotypes for SNPs with heterozygous or homozygous known genotypes grouped by IMPUTE2's Info metric (imputation certainty). Bottom: percent of total imputed calls within each Info group. Data for eight dogs were included for each study. Total imputed SNPs: Border Collie Set 1 n = 325,791, Border Collie Set 2 n = 320,889, and Italian Spinone n = 371,881.

4.6.5 Conclusions

This research has demonstrated and described in detail the successful use of imputation to bring the SNP density of the commonly used Illumina array closer to that of datasets generated using the newer higher-density, and increasingly used, Axiom array. This represents a cost-effective method to make the most use of extant data, without the need to re-genotype all individuals or generate large WGS datasets as would be necessary for imputation up to the density of WGS, which has been the predominant focus of previous literature in the canine field. The present study demonstrates that in-house and publicly available consortium WGS variant datasets can be used to produce multi-breed reference panels large and diverse enough to enable accurate genotype imputation of canine GWAS datasets. This work contributes to building best practice evidence for the optimum size of the breed-specific component of canine reference panels, demonstrating that increasing the number of breed-specific dogs improves accuracy, and providing some initial evidence for the upper threshold after which adding more dogs may have a limited effect. Although the number of breed-specific dogs required may vary significantly between breeds, our analysis of the Border Collie has shown that effective imputation can be carried out in a genetically diverse and numerically large breed using a modest number of breed-specific dogs in the reference panel. As well as investigating imputation in additional breeds, including those with distinct geographically isolated populations, it will be important for future applications to examine regions of gene complexity, such as the major histocompatibility complex (MHC), where imputation accuracy may be highly variable across breeds.

4.7 Acknowledgements

The Petplan Charitable Trust funded this study in the Border Collie (project number S18-688-726). Generation of datasets for the Italian Spinone were funded by the Italian Spinone Club of Great Britain. CAJ, SLR, CSM and ECS are supported by the Kennel Club Charitable Trust.

We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the Axiom array data. Illumina genotyping was conducted by GeneSeek (Neogen Genomics, Lincoln, NE, USA). We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant

reference 090532/Z/09/Z and MRC Hub grant G090074791070) for the whole genomes of 24 dogs. The sequencing of the whole genomes of four dogs was carried out by the University of Cambridge. The whole genomes of 157 other dogs were sequenced by Edinburgh Genomics Laboratories, University of Edinburgh. The whole genome sequencing of 82 of the dogs was part of the Give a Dog a Genome project. We also thank Natasha Olby for sharing whole genome data. We acknowledge all canine researchers who deposited dog whole genome sequencing data into public databases. We thank Dr Janine Lamb, University of Manchester, for critical review of the manuscript.

We would like to thank all owners who provided samples from their dogs, whose data contributed to this study.

4.8 Availability of data

The Italian Spinone and Border Collie Axiom Canine HD array genotype datasets, and the Axiom Canine HD array SNPs extracted from the in-house WGS data, are available for download (doi:10.5061/dryad.prr4xgxkj). The DBVDC variant data are available at the European Variation Archive, project ID PRJEB32865 (https://www.ebi.ac.uk/eva/?eva-study=PRJEB32865).

4.9 References

See chapter 9

4.10 Supplementary tables and figures

See chapter 10

5 Manuscript 3 - A genome-wide association study of paroxysmal dyskinesia in the Norwich Terrier breed of dog

This manuscript has been written in a format to be submitted to Animal Genetics as a short communication.

5.1 Authors and affiliations

Christopher A Jenkins

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Luisa De Risio

Neurology / Neurosurgery Service, Centre for Small Animal Studies, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Current address: Linnaeus Veterinary Ltd, United Kingdom

Elisabeth Dietschi

Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Tosso Leeb

Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Ulrich Rytz

Department of Clinical Veterinary Medicine, Division of Small Animal Surgery, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Peter Schawalder

Department of Clinical Veterinary Medicine, Division of Small Animal Surgery, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Jeff Schoenebeck

The Roslin Institute and Royal (Dick) School for Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom

Cathryn S Mellersh

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Sally L Ricketts

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

5.2 Author contributions

CAJ contributed to study design and funding, performed experimental work and analysis and interpretation of data, and wrote the manuscript. SLR conceived of, designed, and supervised the study, contributed to funding, and provided critical review of the manuscript. LDR provided case definition and diagnosis of cases and clinical expertise, contributed to funding, and provided critical review of the manuscript. CSM contributed to sample collection and provided critical review of the manuscript. TL and JS provided genotype data for an independent set of Norwich Terriers and provided critical review of the manuscript, UR and PS contributed to sample collection, and ED contributed case and control status for a subset of these dogs and provided critical review of the manuscript.

5.3 Summary

Paroxysmal dyskinesia is a neurological disorder in the Norwich Terrier dog breed, with unknown aetiology and a high prevalence, making it a disease of considerable concern for the breed. Previous research has suggested a possible genetic cause. Genome-wide association studies of modest size have had some success in identifying the probable genetic cause of autosomal recessive paroxysmal dyskinesias in other dog breeds. A

preliminary genome-wide association study of paroxysmal dyskinesia was conducted in the Norwich Terrier using 24 cases and 24 controls over the age of six years. Whilst this analysis did not yield any genomic regions reaching genome-wide statistical association, analysis of the genotypes of 44 SNPs with suggestive association (P-value < 1×10^{-3}) in an independent dataset of 232 Norwich Terriers, of which 10 were cases and 45 controls, identified five genomic regions, and genes, for potential future investigation.

5.4 Main text

5.4.1 Background

Paroxysmal dyskinesias (PxDs) are typified by recurring episodes of involuntary abnormal movement without loss of consciousness [101, 188]. PxD is phenotypically heterogeneous in dogs, but characteristics are often shared between breeds and with human PxD [187, 188]. There are canine PxDs with known or suspected single gene genetic causes [189-193]; PxDs considered likely to be inherited [120, 121]; and examples of PxDs influenced by diet or which are drug-induced [194-196]. A pedigree analysis in a previous study of PxD in Norwich Terriers showed clustering of cases and suggested an inherited component, although mode of inheritance could not be ascertained [120]. A PxD prevalence of 13% (95% confidence interval 9% - 18%) was estimated in these Norwich Terriers in the UK, although the questionnaire-based breed-wide study was likely to have a bias towards affected dogs. Typical PxD episodes in the Norwich Terrier are characterised by dystonia of the pelvic limbs, and often the thoracic limbs and trunk, causing difficulty walking and unusual posture [120]. Dogs are neurologically normal between episodes and diagnostic investigations are unremarkable. In about half of cases episodes are triggered by stress, anxiety, excitement, or changes to the dog's normal daily routine [120], suggesting that the disease could be categorised as a paroxysmal nonkinesigenic dyskinesia (PNKD) [188]. Episode duration ranges from less than two minutes to 30 minutes, and the frequency of occurrence has a broad range, from twice a day to only twice a year [120]. Age of onset is also variable, ranging from 0.8 to 5.4 years (median age of onset is three years). Similar to canine epileptoid cramping syndrome (CECS) in Border Terriers [194, 195, 198], PxD in Norwich Terriers is not treated effectively with anticonvulsants and there are anecdotal reports that changing the dog's diet can be beneficial in some cases [120].

In the present study a preliminary genome-wide association study (GWAS) was undertaken, aiming to identify PxD-associated loci in Norwich Terriers and help to assess the probable mode of inheritance of the disorder. Subsequently, the SNPs with the strongest association with PxD were investigated for replication in an independent Norwich Terrier dataset generated for a separate study.

5.4.2 Sample collection and DNA extraction

The study design and sample collection protocols were approved by the Animal Health Trust Clinical Research Ethics Committee (Project No. 35-2017) and University of Cambridge Department of Veterinary Medicine Ethics and Welfare Committee (No. CR527), and samples were obtained following owner informed consent. The GWAS individuals were 24 Norwich Terrier PxD cases and 24 PxD-unaffected Norwich Terriers over the age of six years. All dogs included in the GWAS were living in the United Kingdom at the time of sample submission. All except two of the cases were identified through the previously described questionnaire-based breed-wide study [120]; the remaining two were diagnosed at the Animal Health Trust Neurology unit. The same veterinary neurology specialist (LDR) reviewed the case details of all affected Norwich Terriers. Controls were Norwich Terriers with no owner-reported episode suggestive of PxD or other episodic abnormality of muscle tone, posture and movement and were a minimum age of six years. The six years minimum age threshold for controls was based on the oldest age of onset for an affected Norwich Terrier observed in the previous breed-wide survey (5.4 years) [120]. Where possible, the controls were selected so that each PxD-affected case was matched by at least one related (first, second, or third-degree relative) control. This was to overcome the marked population stratification observed in a GWAS conducted previously in our laboratory (unpublished data). Pedigree information was not available for two of the affected dogs. Samples were in the form of buccal swabs or residual blood samples drawn as part of a veterinary procedure. DNA isolation was carried out for swabs using a QIAamp DNA Blood Midi Kit (Qiagen) and a standard chloroform extraction protocol for blood samples. One sample (a case) with an insufficient DNA quantity was whole-genome amplified using the REPLI-g Single Cell kit (Qiagen).

5.4.3 GWAS

Of the canine PxDs for which the probable underlying genetic cause is known [189-191, 193], PxD in the Soft-Coated Wheaten Terrier is phenotypically most similar to that of the Norwich Terrier. Prior to commencing the GWAS, the putative causal variant (NC_006583.3:g.14705240C>T, *PIGN*:c.398C>T) was genotyped in Norwich Terriers affected by PxD, seizures, or similar clinical signs, including the 24 PxD cases to be included in the GWAS. The variant was not present in the sample set.

Genotype data were generated for the 48 dogs using the Axiom Canine HD array (up to 729,642 SNPs and other DNA variants). The dataset was filtered using PLINK (v1.90) [368]; filtering for individuals with a genotype call rate of under 90% did not exclude any dogs. Population-corrected association analysis (using a linear mixed model) was carried out on the autosomes using GEMMA [54], including only the 230,972 variants with call rate > 97%, minor allele frequency > 0.05, and Hardy-Weinberg equilibrium P-value > 5×10^{-5}. None of the variants passed the Bonferroni-corrected threshold for genome-wide significance ($P \leq 2 \times 10^{-7}$) (**Figure 5.1**). As this was a small preliminary study that was likely to lack statistical power to detect associations for a non-single gene disease, to investigate if any of the strongest-associated variants represented genomic regions with a replicable association with PxD, a threshold of $P \leq 1 \times 10^{-3}$ for suggestive association was implemented for selecting variants for further analysis (**Figure 5.1**). This identified 47 SNPs across eight chromosomes.

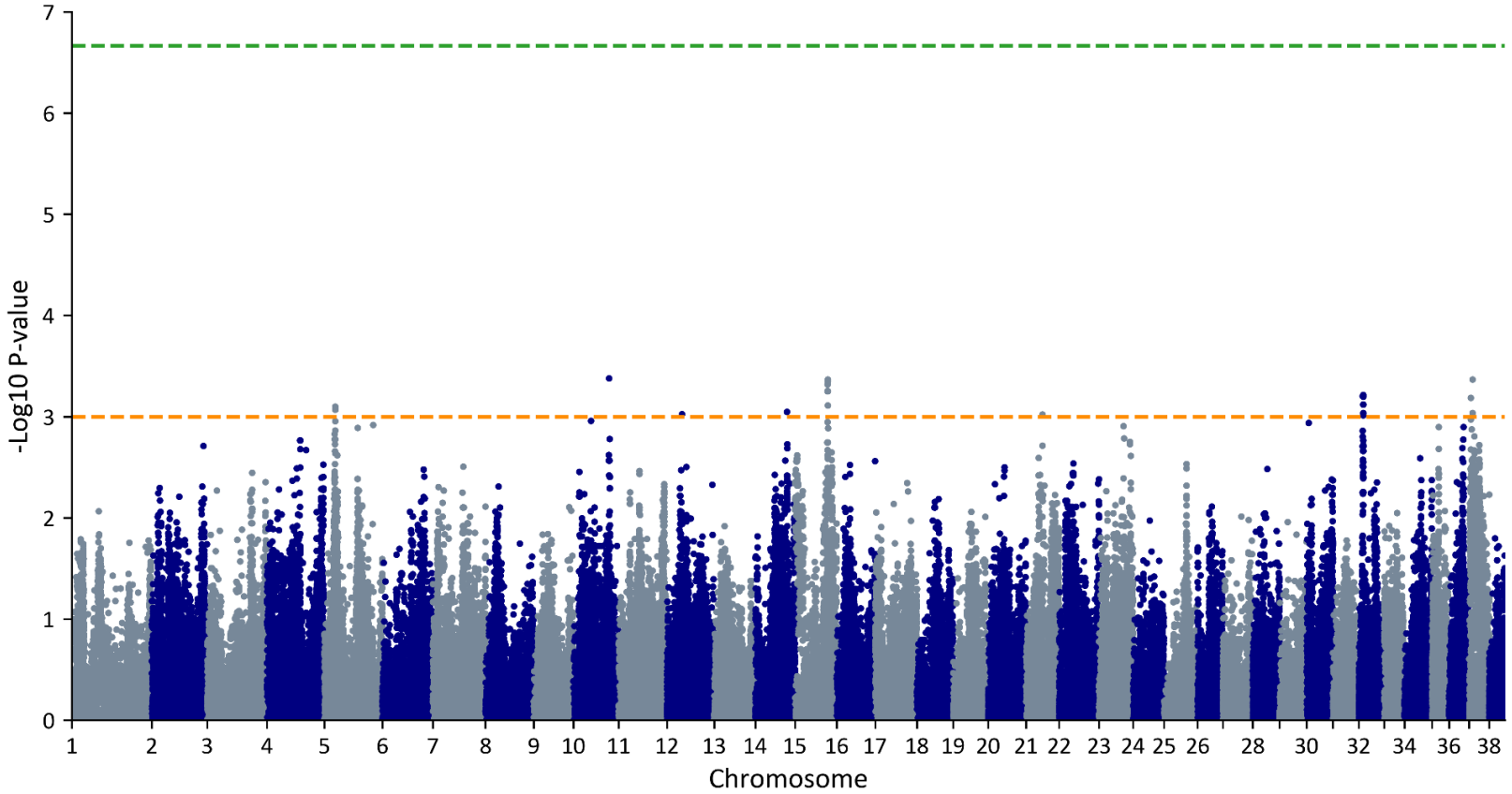


Figure 5.1. Genome-wide association analysis of 24 PxD-affected Norwich Terrier dogs and 24 controls (230,972 SNPs). Plot of negative log (base 10) transformed P-values. X-axis is SNP location by chromosome (left to right, autosomes 1 to 38). Green (upper) line shows Bonferroni-corrected threshold for statistical significance ($P \leq 2 \times 10^{-7}$). Orange (lower) line indicates the empirical threshold for suggestive association ($P \leq 1 \times 10^{-3}$). Genotyped using the Axiom Canine HD array.

5.4.4 Replication in an independent dataset

An independent Norwich Terrier SNP genotype dataset was used for an initial replication study. This publicly available dataset was generated for a GWAS of Norwich Terrier upper airway syndrome [369]. Sample and genotyping details are provided in the published manuscript [369]. The full dataset was initially treated as a population-based control sample of the Norwich Terrier breed but PxD case/control status was subsequently available for a subset of the dogs. Cases (n = 10) were defined in this set as dogs reported by the owner to have had multiple episodes; two of these were reported to have been diagnosed with PxD by a veterinary neurologist. Controls (n = 45) were dogs over the age of seven years not reported to have had an episode suggestive of PxD.

The replication dataset was imputed, using SHAPEIT (v2, r904) [363] for haplotype phasing and IMPUTE2 (IMPUTE v2.3.2) [72, 73] for imputation, using a methodology described in detail previously [370] (see chapter 4). The aim of imputation was to bring the SNP density up to that of the discovery GWAS set and allow analysis of the most significantly associated GWAS SNPs that were not present in the lower density array data (Illumina CanineHD). One dog was removed after filtering for call rate, leaving 232 dogs for imputation and analysis. A reference panel of 421,742 SNPs and 855 individuals was used for imputation, comprising 807 dogs of approximately 159 breeds; 14 mixed breed dogs; 28 Chinese indigenous dogs; and six wolves. The reference panel, which was generated using array genotype data and array SNPs extracted from in-house and publicly available whole-genome sequence variant data [67, 370] (see chapter 4), included 51 Norwich Terriers; 47 of which were part of the discovery GWAS. The mean estimated concordance across the chromosomes, calculated from the concordance statistic (for masked genotypes) provided by IMPUTE2, was 98.8%.

Genotypes for the most significantly associated SNPs identified in the GWAS were extracted from the imputed replication dataset and the PxD GWAS dataset using PLINK; 44 out of 47 SNPs were present in both datasets and therefore available for analysis. Three SNPs were not imputed; one SNP was not present in the whole-genome sequence variant data used to construct the reference panel, the other two were filtered due to missing data. The 44 included the most-highly associated SNP from each of the eight chromosomes. Imputed genotype calls with uncertainty greater than 0.1 were treated as missing. Chi-squared P-values were calculated for each SNP using Stata (Stata 15. College

Station, TX, USA), with and without the inclusion of the additional 10 cases and 45 controls. Risk allele frequencies were also calculated for each SNP in the full population dataset of up to 232 dogs. After the addition of the 10 cases and 45 controls, the Chi-squared P-values were lower for 24 of the 44 SNPs, including three of the eight GWAS SNPs that were the most significantly associated on chromosomes: 5, 14, and 21 (**Table 5.1** and **Appendix iv.i**). Although the association of the SNP on chromosome 14 was stronger after inclusion of the additional dogs, the risk allele frequencies were not directionally consistent in the additional cases and controls. The risk allele for the SNP on chromosome 21 was not present in any of the additional cases and controls and was extremely rare (0.007) in the 232 dogs of the entire population set. The most significantly associated SNP on chromosome 32 did not show an increased level of statistical association with the added dogs; however, an increased statistical significance was observed for two other SNPs on this chromosome. Although genotypes were missing for two of the additional cases and six additional controls for one of these, reanalysis of the genotype data including only dogs without missing genotype data confirmed the same observation (**Appendix iv.ii**). A second SNP on chromosome 37, which was not the top GWAS SNP, became the highest associated after analysis with the second set of dogs (**Table 5.1** and **Appendix iv.i**). The SNPs which showed an increased association identified five potential genes of interest: *SIK3*, *ELMO1*, *FCHSD2*, *TMEM150C*, and *SLC39A10*.

Table 5.1. GWAS data and replication analysis of SNPs with a GWAS P-value < 1 x 10^{-3}

			GWAS (GEMMA-adjusted)					Unadjusted	Replication			
SNP ID	**Genomic Pos. [a]**	**Nearest known gene**	**ca/co [b] (n)**	**P-value**	**Alleles (risk/non-risk)**	**Risk allele freq. (ca/co [b])**	**Odds ratio (95% CI)**	**χ^2 P-value (GWAS dogs)**	**Pop. set risk allele freq. [c]**	**χ^2 P-value [d]**	**Additional ca/co [b] (n)**	**Risk allele freq. (ca/co [b]) [e]**
AX-167401878	5:16564735	*SIK3*	23/24	7.9×10^{-4}	A/G	0.59/0.27	1.48 (1.21 - 1.82)	3.5×10^{-3}	0.16	3.6×10^{-4} *	10/45	0.15/0.13
AX-167923327	10:54532815	*ASB3*	24/24	4.2×10^{-4}	A/C	0.23/0.04	1.74 (1.31 - 2.31)	3.5×10^{-3}	0.23	0.21	10/45	0.50/0.30
AX-167671433	12:22674130	*GFRAL*	24/23	9.4×10^{-4}	A/G	0.35/0.15	1.64 (1.25 - 2.15)	5.6×10^{-3}	0.21	0.14	10/41	0.25/0.24
AX-167677441	14:48725029	*ELMO1*	23/24	8.9×10^{-4}	T/C	0.67/0.44	1.49 (1.17 - 1.89)	0.02	0.33	3.5×10^{-3} *	10/45	0.05/0.24
AX-167780040	15:50486447	*TMEM154*	24/23	4.3×10^{-4}	T/C	0.96/0.76	1.79 (1.33 - 2.42)	2.5×10^{-3}	0.90	0.13	10/45	0.95/0.97
AX-167233313	21:25108139	*FCHSD2*	24/24	9.5×10^{-4}	C/G	0.31/0.06	1.72 (1.33 - 2.24)	3.5×10^{-4}	0.007	5.8×10^{-7} *	10/45	0.00/0.00
AX-167828838	32:6759073	*SEC31A*	24/23	6.1×10^{-4}	T/C	0.73/0.50	1.56 (1.22 – 2.00)	0.02	0.69	0.20	10/45	0.65/0.64
AX-168031858	32:6522283	*TMEM150C* [f]	24/23	9.2×10^{-4}	T/C	0.69/0.37	1.43 (1.17 - 1.76)	6.8×10^{-3}	0.30	1.2×10^{-4} *	10/45	0.30/0.21
AX-167541099	32:6531625	*TMEM150C* [f]	24/24	9.7×10^{-4}	G/A	0.69/0.40	1.46 (1.18 - 1.81)	6.8×10^{-3}	0.32	3.2×10^{-4} *	8/39	0.31/0.24
AX-167857485	37:5161255	*SLC39A10* [g]	23/24	4.3×10^{-4}	G/A	0.61/0.27	1.59 (1.30 - 1.94)	2.5×10^{-3}	0.30	0.07	10/44	0.15/0.31
AX-167666536	37:5073636	*SLC39A10* [g]	24/23	9.2×10^{-4}	T/C	0.46/0.17	1.55 (1.25 - 1.93)	4.9×10^{-3}	0.18	1.4×10^{-3} *	10/44	0.15/0.13

The most significantly associated SNPs on each chromosome are shown, and any SNPs with a higher GWAS P-value but which had a lower P-value in the Chi-squared analysis than the top GWAS SNP. The data for all 44 SNPs are included in **Appendix iv.i**. The SNPs are ordered by chromosome and GWAS P-value. [a] The CanFam3.1 genomic location of the SNP in the format chromosome: bp position. [b] ca/co = case/control. [c] The risk allele frequency in the full population dataset including up to 232 dogs. [d] χ2 P-value for a combined analysis of the dogs included in the GWAS and up to 10 cases and 45 controls. [e] Risk allele frequencies in the additional cases and controls from the population dataset. [f] SNPs correlated at R^2 of 0.98 in the combined GWAS and population datasets. [g] SNPs correlated at R^2 of 0.48 in the combined GWAS and population datasets. * Chi-squared P-value for analysis including additional case and control dogs is lower than the unadjusted P-value when including only those in the GWAS.

5.4.5 Investigating potential differences in study population

To determine if population stratification was present within the GWAS sample set, or between the GWAS set and the replication sample set, two-dimensional multidimensional scaling (MDS) plots based on a matrix of pairwise identity-by-state (IBS) distances were generated using PLINK. Prior to MDS analysis, the datasets were filtered by SNP call rate (> 97%), individual call rate (> 90%), minor allele frequency (> 0.05), and Hardy-Weinberg equilibrium P-value (> 5×10^{-5}). Only SNPs with a IMPUTE2 'Info' statistic (certainty metric) over 0.5 were kept for the imputed dataset. To allow a combined MDS analysis of the GWAS and replication dataset the two genotype datasets were merged using PLINK, keeping only the SNPs present in both. The GWAS MDS plot indicated that efforts to avoid population stratification by selecting controls related to each case were largely successful, although a small number of cases and controls still did not cluster together as tightly as others (**Figure 5.2A**). The MDS plot of the combined datasets showed considerable overlap, and the two sets did not cluster independently, however the majority of the independent replication set did not overlap with the GWAS dataset (**Figure 5.2B**). This indicates some population stratification, likely due to the differences in sample recruitment between studies, and the different methodology used for genotyping, and the imputation, are likely to have introduced some variation through observational error.

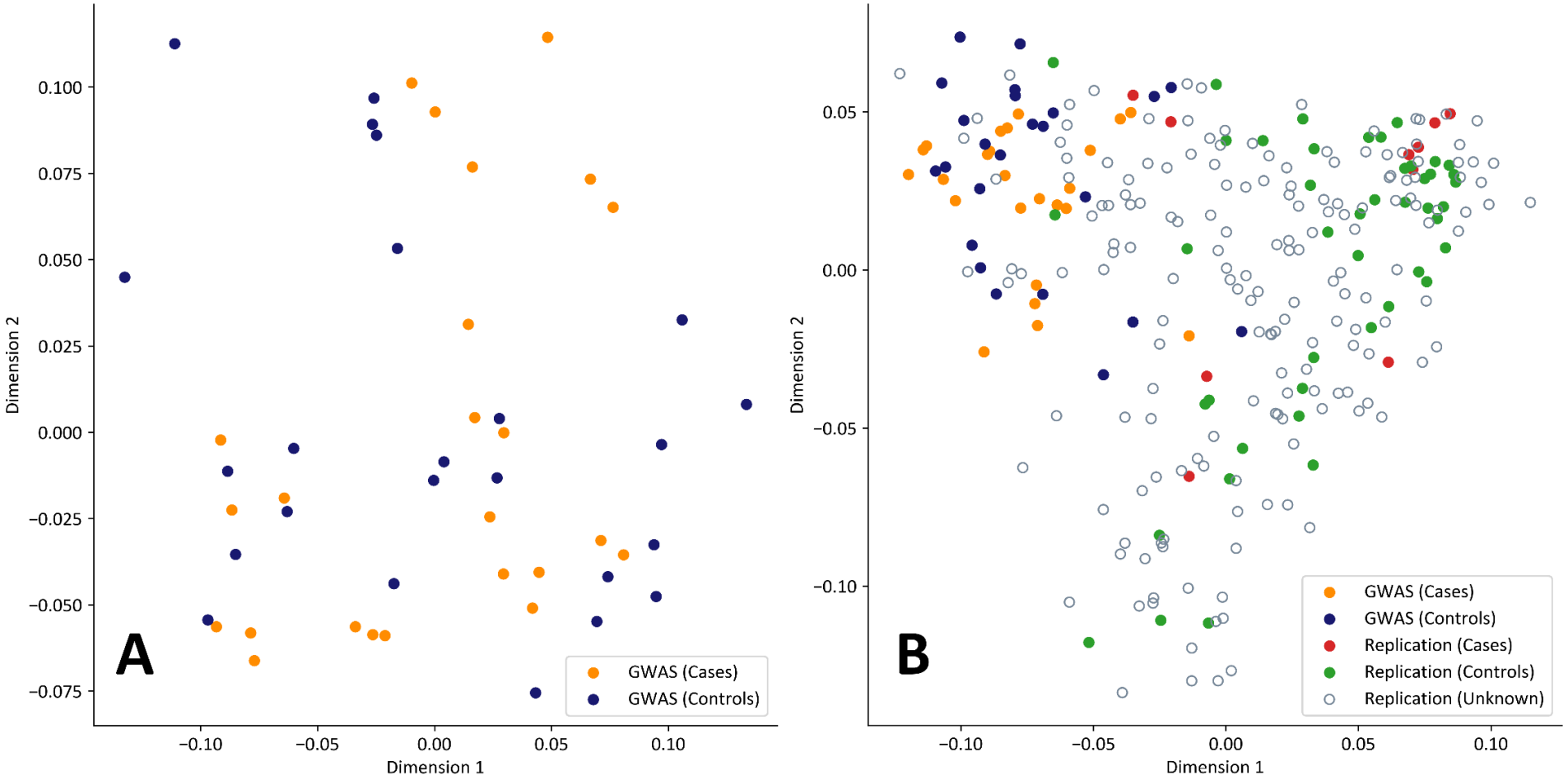


Figure 5.2. MDS plots of Norwich Terrier genotype datasets. The plots show MDS of 24 PxD cases and 24 controls in the GWAS dataset only (A) and MDS comparing the GWAS cases and controls to the replication set of 10 cases, 45 controls, and 177 dogs of unknown phenotype (B).

5.4.6 Discussion

This study of PxD in the Norwich Terrier was a preliminary investigation intended to give an indication of the mode of inheritance; the study aimed to identify potential disease-associated regions if PxD was a single gene disorder in the Norwich Terrier or be a foundation for future studies if PxD had a multigenic or complex mode of inheritance. GWAS of comparable size to that described here, or smaller, have successfully identified regions strongly associated with autosomal recessive PxD and led to identification of putative causal variants in two different breeds [189, 190, 193]. The lack of SNPs with P-values approaching genome-wide significance in the present GWAS suggests that PxD in the Norwich Terrier is not likely to be a single gene disorder, and the underlying genetics may be multigenic or complex. Future research will build on these findings through meta-analysis with a larger GWAS dataset, or by investigating the identified regions and genes of interest. Any future GWAS should include a larger cohort to increase study power and allow further elucidation of the underlying genetics. Despite the limited study power, the current study did identify five candidate regions and genes (*SIK3*, *ELMO1*, *FCHSD2*, *TMEM150C*, and *SLC39A10*) for further study, by analysing the genotypes of the most significantly associated GWAS SNPs in an independent dataset.

The SNP with the strongest statistical association in the replication study is located on chromosome 21 within the *FCHSD2* gene, which encodes FCH and double SH3 domains protein 2 [371]. The risk allele was rare in GWAS controls and in the independent population set in general and absent in the cases and controls used for replication. However, it was comparatively common in the GWAS cases, of which 15 were heterozygotes (no dogs were homozygous). Taken together this suggests that in some Norwich Terrier PxD cases a rare, high effect, variant may underly increased risk of PxD. *FCHSD2* has a role in clathrin-mediated endocytosis (CME) [372], promoting actin polymerization during CME [373], and mediating the endocytosis of membrane receptors [374-376]. *FCHSD2* has also been shown to be essential for the formation and maintenance of actin-based cell protrusions [377], including stereocilia which are important for balance and hearing [378-380]. Strengthening *FCHSD2* as a candidate gene for PxD, the *Drosophila* homologue for *FCHSD2* is Nervous Wreck (*nwk*). A temperature-sensitive mutation within *nwk* was described as causing loss of coordination, seizure-like spasms, and eventual paralysis [381], and nwk has a role in the regulation of synaptic

endocytosis [382]. However, the structure of the domains and roles of the nwk and FCHSD2 proteins have been reported to differ [372].

One of the most strongly associated variants was within the *SIK3* gene, which encodes salt-inducible kinase 3, a member of the AMP-activated protein kinase (AMPK)-related family of kinases [383]. *SIK3* has diverse roles [383], including skeletal development [384, 385]. With potential relevance to PxD, *SIK3* has neurological roles including the regulation of sleep and sleep need; mouse models with a mutation in *SIK3* have been referred to as 'Sleepy' mice [386-388]. *SIK3* has also been demonstrated to be an essential component of a signalling pathway in *Drosophila* glial cells that regulates K^+ and water homeostasis, suppressing neuronal hyperexcitability and seizure susceptibility [389].

Two of the remaining associated SNPs identify genes with neurological functions that could plausibly have an involvement in PxD. One SNP, and a second SNP in near perfect correlation in the combined GWAS and population datasets (R^2 0.98), are both within *TMEM150C* on chromosome 32*,* and *ELMO1* is the closest gene for the SNP on chromosome 14. The *TMEM150C* gene encodes tentonin 3, thought to be either a mechanically activated (MA) ion channel or have a role in the regulation of MA ion channel activity, which mediates slowly adapting currents in mechanosensory neurons and is involved in muscle coordination and sensing changes in arterial pressure [390-395]. The engulfment and cell motility 1 (ELMO1) protein forms a complex with DOCK180 (a guanine nucleotide exchange factor) to mediate RAC1 (a Rho GTPase) signalling which has roles in the regulation of the actin cytoskeleton and is involved in cerebellar development [179, 396-398]. *ELMO1*, which is expressed in the developing mouse brain (including the cerebellum) [399], has been implicated in cell migration [397, 400], phagocytosis of apoptotic neurons by neuronal progenitor cells during neurogenesis [401], and mediation of neurite outgrowth [396, 402, 403].

The remaining gene identified for potential further study is *SLC39A10*, which, to the authors knowledge, has not yet been reported to have neurological roles that suggest a clear candidacy for PxD. *SLC39A10*, the closest known gene to the most significantly associated SNP on chromosome 37, encodes the Zinc (Zn^{2+}) transporter ZIP10 which has roles in the transport of Zn^{2+} in the kidney and intestines [404-406], and is involved in immunity; including the regulation of macrophage and B-cell survival [407, 408], and B-cell antigen receptor signalling [409].

One limitation of this study, in addition to the limited size and therefore power of the GWAS and of the replication case-control set, is the population differences between the GWAS sample set and the replication set. MDS analysis, although showing considerable overlap between sample sets, displayed some population stratification. The replication sample set was collected and genotyped for an unrelated study and population stratification could have resulted in inflation of P-values for the SNPs in the replication study because of differences in allele frequencies in comparison to the dogs in the closely related GWAS set. However, although some of the P-values observed in the replication study were decreased in comparison to the GWAS, the statistical significance of other SNPs was instead reduced, suggesting that the P-values weren't inflated overall. Another potential limitation is that the small number of cases included in the replication set were not as robustly defined as those in the GWAS. To extend our study, we are seeking to collect an additional well-defined independent case-control set in which to follow up these preliminary findings.

Through GWAS and replication in an independent dataset, this preliminary study has identified five genes for future investigation with potential roles in PxD risk. The study provides evidence that PxD is not monogenic autosomal recessive in the Norwich Terrier, although the findings also suggest that rare familial high effect variants might underly increased PxD risk in some dogs. These findings will provide a foundation for future work.

5.5 Acknowledgements

This study was funded by Dogs Trust. CAJ, SLR and CSM are supported by the Kennel Club Charitable Trust. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the Axiom array data. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G090074791070) for the whole genomes of 24 dogs. The sequencing of the whole genomes of four dogs was carried out by the University of Cambridge. The whole genomes of 180 other dogs were sequenced by Edinburgh Genomics Laboratories, University of Edinburgh. Whole genome sequencing of 82 of the dogs was part of the Give a Dog a Genome project. We also thank Natasha Olby for sharing whole genome data. We would like to thank the Dog Biomedical Variant Database Consortium and acknowledge all canine researchers who deposited dog whole

genome sequencing data into public databases. We thank Gary Johnson for providing details of the *PIGN* mutation. We thank Janine Lamb, University of Manchester, for critical review of the manuscript. We thank Donna Foster for support with sample collection and Ellen Schofield for bioinformatics assistance. We would like to sincerely thank all owners and breeders for providing samples from their dogs for this study and the UK Norwich Terrier Club for their support.

5.6 Availability of data

Axiom array data for the 48 Norwich Terriers will be available via the Dryad database upon publication.

5.7 References

See chapter 9

5.8 Supplementary tables and figures

See chapter 10

6 Manuscript 4 - Genome-wide association study of idiopathic epilepsy in the Italian Spinone dog breed

This manuscript has been written to be submitted to PLOS ONE.

6.1 Authors and affiliations

Christopher A Jenkins

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Luisa De Risio

Neurology / Neurosurgery Service, Centre for Small Animal Studies, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Current address: Linnaeus Veterinary Ltd, United Kingdom

Artitaya Lophatananon

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Thomas W Lewis

The Kennel Club, Clarges Street, London, United Kingdom

School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, United Kingdom

Current address: The Guide Dogs for the Blind Association, Reading, United Kingdom

Donna Foster

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Jim Johnson

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

Marjo K Hytönen

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Hannes Lohi

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Cathryn S Mellersh

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Sally L Ricketts

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

6.2 Author contributions

CAJ contributed to study design, performed experimental work and analysis and interpretation of data, and wrote the manuscript. SLR conceived of, designed, and supervised the study, and provided critical review of the manuscript. LDR provided case definition and diagnosis of cases and clinical expertise. AL contributed scripts and expertise for assessing risk score accuracy. TWL performed kinship calculations using the Kennel Club pedigree database. DF and JJ contributed to sample recruitment. HL and MKH contributed to sample collection. CSM contributed to study conception and funding.

6.3 Abstract

Idiopathic epilepsy (IE) has a high prevalence and a severe clinical course in the Italian Spinone breed of dog. A genome-wide association study meta-analysis of 52 cases and 51 controls was conducted to identify genomic regions that may have involvement in the development of IE. Subsequent to the meta-analysis, a set of 175 controls and an independent validation set of 23 cases and 23 controls were genotyped for SNPs showing suggestive association with IE to find variants exhibiting evidence of replicable association and to test the predictiveness of SNPs for IE status when combined in a weighted risk score. Although two regions showed statistically significant association with IE in the GWAS meta-analysis, and additional regions with suggestive association were identified, the findings were not reproduced in the validation sets. This is the first GWAS of IE in the Italian Spinone, and the findings suggest that IE in the breed is not monogenic and demonstrates the challenges when investigating a multigenic or complex inherited disease in a numerically small dog breed.

6.4 Introduction

Epilepsy is common in dogs, with an estimated prevalence of 0.62% - 0.75% in the general dog population [203, 274] and a much higher prevalence reported for some breeds [251, 252, 256, 410, 411]. IE in the Italian Spinone breed of dog has been clinically characterised and defined as two or more epileptic seizures (occurring > 24 hour apart) with an onset between 6 months and 6 years of age in dogs with normal interictal physical and neurologic examinations and unremarkable results of haematology and serum biochemistry [251]. This breed-wide survey of 1,192 UK dogs born between the years 2000 and 2011 estimated the prevalence of IE in the Italian Spinone breed to be 5.3% (95% confidence interval 4.0-6.6%). The age of onset in the 63 cases identified in the study ranged from 11 months to six years (median 2.9 years) and their seizures were found to be typically generalised tonic-clonic, with approximately half of dogs reported to have seizures with a focal onset that became generalised. IE has been reported to be particularly severe in this breed, with cluster seizures reported for the majority of cases; status epilepticus occurring in over a fifth of cases; and a 32% mortality due to epilepsy-related causes [251].

It is thought, based on current evidence, that for common human diseases with an adult-onset, heritability is conferred by many genetic variants acting together that range from

low-frequency with high-effect on disease risk, to common with low-effect on risk [412]. Each single genetic variant does not give a strong indication of an individual's overall disease risk, but a polygenic risk score (PRS) incorporating multiple variants can be used. PRS are typically calculated as the total number of risk alleles that an individual carries, weighted by the measured effect of each of the variants [412]. Following GWAS and replication studies for several complex human diseases, PRS have been developed to help identify individuals at highest risk of disease in the population and who therefore may benefit from clinical and/or lifestyle interventions [412, 413]. PRS have been produced for multiple human diseases, including coronary heart disease [414-416] [412, 415, 416], and some types of cancer [417-419].

The present study sought to identify genetic variants that are reproducibly associated with IE in the Italian Spinone dog breed and to assess their utility as a predictive tool in a genetic risk score in an independent set of Spinoni. We performed a GWAS meta-analysis, utilising genotype imputation, followed by replication testing in additional samples.

6.5 Methods

6.5.1 Sample collection

Idiopathic epilepsy cases were recruited through a breed-wide survey as reported previously [250], owner reported questionnaires, and cases diagnosed by veterinary neurologists in the neurology unit at the Animal Health Trust Centre for Small Animal Studies, Newmarket, UK (led by Luisa De Risio) or Linnaeus referral veterinary hospitals. All Italian Spinoni affected by IE recruited as outlined above were diagnosed by Luisa De Risio (a board-certified veterinary neurologist) through review of the owner questionnaire, medical records, and epileptic seizure video footage when available. The diagnosis of IE was based on the criteria set by the IVETF [250]. In addition to the above, four dogs (two cases and two controls) included in the study were collected by the University of Helsinki, Finland. All controls were dogs over the age of seven years that were reported by their owners to have never had a seizure.

Two GWAS datasets were generated for this study, genotyped four years apart; Set 1 (29 cases, 29 controls) and Set 2 (23 cases, 22 controls). The controls in Set 2 were selected at random from a large pool (approximately 170) of available control samples, and the relatedness of the cases and controls from the UK were compared to the Kennel Club-

registered Italian Spinone population (as outlined below) to ensure that the cohorts did not have an over-representation of closely related individuals. Due to DNA availability, a subset of the UK dogs investigated for relatedness (12 cases, 16 controls) were combined with predominantly non-UK dogs to make up the final Set 2. A set of 175 control dogs over the age of seven years were used as a 'semi-replication' set to investigate the most significantly associated SNPs from the GWAS. These control samples had been accrued over time and via study recruitment including through the Italian Spinone Club of Great Britain (ISCGB). An independent validation set consisted of 23 cases and 23 controls (Set 3). Additional details about the dogs included in the four sets is given in **Appendix v.i** and **Appendix v.ii**.

Samples were collected in the form of buccal swabs or whole blood residual from routine clinical tests (Animal Health Trust Clinical Research Ethics Committee Project No. 73-2016, University of Cambridge Department of Veterinary Medicine Ethics and Welfare Committee No. CR512). DNA was extracted from swabs using a QIAamp DNA Blood Midi Kit (Qiagen), and blood samples using a standard chloroform protocol.

6.5.2 Kinship calculations and selection of controls for Set 2

At the time of genotyping for Set 2, DNA samples of approximately 170 controls were available. A set of controls was selected at random for GWAS from the available dogs; however, we sought to confirm that bias had been avoided in the selection methodology, and that the subset of the potential Set 2 controls (that were UK-Kennel Club registered) were representative of a wider population of Italian Spinoni in a contemporaneous UK Kennel Club-registered population. The relatedness of the potential Set 2 (21 cases and 25 controls that were UK-Kennel Club registered) were compared to 1,000 randomly selected samples of the same number of dogs in each group from the Kennel Club pedigree database during 2017. See **Appendix v.iii** and **Appendix v.iv** for a more detailed description of the methodology used and of the results from this analysis. The mean kinships among and between the cases and controls cohorts were found to be representative of random samples of dogs from similar birth years (**Appendix v.iii** and **Appendix v.iv**).

6.5.3 Array genotyping

The two study sets were genotyped on different arrays: Set 1 on the Illumina CanineHD BeadChip (173,662 SNPs) and Set 2 using the Axiom Canine HD array (up to 729,642 SNPs and other DNA variants). High-quality genotyping data were available for 29 cases and 29 controls (individual call rate > 95%) for Set 1. For Set 2, after quality control using the Axiom Analysis Suite and the Best Practices Workflow, genotype data were available for 23 cases and 22 controls with individual call rate > 90%.

A subset of the independent set used for replication, consisting of 18 cases and 18 controls ('Set 3'), was also subsequently genotyped on the Illumina CanineHD BeadChip and high-quality genotype data was obtained for all dogs.

6.5.4 Genotype imputation and GWAS meta-analysis

Genome-wide genotype imputation was carried out for Set 1, as described previously [370] (see chapter 4), to allow meta-analysis with Set 2 whilst maximising the number of SNPs available for analysis. Imputation increased the SNP density of Set 1 to Axiom-array level; to 476,313 autosomal SNPs (imputation accuracy: mean $R^2 = 0.97$, mean genotype concordance 98.2%) [370] (see chapter 4). Set 3 was imputed to Axiom-array level using the same reference panel used for Set 2. The imputed data was converted to BIMBAM format for downstream analysis.

GWAS analysis was conducted for each dataset independently using GEMMA, with a linear mixed model to correct for population stratification [54]. SNPs were filtered to remove those with a Hardy-Weinberg P-value $< 5 \times 10^{-5}$; those with a SNP call rate < 97%; or with a minor allele frequency (MAF) < 5%. The imputed dataset, Set 1, SNPs were also filtered to exclude those with an IMPUTE2 'Info' statistic (imputation certainty) < 0.5 [72, 73]. After quality control 341,811 and 358,527 SNPs remained for Set 1 and Set 2, respectively.

A fixed-effects meta-analysis was carried out using the statistical software package Stata (Stata 15. College Station, TX, USA), for the SNPs common across the GWAS sets, using the summary statistics produced by GEMMA. Heterogeneity was assessed using the Q statistic.

A two-dimensional multidimensional scaling (MDS) plot was generated for Set 1, Set 2, and Set 3 using PLINK (**Appendix v.v**). For MDS analysis the datasets were filtered by SNP

call rate (< 97%), individual call rate (< 90%), and Hardy-Weinberg equilibrium P-value (< 5 x 10^{-5}). Set 2 and Set 3 were also filtered by the IMPUTE2 'info' statistic (< 0.5). To allow a combined MDS analysis the three genotype datasets were merged using PLINK, keeping only the SNPs present in all.

6.5.5 Genotyping selected GWAS SNPs in additional dogs

The genotyping of individual SNPs in the semi-replication set of 175 control dogs, and Set 3, was carried out using allelic discrimination assays, utilising an ABI StepOne real-time thermal cycler and LUNA Universal Probe qPCR Master Mix (NEB). The sequences of the TaqMan SNP genotyping assay (ThermoFisher) primers and probes (reporters) are in **Appendix v.vi**. A fast ramping speed was used, and the following cycling conditions: 30 seconds pre-PCR read at 25°C, 3 minute 95°C holding stage, 40 cycles of 95°C for 3 seconds and 10 seconds at 60°C, followed by a 25°C post-PCR read stage for 30 seconds.

6.5.6 Chi-square analysis of SNPs in additional control dogs

Genotypes for the most significantly associated SNPs in the 10 regions identified in the meta-analysis were extracted from the array genotype and imputed genotype data for dogs included in the GWAS meta-analysis. Imputed genotype probabilities were converted to hard genotype calls using PLINK (v1.90) [368], and genotype calls with uncertainty > 0.1 were treated as missing. Genotype calls for the 10 SNPs were obtained through allelic discrimination assays, as outlined above, for the semi-replication set of 175 control dogs. Chi-square analysis was caried out using Stata (Stata 15. College Station, TX, USA) with and without the inclusion of the 175 additional controls.

6.5.7 Generating a five-SNP genetic risk score and testing it in an independent study set

Weighted five-SNP risk scores were calculated as follows:

$$5 \times \frac{\beta 1 \times SNP1 + \beta 2 \times SNP2 + \beta 3 \times SNP3 + \beta 4 \times SNP4 + \beta 5 \times SNP5}{\beta 1 + \beta 2 + \beta 3 + \beta 4 + \beta 5}$$

Where 'SNP' represents the count of the risk-conferring allele, and 'β' the beta-coefficient from the meta-analysis, for each of the five SNPs. If the meta-analysis-derived beta-coefficient was calculated for the non-risk allele, and therefore negative, the allele was 'flipped' to produce a positive beta-coefficient coded for the risk allele (× -1).

Genetic risk scores were calculated for dogs included in the GWAS meta-analysis using the array genotype data or imputed genotype data. Imputed genotype probabilities were converted to hard genotype calls as above. Genotype calls from allelic discrimination assays were used to generate genetic risk scores for dogs that were part of Set 3. Individuals were excluded from the analysis if genotype calls were missing for any of the five SNPs.

To assess genetic risk score performance in each set of dogs the receiver operating characteristic (ROC) and area under the curve (AUC) were calculated and plotted using logistic regression, for IE case/control status and the genetic risk score, and using Stata's 'lroc' command. Calibration plots were generated using the 'pmcalplot' Stata module to further evaluate the predictive performance of the genetic risk score.

6.6 Results

6.6.1 GWAS meta-analysis identifies suggestively associated loci on 12 chromosomes

A meta-analysis of two GWAS datasets, Set 1 and Set 2, including a total of 52 cases and 51 controls and comprising data from 328,622 SNPs, identified SNPs in two regions that passed the Bonferroni corrected threshold for statistical significance (1.5×10^{-7}); on chromosomes 2 and 8 (**Figure 6.1** and **Table 6.1**). Suggestive associations also indicated that multiple other regions might have a role in conferring disease risk in this breed. SNPs with P-values $< 1 \times 10^{-5}$ were identified on 10 additional chromosomes (**Figure 6.1** and **Table 6.1**). Two of the 12 SNPs, on chromosomes 33 and 38, were excluded from further analysis due to high heterogeneity (P-value ≤ 0.01) between the two study sets (**Table 6.1** and **Appendix v.vii**).

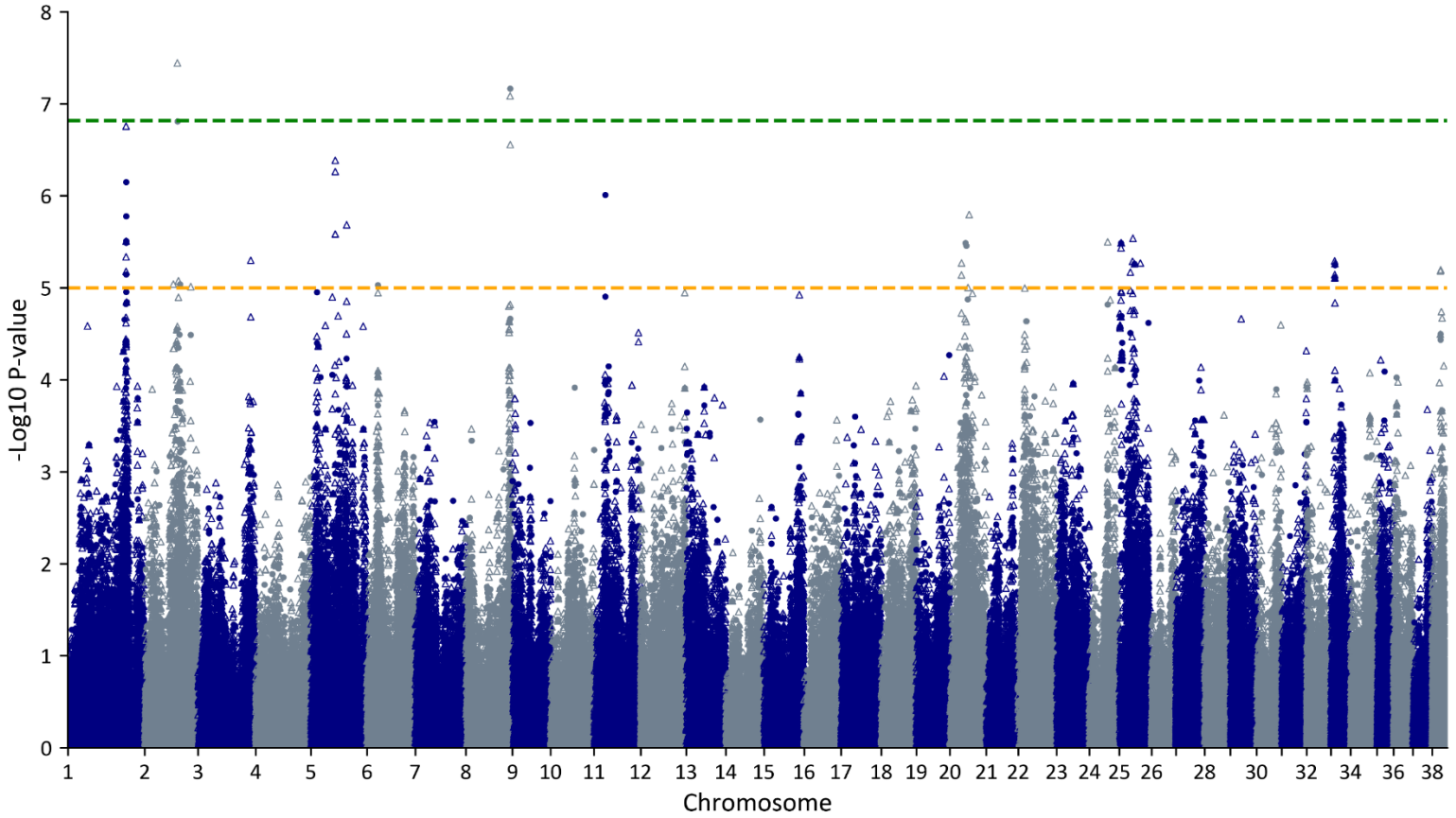


Figure 6.1. GWAS meta-analysis of 52 Italian Spinone idiopathic epilepsy cases and 51 controls (328,622 SNPs). Plot of negative log (base 10) transformed P-values. X-axis is SNP location by chromosome (left to right, autosomes 1 to 38). Solid circles indicate array-genotyped SNPs, hollow triangles denote SNPs imputed for Set 1. Green (upper) line shows the Bonferroni-corrected threshold for statistical significance ($P < 1.5 \times 10^{-7}$). Orange (lower) line indicates the empirical threshold for suggestive association ($P < 1 \times 10^{-5}$). Axiom Canine HD array data, and Illumina CanineHD BeadChip data imputed to Axiom density.

Table 6.1. The most significantly associated GWAS meta-analysis SNPs on each chromosome with P-values < 1×10^{-5}

Genomic pos. [a]	**Nearest known gene**	**ca/co** [b] **(n)**	**Odds ratio (95% CI** [c]**)**	**P-value**	**P-value for het.** [d]	**Alleles (risk/non-risk)**	**Risk allele freq. (ca/co** [b]**)**		
							Set 1	**Set 2**	**Set 3** [e]
1:93123836	*JAK2*	52/51	1.40 (1.24-1.59)	1.8×10^{-7}	0.03	T/C	0.84/0.47	0.67/0.43	0.64/0.50
2:52390106	*MAST4*	52/51	1.38 (1.23-1.54)	3.6×10^{-8}	0.03	C/T	0.50/0.34	0.70/0.16	0.36/0.39
3:84100359	*RBPJ*	52/51	1.41 (1.22-1.63)	5.0×10^{-6}	0.98	C/G	0.34/0.10	0.39/0.11	0.31/0.17
5:38884749	*TEKT3*	52/50	1.52 (1.29-1.79)	4.1×10^{-7}	0.70	T/C	0.27/0.04	0.37/0.10	0.28/0.14
6:18142628	*CORO1A*	52/51	1.41 (1.21-1.64)	9.4×10^{-6}	0.36	G/A	0.48/0.22	0.46/0.25	0.36/0.22
8:70681185	*RCOR1*	52/51	1.42 (1.25-1.62)	6.9×10^{-8}	0.56	G/T	0.83/0.52	0.78/0.52	0.67/0.56
11:17811231	*SLC27A6*	52/51	1.46 (1.26-1.70)	9.8×10^{-7}	0.71	T/C	0.84/0.62	0.91/0.61	0.83/0.69
20:30846012	*FHIT*	51/51	1.39 (1.22-1.59)	1.6×10^{-6}	0.19	A/G	0.69/0.50	0.75/0.36	0.58/0.42
24:29341230	*LPIN3*	52/51	1.55 (1.29-1.87)	3.2×10^{-6}	0.04	T/C	0.24/0.16	0.33/0.05	0.25/0.25
25:22038367	*GALNTL6*	52/51	1.81 (1.41-2.31)	2.9×10^{-6}	0.24	T/A	0.10/0.05	0.22/0.00	0.06/0.11
33:5990467	*COL8A1*	52/51	1.37 (1.20-1.56)	5.1×10^{-6}	8.5×10^{-3}	A/G	0.62/0.46	0.78/0.39	Not tested
38:13596806	*TGFB2*	52/51	1.31 (1.16-1.47)	6.4×10^{-6}	7.4×10^{-3}	G/C	0.48/0.39	0.67/0.18	Not tested

[a] The CanFam3.1 genomic location of the SNP in the format chromosome: base pair position. [b] case / control. [c] 95% confidence intervals. [d] heterogeneity. [e] The risk allele frequencies in the 18 cases and 18 controls in the replication set, which was not part of the GWAS meta-analysis.

6.6.2 Genotyping the most significantly associated SNPs in additional controls

The 10 SNPs identified in the GWAS meta-analysis were initially genotyped in up to 175 additional unaffected Italian Spinone dogs over the age of seven years. Five of the 10 SNPs, on chromosomes 1, 2, 11, 20, and 25 showed stronger statistical associations with IE after the addition of these controls (**Table 6.2**). The other SNPs, on chromosomes 3, 5, 6, 8 and 24, demonstrated a reduced statistical significance when the additional dogs were included, which may suggest a lack of replication of these SNP associations.

Table 6.2. Initial replication analysis of 10 SNPs showing suggestive association with IE in the GWAS

Genomic position*	**GWAS Cases/Controls (n)**	**P-value (GWAS dogs)**	**Control Set (n)**	**P-value (GWAS + controls)**
1:93123836	52/51	2.3×10^{-5}	174	1.7×10^{-6}
2:52390106	52/51	1.1×10^{-4}	174	3.4×10^{-5}
3:84100359	52/51	8.5×10^{-5}	175	2.0×10^{-3}
5:38884749	50/48	3.2×10^{-5}	175	1.1×10^{-3}
6:18142628	52/51	3.4×10^{-4}	175	0.02
8:70681185	52/51	3.0×10^{-5}	173	1.7×10^{-4}
11:17811231	52/51	5.2×10^{-5}	175	1.2×10^{-5}
20:30846012	51/51	8.1×10^{-5}	174	2.3×10^{-5}
24:29341230	52/51	2.8×10^{-3}	175	6.9×10^{-3}
25:22038367	52/51	1.1×10^{-3}	175	8.2×10^{-6}

*The CanFam3.1 genomic location of the SNP in the format chromosome: base pair

6.6.3 Testing a five-SNP genetic risk score in an independent case-control set

To conduct further validation of SNPs showing initial evidence of replication from the above analysis (once new incident cases had become available), and to assess the utility of a genetic risk score to predict a dog's risk of developing IE, the five SNPs that showed an increased statistical association after the addition of the large control set were investigated as a weighted five-SNP genetic risk score in an independent validation set of 23 IE cases and 23 controls (Set 3). Receiver operating characteristic (ROC) curve and calibration plot analysis demonstrated that the genetic risk score had poor predictiveness in Set 3 (area under ROC curve (AUC): 0.58; 95% confidence intervals (CI): 0.41 - 0.75),

failing to replicate that seen in the GWAS discovery sets (AUC: 0.92; 95% CI: 0.87 - 0.97) (**Figure 6.2**). The five SNPs also failed to replicate the statistical associations observed in the GWAS sets, both individually and as a five-SNP genetic risk score (**Appendix v.viii**).

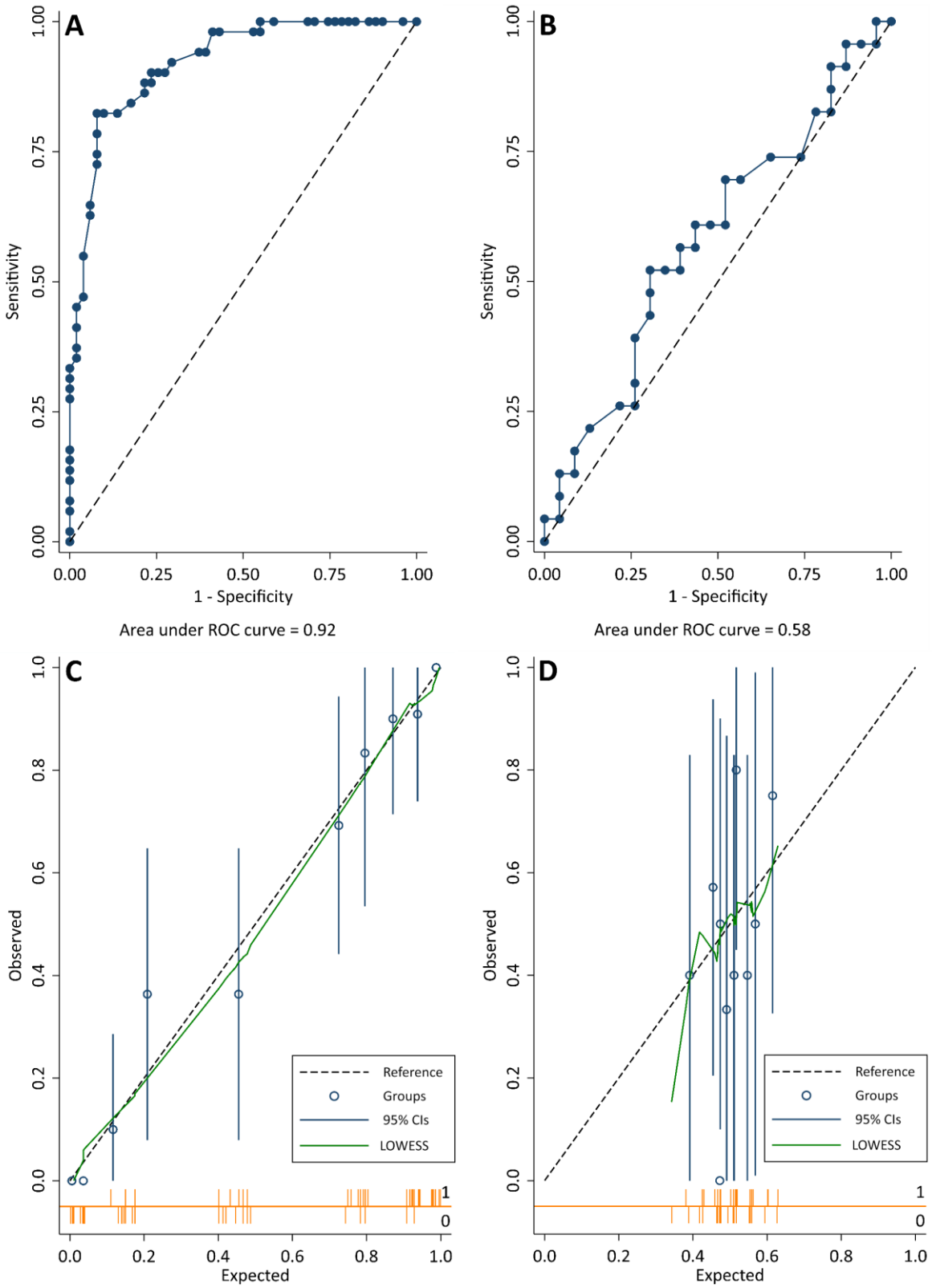


Figure 6.2. Receiver operating characteristic (ROC) curves and calibration plots for a five-SNP genetic risk score for idiopathic epilepsy in the Italian Spinone. Plots A and B are ROC curves for GWAS individuals, and Set 3, respectively. Points represent each potential genetic risk score cut-off for defining cases, from the highest (0,0) to the lowest (1,1). Sensitivity: fraction of cases correctly classified. Specificity: fraction of controls correctly

classified (1 – (minus) specificity is the false-positive fraction). The area under the ROC curve (AUC) is given for each study set. An AUC of 0.5 (indicated by the dashed line) would represent a test unable to discriminate cases from controls. Calibration plots are shown for the GWAS (C), and replication set (D). Points represent ten equally sized groups of individuals divided by predicted risk. Observed: the proportion of cases in each group. Expected: the average (mean) of the predicted probabilities generated from the genetic risk score logistic regression model. The 95% confidence intervals are shown for each group. Orange lines are spike plots indicating the distribution of IE cases (1) and controls (0). The dashed reference line indicates perfect genetic risk score calibration where predicted risk matches the observed proportion of affected dogs within each group. Locally weighted scatterplot smoothing (LOWESS) is displayed in green.

6.6.4 Analysis of GWAS SNPs in the independent case-control set

The most significantly associated GWAS SNP on each chromosome failed to either show evidence of replication when 175 controls were analysed along with GWAS sets, or, for those that did, to demonstrate predictiveness or association with IE in the independent case-control set when combined in a genetic risk score. The focus was therefore expanded to include all SNPs showing suggestive association from the GWAS meta-analysis ($P < 1 \times 10^{-5}$) (excluding chromosomes 33 and 38) to examine whether differences in LD between the study sets could be the cause. Array genotype data were obtained for a replication set of 18 cases and 18 controls (a subset of Set 3). After imputation, genotype data were extracted for these 48 SNPs and a combined analysis conducted using meta-analysis (**Appendix v.ix**). None of the 48 SNPs were statistically associated with IE in Set 3 independently, although 14 showed a stronger association in the meta-analysis after inclusion of Set 3 (**Appendix v.ix**). The statistical significance of the most significantly associated GWAS SNPs on chromosomes 2, 8, 24, 25, and 33 did not increase after the addition of Set 3 to the meta-analysis; two other SNPs on chromosome 2 in a different genomic region did demonstrate a stronger association but not surpassing that of the top GWAS SNP (**Appendix v.ix**).

6.7 Discussion

In the present study we used GWAS meta-analysis of 52 cases and 51 controls to identify regions of the genome associated with IE in the Italian Spinone breed of dog. Genotyping a set of 175 additional controls identified five SNPs for use in a genetic risk score for IE. However, genotyping and testing the genetic risk score in an independent validation set of 23 cases and 23 controls failed to replicate the strong predictiveness seen in the GWAS set. The findings suggest that IE in the Italian Spinone breed is not monogenic autosomal recessive or dominant, and that the mode of inheritance is multigenic and/or complex.

One potential explanation for the observed lack of replication is the differences in year of birth between the four sets of individuals studied meaning that the dogs could be multiple generations apart, potentially allowing genetic drift and selection by dog breeders to affect allele frequencies and changes in LD. The Italian Spinone breed is not numerically large in the UK, with 350-526 (mean 453.9) Kennel Club registrations a year between 2012 and 2021 [420] (https://www.thekennelclub.org.uk/media-centre/breed-registration-statistics/, accessed 06/01/2023), and only a relatively small subset of these registered dogs is likely to be part of the breeding population. Changes in breeding strategies amongst dog breeders, for example avoiding dogs related to epilepsy cases, could therefore have the potential to quickly impact allele frequencies within the UK population. The requirement for controls to be a minimum of seven years old meant that there was only moderate overlap in year of birth between cases and controls for some study sets, Set 2 being the most extreme example of this. This sampling bias meant that cases and controls could be multiple generations apart, again suggesting that allele frequencies could have changed, with the potential of generating false positive results. The selection of 'super controls' with ages greater than the higher end of the range of the reported age of onset, as opposed to age-matched controls which could include dogs that would go on to have IE later in life, was a methodology chosen to increase study power and compensate in part for the small number of cases available [49]. The above factors were, however, likely still exacerbated by the small sample numbers used; the sample size was limited by the small numeric size of the breed which, along with the reportedly severe clinical course and high mortality rate [251], limits the number of IE cases present in the population at any given time. Small sample size would have reduced the study's power to find associations with IE [49]. The set used for validation (Set 3) may have lacked

sufficient power to replicate any true statistical associations with IE that were identified in the discovery GWAS due to the 'winners curse' phenomenon [49, 421]. MDS analysis did not show any clear stratification between the three array-genotyped sample sets, or cases and controls within each set, which would have provided evidence of sampling bias.

The findings demonstrate the importance of replicating associations identified through GWAS in an independent validation set. Two of the regions identified in the GWAS (on CanFam3.1 chromosome 2 and chromosome 8) passed the Bonferroni-adjusted threshold for statistical significance, accounting for multiple testing, and the region on chromosome 2 implicated a gene with compelling candidacy for epilepsy (*MAST4*). Both regions failed to replicate. It is important when intending to develop DNA-based breeding tools that they are based on robust evidence. This study is not the first to identify loci demonstrating a statistically significant association with canine IE in a discovery set that have not replicated in independent study sets. A locus on chromosome 14 demonstrated statistically significant association with IE in Belgian Shepherd dogs in a GWAS of 20 cases and 45 controls after correcting for multiple testing [292]. Subsequent investigation in larger independent case-control sets have failed to replicate the statistically significant association observed in the discovery set [303, 304]. In comparison, a locus on chromosome 37 demonstrates replicable association with IE in the Belgian Shepherd dog and other breeds [275, 276, 291, 292].

The SNP identified in the GWAS with the strongest statistical association, on chromosome 2, that passed the Bonferroni-corrected threshold for statistical significance is located within the *MAST4* gene, which encodes a microtubule associated serine/threonine kinase [422]. *MAST4*, which has been shown to be expressed in the rat brain and to be upregulated in response to electroshock-evoked maximal seizures [422], is a compelling candidate gene that has been associated with epilepsy in humans [423, 424]. In addition to its association with epilepsy, *MAST4* is a convincing candidate gene for neurological development and diseases; *MAST4* was associated with hippocampal volume in a GWAS that included 26,814 individuals [425], and has been observed to be differentially expressed in patients with atypical frontotemporal lobar degeneration. Despite the strong candidacy of this gene, the association failed to replicate. It is possible that this locus represents a higher effect, familial, variant that is present in a small number of dogs which has since been bred out of the population through the avoidance of breeding dogs closely related to cases. The risk allele frequencies suggest that the association observed

in the GWAS meta-analysis was largely driven by Set 2 (**Table 6.1**); however, the validation set (Set 3) showed not only a reduction in risk allele frequency in cases but also an increase in risk allele frequency in controls. WGS analysis of the discovery set dogs, and related individuals, could be used in future work to further examine this hypothesis.

The GWAS approach could not identify very common or fixed (i.e. present in all individuals) variants contributing to the breed's overall increased risk of IE; or rare or *de novo* high-impact variants that may cause IE in an individual dog. The use of large WGS datasets could be used to identify these variant types, complementing the existing GWAS data. This GWAS failed to identify replicable associations with IE, most likely because of small sample sizes, demonstrating that large sample sizes will be required in future work to elucidate the underlying genetic aetiology of IE in this breed. An across-breed analysis, most likely using WGS approaches but also potentially GWAS, could help overcome this limitation by searching for loci, genes, and pathways that are important for IE risk in multiple breeds.

In conclusion, GWAS meta-analysis and replication analysis could not identify genomic loci with replicable association with IE in the Italian Spinone, independently or in a genetic risk score. However, this is the first reported study attempting to elucidate the underlying genetics of IE in this breed and will support future work aiming to gain an understanding of the aetiology of IE in the Italian Spinone and other breeds.

6.8 Acknowledgements

This study was funded by the Italian Spinone Club of Great Britain Health Fund. CAJ, SLR and CSM are supported by the Kennel Club Charitable Trust. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the Axiom array data. We sincerely thank all owners and breeders for providing samples from their dogs for this study and the Italian Spinone Club of Great Britain for their support with sample collection. We thank Janine Lamb, University of Manchester, for critical review of the manuscript. We thank Ellen Schofield for bioinformatics assistance and David Sargan for laboratory support with the project.

6.9 References

See chapter 9

6.10 Supplementary tables and figures

See chapter 10

7 Manuscript 5 - Genome-wide association study of idiopathic epilepsy in the Border Collie dog breed

This manuscript has been written to be submitted to PLOS ONE.

7.1 Authors and affiliations

Christopher A Jenkins

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Luisa De Risio

Neurology / Neurosurgery Service, Centre for Small Animal Studies, Animal Health Trust, Newmarket, Suffolk, United Kingdom

Current address: Linnaeus Veterinary Ltd, United Kingdom

Tiina Heinonen

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Jim Johnson

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

Lorna J Kennedy

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Andrea D Short

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Sofie F M Bhatti

Ghent University, Faculty of Veterinary Medicine, Belgium

Donna Foster

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Peter Leegwater

Department of Clinical Sciences, Utrecht University, Utrecht, Netherlands

Tarja S Jokinen

Department of Equine and Small Animal Medicine, University of Helsinki, Helsinki, Finland

Rowena M A Packer

Department of Clinical Science and Services, Royal Veterinary College, Hatfield, Hertfordshire, UK

Koen Santifort

Evidensia Small Animal Hospital, Arnhem, Netherlands

Paul Mandigers

Evidensia Small Animal Hospital, Arnhem, Netherlands.

Department of Clinical Sciences, Utrecht University, Utrecht, Netherlands

Holger Volk

Department of Small Animal Medicine and Surgery, University of Veterinary Medicine, Hannover, Germany

Cathryn S Mellersh

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Marjo K Hytönen

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Hannes Lohi

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Sally L Ricketts

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom (Formerly at the Animal Health Trust, Newmarket, Suffolk, United Kingdom)

Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

7.2 Authors contributions

CAJ contributed to study design, funding, performed experimental work and analysis and interpretation of data, and wrote the manuscript. SLR conceived of, designed, and supervised the study, contributed to funding, and provided critical review of the manuscript. CSM contributed to study conception and funding. LDR provided case definition and diagnosis of cases, clinical expertise, and contributed to funding. DF and JJ contributed to sample recruitment. TH, TSJ, MH, and HL contributed genotype data for cases and controls (UH GWAS set). MKH and HL also contributed to funding, and HL to study conception. LJK and ADS contributed to funding and provided samples of cases (UM sample set). SFMB, RMAP, and HV contributed samples of cases and controls (RVC sample set). PL, KS, and PM contributed DNA samples of cases and controls (UU sample set).

7.3 Abstract

Border Collies are reported to have an increased prevalence of idiopathic epilepsy (IE) which is particularly severe and has a considerable impact on quality of life. To identify genetic factors that may underly the development of IE, we conducted a genome-wide association study meta-analysis at both array and whole genome level using data from up to 123 cases and 186 controls from three datasets. We subsequently utilised a replication

set of up to 271 cases and 307 controls from four independent studies, identifying three variants that demonstrated evidence of replicable association with IE. When tested as part of a genotypic risk score, these three SNPs show potential for predicting IE status in the Border Collie breed and implicate two genes with previous evidence of roles in the central nervous system or involvement in human neurological disease (*KLF4* and *CSMD1*) and a third gene for which a neurological role has not previously been reported (*TMEM268*). Future studies based on whole genome sequencing will help identify additional risk loci and functional variants underlying these SNP associations and could aid the development of a DNA-based tool for identifying Border Collies that may have a greater genetic risk of developing IE.

7.4 Introduction

Idiopathic epilepsy (IE) is defined by the International Veterinary Epilepsy Task Force (IVETF) as epilepsy with a known or suspected genetic cause, or with an unknown cause and without any identified, or suspected, structural cerebral pathology [240]. Epilepsy, which is relatively common in dogs, has an estimated prevalence of 0.62% - 0.75% in the general dog population [203, 274] and a much higher prevalence has been reported for specific breeds [251, 252, 256, 410, 411]. The increased prevalence within certain breeds, and the evidence for underlying genetic causes for the common epilepsies in humans [217, 221, 222], suggest that there are genetic variants that confer risk of IE in dogs.

Probable genetic causes have been elucidated for canine monogenic neurodegenerative and structural epilepsies, such as Lafora disease putatively caused by repeat expansions in *NHLRC1* in multiple dog breeds [243, 244, 246, 247], and severe early-onset neurodegenerative epileptic encephalopathy with mitochondrial respiratory deficiency putatively caused by a deletion in *PITRM1* in the Parson Russell Terrier [248].

Although the genetics of canine IE is still mostly unknown, genome-wide association studies (GWAS) have had some success both for rare monogenic epilepsies and, to a lesser extent, for more common IE. A SNP in *LGI2* putatively causes remitting benign familial juvenile epilepsy in the Lagotto Romagnolo dog breed [266, 294], and generalised myoclonic epilepsy in Rhodesian Ridgeback dogs is putatively caused by a 4 bp deletion in *DIRAS1* [268]. A common risk haplotype within the *ADAM23* gene is reproducibly associated with IE in the Belgian Shepherd dog and other breeds [276, 290, 291], and in the Belgian Shepherd dog a potential additional associated locus has been found [292,

303, 304]. GWAS have also identified a locus associated with IE in the Irish Wolfhound [80], and a locus showing suggestive association with IE in the Petit Basset Griffon Vendeen [293]. However, the latter three epilepsy-associated regions have not yet been confirmed in additional independent sample sets.

GWAS findings and pedigree studies suggest that for most canine IE a multigenic, or complex, mode of inheritance is likely [80, 236, 290-293], as is hypothesised for the common epilepsies in humans [222, 232]. A small pedigree study of 43 Border Collies with IE in Germany suggested a strong genetic component for IE inheritance [254], and studies have indicated that the Border Collie breed has an increased prevalence of IE and that Border Collie IE cases may have a poorer prognosis than those of other breeds [203, 265, 271, 273, 426].

Two studies have described the phenotype and clinical characteristics of epilepsy in the Border Collie breed: a study of 49 Border Collies diagnosed with IE in Germany [254], and a more recent study of 116 Border Collie IE cases primarily from the Netherlands but including dogs from Germany and Belgium [255]. The most common seizure type reported for Border Collies is generalised (typically tonic-clonic seizures, although those with a focal onset have also been reported) [254, 255]. Cluster seizures are common in Border Collie cases; 94% had cluster seizures in the earlier study and 59% in the larger, more recent, study [254, 255]. Border Collies have also been found to have an increased probability of cluster seizures when compared to the Labrador Retriever breed [259]. Status epilepticus has been reported to occur in 29% - 53% of cases [254, 255]. The median age of onset observed in the two key Border Collie studies were 2.37 and 2.79 years, but age of onset is variable, and dogs with an earlier onset were reported to have a more severe clinical course [254, 255]. IE can have a considerable impact on quality of life in cases, as reported by owners, and cluster seizures and status epilepticus are associated with a poorer quality of life [255]. Most Border Collies diagnosed with idiopathic epilepsy are administered antiepileptic medication (AEM) treatment continually [254], and more dogs are treated with two or more AEMs than just one [254, 255]. Dogs treated with two or more AEMs are frequently resistant to treatment [254], and owners often report that treatment has adverse effects [254, 255]. Studies have investigated the genetics of refractory epilepsy in Border Collies and other breeds [309, 310, 312], however, evidence for replicable associations with refractoriness is limited.

Polygenic risk scores (PRS) encompassing a combined risk allele count, weighted by measured effect, of many variants, or genetic risk scores comprised of smaller numbers of variants, have been used for estimating overall genetic risk for common human diseases with a polygenic mode of inheritance [412, 413]. Subsequent to GWAS and replication studies, risk scores have been used to identify individuals at highest risk of a complex disease to help inform decisions regarding preventative treatment and/or lifestyle adjustments [412, 413].

The present study sought to identify genetic variants that are reproducibly associated with IE in the Border Collie dog breed and to assess their utility as a predictive tool in a genetic risk score in an independent set of Border Collies. GWAS meta-analyses, utilising data imputed to both high-density array and whole genome sequence (WGS) SNP density, were performed to identify regions associated with IE. Replication and genetic risk score testing were carried out for the IE-associated GWAS SNPs in a large independent set of cases and controls; three SNPs demonstrated replicable association with IE and performed well as a genetic risk score in this set of dogs.

7.5 Methods

7.5.1 Sample collection and definition of cases and controls

Three GWAS case-control sets were utilised; two collected by the Kennel Club Genetics Centre (KCGC), UK, and the third by the University of Helsinki (UH), Finland. In addition, four independent sample sets were employed for replication of GWAS findings. The replication sets were collected by the KCGC; the University of Manchester (UM), UK; the Royal Veterinary College (RVC), UK; and the University of Utrecht (UU), Netherlands.

Additional details regarding the sample sets are given in **Appendix vi.i**. The details of case definition varied between sample sets, but for all except the UM set cases were defined based on IVETF criteria and/or were confirmed by a board-certified veterinary neurologist. UM cases were dogs undergoing treatment with phenobarbital for epilepsy. Controls were dogs never reported by their owner to have had a seizure, over the age of eight in the KCGC sets but age at sample recruitment varied between the other sample sets (**Appendix vi.i**).

KCGC GWAS and replication set samples were collected as buccal swabs or residual whole blood from routine clinical tests (Animal Health Trust Clinical Research Ethics Committee

Project No. 73-2016, University of Cambridge Department of Veterinary Medicine Ethics and Welfare Committee No. CR512). The UH GWAS samples were collected as EDTA blood from privately owned pet dogs (approved by the Animal Ethics Committee of State Provincial Office of Southern Finland, ESAVI/25696/2020). RVC cases were recruited through the RVC's Big Brainy Border Collie study (BBBCS) and the Idiopathic Epilepsy and Anxiety Study (IDEAS), and the controls were recruited through the BBBCS. These samples were collected in the form of buccal swabs (RVC Clinical Research Ethical Review Board: URN 2017 1743-2 (IDEAS) and URN 2018 1799-2 (BBBCS)). UU samples were collected in the form of whole blood samples and the DNA extracted as part of a study published previously that characterised the phenotype of IE in the Border Collie [255]. Dog owners gave informed consent, and the study was approved by the departmental board. The UM DNA samples were part of the UK DNA Archive for Companion Animals held at the University of Manchester [427]. All dogs in this set were cases, residual EDTA whole blood samples were received from a UK-based diagnostic company providing routine screening of serum phenobarbital concentrations in dogs with epilepsy undergoing treatment [273]. All samples were collected with full informed owner consent.

DNA was extracted from buccal swabs for the KCGC GWAS sets, the KCGC replication set, and the RVC samples, using a QIAamp DNA Blood Midi Kit (Qiagen). DNA from KCGC whole blood samples was extracted using a standard chloroform protocol or extracted with the UM residual EDTA blood samples by the service provider used for genotyping the replication set (LGC Genomics Ltd., UK).

7.5.2 Array genotyping

KCGC GWAS Sets 1 and 2, and the UH GWAS set, were genotyped using the Illumina CanineHD BeadChip (173,662 SNPs). Genotype data (individual call rate > 90%) were obtained for 162 dogs (73 cases, 89 controls) in KCGC GWAS Set 1 and 93 dogs (29 cases, 64 controls) in KCGC GWAS Set 2. Data for 55 dogs (21 cases, 34 controls) were utilised from the UH GWAS set.

Axiom Canine HD array genotype data (729,642 SNPs/variants) was acquired for 47 dogs. These were dogs originally genotyped on the Illumina CanineHD BeadChip and part of KCGC GWAS Sets 1 and 2 (33 dogs from KCGC GWAS Set 1, 19 cases and 14 controls; 14 from KCGC GWAS Set 2, 5 cases and 9 controls). These were genotyped at the higher density to be used as part of a reference panel for imputation, or to be used for

calculations of imputation accuracy for KCGC GWAS Sets 1 and 2, as described previously [370] (see chapter 4).

7.5.3 Genotype imputation to higher density array and whole genome sequence level

Genotype imputation was performed twice for each GWAS dataset; to increase SNP density to array and WGS level. The imputation of KCGC GWAS Set 1 and Set 2 to Axiom Canine HD array density using a multi-breed reference panel, SHAPEIT (v2, r904) [363], and IMPUTE2 (IMPUTE v2.3.2) [72, 73], has been described previously [370] (see chapter 4). When imputing the two KCGC GWAS sets up to the array density, dogs overlapping with the reference panel were excluded from the study sets as described previously. The same imputation methodology was used for the UH GWAS set, except for the inclusion of 10 additional Border Collies (89 total, up from 79) in the reference panel.

The three GWAS datasets were imputed to WGS-level using a higher density reference panel that included all SNPs (that passed quality control filtering) present in the WGS data. The WGS data used for imputation included the same in-house WGS and publicly available WGS from the Dog Biomedical Variant Database Consortium (DBVDC) that was used for the multibreed component of the Axiom Canine HD array-level imputation [370] (see chapter 4) [67]. Additional in-house WGS acquired since the construction of the reference panel for the array-level imputation, enabled inclusion of up to 33 additional dogs and representation of a greater number of breeds (**Appendix vi.ii**). A separate reference panel was produced for each chromosome independently due to the size of the datasets. As described previously [370] (see chapter 4), SNPs were filtered to keep only biallelic SNPs with a quality score > 20, genotyped in > 97% of the WGS dogs, and with minor allele frequency (MAF) > 1%. SNPs for which the strand was ambiguous (i.e., those between complementary bases, A to T and C to G) were excluded. Individuals lacking data for > 10% of SNPs on a chromosome were excluded from the reference panel for that chromosome. The reference panels were produced using a pool of individuals, with different numbers for each chromosome due to missingness filtering; however, all chromosome-specific reference panels for each study set included the same Border Collies. The same chromosome-specific reference panels were used for both KCGC GWAS sets. The total number of SNPs across all chromosomes, and the number of individuals included, for each reference panel are shown in **Appendix vi.iii**. The KCGC and UH GWAS

reference panels included dogs of 160 breeds, 15 mixed breed dogs, 28 Chinese indigenous dogs, and 8 wolves.

Haplotype phasing of study sets and reference panels for both array and WGS SNP density was carried out using SHAPEIT, and imputation was carried out using IMPUTE2, utilising a publicly available canine genetic map [38, 370] (see chapter 4). A window size of 2 Mb was employed for haplotype phasing of the study sets and the array-density reference panels, whereas 0.5 Mb windows were used for those with WGS SNP-density [76]. The effective population size was set at 200 for both phasing and imputation [76].

The accuracy of the array-density imputation for KCGC GWAS Sets 1 and 2 was assessed previously [370] (see chapter 4). IMPUTE2 provides an estimation of imputation accuracy by masking known genotypes before imputing them, then comparing the imputed genotypes to the original data and producing a concordance statistic. For the array-density imputed UH GWAS dataset the mean estimated concordance across chromosomes was 98.1% (range 97.2% to 98.7% per chromosome). A summary of the IMPUTE2 concordance statistics for the WGS-density imputation is given in **Appendix vi.iv**.

7.5.4 GWAS meta-analyses

GWAS analysis with a linear mixed model to correct for population stratification was conducted independently for each dataset using GEMMA [54]. SNPs were filtered to remove those with a Hardy-Weinberg P-value < 5×10^{-5}, those with a SNP call rate < 97%, or with a minor allele frequency (MAF) < 5%. SNPs were also filtered to exclude those with an IMPUTE2 'Info' statistic (imputation certainty) < 0.5 [72, 73]. A summary of the dogs and SNPs remaining in each dataset after filtering is provided in **Appendix vi.v**. Sex was included in the analyses as a covariate to adjust for an observed bias towards male cases.

A fixed-effects meta-analysis was carried out using the statistical software package Stata (Stata 15. College Station, TX, USA), for the SNPs common to the three GWAS sets, using the summary statistics produced by GEMMA. Heterogeneity was assessed using the Q statistic. The array SNP density meta-analysis included 104 cases, 167 controls, and 291,450 SNPs; the WGS SNP density meta-analysis included 123 cases, 186 controls, and 5,993,069 SNPs.

To identify any independent associations, genome-wide association analysis at array density as described above was repeated, conditioning on the 16 SNPs identified in the initial analysis, with these SNPs included as covariates. The meta-analysis was then repeated with the summary statistics generated.

7.5.5 Validation of GWAS associations in independent replication sets

SNPs identified in the GWAS meta-analyses were genotyped in the independent replication sets by LGC Genomics Ltd., UK using competitive allele-specific PCR (KASP) [428]. Quality control and analyses were carried out using Stata (Stata 15. College Station, TX, USA). Genotype clusters were examined for SNPs with a Hardy-Weinberg P-value < 0.05 to confirm assay quality. Logistic regression analysis was carried out for each SNP, for the GWAS (to obtain a P-value not corrected for population effects for direct comparison) and replication sets independently. Imputed genotype probabilities for the GWAS dogs were converted to hard genotype calls using PLINK (v1.90) [368] and genotype calls with uncertainty > 0.1 were treated as missing. Logistic regression analysis of the replication set was adjusted for study as a categorical variable to account for any between-study variability, with the UM cases as part of the KCGC replication set.

7.5.6 Generating a three-SNP risk score and testing it in the replication set

Weighted three-SNP risk scores were calculated for each dog in the replication set as follows:

$$3 \times \frac{\beta 1 \times SNP1 + \beta 2 \times SNP2 + \beta 3 \times SNP3}{\beta 1 + \beta 2 + \beta 3}$$

'β' represents the beta-coefficient from the meta-analysis, 'SNP' the number of copies of the risk-conferring allele, for each SNP. If the beta-coefficient was negative (calculated for the non-risk allele) a positive beta-coefficient coded for the risk allele was calculated (× -1).

To assess risk score performance in the replication set the receiver operating characteristic (ROC) analysis was performed as described previously (see chapter 6).

7.6 Results

7.6.1 GWAS meta-analyses

None of the SNPs in the array-density, or the secondary WGS-density, IE GWAS meta-analyses passed the Bonferroni-corrected threshold for statistical significance (Array-density $P < 1.7 \times 10^{-7}$, WGS-density $P < 8.3 \times 10^{-9}$) (**Figure 7.1** and **Figure 7.2**). As this could be due to a lack of statistical power, thresholds of $P < 1 \times 10^{-4}$ and $P < 1 \times 10^{-5}$ were used for the array-density and WGS-density imputed meta-analyses respectively, to select SNPs for follow-up. The array-density analysis identified at least 16 genomic regions, each on a different chromosome, with SNPs passing the threshold (**Figure 7.1**). The most significantly associated SNPs from each of these 16 regions was included as a covariate in a second, conditional, analysis at array density to identify any independent associations on these 16 chromosomes. An additional SNP independently associated with IE was identified on each of chromosomes 1, 8, 11, and 37 (**Appendix vi.vi**). An additional 12 associated SNPs passing the empirical threshold were identified in the WGS-density meta-analysis (**Figure 7.2**). In total, 32 SNPs from the GWAS meta-analyses were considered to be taken forward for genotyping in the replication sets (**Appendix vi.vii**). The SNP on chromosome 6 was excluded from further analysis due to high heterogeneity across study sets ($P < 0.01$).

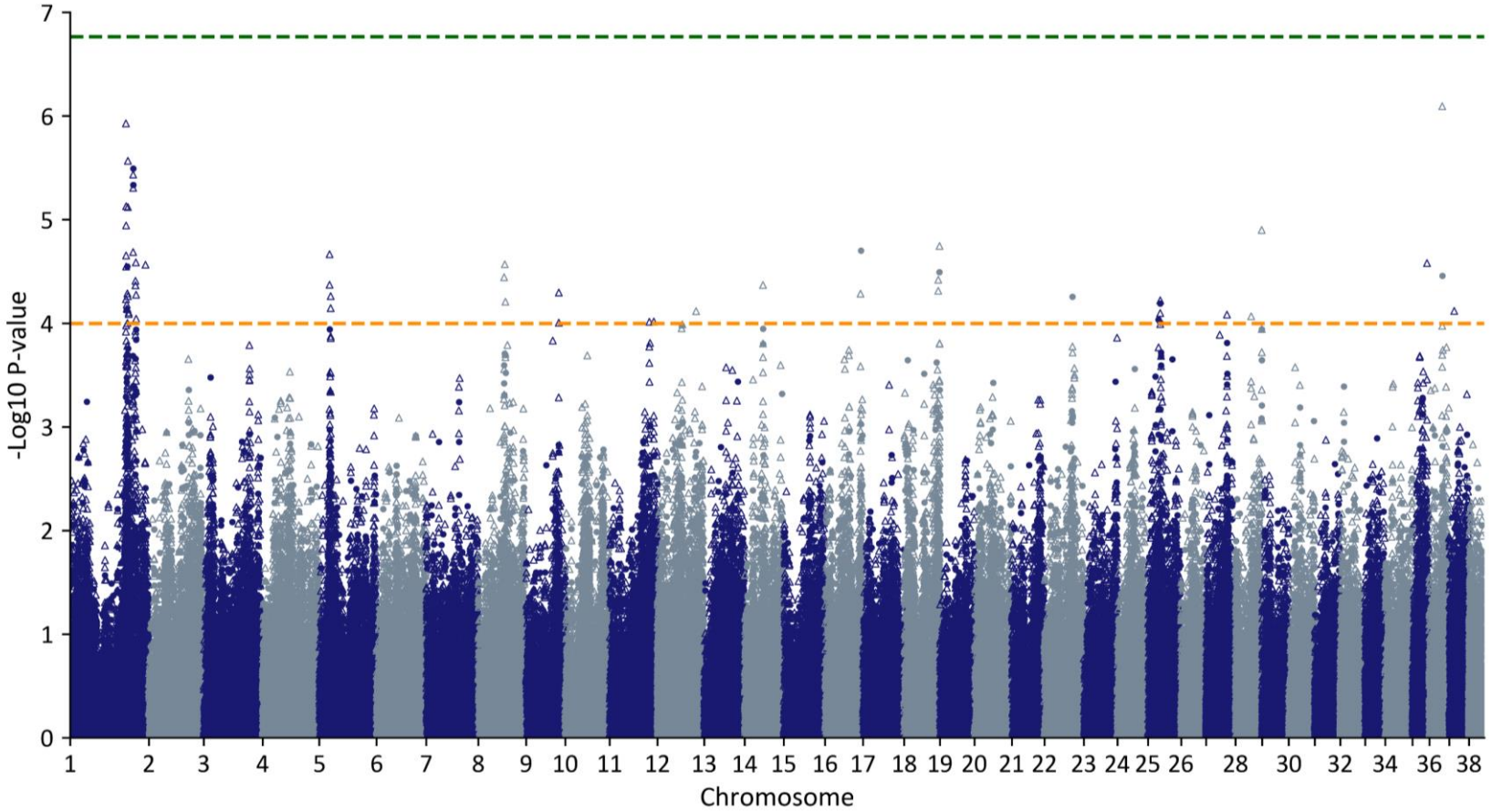


Figure 7.1. Array SNP density imputed genome-wide association meta-analysis of 104 Border Collie IE cases and 167 controls (291,450 SNPs). Plot of negative log (base 10) transformed P-values. X-axis is SNP location by chromosome (left to right, autosomes 1 to 38). Solid circles indicate array genotyped SNPs, hollow triangles denote SNPs that were imputed for any of the three datasets. Green (upper) line shows Bonferroni corrected threshold for statistical significance (1.7×10^{-7}). Orange (lower) line indicates the empirical threshold for suggestive association (1×10^{-4}).

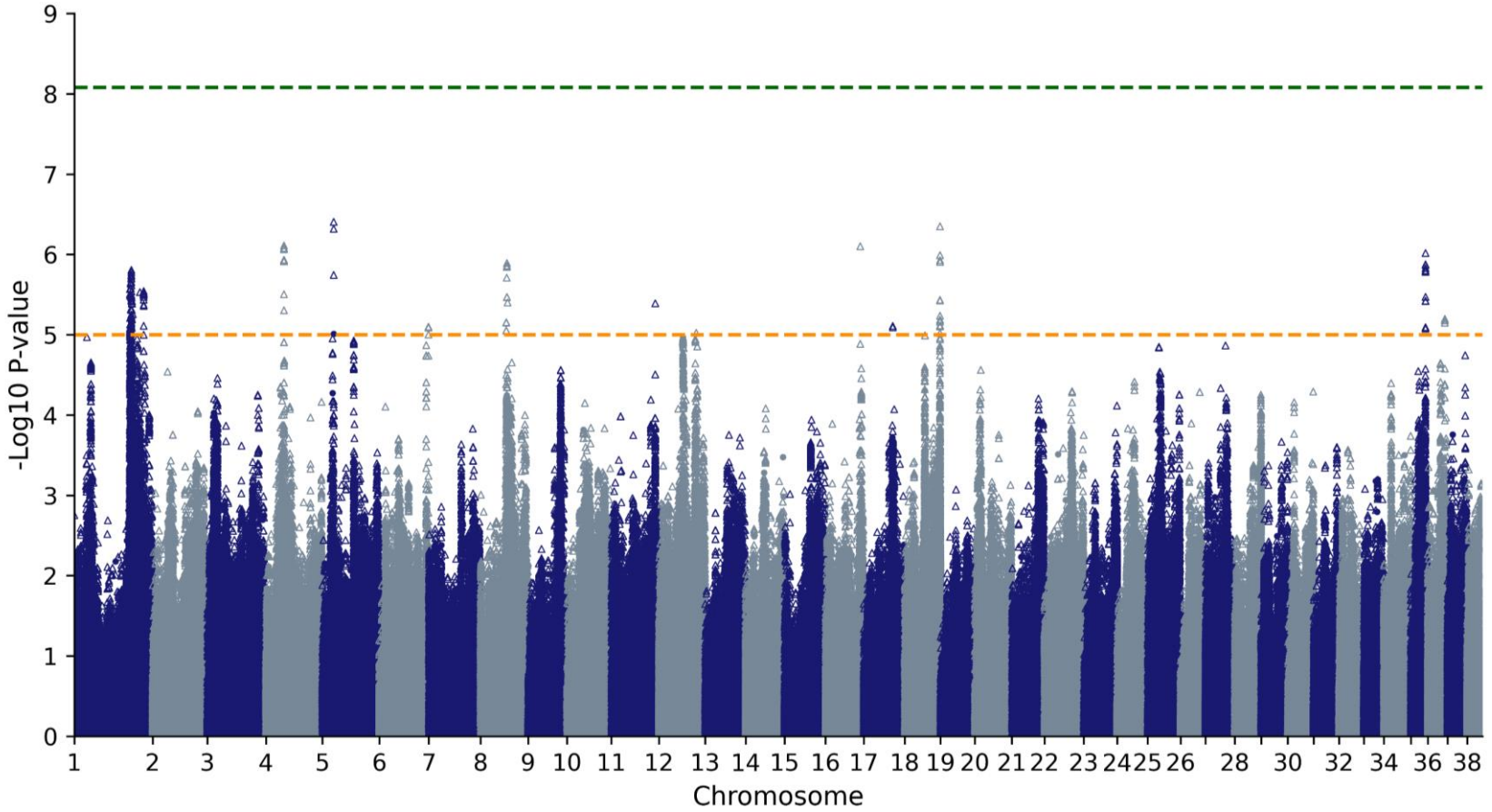


Figure 7.2. WGS SNP density imputed genome-wide association meta-analysis of 123 Border Collie IE cases and 186 controls (5,993,069 SNPs). Plot of negative log (base 10) transformed P-values. X-axis is SNP location by chromosome (left to right, autosomes 1 to 38). Solid circles indicate array genotyped SNPs, hollow triangles denote SNPs that were imputed for any of the three datasets. Green (upper) line shows Bonferroni corrected threshold for statistical significance (8.3×10^{-9}). Orange (lower) line indicates the empirical threshold for suggestive association (1×10^{-5}).

7.6.2 Three SNPs show reproducible association with IE risk in independent sample sets

Genotype data were obtained for 27 SNPs, across 16 chromosomes, in the replication set. The second most associated SNP on chromosome 4 (CanFam3.1 4:28195881), which is 5,970 bp upstream and in linkage disequilibrium (LD) (R^2 = 0.93) with the top SNP, was genotyped as a proxy for CanFam3.1 4:28189911. Three SNPs were excluded from analysis because they failed genotyping or an assay could not be developed successfully; CanFam3.1 8:41418371, CanFam3.1 16:54266694, and CanFam3.1 27:32560742. The most strongly associated SNP on chromosome 17 (CanFam3.1 17:45586797) is yet to be genotyped.

While none of the SNPs showed an association with IE in the replication set as strong as in the GWAS; four SNPs showed directional consistency; were close to statistical association in the replication set (P-value < 0.1); and their odds ratio point estimates were not outside the confidence intervals of the discovery set (which would indicate high heterogeneity) (Appendix vi.viii and Table 7.1). Two of these SNPs (CanFam3.1 11:68783067 and 11:68785445) were located within the same gene on chromosome 11 (*TMEM268*, also referred to as *C9orf91*); 2,378 bp apart, and in LD (R^2 = 0.98). The SNP identified through the array-density analysis, CanFam3.1 11:68785445, was selected for the subsequent risk score analysis due to the slightly stronger statistical association with IE in the replication set.

Table 7.1. Results of the Border Collie IE replication study for the four SNPs that showed evidence of replication.

	Three GWAS sets			**Replication sets**		
Genomic pos. [a]	ca/co [b] (n)	Odds Ratio (95% CI)	P-value	ca/co [b] (n)	Odds Ratio (95% CI)	P-value
11:62873725*	102/160	1.74 (1.19-2.54)	3.7×10^{-3}	268/304	1.25 (0.97-1.61)	0.09
11:68783067	94/132	1.79 (1.23-2.62)	2.1×10^{-3}	270/306	1.28 (0.99-1.66)	0.05
11:68785445*	104/164	1.86 (1.28-2.71)	9.1×10^{-4}	271/306	1.29 (1.00-1.67)	0.05
16:55920429*	104/167	1.97 (1.32-2.94)	8.1×10^{-4}	271/306	1.39 (1.04-1.84)	0.02

Results from logistic regression analysis of the three GWAS sets and the replication sets. [a] The CanFam3.1 genomic location of the SNP in the format chromosome: bp position. [b] case/control. * SNPs taken forward to be tested as a weighted risk score. CI: confidence interval.

7.6.3 A three-SNP risk score is predictive of IE status in replication sets

Three SNPs that showed evidence of replication (**Table 7.1**), and that represented independent genomic regions, were included in a three-SNP weighted risk score for IE. The risk score was tested in the replication set (268 cases and 302 controls) to investigate the feasibility of developing a genetic tool predictive of the risk of a dog developing IE. Analysis using a ROC curve and calibration plot suggested potential utility for the SNPs as a risk score for predicting IE risk. The ROC AUC of 0.71 (95% CI: 0.67-0.75) indicates that the risk score could discriminate cases from controls in the replication set. On the calibration plot, the points for groups and the locally weighted scatterplot smoothing (LOWESS) line closely followed the reference line, suggesting that the predicted risk reflected the observed proportion of affected dogs within each group (**Figure 7.3**). The risk score had a stronger association with IE than each SNP independently in the replication set (odds ratio: 1.31; 95% CI: 1.12-1.54; $P = 5.3 \times 10^{-4}$). The unweighted count of risk alleles was also tested for predictiveness and association with IE and gave similar results (AUC = 0.72, 95% CI: 0.67 - 0.76; odds ratio 1.31; 95% CI: 1.12-1.54; $P = 5.7 \times 10^{-4}$) (**Appendix vi.ix**). The risk allele frequencies in each set of cases and controls, and in the GWAS and replication sets independently and combined, are shown in **Table 7.2**.

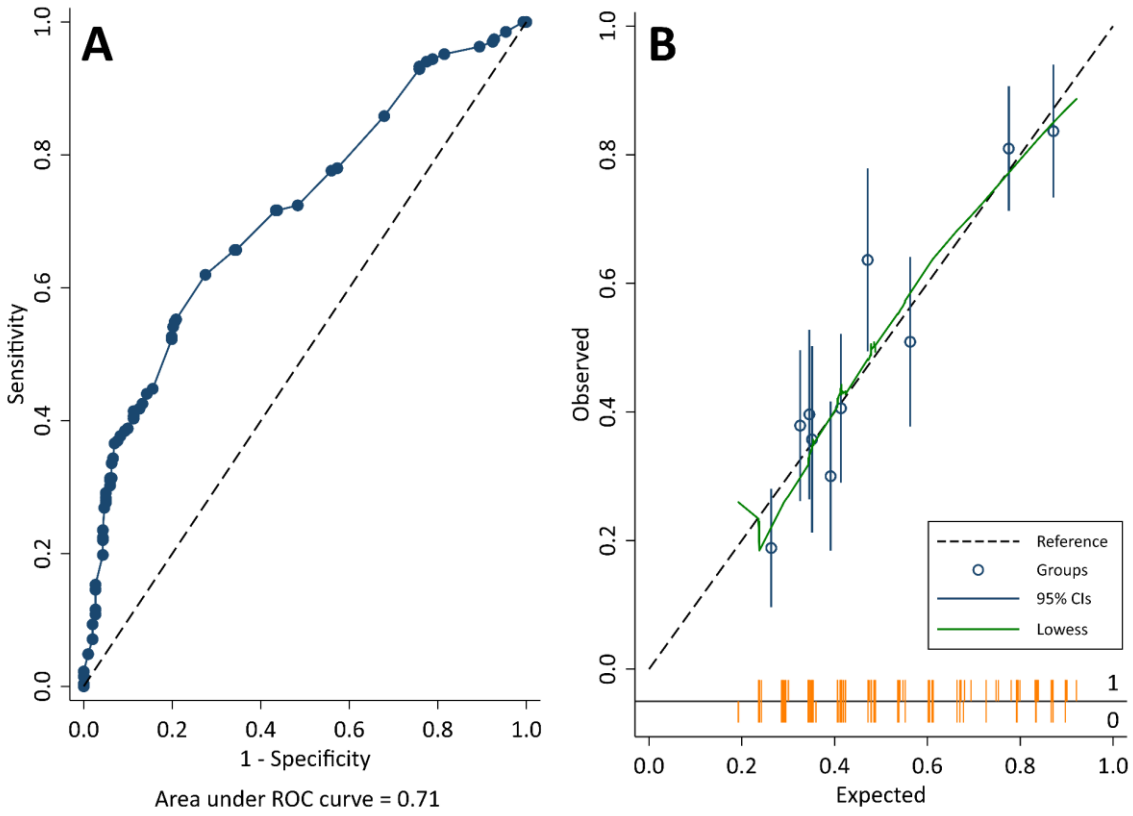


Figure 7.3. Receiver operating characteristic (ROC) curve and calibration plot for a three-SNP risk score for idiopathic epilepsy in Border Collie replication sets. Plot A is a ROC curve; points represent each potential risk score cut off for defining cases, from the highest (0,0) to the lowest (1,1). Sensitivity: fraction of cases correctly classified. Specificity: fraction of controls correctly classified (1 – (minus) specificity is the false-positive fraction). The area under the ROC curve (AUC) is given below the plot. An AUC of 0.5 (indicated by the dashed line) would represent a test unable to discriminate cases from controls. On the calibration plot (B) points represent ten equally sized groups of individuals divided by predicted risk. Observed: the proportion of cases in each group. Expected: the average (mean) of the predicted probabilities generated from the risk score logistic regression model. The 95% confidence intervals are shown for each group. The dashed reference line indicates perfect risk score calibration where predicted risk matches the observed proportion of affected dogs within each group. Locally weighted scatterplot smoothing (LOWESS) is displayed in green. The orange lines at the base of the graph are a spike plot indicating the distribution of IE cases (1) and controls (0).

Table 7.2. Risk allele frequency for three risk score SNPs in cases and controls across the sample sets used for GWAS and replication

	11:68785445		**11:62873725**		**16:55920429**	
Study	**ca/co (n)**	**Risk allele freq. (ca/co)**	**ca/co (n)**	**Risk allele freq. (ca/co)**	**ca/co (n)**	**Risk allele freq. (ca/co)**
KCGC GWAS Set 1	57/72	0.77/0.61	56/71	0.71/0.56	57/73	0.30/0.21
KCGC GWAS Set 2	26/60	0.73/0.66	25/57	0.68/0.60	26/60	0.48/0.20
UH GWAS	21/32	0.79/0.63	21/32	0.79/0.69	21/34	0.26/0.25
Combined GWAS	104/164	0.76/0.63	102/160	0.72/0.60	104/167	0.34/0.21
KCGC and UM Replication	155/267	0.75/0.67	153/265	0.70/0.62	155/266	0.32/0.25
RVC	37/24	0.72/0.63	36/24	0.63/0.65	37/25	0.27/0.24
UU	79/15	0.70/0.90	79/15	0.62/0.70	79/15	0.28/0.20
Combined Replication	271/306	0.73/0.67	268/304	0.67/0.63	271/306	0.30/0.25
Combined GWAS and Replication	375/470	0.74/0.66	370/464	0.68/0.62	375/473	0.31/0.24

7.6.4 Genomic regions in LD with risk score SNPs implicate multiple genes with potential involvement in conferring IE risk

Regions of LD around the three SNPs included in the risk score were identified using the WGS SNP density datasets (**Figure 7.4**). The regions were defined by identifying SNPs 'tagging' the target SNPs (i.e. those with an $R^2 \geq 0.8$). A 165.1 kb region of LD was found for CanFam3.1 11:62873725 (CanFam3.1 11:62,830,376-62,995,472). The region includes the *KLF4* gene, *U6* spliceosomal RNA gene, and a gene annotated with only an ENSCAF ID (**Figure 7.4A**). Two genes were annotated for the 27.9 kb LD region (CanFam3.1 11:68,777,648-68,805,534) for the second chromosome 11 SNP, CanFam3.1 11:68785445, *TMEM268* (also referred to as *C9orf91*) and *TEX48* (**Figure 7.4B**). Most tagging SNPs were located within *TMEM268*. An 82.2 kb region within an intron of the *CSMD1* gene (CanFam3.1 16:55,916,800-55,998,962) was found to be in LD with CanFam3.1 16:55920429 (**Figure 7.4C**).

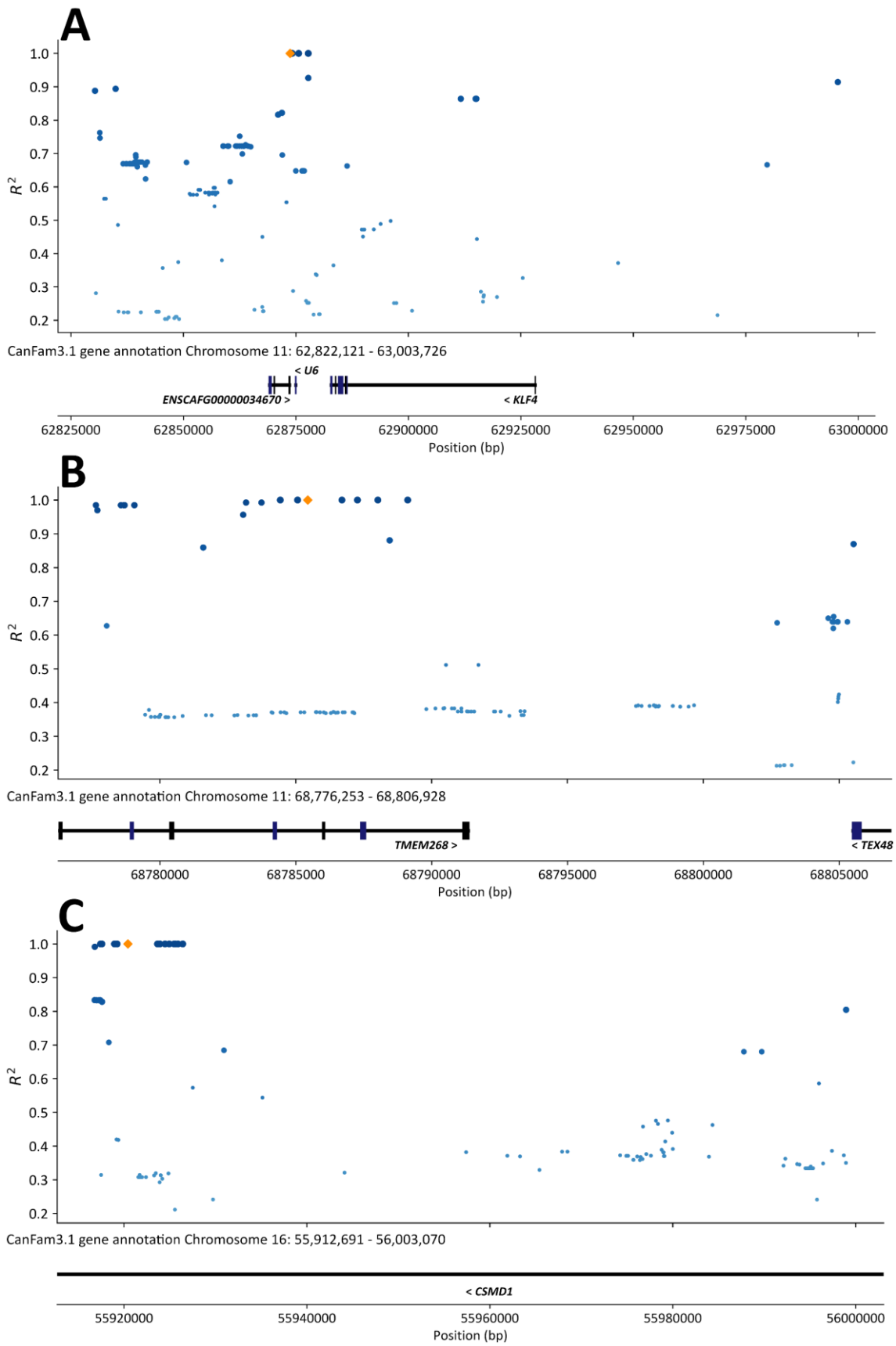


Figure 7.4. Regional plots of three loci that show evidence of reproducible association with IE in the Border Collie. Plots A, B, and C depict the SNPs in the genomic regions of SNPs CanFam3.1 11:62873725, CanFam3.1 11:68785445, and CanFam3.1 16:55920429, respectively (represented by orange diamonds). Each point represents a SNP, with the size

an additional indictor of the extent of linkage disequilibrium with the target SNP (largest circles $R^2 = 1$; then successive sizes for $R^2 \geq 0.8$, $0.6 \leq R^2 < 0.8$, and $R^2 < 0.6$). R^2 statistics were calculated within the WGS SNP density imputed datasets using PLINK, and plots include only those SNPs with an $R^2 \geq 0.2$ that were present in all three study sets and that were included in the meta-analysis. LD regions were defined by the lowest and highest base pair (bp) position of SNPs with an $R^2 \geq 0.8$ in relation to the target SNP (limited to SNPs within 250 kilobase (kb)). The genes shown below the regional R^2 plots are the CanFam3.1 gene annotations for the region, plotted against the genomic bp position. Exons are indicated by wider lines. Genes labelled with '>' are encoded on the positive/ forward strand, those with '<' are on the negative/ reverse strand. The gene annotations and regional plot of LD share an X-axis.

7.7 Discussion

Through GWAS meta-analysis and replication in an independent sample set, this investigation has identified three SNPs, representing three genomic regions across two chromosomes, that exhibit evidence of a reproducible association with IE in the Border Collie. When grouped together in a genotypic risk score the SNPs demonstrate potential utility as a predictive tool for IE in the breed. To validate the risk score's ability to predict IE risk it will be necessary to conduct testing in an independent set of cases and controls using a weighted risk score generated from effect estimates of the replication set. In the current analysis the weighted risk score is comprised of effect estimates from the discovery set and is therefore susceptible to the 'winners curse' phenomenon [421]. However, a non-weighted analysis based on genotypic counts for the three SNPs was not materially different to a weighted score, suggesting that the findings are robust.

Two SNP associations identified have regions of LD containing compelling candidate genes for involvement in idiopathic epilepsy. Chromosome 11 SNP CanFam3.1 11:62873725 is located less than 8 kb away from the Krüppel-like factor 4 (*KLF4*) gene. This zinc finger-containing transcription factor is involved in the regulation of a diverse range of genes and cellular processes [429, 430], and has been of considerable scientific interest due to its use in the induction of pluripotency in cells [431-433]. Studies have demonstrated functions for *KLF4* in the central nervous system (CNS) where it has been shown to repress axon growth [434], and to have roles in pathways that regulate neurogenesis (the generation of neurons from neural progenitor cells) and synaptogenesis (the formation of

synapses) [435-437]. *KLF4* has also been suggested to have a role in the regulation of neuroinflammation [438, 439], and to be involved in the pathways behind sedation [440]. The downregulation of *KLF4* has also been implicated in epileptogenesis in a mouse model of epilepsy, and its overexpression demonstrated to suppress seizures and synaptic plasticity in the model [441]. A *U6* spliceosomal RNA gene is also annotated within the LD region found for CanFam3.1 11:62873725. U6 is an important component of the spliceosome which is a ribonucleoprotein complex that facilitates gene splicing, the process of removing introns, and ligating exons, in precursor messenger RNA [442]. Mammalian genomes contain many copies of the U6 gene [443], the majority of which are thought to be pseudogenes [444].

The second SNP association that is a compelling candidate for involvement in idiopathic epilepsy is chromosome 16 SNP CanFam3.1 16:55920429. This SNP, and the 82.2 kb region of LD identified, is located within an intron of the *CSMD1* gene. CUB and Sushi multiple domains 1 (*CSMD1*) is highly expressed in epithelial tissues and the CNS during development, most notably in the nerve growth cone (which is the leading edge of the growing neuron), and is expressed in the adult brain including the cerebral cortex, cerebellum, and white matter [445, 446]. *CSMD1* inhibits the complement pathway, a part of the innate immune system which promotes inflammation and can have both beneficial and damaging effects on the CNS including neuropsychiatric and neurological diseases [445-447]. Neuroinflammation has been linked to epilepsy and epileptogenesis [447], and a difference in plasma complement biomarkers between epilepsy patients and controls, and patients with controlled or uncontrolled seizures, has been observed [448]. Variants within the *CSMD1* gene have been implicated in neurodevelopmental and neuropsychiatric disorders, and neurological disease; including schizophrenia [449], bipolar disorder [450], autism spectrum disorder [451], familial Parkinson disease [452], ataxia with cerebellar hypoplasia [453], and cerebellar agenesis [454]. Studies that identified a translocation interrupting the coding sequence of *CSMD1* [455], and an 18.5 kb deletion in the region containing *CSMD1* [456], have linked the gene to epilepsy in humans, further strengthening its candidacy as a gene with involvement in canine IE.

Chromosome 11 SNP CanFam3.1 11:68785445, and most tagging SNPs in the LD analysis, are located within *C9orf91*; in recent years identified in humans to be the transmembrane protein 268 (*TMEM268*) gene [457]. There is a scarcity of publications regarding the gene, reflecting its relatively recent characterisation. TMEM268 has been found to be mostly

present at the endoplasmic reticulum and plasma membrane, interacts with integrin subunit β4 (ITGB4), and is thought to have a role in cell growth and cell adhesion [457].

The SNPs associated with IE risk identified here are array-based SNPs and are unlikely to be the variants that confer disease risk themselves. Instead, it is likely that they are in LD with, or 'tag', the genomic variants that cause individual Border Collies to have an increased risk of developing IE. The three genomic regions implicated in IE risk by the SNPs identified in this study will be interrogated in future WGS-based studies to identify potential functional variants using *in-silico* approaches. The statistical association of the individual SNPs that we identified from the replication study was weak, not surviving correction for multiple testing. One potential explanation for this is that the risk allele frequencies were high in controls, particularly for the two SNPs located on chromosome 11, indicating that risk alleles are common in the general population. In addition, the GWAS approach used in this study is unable to identify variants that are very common or fixed (i.e. present in all individuals) in the breed and which therefore contribute to the breed's overall increased risk of IE; or rare or *de novo* high-impact variants that may cause IE in an individual dog. Future work utilising WGS approaches and large comparative datasets could be used to overcome this limitation, and this will be facilitated by the gathering of larger WGS-based datasets to complement existing case-control studies.

In conclusion, through GWAS meta-analysis and replication in a large independent set of cases and controls, the present study identified three variants that demonstrate evidence of reproducible association with IE in the Border Collie and are predictive of IE status when included in a genotypic risk score. The implicated genomic regions include genes which represent provocative candidates for IE risk. These findings represent a foundation on which to build a better understanding of the underlying genetics of IE in the Border Collie dog breed.

7.8 Acknowledgements

This study was funded by the PetPlan Charitable Trust (project numbers S13.11, S18-688-726 and S21-959-998). CAJ, SLR and CSM are supported by the Kennel Club Charitable Trust. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the Axiom array data. We sincerely thank all owners and breeders for providing samples from their dogs for this study. We thank Janine Lamb, University of

Manchester, for critical review of the manuscript. We thank Artitaya Lophatananon for advice regarding the assessment of risk score accuracy. We thank Ellen Schofield for bioinformatics assistance. We also thank our collaborators at Linneaus practices for help with sample collection.

7.9 References

See chapter 9

7.10 Supplementary tables and figures

See chapter 10

8 General discussion

This PhD aimed to identify genetic factors contributing to the risk of developing IE and movement disorders in the dog, using methodology tailored to the hypothesised modes of inheritance to investigate these distinctive but sometimes phenotypically overlapping neurological disorders. The PhD utilised a broad range of genetic, molecular, and computational approaches to meet this aim; exploring and implementing methodology that was novel to the research group and developing tools that will have utility for future research.

The PhD met all its objectives:

1. To whole genome sequence two Norwegian Buhund siblings affected by cerebellar ataxia and perform analyses to identify and characterise potentially causal variants.
2. To validate and implement genome-wide genotype imputation to impute Illumina CanineHD datasets for the Border Collie and Italian Spinone up to the genotype density possible through the Axiom Canine HD array, assessing the effect of breed and reference panel size on imputation accuracy.
3. To carry out a GWAS of PxD in the Norwich Terrier and to investigate any disease-associated loci in a large population-based dataset.
4. To carry out a GWAS of IE in the Italian Spinone and to investigate any disease-associated loci in independent sample sets.
5. To carry out a GWAS and meta-analysis of IE in the Border Collie and to investigate any disease-associated loci in a large independent case-control set.

Whole genome sequencing of two Norwegian Buhund siblings affected by cerebellar ataxia successfully identified a variant associated with the disease in a gene not previously associated with ataxia in any species (chapter 3). Emerging publicly available WGS variant datasets were essential to limit the number of candidate genetic variants to a number practical for follow up, without filtering by candidacy, to allow the variant within the novel ataxia gene to be identified. A combination of molecular and *in silico* approaches were used to support the findings.

Genome-wide genotype imputation of Illumina CanineHD datasets to the genotype density of the Axiom Canine HD array was validated (chapter 4). This approach was first

implemented for the Border Collie and Italian Spinone, and the effect of breed and reference panel size on accuracy was assessed. Imputation accuracy was found to be high, and imputation was found to be an effective tool for canine genetic research, allowing meta-analysis of genotype data generated using different arrays. Within this PhD project, the array density imputed data was subsequently used for IE GWAS in the Italian Spinone and Border Collie (chapters 6 and 7); imputation to array density was also utilised for the Norwich Terrier (chapter 5), and imputation to WGS SNP density was carried out for the Border Collie (chapter 7).

This PhD project used a GWAS approach to investigate neurological diseases in three breeds: PxD in the Norwich Terrier (chapter 5), and IE in the Italian Spinone and Border Collie (chapters 6 and 7). The GWAS of PxD in the Norwich Terrier, including 24 cases and 24 controls, identified disease-associated loci which were then investigated in a population-based dataset of 232 Norwich Terriers (including 10 known cases and 45 confirmed controls), implicating five genomic regions for future investigation. A GWAS meta-analysis of IE in the Italian Spinone (52 cases and 51 controls) identified disease-associated loci; however, investigation in independent sample sets was unable to identify variants with replicable association with the disease. GWAS meta-analysis of IE in the Border Collie (up to 123 cases and 186 controls) and investigation of disease-associated loci in a large independent case-control set (up to 271 cases and 307 controls) identified three SNPs demonstrating evidence of replicable association with IE and potential for predicting IE status when included in a weighted risk score.

In the remainder of this chapter the wider implications of this PhD research will be discussed. The strengths and limitations of the methodology that was used will be considered, along with the potential impact of the findings on the genetics of canine and human neurological disease, and what the results suggest should be taken into consideration for future research.

8.1 *KCNIP4*: new evidence suggests a role for the gene in human neurological disease

Since publication of this PhD project's findings proposing *KCNIP4* as a candidate gene in cerebellar ataxia in dogs and other species, the gene has been implicated in a movement disorder in humans [458]. WGS of patients with unexplained developmental and epileptic encephalopathies found that a patient with a previously reported rare pathogenic variant

in *KCNA2* also had a *de novo* nonsense variant in *KCNIP4*. The patient had early onset seizures, intellectual disability, dystonia, and ataxia. The variant within *KCNIP4* was suggested to underly the more severe intellectual disability and ataxia seen in this patient in comparison to patients with only a variant within *KCNA2* [458]. This validates *KCNIP4* as a candidate gene for cerebellar ataxia and demonstrates the wider relevance of movement disorders in the dog as a naturally occurring disease model. Two other recent studies have associated *KCNIP4* with epilepsy [459, 460].

The evidence supporting a role for *KCNIP4* and voltage-gated potassium channels in canine and human neurological diseases raises the possibility of therapeutic implications in both species [326, 458-461]. Compounds that target voltage-gated potassium channels and KCNIP proteins and modulate A-type potassium currents have been demonstrated [462, 463], suggesting that these proteins could be pharmacological targets for canine and human cerebellar ataxia and other neurological diseases in which *KCNIP4* has been implicated.

8.2 DNA-based tools for dog breeders and potential for future clinical use

One of the key goals for most research investigating the genetics of canine inherited diseases is that it will eventually lead to DNA-based breeding tools that will help reduce disease incidence. This is the most straightforward in the case of fully penetrant monogenic disorders, particularly those which are autosomal recessive or dominant, where the objective of DNA-testing is to allow dog breeders to make informed breeding decisions to avoid producing affected dogs and to reduce the disease-causing allele frequency within the population [26, 33].

Between the launch of the DNA test for cerebellar ataxia in the Norwegian Buhund on 4th December 2017, and the closure of the Animal Health Trust in July 2020, the Animal Health Trust DNA testing service tested only 29 Norwegian Buhunds for the *KCNIP4* variant. The Norwegian Buhund is a numerically small breed (7-45 UK Kennel Club registrations a year 2012-2021, median 20.5) [316] (https://www.thekennelclub.org.uk/media-centre/breed-registration-statistics/, accessed 16/01/2023), and other DNA testing providers were available. Two of the 29 dogs tested were heterozygous for the variant; the remaining 27 were all homozygous for the reference allele. The results suggested that although the variant was still present in the UK Buhund population at the time, it was not common. As of June 2022 our collaborators

at Wisdom Health had tested 1,731,985 DNA samples, mostly from mixed breed dogs and dogs of other breeds but including six from Norwegian Buhunds, and none of the dogs had any copies of the variant (Oliver Forman, personal correspondence, June 2022). This suggests that the putatively disease-causing allele is rare and most likely specific to the Norwegian Buhund breed, supporting the findings of the study that is part of this PhD. Despite the huge number of dogs that have now been tested, the only dogs to have ever been found to be homozygous for the ataxia-associated allele are the four cerebellar ataxia cases. The availability of the DNA test, the apparent low frequency of the putative disease-causing allele, and the absence of the mutation from other breeds, means that the disease could feasibly be quickly excluded from the population through the avoidance of breeding carriers.

The findings from this PhD project's research focusing on PxD and IE indicate that, in the breeds investigated, these disorders are unlikely to be monogenic. Five potentially PxD-associated genomic regions were identified for the Norwich Terrier for future investigation, and three SNPs in distinct genomic regions show evidence of replicable association with IE in the Border Collie. So far, attempts to identify variants reproducibly associated with IE in the Italian Spinone have been unsuccessful; however, this provides evidence that IE is unlikely to be a single gene autosomal recessive or dominant disease in the breed, for which the study was expected to have had sufficient power to detect associated variants. One potential DNA-test based methodology for providing breeding advice to dog breeders for oligogenic or polygenic inherited canine disorders could be the use of weighted risk scores. Dogs with an increased risk allele load, above a threshold which would need to be determined through research, could be either avoided when making breeding decisions or only mated with dogs with a risk allele load below a predetermined threshold. Such an approach could be used to reduce risk allele frequency within a population over time.

The risk alleles for two of the three IE-associated SNPs identified for the Border Collie were observed to be common in controls as well as cases (chapter 7). The risk allele frequencies for CanFam3.1 11:68785445 (G) and 11:62873725 (T) were 0.66 and 0.62 respectively in the combined sets of controls included in the study. This represents a challenge for any future attempts to reduce allele frequency within the breed through DNA-testing. A cautious approach using a weighted risk score, as outlined above, that slowly worked to reduce risk-allele frequencies over time might be most effective.

However, dog breeders are much more familiar with DNA test results for autosomal recessive diseases and are accustomed to receiving the advice to avoid mating two dogs which both carry a disease-causing variant. Such an approach should obviously be avoided for the risk-conferring variants underlying the SNPs identified in the Border Collie study; if it was applied, a large proportion of the breeding population would be excluded, this added selection likely having an undesirable impact on genetic diversity [11]. The resultant loss of genetic diversity could have adverse consequences, such as those discussed elsewhere in this thesis regarding the effect of genetic bottlenecks and selection including extensive homozygosity and the increased frequency of other disease conferring variants. The Border Collie breed is not small numerically (1679-2144 registrations a year 2012-2021, median 2070.5) [316] (https://www.thekennelclub.org.uk/media-centre/breed-registration-statistics/, accessed 16/01/2023), although the effective population size (i.e. the breeding population) of breeds is not necessarily proportional to registration numbers [11]. Increased inbreeding through drastic selection away from common risk conferring variants would still be a matter of concern, and if it occurred in a numerically small breed such as the Italian Spinone the introduction of dogs from a different population may be required to improve genetic variation [11]. Breeding advice would therefore need to be cautious and efforts made to educate dog breeders and breed clubs so that they can make informed breeding decisions.

The three-SNP risk score tested for Border Collie IE showed evidence of ability to predict disease status (chapter 7). Such a test could be used to identify dogs with a higher risk of developing IE and could be an additional diagnostic test to assist veterinary neurologists with diagnosis which is currently typically a process of elimination by excluding other potential causes of seizures. A study of IE in the Belgian Shepherd dog found that a two-loci model combining associated SNPs on canine chromosomes 14 and 37 showed evidence of being predictive of IE risk [292]. However, subsequent studies have been unable to replicate the association of chromosome 14 [303, 304]. A patented risk score comprising 17 variants reported to be associated with canine hip dysplasia in German Shepherd dogs did not demonstrate predictiveness in a study that evaluated it independently [464]. To the authors knowledge, there have been no previous publications describing the use of a combined, weighted, genetic risk score for canine inherited disease for which evidence of predictiveness has been demonstrated using ROC

analysis. The three-SNP risk score for IE in the Border Collie is therefore likely to be the first canine genetic risk score to have been found to have evidence of predictiveness in a large replication set. Genetic risk scores based on relatively small numbers of SNPs associated with disease [416], and PRS that incorporate up to many thousands of array SNPs based on data from GWAS, often using the combination of SNPs yielding the best predictive model [465], have been developed for complex human diseases to help identify individuals at increased risk of disease who could benefit from clinical and/or lifestyle interventions [412, 413]. The largest human GWAS meta-analysis of common epilepsies to date identified 16 associated loci [231], and has enabled studies to use PRS to investigate the underlying heritability of epilepsy [466-469]. However, more research is required, using larger study cohorts to investigate the PRS accuracy relating to epilepsy subtypes and individuals of non-European ancestry, before epilepsy PRS can be used to make clinical decisions. PRS have also been investigated for various psychological disorders [470], and shared heritability has been found across disorders [471]. A PRS for type 2 diabetes has been assessed using data derived from meta-analysis of GWAS [472], and may be able to identify individuals at increased risk. Genetic risk scores based on 13 [414], 27 [415], and 50 [416] SNPs derived from GWAS have been demonstrated to be capable of identifying individuals at an increased risk of coronary heart disease. Preventative treatment of coronary heart disease using statin therapy, and guidance for healthy lifestyle decisions, can be informed by genetic risk scores [412, 415]. Genetic risk scores have also been developed for colorectal cancer (27 SNPs) [417], prostate cancer (54 SNPs) [418], and breast cancer (92 SNPs) [419], and have the potential to inform the decision to undergo screening for the respective diseases or to advise healthy lifestyle choices in those at high risk.

A genetic risk score for IE in the Border Collie would need further validation and optimisation to establish that it is robust, before it could be implemented as a clinical tool, or as a tool for dog breeders. Increasing the number of SNPs included in the genetic risk score has been shown to improve prediction of coronary heart disease risk [416], and would likely have the same affect for canine genetic risk scores. Additional variants discovered in future GWAS or WGS analysis could be added to the three SNPs that have been included so far, potentially improving predictiveness. The use of multi-ethnic association data when developing PRS has been suggested to improve prediction accuracy for disorders such as type 2 diabetes and inflammatory bowel disease in humans [473,

474]. Performing across-breed GWAS or WGS analysis to identify loci or variants associated with risk common to multiple breeds, and inclusion of these within a genetic risk score could improve accuracy and utility, potentially allowing development of a risk score applicable across breeds. It will also be important to ensure that analyses include subpopulations within breeds, such as 'working' and 'show' dogs, and dogs in different geographical regions or countries. A human PRS for generalized and focal epilepsy derived from a European-ancestry GWAS showed poor predictiveness in Japanese-ancestry individuals [466], demonstrating the importance of including individuals from a wide range of populations within discovery analyses, or else limiting the application of a risk score to the population of the discovery set. For this reason it is potentially advantageous that the Border Collie study included the UH GWAS (Finland) and UU replication (Netherlands) study sets from outside of the UK.

8.3 The importance of replication

The findings of this PhD project, and the Italian Spinone IE study in particular, demonstrate the importance of using an independent validation cohort to attempt to replicate any GWAS findings when developing breeding tools. Two SNPs passed the threshold for genome-wide significance in the Italian Spinone study, but later failed to replicate in a validation set (chapter 6). The threshold for significance used was the Bonferroni-adjusted P-value, correcting for multiple testing. This threshold is sometimes considered over stringent for canine genetic research because the extensive LD of the canine genome causes the true number of independent tests to be considerably lower than the number of SNPs included in a GWAS. Despite this, the two loci identified appear to be false positives or potentially specific to the subpopulation tested as a result of sampling bias or generational effects. In the Border Collie study, the SNP with the strongest association with IE in the GWAS, on canine chromosome 1, did not replicate in the independent validation set. However, three other SNPs on other chromosomes demonstrated evidence of replicable association with IE. This again demonstrates the importance of replication. The reported association on canine chromosome 14 for IE in the Belgian Shepherd dog, also discussed elsewhere in this thesis, is another example of an IE GWAS finding for which subsequent replication studies have reported negative results [292, 303, 304]. The role of variants within the *ABCB1* gene, also known as *MDR1*, in refractory epilepsy in Border Collies and other breeds is unclear as the initial findings

have so far not replicated in independent sets, although sample sizes have been small [309, 310, 312]. In comparison, a common risk haplotype within the *ADAM23* gene has been demonstrated to have a replicable association with IE in the Belgian Shepherd dog and other breeds [276, 290, 291]. The importance of replication is not limited to GWAS, complex diseases, or canine genetic research; for example, variants within the canine cyclooxgenase-2 gene promoter associated with renal dysplasia in dogs identified through candidate gene sequencing [475, 476], and an intronic variant in *TNNT2* linked with feline cardiomyopathy identified through WGS [477, 478], both failed to replicate in independent studies.

It is not unusual for GWAS results to be published without first validating the findings in an independent set. For example, the GWAS study including 34 cases and 168 controls that identified a risk haplotype for IE in the Irish Wolfhound did not attempt to replicate the association or narrow the haplotype [80]. The study included 4,224 dogs from 150 breeds, mixed breed, and village dogs and investigated 12 different phenotypes including complex disease and morphology. The authors aimed to demonstrate the power of such an approach to map complex traits, reporting loci to which each phenotype was mapped without exploring further. It is important that researchers bear the need for replication in mind when drawing conclusions and attempt to replicate findings before proceeding with further research with hypotheses based on GWAS findings. This is likely to be of particular importance for canine complex disease research in which risk conferring variants will have lower effect than those of Mendelian disorders, causing GWAS, and validation sets, to be potentially underpowered [49]. For this reason, it is also necessary to consider if a sample set for validation is of sufficient size to replicate the discovery study's findings before dismissing an associated region based on a small replication study.

When a replication study utilises a sample set from a population distinct from that of the discovery set, lack of replication does not necessarily indicate that the initial study's findings were false-positive. An illustrative example of this is that different genetic risk factors that have been identified for mast cell tumours in European and US Golden Retrievers [366, 367]. In numerically small breeds in particular, heterogeneity in genetic risk factors between different sample sets could confound GWAS and impede replication. However, the identification of genetic variants within particular genes and pathways conferring disease could allow inferences about disease aetiology in the wider breed population and dogs in general.

8.4 Approaches used

This section will discuss the approaches used, the rationale for their use and their effectiveness, and the limitations of the studies in this PhD project.

8.4.1 GWAS

This PhD used the GWAS approach to investigate PxD in the Norwich Terrier and IE in the Italian Spinone and Border Collie breeds (chapters 5, 6, and 7).

The Norwich Terrier PxD GWAS did not identify any SNPs approaching, or surpassing, (Bonferroni-corrected) genome-wide statistical significance. This study is preliminary, and the GWAS design, including only 24 cases and 24 controls, reflects the unknown mode of inheritance for PxD in the Norwich Terrier breed at the time of study conception. Analysis of pedigrees had been inconclusive [120], examples of PxD with autosomal recessive or dominant inheritance have been described [189-193], and GWAS of comparable size or smaller have been used to successfully identify loci containing candidate causal variants [189, 190, 193]. The lack of genome-wide significant findings suggest the PxD in the Norwich Terrier is not single-gene autosomal recessive or dominant and may be oligogenic or multigenic. The mode of inheritance in the Norwich Terrier could be similar to that of Border Terrier PxD, for which a much larger GWAS was unable to identify genome-wide significant associations [121]. In the Norwich Terrier PxD GWAS, where possible, at least one control dog related to each case was included to reduce population stratification that was observed in a previous, unpublished, GWAS conducted in our laboratory. It is possible that the low study power that was a result of the small sample size was exacerbated by the use of closely related controls. After setting a much lower threshold to select SNPs for follow up, analysis with 10 cases and 45 controls from a large independent dataset of 232 Norwich Terriers identified five genomic regions associated with PxD. This suggests that the use of a larger GWAS, with controls not selected for close relatedness to the cases, may have sufficient power to find regions associated with PxD in this breed. PxD in the Norwich Terrier may therefore have underlying genetic risk factors with a higher effect than those that cause increased risk in the Border Terrier, or the GWAS in the Border Terrier could have been hampered by phenotyping of insufficient stringency causing multiple conditions with similar clinical signs to be included within the case definition.

The Italian Spinone IE study used a large independent set of 175 controls as an initial semi-replication set to identify GWAS SNPs with evidence of replicable association with disease. The five SNPs that were identified using this approach in the Italian Spinone IE study failed to replicate when genotyped in a case-control set. The case-control set used was small and may have had insufficient power to detect the small to moderate effect sizes indicated by the GWAS. However, the lack of replicable association for the SNPs identified in this manner in the Italian Spinone study suggests that caution is necessary when drawing conclusions about loci implicated using this method. Future research should use validation sets which include both cases and controls; the findings here indicate that including only controls is insufficient to determine the reproducibility of an association found through GWAS.

As for the Norwich Terrier, the Italian Spinone GWAS included only moderate numbers of cases and controls; the mode of inheritance was unknown and studies have found monogenic forms of idiopathic epilepsy in the dog [268, 294]. The GWAS that identified the locus on canine chromosome 37 associated with IE with genome-wide significance in the Belgian Shepherd dog breed included 40 cases and 44 controls [276], sample numbers smaller than the Italian Spinone GWAS meta-analysis. This suggested that such an approach could be fruitful for the Italian Spinone breed; the high prevalence of IE (5.3%) and severe clinical course led us to hypothesise that the risk conferring variants were not rare and would be of moderate to high effect [251]. The Italian Spinone GWAS meta-analysis did find two SNPs surpassing the threshold for genome-wide significance, however they did not replicate in the subsequent validation set, as previously discussed. No SNPs reached genome-wide significance in a smaller GWAS of IE in the Petit Basset Griffon Vendeen that included 23 cases and 30 controls [293]. A larger sample could be used to increase study power, although as discussed in chapter 6, sample numbers are mostly limited by the numerically small population size for the Italian Spinone breed. Expanding the study to include more dogs from outside the UK could compensate for this, although caution would be needed to avoid population stratification and a validation set of sufficient size would also be necessary. The study used the meta-analysis approach to do a combined analysis of Set 1 and Set 2. Set 2 was genotyped four years after Set 1, once DNA samples became available for additional cases and controls following a drive for study recruitment. The decision to perform a meta-analysis of summary statistics from GWAS, as opposed to a mega-analysis pooling the genotype data into a single GWAS, was

to compensate for any population stratification caused by sampling bias and, primarily, to allow the use of a higher density genotyping array and genotype imputation to improve the resolution of the GWAS. Meta-analyses can however have reduced power when compared to mega-analyses [479], and the mega-analysis approach was the chosen methodology for the largest GWAS performed so far for epilepsy in humans that successfully identified 16 genome-wide significant loci [231].

The Border Collie GWAS meta-analysis is by far the largest of the three GWAS in this PhD research. There has not previously been any published GWAS or other study investigating the underlying genetic aetiology of IE in the Border Collie, with the exception of those looking at the *ABCB1* gene and refractory epilepsy. However, an unpublished GWAS for IE in the Border Collie was conducted as part of the LUPA consortium [3], and was unable to identify any loci associated with IE. Although the results have not been published, the lack of findings from the LUPA GWAS suggested that the disease was unlikely to be monogenic in the Border Collie. The study carried out as part of this PhD aimed to use large, well phenotyped, study sets to identify SNPs associated with IE in the Border Collie, and to utilise a validation set of sufficient size and power to find loci with evidence of replicable association with IE. Although no SNPs reached genome-wide significance in the meta-analysis, analysis of the most strongly associated SNPs in the independent validation sets did replicate the association of three loci. This study demonstrated the importance of sufficient sample size in GWAS of IE in the dog, and provides further evidence that the GWAS approach does have utility for research into this disease. Based on this, and the results of the other two GWAS in this PhD project, it is advisable that future research intending to use the GWAS approach to investigate canine complex diseases such as IE ensures that samples of sufficient size can be obtained. Other methodology, such as across-breed GWAS or WGS as described in more detail below in study limitations and future work, may be advisable for numerically small breeds for which the number of incident cases are likely to be smaller than the required sample size for both GWAS and validation.

8.4.2 Genotype imputation

Genome-wide genotype imputation was utilised for all three of the GWAS-based studies described here. This approach is discussed in detail in chapter 4 [370]. The research outlined in chapter 4 built on our understanding of the optimal number of breed specific

individuals in a reference panel for imputation of canine genotype, indicating that a moderate number is required for high accuracy (e.g. greater than 20 for the Border Collie breed). The resulting publication was also of particular value for the field because it described the methodology in detail to encourage use of the technique [370], building on previous work which typically included a briefer description of the preparation of data for genotype imputation [76-78]. The focus on imputation up to array density as opposed to WGS variant density is also likely to be valuable as a cost-effective alternative to generating the required number of breed-specific WGS for a reference panel.

The Norwich Terrier PxD GWAS utilised Axiom array genotyped data, and genotype imputation allowed the most strongly associated SNPs to be assessed in the genotype dataset of 232 Norwich Terriers which had been genotyped for an unrelated study using the lower density Illumina array. Imputation allowed the lower density Illumina CanineHD array genotyped GWAS data to be meta-analysed with Axiom Canine HD array data in the Italian Spinone study, and also allowed analysis of Axiom array SNPs in the validation set which was genotyped on the lower-density array. All three sets in the Border Collie GWAS meta-analysis were initially genotyped on the lower density array, and imputation allowed analysis at the higher density without re-genotyping all individuals. The Border Collie data was also imputed to the density of WGS SNPs, allowing yet more variants to be investigated. This was possible for this breed because of the availability of 40 Border Collie WGS that were part of in-house and publicly available WGS variant datasets, allowing imputation to be of a high accuracy. The investigations as part of this PhD project (chapter 4) [370], and other published research [76, 78], have demonstrated that including sufficient numbers of breed specific individuals within a reference panel are necessary to achieve high imputation accuracy. Imputation accuracy is likely to be particularly important when investigating complex canine diseases where error could potentially reduce a study's power to identify associations where allele differences are relatively small between cases and controls. For this reason, imputation up to WGS SNP level was not carried out for the Italian Spinone or Norwich Terrier Studies, for which breeds only low numbers ($n \leq 5$) of WGS were available. These studies therefore further demonstrate the utility of imputation in canine genetic research, and the approach is now becoming widely used in the field [79, 480-483].

8.4.3 Whole genome sequencing

WGS of two siblings proved an effective methodology when investigating the rare, autosomal recessive, cerebellar ataxia in the Norwegian Buhund breed (chapter 3) [461]. This is demonstrative of the drastically smaller sample numbers required in comparison to common and multigenic or complex neurological diseases such as PxD in the Norwich Terrier and IE in the Italian Spinone and Border Collie breeds. Such studies do however still require a sample set for validation of the findings, but pedigree information can facilitate the use of smaller numbers of cases. Cerebellar ataxia is rare in the breed, the breed is numerically small, and DNA from only four cases were available for this study; this made the use of pedigree information to identify obligate carriers extremely valuable. The biobank of > 40,000 canine DNA samples at Kennel Club Genetic Centre is an important resource for canine genetic research, and allowed the validation set of Norwegian Buhunds to be supplemented with a large multi-breed panel (359 dogs, 122 breeds).

The Norwegian Buhund study identified a novel gene for cerebellar ataxia. WGS could be used to identify a novel gene for cerebellar ataxia in the Norwegian Buhund because of the large number of 'control' genomes that could be used for variant filtering, and the availability of tissue from the affected dogs. The rarity of the disease, and the likely-causal variant's low frequency within the dog population and apparent breed-specificity, meant that filtering using publicly available and in-house WGS datasets reduced the number of potential variants identified to a number manageable for follow up genotyping without filtering by gene candidacy, as was previously typically necessary for studies utilising WGS of small numbers of cases [65, 66]. Variants in genes that are novel for a disease require robust evidence that they are causal. The availability of tissue samples from the cases allowed the use of RT-qPCR, Western blot, and immunohistochemistry to support the findings, along with *in silico* tools that can be used in the absence of tissue. The absence of tissue samples for many inherited canine diseases that are manageable through treatment, are not life threatening, or do not cause a quality of life concern that leads to euthanasia, will limit the ability of studies to provide strong evidence that a variant is causal when it affects a novel gene. The lack of availability of healthy control tissue is also often a limitation. However, projects such as the dog genome annotation project (DoGA), which aims to improve functional annotation of regulatory elements of the canine genome, and is collecting a tissue biobank and gene expression profile data of different

regions of the canine brain, will be extremely valuable [484, 485] (www.doggenomeannotation.org, accessed 22/01/2023).

The large publicly available WGS variant dataset that facilitated the Norwegian Buhund study was also essential for building a multi-breed reference panel for accurate genotype imputation. The number of publicly available canine WGS is growing all the time, and is expected to grow more quickly as a result of genome-sequencing consortia such as the DBVDC and Dog10K [67-69]. This is another exciting emerging resource for canine genetics, and will be of great use to future research. As the cost of NGS continues to decrease, it is also likely that even smaller research groups will be able to generate their own in-house genome banks which can also be made publicly available for use by the field in general. The collaborative nature of the canine genetics international community, and the willingness of researchers to share data, contributes greatly to its success and is a key benefit to conducting research in the field.

8.5 Study limitations and future work

8.5.1 Copy number and structural variants

A limitation of the WGS analysis pipeline utilised to investigate cerebellar ataxia in the Norwegian Buhund is that it was unable to detect structural variants (SVs) or insertions and deletions larger than approximately 75 bp. Tools have been developed and used for detecting SVs in human genetic research and clinical sequencing [486, 487], and include Pindel [488], SVMerge [489], DELLY [490], Lumpy [491], and Manta [492]. SVs are detected by identifying read pairs mapping in an unexpected orientation or distance, split reads (where a read shows broken alignment, with the single read aligning to two regions), and increases or decreases in read depth [486, 487, 493]. SVs can also be detected through long-read sequence assembly [487, 493]. SV detection remains challenging, however, due to the short reads used for most NGS, the variable coverage of WGS data, and the localisation of SVs at difficult to sequence repetitive or duplicated regions [486, 487, 493].

SVs have been associated with canine inherited diseases and other phenotypes [494-497]. The detection of SVs in the dog has often relied on manual inspection of the sequence in a gene of interest using software such as the Integrative Genomics Viewer (IGV) [498], such as was carried out to identify the complex structural variant in *KRT71* in curly-coated

dogs [496], the 2.46 KB deletion in *GP9* putatively causing Bernard-Soulier syndrome in Cocker Spaniels [497], and the > 5 MB deletion on the canine X chromosome that putatively causes dystrophin-deficient muscular dystrophy in Miniature Poodles [495]. This approach relies on a GWAS or candidate gene approach to be used first to identify a region of a practical size for manual inspection. Once a region of interest has been identified, SV discovery tools can also be used to find SVs which can then be compared between cases to controls. This approach, utilising DELLY [490], was used to identify the complex structural variant associated with non-syndromic canine retinal degeneration (with incomplete penetrance) in Miniature Schnauzers [494].

Copy number variation (CNV) is another type of SV that is of interest, and which can also be challenging to detect. CNVs may have roles in canine complex disease; for example, a CNV showed evidence of association with risk for canine squamous cell carcinoma in Standard Poodles [499]. Read depth differences can be used to detect CNVs in WGS data, and sophisticated custom-made pipelines have been used to detect canine CNVs using WGS read depth, including a CNV GWAS of breed-specific morphology and phenotypes [500].

The above examples demonstrate that structural variants do have roles in canine inherited disease. As WGS of small numbers of cases is increasingly common practice for investigating canine autosomal recessive inherited diseases [65, 66, 461] (chapter 3), in a manner not entirely unlike clinical sequencing in humans [486], it becomes increasingly important that SVs are not overlooked. SVs may also have a role in canine IE; the Genomics England PanelApp genetic epilepsy syndromes gene panel (version: 3.0) includes 15 CNVs that have a diagnostic level of supporting evidence for disease association in humans [223, 224]. To date, most genetic studies in the dog have focused on analysis of SNPs and smaller insertions and deletions [67, 500]. The development of genome analysis pipelines that can perform genome-wide analysis of SVs will be important to allow investigation of the full extent of genetic variation in dogs and how this relates to disease.

8.5.2 Common or fixed variants

Case-control GWAS including only individuals from the same breed, such as those carried out in this PhD project, are unable to identify associations for risk-conferring variants which are extremely common or fixed (i.e. present in all dogs) within a breed population.

This is because the variants will be as frequent, or nearly as frequent, within the controls as they are in the cases. In the Border Collie IE GWAS, risk allele frequencies for CanFam3.1 11:68785445 (G) and 11:62873725 (T) were 0.63 and 0.60 respectively in GWAS meta-analysis controls, with meta-analysis odds ratio point estimates of 1.86 and 1.74 indicating modest effect (chapter 7). Neither association reached Bonferroni-adjusted genome-wide statistical significance but they did demonstrate evidence of replication in a validation set approximately twice the size of the discovery set, suggesting that these allele frequencies and effect sizes are close to the limit of the study's power to detect.

It is possible that the genetic bottlenecks and selection pressures that were part of breed formation and a result of selection to meet breed standards have led to variants that increase risk being very common or fixed in the breed. Risk-conferring variants that are fixed would contribute to the overall increased risk of a disease in the breed, with other less common variants and environmental factors causing the higher risk seen in some individuals. Selection for a desirable trait, for example herding behaviour in the Border Collie and other shepherding breeds where IE is prevalent, could have increased the frequency of the genetic variants that underly the desired trait but which also cause increased risk of a disease such as IE. This process is hypothesised to have occurred with selection for a copy of a *FGF4* retro-gene that is associated with Hansen's type I intervertebral disk disease [15, 16], and variants within *ADAMTS17* associated with primary open angle glaucoma [17, 18], both of which have also been associated with short stature which can be considered a desirable trait in some breeds [15, 16, 19]. The genetic bottlenecks that are a result of relatively recent breed formation and closed stud books have led to extensive LD in dogs, and this, along with widespread inbreeding, have resulted in long runs of homozygosity [5]. Selection for variants that underly a trait can therefore lead not just to that variant becoming fixed or common, but also a wide genomic region around it (a selective sweep) which can contain variants conferring disease risk [21, 22].

A potential approach to find very common or fixed variants underlying IE could be WGS of IE cases, and variant filtering using control dogs of other breeds that are not known to have a high prevalence of IE. An across breeds GWAS could also have utility for identifying regions of the genome that have been under selective pressure in a breed. Both approaches are potentially likely to lead to identification of variants that underly other

breed-specific traits not associated with IE. However, pathway and gene ontology database analysis may enable variants to be prioritised for further evaluation based on candidacy.

As discussed previously, identifying fixed or very common variants underlying risk of disease would represent challenges for dog breeders because they could not be bred out of the population without introducing dogs from outside the breed into the breeding population. However, findings could lead to an understanding of the biology that underlies IE and potentially inform treatments.

8.5.3 Rare or *de novo* variants

The GWAS approaches used in the studies described in this PhD thesis would have also been unable to identify regions containing very rare variants conferring risk of disease [49]. The GWAS would have lacked power to identify very rare variants, and standard practice was to filter out SNPs with a MAF lower than 5% for this reason. Numerous rare moderate effect variants could be contributing to disease risk within a breed [49, 412], but GWAS would not be able to identify the regions in which they are located. Many human epileptic encephalopathies can be caused by *de novo* mutations [222, 228-230]. *De novo* mutations, if they are involved in canine epilepsy or PxD, would by their nature be individually extremely rare (potentially found in only a single dog), and would therefore also not be found via GWAS. Rare and *de novo* variants would have limited use within a genetic risk score or a breeding tool because such a tool would need to be applicable to the whole breed population. However, understanding the underlying aetiology and the genes and pathways involved in canine complex diseases could lead to improved treatment. Building a list of genes implicated in a canine complex disease such as IE may facilitate clinical sequencing, such as is carried out for human disease [223, 224, 486], and to support diagnosis.

The approach likely to have the most success identifying rare variants of moderate effect is WGS of cases and filtering variants using control dogs of multiple breeds not known to have high risk of the disease being studied. *De novo* variants can be identified by WGS of parent and affected offspring trios [229, 230], filtering for variants present in the affected offspring but not in either parent. These approaches are likely to produce large numbers of potential variants. Variants could be prioritised by predicted consequence using tools such as SIFT and PolyPhen-2 [319, 501]. However, the rare variants underlying risk of a

complex disease may not be high-effect variants that directly affect a protein-coding sequence or disrupt a transcript. Variant filtering will therefore need to be more inclusive than would be used to investigate an autosomal recessive (or dominant) disease such as in the Norwegian Buhund cerebellar ataxia study (chapter 3) [461]. Improved annotation of regulatory elements in the canine genome would allow discovery of risk variants that affect gene expression [484, 485] (www.doggenomeannotation.org, accessed 22/01/2023), and allow the variants to be filtered based on the functional information. The filtered variants could then be interrogated using gene ontology databases to identify those most likely to confer risk, and functional relationships between the implicated genes could be investigated to identify pathways that are important for disease risk. Large numbers of WGS may be necessary to gain a good understanding of the rare variants, or *de novo* variants, underlying IE. Low pass sequencing and imputation would allow generation of large datasets [79], but low imputation accuracy for rare variants, or their absence from reference panels, would make such an approach unsuitable (chapter 4) [370].

8.5.4 Investigating potential PxD-associated regions in the Norwich Terrier

Five potentially PxD-associated regions were identified in the Norwich Terrier GWAS and replication analysis (chapter 5). The regions include genes with good candidacy for PxD. However, only 10 cases were available for replication analysis, and these were not as well defined as for the GWAS. Future work should include a large replication set including a larger number of well-defined cases and controls than were included in the GWAS. If the subsequent analysis confirms that any of the regions have replicable association with PxD, WGS could be utilised to search the identified genomic regions for variants conferring risk of PxD.

8.5.5 Future investigations of IE in the Italian Spinone

The Italian Spinone GWAS identified regions passing the threshold for statistical significance that did not replicate in an independent set of cases and controls (chapter 6). One of the associations implicated *MAST4*, a compelling candidate for epilepsy. This could represent a rare variant associated with IE in a small subset of cases. Future work could WGS dogs within the subset of individuals in which this association was strongest, and related dogs, to identify potentially causal or risk conferring variants within and in the region of this gene and determine if the association represents a rare familial epilepsy

within the breed. The GWAS would have been unable to identify rare or very common variants, and WGS analysis as described above should be used to explore the potential involvement of these variants in risk for IE in this breed. It is notable that the replication set used to follow up the GWAS associations included fewer individuals than the GWAS, and therefore had reduced power to detect associations with disease. Investigation of the GWAS associated regions in a new independent replication set, which should include more individuals than the GWAS, could be used to confirm the study's findings. However, as outlined above, study size is limited by the small population and the limited availability of well-defined cases.

8.5.6 Investigating Border Collie IE-associated regions

GWAS meta-analysis of IE in the Border Collie and genotyping a large independent case-control set identified three SNPs demonstrating evidence of replicable association with IE and potential for predicting IE status when included in a weighted risk score (chapter 7). The array SNPs are unlikely to be the variants conferring IE risk but instead will be in LD with the genomic variants that are affecting risk of disease. Future work will look for risk-conferring variants by generating WGS data for Border Collies with the identified risk alleles for the array SNPs and comparing with dogs with the alternate (non-risk) allele. The variants in the regions will be prioritised based on gene candidacy and predicted effect on protein coding sequences or gene regulation and expression and genotyped in sets of dogs to test if they are more predictive of IE status than the array SNPs.

8.6 Conclusions

Canine neurological diseases are fascinating naturally occurring models of human disease, and are major welfare concerns for dog breeds. This PhD project has led to the identification of the novel gene *KCNIP4* as cerebellar ataxia-associated, with subsequent research indicating potential relevance to human disease. The identification of the putative causal variant for cerebellar ataxia in the Norwegian Buhund has also allowed the development of a DNA test to allow dog breeders to avoid producing affected dogs. The studies of canine complex neurological diseases identified genomic regions with evidence of replicable association with PxD in the Norwich Terrier, and IE in the Border Collie. This PhD has built evidence indicating the involvement of both rare variants with higher effect and common lower effect variants in these canine neurological diseases, expanding our understanding of the underlying genetics of these diseases in dogs. A

three-SNP genetic risk score for IE was tested in the Border Collie breed and demonstrated that future refinement of this risk score may have utility for predicting disease status. This PhD project utilised a range of approaches to investigate the underlying genetics of canine neurological disease, including implementing methodology novel to the research group leading to the development of valuable tools. The studies are demonstrative of the strengths and weaknesses of the various approaches used, and will inform the choice of methodology for future studies.

9 References

1. Fogle, B., *The Encyclopedia of the Dog*. 1995: Dorling Kindersley.
2. Freedman, A.H. and R.K. Wayne, *Deciphering the Origin of Dogs: From Fossils to Genomes.* Annu Rev Anim Biosci, 2017. **5**: p. 281-307.
3. Lequarre, A.S., L. Andersson, C. Andre, M. Fredholm, C. Hitte, T. Leeb, et al., *LUPA: a European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs.* Vet J, 2011. **189**(2): p. 155-9.
4. Patterson, E.E., *Canine epilepsy: an underutilized model.* ILAR J, 2014. **55**(1): p. 182-6.
5. Lindblad-Toh, K., C.M. Wade, T.S. Mikkelsen, E.K. Karlsson, D.B. Jaffe, M. Kamal, et al., *Genome sequence, comparative analysis and haplotype structure of the domestic dog.* Nature, 2005. **438**(7069): p. 803-19.
6. Bergstrom, A., L. Frantz, R. Schmidt, E. Ersmark, O. Lebrasseur, L. Girdland-Flink, et al., *Origins and genetic legacy of prehistoric dogs.* Science, 2020. **370**(6516): p. 557-564.
7. Vila, C., P. Savolainen, J.E. Maldonado, I.R. Amorim, J.E. Rice, R.L. Honeycutt, et al., *Multiple and ancient origins of the domestic dog.* Science, 1997. **276**(5319): p. 1687-9.
8. Freedman, A.H., I. Gronau, R.M. Schweizer, D. Ortega-Del Vecchyo, E. Han, P.M. Silva, et al., *Genome sequencing highlights the dynamic early history of dogs.* PLoS Genet, 2014. **10**(1): p. e1004016.
9. Parker, H.G., A.L. Shearin, and E.A. Ostrander, *Man's best friend becomes biology's best in show: genome analyses in the domestic dog.* Annu Rev Genet, 2010. **44**: p. 309-36.
10. Vonholdt, B.M., J.P. Pollinger, K.E. Lohmueller, E. Han, H.G. Parker, P. Quignon, et al., *Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication.* Nature, 2010. **464**(7290): p. 898-902.
11. Lewis, T.W., B.M. Abhayaratne, and S.C. Blott, *Trends in genetic diversity for all Kennel Club registered pedigree dog breeds.* Canine Genet Epidemiol, 2015. **2**: p. 13.
12. Hendricks, J.C., *Brachycephalic airway syndrome.* Vet Clin North Am Small Anim Pract, 1992. **22**(5): p. 1145-53.
13. Packer, R.M., A. Hendricks, M.S. Tivers, and C.C. Burn, *Impact of Facial Conformation on Canine Health: Brachycephalic Obstructive Airway Syndrome.* PLoS One, 2015. **10**(10): p. e0137496.
14. Liu, N.C., E.L. Troconis, L. Kalmar, D.J. Price, H.E. Wright, V.J. Adams, et al., *Conformational risk factors of brachycephalic obstructive airway syndrome (BOAS) in pugs, French bulldogs, and bulldogs.* PLoS One, 2017. **12**(8): p. e0181928.
15. Brown, E.A., P.J. Dickinson, T. Mansour, B.K. Sturges, M. Aguilar, A.E. Young, et al., *FGF4 retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs.* Proc Natl Acad Sci U S A, 2017. **114**(43): p. 11476-11481.
16. Batcher, K., P. Dickinson, M. Giuffrida, B. Sturges, K. Vernau, M. Knipe, et al., *Phenotypic Effects of FGF4 Retrogenes on Intervertebral Disc Disease in Dogs.* Genes (Basel), 2019. **10**(6).
17. Forman, O.P., L. Pettitt, A.M. Komaromy, P. Bedford, and C. Mellersh, *A Novel Genome-Wide Association Study Approach Using Genotyping by Exome*

Sequencing Leads to the Identification of a Primary Open Angle Glaucoma Associated Inversion Disrupting ADAMTS17. PLoS One, 2015. **10**(12): p. e0143546.
18. Oliver, J.A.C., S. Rustidge, L. Pettitt, C.A. Jenkins, F.H.G. Farias, E.A. Giuliano, et al., *Evaluation of ADAMTS17 in Chinese Shar-Pei with primary open-angle glaucoma, primary lens luxation, or both.* Am J Vet Res, 2018. **79**(1): p. 98-106.
19. Jeanes, E.C., J.A.C. Oliver, S.L. Ricketts, D.J. Gould, and C.S. Mellersh, *Glaucoma-causing ADAMTS17 mutations are also reproducibly associated with height in two domestic dog breeds: selection for short stature may have contributed to increased prevalence of glaucoma.* Canine Genet Epidemiol, 2019. **6**: p. 5.
20. Morales, J., L. Al-Sharif, D.S. Khalil, J.M. Shinwari, P. Bavi, R.A. Al-Mahrouqi, et al., *Homozygous mutations in ADAMTS10 and ADAMTS17 cause lenticular myopia, ectopia lentis, glaucoma, spherophakia, and short stature.* Am J Hum Genet, 2009. **85**(5): p. 558-68.
21. Vaysse, A., A. Ratnakumar, T. Derrien, E. Axelsson, G. Rosengren Pielberg, S. Sigurdsson, et al., *Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping.* PLoS Genet, 2011. **7**(10): p. e1002316.
22. Marsden, C.D., D. Ortega-Del Vecchyo, D.P. O'Brien, J.F. Taylor, O. Ramirez, C. Vila, et al., *Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs.* Proc Natl Acad Sci U S A, 2016. **113**(1): p. 152-7.
23. Woolliams, J.A., T.W. Lewis, and S.C. Blott, *Canine hip and elbow dysplasia in UK Labrador retrievers.* Vet J, 2011. **189**(2): p. 169-76.
24. Crispin, S., *Hereditary eye disease and the BVA/KC/ISDS Eye Scheme.* In Practice, 1995. **17**(6): p. 254-264.
25. British Veterinary Association. *Eye Scheme for Dogs*. 2023 [cited 2023 24/03/23]; BVA/KC/ISDS Eye Scheme]. Available from: https://www.bva.co.uk/Canine-Health-Schemes/Eye-scheme/.
26. Donner, J., H. Anderson, S. Davison, A.M. Hughes, J. Bouirmane, J. Lindqvist, et al., *Frequency and distribution of 152 genetic disease variants in over 100,000 mixed breed and purebred dogs.* PLoS Genet, 2018. **14**(4): p. e1007361.
27. Wang, S., D. Laloe, F.M. Missant, S. Malm, T. Lewis, E. Verrier, et al., *Breeding policies and management of pedigree dogs in 15 national kennel clubs.* Vet J, 2018. **234**: p. 130-135.
28. Lewis, T.W., S.C. Blott, and J.A. Woolliams, *Comparative analyses of genetic trends and prospects for selection against hip and elbow dysplasia in 15 UK dog breeds.* BMC Genet, 2013. **14**: p. 16.
29. Lewis, T.W., S.C. Blott, and J.A. Woolliams, *Genetic evaluation of hip score in UK Labrador Retrievers.* PLoS One, 2010. **5**(10): p. e12797.
30. O'Neill, D.G., D.B. Church, P.D. McGreevy, P.C. Thomson, and D.C. Brodbelt, *Approaches to canine health surveillance.* Canine Genet Epidemiol, 2014. **1**: p. 2.
31. Lindblad-Toh, K., *Canine Genomics*, in *The Genetics of the Dog*, E.A. Ostrander and A. Ruvinsky, Editors. 2012, CABI. p. 255-274.
32. Sydney School of Veterinary Science. *Online Mendelian Inheritance in Animals, OMIA*. 2023 [cited 2023 24/03/2023]; Available from: http://omia.org/.
33. Lewis, T.W. and C.S. Mellersh, *Changes in mutation frequency of eight Mendelian inherited disorders in eight pedigree dog populations following introduction of a commercial DNA test.* PLoS One, 2019. **14**(1): p. e0209864.
34. Mellersh, C.S., A.A. Langston, G.M. Acland, M.A. Fleming, K. Ray, N.A. Wiegand, et al., *A linkage map of the canine genome.* Genomics, 1997. **46**(3): p. 326-36.

35. Li, R., E. Mignot, J. Faraco, H. Kadotani, J. Cantanese, B. Zhao, et al., *Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library.* Genomics, 1999. **58**(1): p. 9-17.
36. Breen, M., C.F. Langford, N.P. Carter, N.G. Holmes, H.F. Dickens, R. Thomas, et al., *FISH mapping and identification of canine chromosomes.* J Hered, 1999. **90**(1): p. 27-30.
37. Neff, M.W., K.W. Broman, C.S. Mellersh, K. Ray, G.M. Acland, G.D. Aguirre, et al., *A second-generation genetic linkage map of the domestic dog, Canis familiaris.* Genetics, 1999. **151**(2): p. 803-20.
38. Wong, A.K., A.L. Ruhe, B.L. Dumont, K.R. Robertson, G. Guerrero, S.M. Shull, et al., *A comprehensive linkage map of the dog genome.* Genetics, 2010. **184**(2): p. 595-605.
39. Auton, A., Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel, J.K. Holloway, et al., *Genetic recombination is targeted towards gene promoter regions in dogs.* PLoS Genet, 2013. **9**(12): p. e1003984.
40. Klug, W.S., *Concepts of genetics*. 8th ed. / Willaim S. Klug, Michael R. Cummings, Charlotte A. Spencer. ed, ed. C. Spencer and M.R. Cummings. 2006, Upper Saddle River, NJ: Upper Saddle River, NJ : Pearson Prentice Hall.
41. Kirkness, E.F., V. Bafna, A.L. Halpern, S. Levy, K. Remington, D.B. Rusch, et al., *The dog genome: survey sequencing and comparative analysis.* Science, 2003. **301**(5641): p. 1898-903.
42. Ewing, B., L. Hillier, M.C. Wendl, and P. Green, *Base-calling of automated sequencer traces using phred. I. Accuracy assessment.* Genome Res, 1998. **8**(3): p. 175-85.
43. Ewing, B. and P. Green, *Base-calling of automated sequencer traces using phred. II. Error probabilities.* Genome Res, 1998. **8**(3): p. 186-94.
44. Wang, C., O. Wallerman, M.L. Arendt, E. Sundstrom, A. Karlsson, J. Nordin, et al., *A novel canine reference genome resolves genomic architecture and uncovers transcript complexity.* Commun Biol, 2021. **4**(1): p. 185.
45. Halo, J.V., A.L. Pendleton, F. Shen, A.J. Doucet, T. Derrien, C. Hitte, et al., *Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes.* Proc Natl Acad Sci U S A, 2021. **118**(11).
46. Jagannathan, V., C. Hitte, J.M. Kidd, P. Masterson, T.D. Murphy, S. Emery, et al., *Dog10K_Boxer_Tasha_1.0: A Long-Read Assembly of the Dog Reference Genome.* Genes (Basel), 2021. **12**(6).
47. Sherman, R.M. and S.L. Salzberg, *Pan-genomics in the human genome era.* Nat Rev Genet, 2020. **21**(4): p. 243-254.
48. Acland, G.M., K. Ray, C.S. Mellersh, W. Gu, A.A. Langston, J. Rine, et al., *Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 3048-53.
49. McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P. Ioannidis, et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.
50. Uffelmann, E., Q.Q. Huang, N.S. Munung, J. de Vries, Y. Okada, A.R. Martin, et al., *Genome-wide association studies.* Nature Reviews Methods Primers, 2021. **1**(1): p. 59.

51. Duncan, L.E., M. Ostacher, and J. Ballon, *How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete.* Neuropsychopharmacology, 2019. **44**(9): p. 1518-1523.
52. Johnson, R.C., G.W. Nelson, J.L. Troyer, J.A. Lautenberger, B.D. Kessing, C.A. Winkler, et al., *Accounting for multiple comparisons in a genome-wide association study (GWAS).* BMC Genomics, 2010. **11**: p. 724.
53. Astle, W. and D.J. Balding, *Population Structure and Cryptic Relatedness in Genetic Association Studies.* Statist. Sci., 2009. **24**(4): p. 451-471.
54. Zhou, X. and M. Stephens, *Genome-wide efficient mixed-model analysis for association studies.* Nat Genet, 2012. **44**(7): p. 821-4.
55. Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
56. Aguirre-Hernandez, J. and D.R. Sargan, *Evaluation of candidate genes in the absence of positional information: a poor bet on a blind dog!* J Hered, 2005. **96**(5): p. 475-84.
57. Jenkins, C. and O. Forman, *Identification of a novel frameshift mutation in the DMD gene as the cause of muscular dystrophy in a Norfolk terrier dog.* Canine Genetics and Epidemiology, 2015. **2**(1): p. 7.
58. Mellersh, C.S., L. Pettitt, O.P. Forman, M. Vaudin, and K.C. Barnett, *Identification of mutations in HSF4 in dogs of three different breeds with hereditary cataracts.* Vet Ophthalmol, 2006. **9**(5): p. 369-78.
59. Tabor, H.K., N.J. Risch, and R.M. Myers, *Candidate-gene approaches for studying complex genetic traits: practical considerations.* Nat Rev Genet, 2002. **3**(5): p. 391-7.
60. Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
61. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1135-45.
62. Forman, O.P., L. De Risio, J. Stewart, C.S. Mellersh, and E. Beltran, *Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated SPTBN2 mutation.* BMC Genet, 2012. **13**: p. 55.
63. 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): p. e64627.
64. Forman, O.P., L. De Risio, K. Matiasek, S. Platt, and C. Mellersh, *Spinocerebellar ataxia in the Italian Spinone dog is associated with an intronic GAA repeat expansion in ITPR1.* Mamm Genome, 2015. **26**(1-2): p. 108-17.
65. Fenn, J., M. Boursnell, R.J. Hitti, C.A. Jenkins, R.L. Terry, S.L. Priestnall, 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 Genet, 2016. **17**(1): p. 123.
66. Gilliam, D., D.P. O'Brien, J.R. Coates, G.S. Johnson, G.C. Johnson, T. Mhlanga-Mutangadura, et al., *A homozygous KCNJ10 mutation in Jack Russell Terriers and related breeds with spinocerebellar ataxia with myokymia, seizures, or both.* J Vet Intern Med, 2014. **28**(3): p. 871-7.

67. Jagannathan, V., C. Drogemuller, T. Leeb, and C. Dog Biomedical Variant Database, *A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves.* Anim Genet, 2019. **50**(6): p. 695-704.
68. Wang, G.D., G. Larson, J.M. Kidd, B.M. vonHoldt, E.A. Ostrander, and Y.P. Zhang, *Dog10K: the International Consortium of Canine Genome Sequencing.* Natl Sci Rev, 2019. **6**(4): p. 611-613.
69. Ostrander, E.A., G.D. Wang, G. Larson, B.M. vonHoldt, B.W. Davis, V. Jagannathan, et al., *Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health.* Natl Sci Rev, 2019. **6**(4): p. 810-824.
70. Plassais, J., J. Kim, B.W. Davis, D.M. Karyadi, A.N. Hogan, A.C. Harris, et al., *Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology.* Nat Commun, 2019. **10**(1): p. 1489.
71. Marchini, J., B. Howie, S. Myers, G. McVean, and P. Donnelly, *A new multipoint method for genome-wide association studies by imputation of genotypes.* Nat Genet, 2007. **39**(7): p. 906-13.
72. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.
73. Howie, B., J. Marchini, and M. Stephens, *Genotype imputation with thousands of genomes.* G3 (Bethesda), 2011. **1**(6): p. 457-70.
74. Zeggini, E. and J.P. Ioannidis, *Meta-analysis in genome-wide association studies.* Pharmacogenomics, 2009. **10**(2): p. 191-201.
75. Rubinacci, S., D.M. Ribeiro, R.J. Hofmeister, and O. Delaneau, *Efficient phasing and imputation of low-coverage sequencing data using large reference panels.* Nat Genet, 2021. **53**(1): p. 120-126.
76. Friedenberg, S.G. and K.M. Meurs, *Genotype imputation in the domestic dog.* Mamm Genome, 2016. **27**(9-10): p. 485-94.
77. Friedrich, J., R. Antolin, S.M. Edwards, E. Sanchez-Molano, M.J. Haskell, J.M. Hickey, et al., *Accuracy of genotype imputation in Labrador Retrievers.* Anim Genet, 2018. **49**(4): p. 303-311.
78. Hayward, J.J., M.E. White, M. Boyle, L.M. Shannon, M.L. Casal, M.G. Castelhano, et al., *Imputation of canine genotype array data using 365 whole-genome sequences improves power of genome-wide association studies.* PLoS Genet, 2019. **15**(9): p. e1008003.
79. Buckley, R.M., A.C. Harris, G.D. Wang, D.T. Whitaker, Y.P. Zhang, and E.A. Ostrander, *Best practices for analyzing imputed genotypes from low-pass sequencing in dogs.* Mamm Genome, 2021.
80. Hayward, J.J., M.G. Castelhano, K.C. Oliveira, E. Corey, C. Balkman, T.L. Baxter, et al., *Complex disease and phenotype mapping in the domestic dog.* Nat Commun, 2016. **7**: p. 10460.
81. Hytonen, M.K. and H. Lohi, *Canine models of human rare disorders.* Rare Dis, 2016. **4**(1): p. e1241362.
82. Gan-Or, Z., N. Bouslam, N. Birouk, A. Lissouba, D.B. Chambers, J. Veriepe, et al., *Mutations in CAPN1 Cause Autosomal-Recessive Hereditary Spastic Paraplegia.* Am J Hum Genet, 2016. **98**(6): p. 1271.
83. Mereaux, J.L., C. Firanescu, G. Coarelli, M. Kvarnung, R. Rodrigues, E. Pegoraro, et al., *Increasing involvement of CAPN1 variants in spastic ataxias and phenotype-genotype correlations.* Neurogenetics, 2021. **22**(1): p. 71-79.
84. McKusick-Nathans Institute of Genetic Medicine. *Online Mendelian Inheritance in Man, OMIM®*. 2023 [cited 2023 24/03/23]; Available from: https://omim.org/.

85. Grone, B.P. and S.C. Baraban, *Animal models in epilepsy research: legacies and new directions.* Nat Neurosci, 2015. **18**(3): p. 339-43.
86. Kline, K.L., *Feline epilepsy.* Clin Tech Small Anim Pract, 1998. **13**(3): p. 152-8.
87. Erlen, A., H. Potschka, H.A. Volk, C. Sauter-Louis, and D.G. O'Neill, *Seizures in dogs under primary veterinary care in the United Kingdom: Etiology, diagnostic testing, and clinical management.* J Vet Intern Med, 2020. **34**(6): p. 2525-2535.
88. Fahn, S., *Classification of movement disorders.* Mov Disord, 2011. **26**(6): p. 947-57.
89. Stoessl, A.J. and M.J. McKeown, *Movement disorders.* Handb Clin Neurol, 2016. **136**: p. 957-69.
90. Fahn, S., J. Jankovic, and M. Hallett, *Chapter 1 - Clinical overview and phenomenology of movement disorders*, in *Principles and Practice of Movement Disorders (Second Edition)*. 2011, W.B. Saunders: Edinburgh. p. 1-35.
91. Klockgether, T., C. Mariotti, and H.L. Paulson, *Spinocerebellar ataxia.* Nat Rev Dis Primers, 2019. **5**(1): p. 24.
92. Akbar, U. and T. Ashizawa, *Ataxia.* Neurol Clin, 2015. **33**(1): p. 225-48.
93. Synofzik, M. and A.H. Nemeth, *Recessive ataxias.* Handb Clin Neurol, 2018. **155**: p. 73-89.
94. Zanni, G. and E. Bertini, *X-linked ataxias.* Handb Clin Neurol, 2018. **155**: p. 175-189.
95. Rossi, M., M. Anheim, A. Durr, C. Klein, M. Koenig, M. Synofzik, et al., *The genetic nomenclature of recessive cerebellar ataxias.* Mov Disord, 2018. **33**(7): p. 1056-1076.
96. Ruano, L., C. Melo, M.C. Silva, and P. Coutinho, *The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies.* Neuroepidemiology, 2014. **42**(3): p. 174-83.
97. Cossee, M., M. Schmitt, V. Campuzano, L. Reutenauer, C. Moutou, J.L. Mandel, et al., *Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7452-7.
98. Galea, C.A., A. Huq, P.J. Lockhart, G. Tai, L.A. Corben, E.M. Yiu, et al., *Compound heterozygous FXN mutations and clinical outcome in friedreich ataxia.* Ann Neurol, 2016. **79**(3): p. 485-95.
99. Delatycki, M.B. and L.A. Corben, *Clinical features of Friedreich ataxia.* J Child Neurol, 2012. **27**(9): p. 1133-7.
100. Garone, G., A. Capuano, L. Travaglini, F. Graziola, F. Stregapede, G. Zanni, et al., *Clinical and Genetic Overview of Paroxysmal Movement Disorders and Episodic Ataxias.* Int J Mol Sci, 2020. **21**(10).
101. Bhatia, K.P., *Paroxysmal dyskinesias.* Movement Disorders, 2011. **26**(6): p. 1157-1165.
102. Erro, R. and K.P. Bhatia, *Unravelling of the paroxysmal dyskinesias.* J Neurol Neurosurg Psychiatry, 2019. **90**(2): p. 227-234.
103. Erro, R., K.P. Bhatia, A.J. Espay, and P. Striano, *The epileptic and nonepileptic spectrum of paroxysmal dyskinesias: Channelopathies, synaptopathies, and transportopathies.* Mov Disord, 2017. **32**(3): p. 310-318.
104. Demirkiran, M. and J. Jankovic, *Paroxysmal dyskinesias: clinical features and classification.* Ann Neurol, 1995. **38**(4): p. 571-9.
105. Gardiner, A.R., F. Jaffer, R.C. Dale, R. Labrum, R. Erro, E. Meyer, et al., *The clinical and genetic heterogeneity of paroxysmal dyskinesias.* Brain, 2015. **138**(Pt 12): p. 3567-80.

106. Erro, R., U.M. Sheerin, and K.P. Bhatia, *Paroxysmal dyskinesias revisited: a review of 500 genetically proven cases and a new classification.* Mov Disord, 2014. **29**(9): p. 1108-16.
107. Liao, J.Y., P.A. Salles, U.A. Shuaib, and H.H. Fernandez, *Genetic updates on paroxysmal dyskinesias.* J Neural Transm (Vienna), 2021. **128**(4): p. 447-471.
108. Tan, G.H., Y.Y. Liu, L. Wang, K. Li, Z.Q. Zhang, H.F. Li, et al., *PRRT2 deficiency induces paroxysmal kinesigenic dyskinesia by regulating synaptic transmission in cerebellum.* Cell Res, 2018. **28**(1): p. 90-110.
109. Shen, Y., W.P. Ge, Y. Li, A. Hirano, H.Y. Lee, A. Rohlmann, et al., *Protein mutated in paroxysmal dyskinesia interacts with the active zone protein RIM and suppresses synaptic vesicle exocytosis.* Proc Natl Acad Sci U S A, 2015. **112**(10): p. 2935-41.
110. Weber, Y.G., A. Storch, T.V. Wuttke, K. Brockmann, J. Kempfle, S. Maljevic, et al., *GLUT1 mutations are a cause of paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak.* J Clin Invest, 2008. **118**(6): p. 2157-68.
111. Seidner, G., M.G. Alvarez, J.I. Yeh, K.R. O'Driscoll, J. Klepper, T.S. Stump, et al., *GLUT-1 deficiency syndrome caused by haploinsufficiency of the blood-brain barrier hexose carrier.* Nat Genet, 1998. **18**(2): p. 188-91.
112. Klepper, J. and B. Leiendecker, *Glut1 deficiency syndrome and novel ketogenic diets.* J Child Neurol, 2013. **28**(8): p. 1045-8.
113. Mochel, F., E. Hainque, D. Gras, I.M. Adanyeguh, S. Caillet, B. Heron, et al., *Triheptanoin dramatically reduces paroxysmal motor disorder in patients with GLUT1 deficiency.* J Neurol Neurosurg Psychiatry, 2016. **87**(5): p. 550-3.
114. Urkasemsin, G. and N.J. Olby, *Canine hereditary ataxia.* Vet Clin North Am Small Anim Pract, 2014. **44**(6): p. 1075-89.
115. Urkasemsin, G. and N.J. Olby, *Canine paroxysmal movement disorders.* Vet Clin North Am Small Anim Pract, 2014. **44**(6): p. 1091-102.
116. Richter, A., M. Hamann, J. Wissel, and H.A. Volk, *Dystonia and Paroxysmal Dyskinesias: Under-Recognized Movement Disorders in Domestic Animals? A Comparison with Human Dystonia/Paroxysmal Dyskinesias.* Front Vet Sci, 2015. **2**: p. 65.
117. Urkasemsin, G., K.E. Linder, J.S. Bell, A. de Lahunta, and N.J. Olby, *Hereditary cerebellar degeneration in Scottish terriers.* J Vet Intern Med, 2010. **24**(3): p. 565-70.
118. Olby, N., S. Blot, J.L. Thibaud, J. Phillips, D.P. O'Brien, J. Burr, et al., *Cerebellar cortical degeneration in adult American Staffordshire Terriers.* J Vet Intern Med, 2004. **18**(2): p. 201-8.
119. Karamatic, S.L., G.A. Anderson, B.W. Parry, R.F. Slocombe, and C.S. Mansfield, *Prevalence and risk factors for medical events following exercise at Australian Greyhound race meetings.* Aust Vet J, 2018. **96**(4): p. 120-126.
120. De Risio, L., O.P. Forman, C.S. Mellersh, and J. Freeman, *Paroxysmal Dyskinesia in Norwich Terrier Dogs.* Movement Disorders Clinical Practice, 2016. **3**(6): p. 573-579.
121. Stassen, Q.E.M., L.L.E. Koskinen, F.G. van Steenbeek, E.H. Seppala, T.S. Jokinen, P.G.M. Prins, et al., *Paroxysmal Dyskinesia in Border Terriers: Clinical, Epidemiological, and Genetic Investigations.* J Vet Intern Med, 2017. **31**(4): p. 1123-1131.
122. Kyostila, K., S. Cizinauskas, E.H. Seppala, E. Suhonen, J. Jeserevics, A. Sukura, et al., *A SEL1L mutation links a canine progressive early-onset cerebellar ataxia to the endoplasmic reticulum-associated protein degradation (ERAD) machinery.* PLoS Genet, 2012. **8**(6): p. e1002759.

123. Agler, C., D.M. Nielsen, G. Urkasemsin, A. Singleton, N. Tonomura, S. Sigurdsson, et al., *Canine hereditary ataxia in old english sheepdogs and gordon setters is associated with a defect in the autophagy gene encoding RAB24.* PLoS Genet, 2014. **10**(2): p. e1003991.
124. Urkasemsin, G., K.E. Linder, J.S. Bell, A. de Lahunta, and N.J. Olby, *Mapping of Purkinje neuron loss and polyglucosan body accumulation in hereditary cerebellar degeneration in Scottish terriers.* Vet Pathol, 2012. **49**(5): p. 852-9.
125. Steinberg, H.S., J.C. Troncoso, L.C. Cork, and D.L. Price, *Clinical features of inherited cerebellar degeneration in Gordon setters.* J Am Vet Med Assoc, 1981. **179**(9): p. 886-90.
126. Hartley, W.J. and A.C. Palmer, *Ataxia in Jack Russell terriers.* Acta Neuropathol, 1973. **26**(1): p. 71-4.
127. Rohdin, C., L. Ludtke, P. Wohlsein, and K.H. Jaderlund, *New aspects of hereditary ataxia in smooth-haired fox terriers.* Vet Rec, 2010. **166**(18): p. 557-60.
128. Wessmann, A., T. Goedde, A. Fischer, P. Wohlsein, H. Hamann, O. Distl, et al., *Hereditary ataxia in the Jack Russell Terrier--clinical and genetic investigations.* J Vet Intern Med, 2004. **18**(4): p. 515-21.
129. Kleiter, M., S. Hogler, S. Kneissl, A. Url, and M. Leschnik, *Spongy degeneration with cerebellar ataxia in Malinois puppies: a hereditary autosomal recessive disorder?* J Vet Intern Med, 2011. **25**(3): p. 490-6.
130. Mauri, N., M. Kleiter, E. Dietschi, M. Leschnik, S. Hogler, M. Wiedmer, et al., *A SINE Insertion in ATP1B2 in Belgian Shepherd Dogs Affected by Spongy Degeneration with Cerebellar Ataxia (SDCA2).* G3 (Bethesda), 2017. **7**(8): p. 2729-2737.
131. Mauri, N., M. Kleiter, M. Leschnik, S. Hogler, E. Dietschi, M. Wiedmer, et al., *A Missense Variant in KCNJ10 in Belgian Shepherd Dogs Affected by Spongy Degeneration with Cerebellar Ataxia (SDCA1).* G3 (Bethesda), 2017. **7**(2): p. 663-669.
132. Van Poucke, M., K. Stee, L. Sonck, E. Stock, L. Bosseler, J. Van Dorpe, et al., *Truncating SLC12A6 variants cause different clinical phenotypes in humans and dogs.* Eur J Hum Genet, 2019.
133. Christen, M., S. Hogler, M. Kleiter, M. Leschnik, C. Weber, D. Thaller, et al., *Deletion of the SELENOP gene leads to CNS atrophy with cerebellar ataxia in dogs.* PLoS Genet, 2021. **17**(8): p. e1009716.
134. Letko, A., E. Dietschi, M. Nieburg, V. Jagannathan, C. Gurtner, A. Oevermann, et al., *A Missense Variant in SCN8A in Alpine Dachsbracke Dogs Affected by Spinocerebellar Ataxia.* Genes (Basel), 2019. **10**(5).
135. Bellamy, K.K.L., F.S. Skedsmo, J. Hultman, E.F. Arnet, O.A. Guttersrud, H.K. Skogmo, et al., *A 1 bp deletion in HACE1 causes ataxia in Norwegian elkhound, black.* PLoS One, 2022. **17**(1): p. e0261845.
136. Coates, J.R., D.P. O'Brien, K.L. Kline, R.W. Storts, G.C. Johnson, G.D. Shelton, et al., *Neonatal cerebellar ataxia in Coton de Tulear dogs.* J Vet Intern Med, 2002. **16**(6): p. 680-9.
137. Bernardino, F., K. Rentmeister, M.J. Schmidt, A. Bruehschwein, K. Matiasek, L.A. Matiasek, et al., *Inferior cerebellar hypoplasia resembling a Dandy-Walker-like malformation in purebred Eurasier dogs with familial non-progressive ataxia: a retrospective and prospective clinical cohort study.* PLoS One, 2015. **10**(2): p. e0117670.
138. Spennato, P., G. Mirone, A. Nastro, M.C. Buonocore, C. Ruggiero, V. Trischitta, et al., *Hydrocephalus in Dandy-Walker malformation.* Childs Nerv Syst, 2011. **27**(10): p. 1665-81.

139. Kyostila, K., P. Syrja, V. Jagannathan, G. Chandrasekar, T.S. Jokinen, E.H. Seppala, et al., *A missense change in the ATG4D gene links aberrant autophagy to a neurodegenerative vacuolar storage disease.* PLoS Genet, 2015. **11**(4): p. e1005169.
140. Urkasemsin, G., D.M. Nielsen, A. Singleton, S. Arepalli, D. Hernandez, C. Agler, et al., *Genetics of Hereditary Ataxia in Scottish Terriers.* J Vet Intern Med, 2017. **31**(4): p. 1132-1139.
141. Zeng, R., F.H. Farias, G.S. Johnson, S.D. McKay, R.D. Schnabel, J.E. Decker, et al., *A truncated retrotransposon disrupts the GRM1 coding sequence in Coton de Tulear dogs with Bandera's neonatal ataxia.* J Vet Intern Med, 2011. **25**(2): p. 267-72.
142. Watson, L.M., E. Bamber, R.P. Schnekenberg, J. Williams, C. Bettencourt, S. Jayawant, et al., *Dominant Mutations in GRM1 Cause Spinocerebellar Ataxia Type 44.* Am J Hum Genet, 2017. **101**(4): p. 638.
143. Guergueltcheva, V., D.N. Azmanov, D. Angelicheva, K.R. Smith, T. Chamova, L. Florez, et al., *Autosomal-recessive congenital cerebellar ataxia is caused by mutations in metabotropic glutamate receptor 1.* Am J Hum Genet, 2012. **91**(3): p. 553-64.
144. Elsayed, S.M., R. Heller, M. Thoenes, M.S. Zaki, D. Swan, E. Elsobky, et al., *Autosomal dominant SCA5 and autosomal recessive infantile SCA are allelic conditions resulting from SPTBN2 mutations.* Eur J Hum Genet, 2014. **22**(2): p. 286-8.
145. Scholl, U.I., M. Choi, T. Liu, V.T. Ramaekers, M.G. Hausler, J. Grimmer, et al., *Seizures, sensorineural deafness, ataxia, mental retardation, and electrolyte imbalance (SeSAME syndrome) caused by mutations in KCNJ10.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5842-7.
146. Celmina, M., I. Micule, I. Inashkina, M. Audere, S. Kuske, J. Pereca, et al., *EAST/SeSAME syndrome: Review of the literature and introduction of four new Latvian patients.* Clin Genet, 2019. **95**(1): p. 63-78.
147. Rohdin, C., D. Gilliam, C.A. O'Leary, D.P. O'Brien, J.R. Coates, G.S. Johnson, et al., *A KCNJ10 mutation previously identified in the Russell group of terriers also occurs in Smooth-Haired Fox Terriers with hereditary ataxia and in related breeds.* Acta Vet Scand, 2015. **57**: p. 26.
148. van de Leemput, J., J. Chandran, M.A. Knight, L.A. Holtzclaw, S. Scholz, M.R. Cookson, et al., *Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans.* PLoS Genet, 2007. **3**(6): p. e108.
149. Iwaki, A., Y. Kawano, S. Miura, H. Shibata, D. Matsuse, W. Li, et al., *Heterozygous deletion of ITPR1, but not SUMF1, in spinocerebellar ataxia type 16.* J Med Genet, 2008. **45**(1): p. 32-5.
150. Huang, L., J.W. Chardon, M.T. Carter, K.L. Friend, T.E. Dudding, J. Schwartzentruber, et al., *Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia.* Orphanet J Rare Dis, 2012. **7**: p. 67.
151. Gerber, S., K.J. Alzayady, L. Burglen, D. Bremond-Gignac, V. Marchesin, O. Roche, et al., *Recessive and Dominant De Novo ITPR1 Mutations Cause Gillespie Syndrome.* Am J Hum Genet, 2016. **98**(5): p. 971-980.
152. Gerber, M., A. Fischer, V. Jagannathan, M. Drogemuller, C. Drogemuller, M.J. Schmidt, et al., *A deletion in the VLDLR gene in Eurasier dogs with cerebellar hypoplasia resembling a Dandy-Walker-like malformation (DWLM).* PLoS One, 2015. **10**(2): p. e0108917.

153. Boycott, K.M., S. Flavelle, A. Bureau, H.C. Glass, T.M. Fujiwara, E. Wirrell, et al., *Homozygous deletion of the very low density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification.* Am J Hum Genet, 2005. **77**(3): p. 477-83.
154. Thomas, A.C., H. Williams, N. Seto-Salvia, C. Bacchelli, D. Jenkins, M. O'Sullivan, et al., *Mutations in SNX14 cause a distinctive autosomal-recessive cerebellar ataxia and intellectual disability syndrome.* Am J Hum Genet, 2014. **95**(5): p. 611-21.
155. Hollstein, R., D.A. Parry, L. Nalbach, C.V. Logan, T.M. Strom, V.L. Hartill, et al., *HACE1 deficiency causes an autosomal recessive neurodevelopmental syndrome.* J Med Genet, 2015. **52**(12): p. 797-803.
156. Djukic, B., K.B. Casper, B.D. Philpot, L.S. Chin, and K.D. McCarthy, *Conditional knock-out of Kir4.1 leads to glial membrane depolarization, inhibition of potassium and glutamate uptake, and enhanced short-term synaptic potentiation.* J Neurosci, 2007. **27**(42): p. 11354-65.
157. Berridge, M.J., *Inositol trisphosphate and calcium signalling.* Nature, 1993. **361**(6410): p. 315-25.
158. Dellis, O., S.G. Dedos, S.C. Tovey, R. Taufiq Ur, S.J. Dubel, and C.W. Taylor, *Ca2+ entry through plasma membrane IP3 receptors.* Science, 2006. **313**(5784): p. 229-33.
159. Hsieh, C.L., A. Cheng-Deutsch, S. Gloor, M. Schachner, and U. Francke, *Assignment of Amog (adhesion molecule on glia) gene to mouse chromosome 11 near Zfp-3 and Asgr-1,2 and to human chromosome 17.* Somat Cell Mol Genet, 1990. **16**(4): p. 401-5.
160. Gloor, S., H. Antonicek, K.J. Sweadner, S. Pagliusi, R. Frank, M. Moos, et al., *The adhesion molecule on glia (AMOG) is a homologue of the beta subunit of the Na,K-ATPase.* J Cell Biol, 1990. **110**(1): p. 165-74.
161. Larsen, B.R., A. Stoica, and N. MacAulay, *Managing Brain Extracellular K(+) during Neuronal Activity: The Physiological Role of the Na(+)/K(+)-ATPase Subunit Isoforms.* Front Physiol, 2016. **7**: p. 141.
162. Garneau, A.P., A.A. Marcoux, R. Frenette-Cotton, F. Mac-Way, J.L. Lavoie, and P. Isenring, *Molecular insights into the normal operation, regulation, and multisystemic roles of K(+)-Cl(-) cotransporter 3 (KCC3).* Am J Physiol Cell Physiol, 2017. **313**(5): p. C516-C532.
163. Howard, H.C., D.B. Mount, D. Rochefort, N. Byun, N. Dupre, J. Lu, et al., *The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum.* Nat Genet, 2002. **32**(3): p. 384-92.
164. Hu, W., C. Tian, T. Li, M. Yang, H. Hou, and Y. Shu, *Distinct contributions of Na(v)1.6 and Na(v)1.2 in action potential initiation and backpropagation.* Nat Neurosci, 2009. **12**(8): p. 996-1002.
165. Kearney, J.A., D.A. Buchner, G. De Haan, M. Adamska, S.I. Levin, A.R. Furay, et al., *Molecular and pathological effects of a modifier gene on deficiency of the sodium channel Scn8a (Na(v)1.6).* Hum Mol Genet, 2002. **11**(22): p. 2765-75.
166. Johannesen, K.M., E. Gardella, A.C. Encinas, A.E. Lehesjoki, T. Linnankivi, M.B. Petersen, et al., *The spectrum of intermediate SCN8A-related epilepsy.* Epilepsia, 2019. **60**(5): p. 830-844.
167. Trudeau, M.M., J.C. Dalton, J.W. Day, L.P. Ranum, and M.H. Meisler, *Heterozygosity for a protein truncation mutation of sodium channel SCN8A in a patient with cerebellar atrophy, ataxia, and mental retardation.* J Med Genet, 2006. **43**(6): p. 527-30.

168. Yamasaki, M., A. Aiba, M. Kano, and M. Watanabe, *mGluR1 signaling in cerebellar Purkinje cells: Subcellular organization and involvement in cerebellar function and disease.* Neuropharmacology, 2021. **194**: p. 108629.
169. Jackson, M., W. Song, M.Y. Liu, L. Jin, M. Dykes-Hoberg, C.I. Lin, et al., *Modulation of the neuronal glutamate transporter EAAT4 by two interacting proteins.* Nature, 2001. **410**(6824): p. 89-93.
170. Armbrust, K.R., X. Wang, T.J. Hathorn, S.W. Cramer, G. Chen, T. Zu, et al., *Mutant beta-III spectrin causes mGluR1alpha mislocalization and functional deficits in a mouse model of spinocerebellar ataxia type 5.* J Neurosci, 2014. **34**(30): p. 9891-904.
171. Yla-Anttila, P. and E.L. Eskelinen, *Roles for RAB24 in autophagy and disease.* Small GTPases, 2018. **9**(1-2): p. 57-65.
172. Hara, T., K. Nakamura, M. Matsui, A. Yamamoto, Y. Nakahara, R. Suzuki-Migishima, et al., *Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice.* Nature, 2006. **441**(7095): p. 885-9.
173. Komatsu, M., S. Waguri, T. Chiba, S. Murata, J. Iwata, I. Tanida, et al., *Loss of autophagy in the central nervous system causes neurodegeneration in mice.* Nature, 2006. **441**(7095): p. 880-4.
174. Sullivan, R., W.Y. Yau, E. O'Connor, and H. Houlden, *Spinocerebellar ataxia: an update.* J Neurol, 2019. **266**(2): p. 533-544.
175. Tamargo-Gomez, I., G.G. Martinez-Garcia, M.F. Suarez, V. Rey, A. Fueyo, H. Codina-Martinez, et al., *ATG4D is the main ATG8 delipidating enzyme in mammalian cells and protects against cerebellar neurodegeneration.* Cell Death Differ, 2021. **28**(9): p. 2651-2672.
176. Mueller, B., E.J. Klemm, E. Spooner, J.H. Claessen, and H.L. Ploegh, *SEL1L nucleates a protein complex required for dislocation of misfolded glycoproteins.* Proc Natl Acad Sci U S A, 2008. **105**(34): p. 12325-30.
177. Anglesio, M.S., V. Evdokimova, N. Melnyk, L. Zhang, C.V. Fernandez, P.E. Grundy, et al., *Differential expression of a novel ankyrin containing E3 ubiquitin-protein ligase, Hace1, in sporadic Wilms' tumor versus normal kidney.* Hum Mol Genet, 2004. **13**(18): p. 2061-74.
178. Torrino, S., O. Visvikis, A. Doye, L. Boyer, C. Stefani, P. Munro, et al., *The E3 ubiquitin-ligase HACE1 catalyzes the ubiquitylation of active Rac1.* Dev Cell, 2011. **21**(5): p. 959-65.
179. Mulherkar, S., M.D. Uddin, A.D. Couvillon, R.V. Sillitoe, and K.F. Tolias, *The small GTPases RhoA and Rac1 regulate cerebellar development by controlling cell morphogenesis, migration and foliation.* Dev Biol, 2014. **394**(1): p. 39-53.
180. Jossin, Y., *Reelin Functions, Mechanisms of Action and Signaling Pathways During Brain Development and Maturation.* Biomolecules, 2020. **10**(6).
181. Baudry, M. and X. Bi, *Calpain-1 and Calpain-2: The Yin and Yang of Synaptic Plasticity and Neurodegeneration.* Trends Neurosci, 2016. **39**(4): p. 235-245.
182. Huang, H.S., B.J. Yoon, S. Brooks, R. Bakal, J. Berrios, R.S. Larsen, et al., *Snx14 regulates neuronal excitability, promotes synaptic transmission, and is imprinted in the brain of mice.* PLoS One, 2014. **9**(5): p. e98383.
183. Bryant, D., Y. Liu, S. Datta, H. Hariri, M. Seda, G. Anderson, et al., *SNX14 mutations affect endoplasmic reticulum-associated neutral lipid metabolism in autosomal recessive spinocerebellar ataxia 20.* Hum Mol Genet, 2018. **27**(11): p. 1927-1940.
184. Zhang, H., Y. Hong, W. Yang, R. Wang, T. Yao, J. Wang, et al., *SNX14 deficiency-induced defective axonal mitochondrial transport in Purkinje cells underlies*

cerebellar ataxia and can be reversed by valproate. Natl Sci Rev, 2021. **8**(7): p. nwab024.
185. Akizu, N., V. Cantagrel, M.S. Zaki, L. Al-Gazali, X. Wang, R.O. Rosti, et al., *Biallelic mutations in SNX14 cause a syndromic form of cerebellar atrophy and lysosome-autophagosome dysfunction.* Nat Genet, 2015. **47**(5): p. 528-34.
186. Solovyev, N.D., *Importance of selenium and selenoprotein for brain function: From antioxidant protection to neuronal signalling.* J Inorg Biochem, 2015. **153**: p. 1-12.
187. Lowrie, M. and L. Garosi, *Classification of involuntary movements in dogs: Paroxysmal dyskinesias.* Vet J, 2017. **220**: p. 65-71.
188. Cerda-Gonzalez, S., R.A. Packer, L. Garosi, M. Lowrie, P.J.J. Mandigers, D.P. O'Brien, et al., *International veterinary canine dyskinesia task force ECVN consensus statement: Terminology and classification.* J Vet Intern Med, 2021. **35**(3): p. 1218-1230.
189. Forman, O.P., J. Penderis, C. Hartley, L.J. Hayward, S.L. Ricketts, and C.S. Mellersh, *Parallel mapping and simultaneous sequencing reveals deletions in BCAN and FAM83H associated with discrete inherited disorders in a domestic dog breed.* PLoS Genet, 2012. **8**(1): p. e1002462.
190. Gill, J.L., K.L. Tsai, C. Krey, R.E. Noorai, J.F. Vanbellinghen, L.S. Garosi, et al., *A canine BCAN microdeletion associated with episodic falling syndrome.* Neurobiol Dis, 2012. **45**(1): p. 130-6.
191. Kolicheski, A.L., G.S. Johnson, T. Mhlanga-Mutangadura, J.F. Taylor, R.D. Schnabel, T. Kinoshita, et al., *A homozygous PIGN missense mutation in Soft-Coated Wheaten Terriers with a canine paroxysmal dyskinesia.* Neurogenetics, 2017. **18**(1): p. 39-47.
192. Nessler, J., P. Hug, P.J.J. Mandigers, P.A.J. Leegwater, V. Jagannathan, A.M. Das, et al., *Mitochondrial PCK2 Missense Variant in Shetland Sheepdogs with Paroxysmal Exercise-Induced Dyskinesia (PED).* Genes (Basel), 2020. **11**(7).
193. Mandigers, P.J.J., F.G. Van Steenbeek, W. Bergmann, M. Vos-Loohuis, and P.A. Leegwater, *A knockout mutation associated with juvenile paroxysmal dyskinesia in Markiesje dogs indicates SOD1 pleiotropy.* Hum Genet, 2021. **140**(11): p. 1547-1552.
194. Lowrie, M., O.A. Garden, M. Hadjivassiliou, R.J. Harvey, D.S. Sanders, R. Powell, et al., *The Clinical and Serological Effect of a Gluten-Free Diet in Border Terriers with Epileptoid Cramping Syndrome.* J Vet Intern Med, 2015. **29**(6): p. 1564-8.
195. Lowrie, M., O.A. Garden, M. Hadjivassiliou, D.S. Sanders, R. Powell, and L. Garosi, *Characterization of Paroxysmal Gluten-Sensitive Dyskinesia in Border Terriers Using Serological Markers.* J Vet Intern Med, 2018. **32**(2): p. 775-781.
196. Kube, S.A., K.M. Vernau, and R.A. LeCouteur, *Dyskinesia associated with oral phenobarbital administration in a dog.* J Vet Intern Med, 2006. **20**(5): p. 1238-40.
197. Herrtage, M.E. and A.C. Palmer, *Episodic falling in the cavalier King Charles spaniel.* Vet Rec, 1983. **112**(19): p. 458-9.
198. Black, V., L. Garosi, M. Lowrie, R.J. Harvey, and J. Gale, *Phenotypic characterisation of canine epileptoid cramping syndrome in the Border terrier.* J Small Anim Pract, 2014. **55**(2): p. 102-7.
199. Packer, R.A., I. Wachowiak, S.A. Thomovsky, J. Berg, L. Vasquez, and D.P. O'Brien, *Phenotypic characterization of PIGN-associated paroxysmal dyskinesia in Soft-coated wheaten terriers and preliminary response to acetazolamide therapy.* Vet J, 2021. **269**: p. 105606.
200. Riviere, J.E. and M.G. Papich, *Veterinary Pharmacology and Therapeutics*. Tenth Edition ed. 2018: Wiley.

201. Brakebusch, C., C.I. Seidenbecher, F. Asztely, U. Rauch, H. Matthies, H. Meyer, et al., *Brevican-deficient mice display impaired hippocampal CA1 long-term potentiation but show no obvious deficits in learning and memory.* Mol Cell Biol, 2002. **22**(21): p. 7417-27.
202. Hu, B., L.L. Kong, R.T. Matthews, and M.S. Viapiano, *The proteoglycan brevican binds to fibronectin after proteolytic cleavage and promotes glioma cell motility.* J Biol Chem, 2008. **283**(36): p. 24848-59.
203. Kearsley-Fleet, L., D.G. O'Neill, H.A. Volk, D.B. Church, and D.C. Brodbelt, *Prevalence and risk factors for canine epilepsy of unknown origin in the UK.* Vet Rec, 2013. **172**(13): p. 338.
204. Orlean, P. and A.K. Menon, *Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids.* J Lipid Res, 2007. **48**(5): p. 993-1011.
205. Bellai-Dussault, K., T.T.M. Nguyen, N.V. Baratang, D.A. Jimenez-Cruz, and P.M. Campeau, *Clinical variability in inherited glycosylphosphatidylinositol deficiency disorders.* Clin Genet, 2019. **95**(1): p. 112-121.
206. Bayat, A., G. de Valles-Ibanez, M. Pendziwiat, A. Knaus, K. Alt, E. Biamino, et al., *PIGN encephalopathy: Characterizing the epileptology.* Epilepsia, 2022.
207. Bluemel, G., M. Planque, C.T. Madreiter-Sokolowski, T. Haitzmann, A. Hrzenjak, W.F. Graier, et al., *PCK2 opposes mitochondrial respiration and maintains the redox balance in starved lung cancer cells.* Free Radic Biol Med, 2021. **176**: p. 34-45.
208. Hommes, F.A., K. Bendien, J.D. Elema, H.J. Bremer, and I. Lombeck, *Two cases of phosphoenolpyruvate carboxykinase deficiency.* Acta Paediatr Scand, 1976. **65**(2): p. 233-40.
209. Xu, M., J. Yang, J. Sun, X. Xing, Z. Liu, and T. Liu, *A novel mutation in PCK2 gene causes primary angle-closure glaucoma.* Aging (Albany NY), 2021. **13**(19): p. 23338-23347.
210. Eleutherio, E.C.A., R.S. Silva Magalhaes, A. de Araujo Brasil, J.R. Monteiro Neto, and L. de Holanda Paranhos, *SOD1, more than just an antioxidant.* Arch Biochem Biophys, 2021. **697**: p. 108701.
211. Awano, T., G.S. Johnson, C.M. Wade, M.L. Katz, G.C. Johnson, J.F. Taylor, et al., *Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis.* Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2794-9.
212. Zeng, R., J.R. Coates, G.C. Johnson, L. Hansen, T. Awano, A. Kolicheski, et al., *Breed distribution of SOD1 alleles previously associated with canine degenerative myelopathy.* J Vet Intern Med, 2014. **28**(2): p. 515-21.
213. Qi, Y., P. Montague, C. Loney, C. Campbell, I.N.F. Shafie, T.J. Anderson, et al., *In vitro evidence consistent with an interaction between wild-type and mutant SOD1 protein associated with canine degenerative myelopathy.* Eur J Neurosci, 2019. **50**(12): p. 3896-3905.
214. Ekhtiari Bidhendi, E., J. Bergh, P. Zetterstrom, K. Forsberg, B. Pakkenberg, P.M. Andersen, et al., *Mutant superoxide dismutase aggregates from human spinal cord transmit amyotrophic lateral sclerosis.* Acta Neuropathol, 2018. **136**(6): p. 939-953.
215. Fisher, R.S., C. Acevedo, A. Arzimanoglou, A. Bogacz, J.H. Cross, C.E. Elger, et al., *ILAE official report: a practical clinical definition of epilepsy.* Epilepsia, 2014. **55**(4): p. 475-82.

216. Fisher, R.S., J.H. Cross, J.A. French, N. Higurashi, E. Hirsch, F.E. Jansen, et al., *Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology.* Epilepsia, 2017. **58**(4): p. 522-530.
217. Devinsky, O., A. Vezzani, T.J. O'Brien, N. Jette, I.E. Scheffer, M. de Curtis, et al., *Epilepsy.* Nat Rev Dis Primers, 2018. **4**: p. 18024.
218. Stafstrom, C.E. and L. Carmant, *Seizures and epilepsy: an overview for neuroscientists.* Cold Spring Harb Perspect Med, 2015. **5**(6).
219. Berg, A.T. and J.J. Millichap, *The 2010 revised classification of seizures and epilepsy.* Continuum (Minneap Minn), 2013. **19**(3 Epilepsy): p. 571-97.
220. Scheffer, I.E., S. Berkovic, G. Capovilla, M.B. Connolly, J. French, L. Guilhoto, et al., *ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology.* Epilepsia, 2017. **58**(4): p. 512-521.
221. Ellis, C.A., S. Petrovski, and S.F. Berkovic, *Epilepsy genetics: clinical impacts and biological insights.* Lancet Neurol, 2020. **19**(1): p. 93-100.
222. Perucca, P., M. Bahlo, and S.F. Berkovic, *The Genetics of Epilepsy.* Annu Rev Genomics Hum Genet, 2020. **21**: p. 205-230.
223. Martin, A.R., E. Williams, R.E. Foulger, S. Leigh, L.C. Daugherty, O. Niblock, et al., *PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels.* Nat Genet, 2019. **51**(11): p. 1560-1565.
224. Genomics England PanelApp. *Genetic epilepsy syndromes (version: 3.0)* 2022 [cited 2023 23/01/2023]; Available from: https://panelapp.genomicsengland.co.uk.
225. Myers, C.T. and H.C. Mefford, *Genetic investigations of the epileptic encephalopathies: Recent advances.* Prog Brain Res, 2016. **226**: p. 35-60.
226. Berg, A.T., S.F. Berkovic, M.J. Brodie, J. Buchhalter, J.H. Cross, W. van Emde Boas, et al., *Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009.* Epilepsia, 2010. **51**(4): p. 676-85.
227. Claes, L., J. Del-Favero, B. Ceulemans, L. Lagae, C. Van Broeckhoven, and P. De Jonghe, *De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy.* Am J Hum Genet, 2001. **68**(6): p. 1327-32.
228. Marini, C., I.E. Scheffer, R. Nabbout, A. Suls, P. De Jonghe, F. Zara, et al., *The genetics of Dravet syndrome.* Epilepsia, 2011. **52 Suppl 2**: p. 24-9.
229. Epi4K Consortium, Epilepsy Phenome/Genome Project, A.S. Allen, S.F. Berkovic, P. Cossette, N. Delanty, et al., *De novo mutations in epileptic encephalopathies.* Nature, 2013. **501**(7466): p. 217-21.
230. Epi4K Consortium, *De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies.* Am J Hum Genet, 2016. **99**(2): p. 287-98.
231. International League Against Epilepsy Consortium on Complex Epilepsies, *Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies.* Nat Commun, 2018. **9**(1): p. 5269.
232. Koeleman, B.P.C., *What do genetic studies tell us about the heritable basis of common epilepsy? Polygenic or complex epilepsy?* Neurosci Lett, 2018. **667**: p. 10-16.
233. Epi4K consortium and Epilepsy Phenome/Genome Project, *Ultra-rare genetic variation in common epilepsies: a case-control sequencing study.* Lancet Neurol, 2017. **16**(2): p. 135-143.

234. Epi25 Collaborative, *Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals.* Am J Hum Genet, 2019. **105**(2): p. 267-282.
235. International League Against Epilepsy Consortium on Complex Epilepsies, *Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies.* Lancet Neurol, 2014. **13**(9): p. 893-903.
236. Casal, M.L., R.M. Munuve, M.A. Janis, P. Werner, and P.S. Henthorn, *Epilepsy in Irish Wolfhounds.* J Vet Intern Med, 2006. **20**(1): p. 131-5.
237. Packer, R.M. and H.A. Volk, *Epilepsy beyond seizures: a review of the impact of epilepsy and its comorbidities on health-related quality of life in dogs.* Vet Rec, 2015. **177**(12): p. 306-15.
238. De Risio, L., J. Freeman, and A. Shea, *Evaluation of quality of life of carers of Italian spinoni with idiopathic epilepsy.* Vet Rec Open, 2016. **3**(1): p. e000174.
239. Fredso, N., N. Toft, A. Sabers, and M. Berendt, *A prospective observational longitudinal study of new-onset seizures and newly diagnosed epilepsy in dogs.* BMC Vet Res, 2017. **13**(1): p. 54.
240. Berendt, M., R.G. Farquhar, P.J. Mandigers, A. Pakozdy, S.F. Bhatti, L. De Risio, et al., *International veterinary epilepsy task force consensus report on epilepsy definition, classification and terminology in companion animals.* BMC Vet Res, 2015. **11**: p. 182.
241. Hamamoto, Y., D. Hasegawa, S. Mizoguchi, Y. Yu, M. Wada, T. Kuwabara, et al., *Retrospective epidemiological study of canine epilepsy in Japan using the International Veterinary Epilepsy Task Force classification 2015 (2003-2013): etiological distribution, risk factors, survival time, and lifespan.* BMC Vet Res, 2016. **12**(1): p. 248.
242. Hall, R., J. Labruyere, H. Volk, and T.J. Cardy, *Estimation of the prevalence of idiopathic epilepsy and structural epilepsy in a general population of 900 dogs undergoing MRI for epileptic seizures.* Vet Rec, 2020. **187**(10): p. e89.
243. Lohi, H., E.J. Young, S.N. Fitzmaurice, C. Rusbridge, E.M. Chan, M. Vervoort, et al., *Expanded repeat in canine epilepsy.* Science, 2005. **307**(5706): p. 81.
244. von Klopmann, T., S. Ahonen, I. Espadas-Santiuste, K. Matiasek, D. Sanchez-Masian, S. Rupp, et al., *Canine Lafora Disease: An Unstable Repeat Expansion Disorder.* Life (Basel), 2021. **11**(7).
245. Duran, J., M.F. Tevy, M. Garcia-Rocha, J. Calbo, M. Milan, and J.J. Guinovart, *Deleterious effects of neuronal accumulation of glycogen in flies and mice.* EMBO Mol Med, 2012. **4**(8): p. 719-29.
246. Hajek, I., F. Kettner, V. Simerdova, C. Rusbridge, P. Wang, B.A. Minassian, et al., *NHLRC1 repeat expansion in two beagles with Lafora disease.* J Small Anim Pract, 2016. **57**(11): p. 650-652.
247. Barrientos, L., A. Maiolini, A. Hani, V. Jagannathan, and T. Leeb, *NHLRC1 dodecamer repeat expansion demonstrated by whole genome sequencing in a Chihuahua with Lafora disease.* Anim Genet, 2019. **50**(1): p. 118-119.
248. Hytonen, M.K., R. Sarviaho, C.B. Jackson, P. Syrja, T. Jokinen, K. Matiasek, et al., *In-frame deletion in canine PITRM1 is associated with a severe early-onset epilepsy, mitochondrial dysfunction and neurodegeneration.* Hum Genet, 2021. **140**(11): p. 1593-1609.
249. Hulsmeyer, V.I., A. Fischer, P.J. Mandigers, L. DeRisio, M. Berendt, C. Rusbridge, et al., *International Veterinary Epilepsy Task Force's current understanding of idiopathic epilepsy of genetic or suspected genetic origin in purebred dogs.* BMC Vet Res, 2015. **11**: p. 175.

250. De Risio, L., S. Bhatti, K. Munana, J. Penderis, V. Stein, A. Tipold, et al., *International veterinary epilepsy task force consensus proposal: diagnostic approach to epilepsy in dogs.* BMC Vet Res, 2015. **11**: p. 148.
251. De Risio, L., R. Newton, J. Freeman, and A. Shea, *Idiopathic epilepsy in the Italian Spinone in the United Kingdom: prevalence, clinical characteristics, and predictors of survival and seizure remission.* J Vet Intern Med, 2015. **29**(3): p. 917-24.
252. Berendt, M., C.H. Gulløv, S.L. Christensen, H. Gudmundsdottir, H. Gredal, M. Fredholm, et al., *Prevalence and characteristics of epilepsy in the Belgian shepherd variants Groenendael and Tervueren born in Denmark 1995-2004.* Acta Vet Scand, 2008. **50**: p. 51.
253. Berendt, M., C.H. Gulløv, and M. Fredholm, *Focal epilepsy in the Belgian shepherd: evidence for simple Mendelian inheritance.* J Small Anim Pract, 2009. **50**(12): p. 655-61.
254. Hülsmeyer, V., R. Zimmermann, C. Brauer, C. Sauter-Louis, and A. Fischer, *Epilepsy in Border Collies: clinical manifestation, outcome, and mode of inheritance.* J Vet Intern Med, 2010. **24**(1): p. 171-8.
255. Santifort, K.M., E. Bertijn, S.F.M. Bhatti, P. Leegwater, A. Fischer, and P.J.J. Mandigers, *Phenotypic Characterization of Idiopathic Epilepsy in Border Collies.* Front Vet Sci, 2022. **9**: p. 880318.
256. Gulløv, C.H., N. Toft, M.M. Baadsager, and M. Berendt, *Epilepsy in the Petit Basset Griffon Vendeen: prevalence, semiology, and clinical phenotype.* J Vet Intern Med, 2011. **25**(6): p. 1372-8.
257. Monteiro, R., V. Adams, D. Keys, and S.R. Platt, *Canine idiopathic epilepsy: prevalence, risk factors and outcome associated with cluster seizures and status epilepticus.* J Small Anim Pract, 2012. **53**(9): p. 526-30.
258. Gulløv, C.H., N. Toft, and M. Berendt, *A longitudinal study of survival in Belgian Shepherds with genetic epilepsy.* J Vet Intern Med, 2012. **26**(5): p. 1115-20.
259. Packer, R.M., N.K. Shihab, B.B. Torres, and H.A. Volk, *Risk factors for cluster seizures in canine idiopathic epilepsy.* Res Vet Sci, 2016. **105**: p. 136-8.
260. den Hertog, E., *Seizure-precipitating factors in dogs with idiopathic epilepsy.* J Vet Intern Med, 2019. **33**(3): p. 1120.
261. Forsgard, J., L. Metsahonkala, A.M. Kiviranta, S. Cizinauskas, J.J.T. Junnila, O. Laitinen-Vapaavuori, et al., *Response to letter to editor regarding Seizure-precipitating factors in dogs with idiopathic epilepsy.* J Vet Intern Med, 2019. **33**(3): p. 1121-1122.
262. Forsgard, J.A., L. Metsahonkala, A.M. Kiviranta, S. Cizinauskas, J.J.T. Junnila, O. Laitinen-Vapaavuori, et al., *Seizure-precipitating factors in dogs with idiopathic epilepsy.* J Vet Intern Med, 2019. **33**(2): p. 701-707.
263. Finnegan, S.L., H.A. Volk, L. Asher, M. Daley, and R.M.A. Packer, *Investigating the potential for seizure prediction in dogs with idiopathic epilepsy: owner-reported prodromal changes and seizure triggers.* Vet Rec, 2020. **187**(4): p. 152.
264. Fischer, A., *Idiopathic epilepsy in dogs: insights into factors that may predict upcoming seizure activity.* Vet Rec, 2020. **187**(4): p. 149-151.
265. Arrol, L., J. Penderis, L. Garosi, P. Cripps, R. Gutierrez-Quintana, and R. Goncalves, *Aetiology and long-term outcome of juvenile epilepsy in 136 dogs.* Vet Rec, 2012. **170**(13): p. 335.
266. Jokinen, T.S., L. Metsahonkala, L. Bergamasco, R. Viitmaa, P. Syrja, H. Lohi, et al., *Benign familial juvenile epilepsy in Lagotto Romagnolo dogs.* J Vet Intern Med, 2007. **21**(3): p. 464-71.

267. Jokinen, T.S., K. Tiira, L. Metsahonkala, E.H. Seppala, A. Hielm-Bjorkman, H. Lohi, et al., *Behavioral Abnormalities in Lagotto Romagnolo Dogs with a History of Benign Familial Juvenile Epilepsy: A Long-Term Follow-Up Study.* J Vet Intern Med, 2015. **29**(4): p. 1081-7.
268. Wielaender, F., R. Sarviaho, F. James, M.K. Hytonen, M.A. Cortez, G. Kluger, et al., *Generalized myoclonic epilepsy with photosensitivity in juvenile dogs caused by a defective DIRAS family GTPase 1.* Proc Natl Acad Sci U S A, 2017. **114**(10): p. 2669-2674.
269. Wielaender, F., F.M.K. James, M.A. Cortez, G. Kluger, J.N. Nessler, A. Tipold, et al., *Absence Seizures as a Feature of Juvenile Myoclonic Epilepsy in Rhodesian Ridgeback Dogs.* J Vet Intern Med, 2018. **32**(1): p. 428-432.
270. Murray, J.K., R.H. Kinsman, M.S. Lord, R.E.P. Da Costa, J.L. Woodward, S.C. Owczarczak-Garstecka, et al., *'Generation Pup' - protocol for a longitudinal study of dog behaviour and health.* BMC Vet Res, 2021. **17**(1): p. 1.
271. Erlen, A., H. Potschka, H.A. Volk, C. Sauter-Louis, and D.G. O'Neill, *Seizure occurrence in dogs under primary veterinary care in the UK: prevalence and risk factors.* J Vet Intern Med, 2018. **32**(5): p. 1665-1676.
272. Van Meervenne, S.A., H.A. Volk, K. Matiasek, and L.M. Van Ham, *The influence of sex hormones on seizures in dogs and humans.* Vet J, 2014. **201**(1): p. 15-20.
273. Short, A.D., A. Dunne, H. Lohi, S. Boulton, S.D. Carter, D. Timofte, et al., *Characteristics of epileptic episodes in UK dog breeds: an epidemiological approach.* Vet Rec, 2011. **169**(2): p. 48.
274. Heske, L., A. Nodtvedt, K.H. Jaderlund, M. Berendt, and A. Egenvall, *A cohort study of epilepsy among 665,000 insured dogs: incidence, mortality and survival after diagnosis.* Vet J, 2014. **202**(3): p. 471-6.
275. Oberbauer, A.M., J.M. Belanger, D.I. Grossman, K.R. Regan, and T.R. Famula, *Genome-wide linkage scan for loci associated with epilepsy in Belgian shepherd dogs.* BMC Genet, 2010. **11**: p. 35.
276. Seppala, E.H., L.L. Koskinen, C.H. Gullov, P. Jokinen, P. Karlskov-Mortensen, L. Bergamasco, et al., *Identification of a novel idiopathic epilepsy locus in Belgian Shepherd dogs.* PLoS One, 2012. **7**(3): p. e33549.
277. Van Meervenne, S., H.A. Volk, P.S. Verhoeven, L. Van Ham, and D.G. O'Neill, *Associations between neutering and idiopathic epilepsy in Labrador retrievers and Border collies under primary veterinary care in the UK.* Vet J, 2019. **252**: p. 105354.
278. Norton, E.M., K.M. Minor, S.M. Taylor, M.E. McCue, and J.R. Mickelson, *Heritability and Genomic Architecture of Episodic Exercise-Induced Collapse in Border Collies.* Genes (Basel), 2021. **12**(12).
279. Bhatti, S.F., L. De Risio, K. Munana, J. Penderis, V.M. Stein, A. Tipold, et al., *International Veterinary Epilepsy Task Force consensus proposal: medical treatment of canine epilepsy in Europe.* BMC Vet Res, 2015. **11**: p. 176.
280. Faingold, C.L. and R.A. Browning, *Mechanisms of anticonvulsant drug action. I. Drugs primarily used for generalized tonic-clonic and partial epilepsies.* Eur J Pediatr, 1987. **146**(1): p. 2-7.
281. Rundfeldt, C. and W. Loscher, *The pharmacology of imepitoin: the first partial benzodiazepine receptor agonist developed for the treatment of epilepsy.* CNS Drugs, 2014. **28**(1): p. 29-43.
282. Charalambous, M., S.K. Shivapour, D.C. Brodbelt, and H.A. Volk, *Antiepileptic drugs' tolerability and safety - a systematic review and meta-analysis of adverse effects in dogs.* BMC Vet Res, 2016. **12**: p. 79.

283. Leppik, I.E., *Zonisamide: chemistry, mechanism of action, and pharmacokinetics.* Seizure, 2004. **13 Suppl 1**: p. S5-9; discussion S10.
284. Perucca, E., *A pharmacological and clinical review on topiramate, a new antiepileptic drug.* Pharmacol Res, 1997. **35**(4): p. 241-56.
285. Taylor, C.P., T. Angelotti, and E. Fauman, *Pharmacology and mechanism of action of pregabalin: the calcium channel alpha2-delta (alpha2-delta) subunit as a target for antiepileptic drug discovery.* Epilepsy Res, 2007. **73**(2): p. 137-50.
286. Law, T.H., E.S. Davies, Y. Pan, B. Zanghi, E. Want, and H.A. Volk, *A randomised trial of a medium-chain TAG diet as treatment for dogs with idiopathic epilepsy.* Br J Nutr, 2015. **114**(9): p. 1438-47.
287. Packer, R.M., T.H. Law, E. Davies, B. Zanghi, Y. Pan, and H.A. Volk, *Effects of a ketogenic diet on ADHD-like behavior in dogs with idiopathic epilepsy.* Epilepsy Behav, 2016. **55**: p. 62-8.
288. Berk, B.A., T.H. Law, R.M.A. Packer, A. Wessmann, A. Bathen-Nothen, T.S. Jokinen, et al., *A multicenter randomized controlled trial of medium-chain triglyceride dietary supplementation on epilepsy in dogs.* J Vet Intern Med, 2020. **34**(3): p. 1248-1259.
289. Berk, B.A., R.M.A. Packer, T.H. Law, A. Wessmann, A. Bathen-Nothen, T.S. Jokinen, et al., *Medium-chain triglycerides dietary supplement improves cognitive abilities in canine epilepsy.* Epilepsy Behav, 2021. **114**(Pt A): p. 107608.
290. Koskinen, L.L., E.H. Seppala, J.M. Belanger, M. Arumilli, O. Hakosalo, P. Jokinen, et al., *Identification of a common risk haplotype for canine idiopathic epilepsy in the ADAM23 gene.* BMC Genomics, 2015. **16**: p. 465.
291. Koskinen, L.L.E., E.H. Seppälä, J. Weissl, T.S. Jokinen, R. Viitmaa, R.L. Hänninen, et al., *ADAM23 is a common risk gene for canine idiopathic epilepsy.* BMC Genetics, 2017. **18**(1): p. 8.
292. Belanger, J.M., T.R. Famula, L.C. Gershony, M.K. Palij, and A.M. Oberbauer, *Genome-wide association analysis of idiopathic epilepsy in the Belgian shepherd.* Canine Medicine and Genetics, 2020. **7**(1): p. 12.
293. Deschain, T., J. Fabricius, M. Berendt, M. Fredholm, and P. Karlskov-Mortensen, *The first genome-wide association study concerning idiopathic epilepsy in Petit Basset Griffon Vendeen.* Anim Genet, 2021. **52**(5): p. 762-766.
294. Seppala, E.H., T.S. Jokinen, M. Fukata, Y. Fukata, M.T. Webster, E.K. Karlsson, et al., *LGI2 truncation causes a remitting focal epilepsy in dogs.* PLoS Genet, 2011. **7**(7): p. e1002194.
295. Kegel, L., E. Aunin, D. Meijer, and J.R. Bermingham, *LGI proteins in the nervous system.* ASN Neuro, 2013. **5**(3): p. 167-81.
296. Kalachikov, S., O. Evgrafov, B. Ross, M. Winawer, C. Barker-Cummings, F. Martinelli Boneschi, et al., *Mutations in LGI1 cause autosomal-dominant partial epilepsy with auditory features.* Nat Genet, 2002. **30**(3): p. 335-41.
297. Morante-Redolat, J.M., A. Gorostidi-Pagola, S. Piquer-Sirerol, A. Saenz, J.J. Poza, J. Galan, et al., *Mutations in the LGI1/Epitempin gene on 10q24 cause autosomal dominant lateral temporal epilepsy.* Hum Mol Genet, 2002. **11**(9): p. 1119-28.
298. Ottman, R., M.R. Winawer, S. Kalachikov, C. Barker-Cummings, T.C. Gilliam, T.A. Pedley, et al., *LGI1 mutations in autosomal dominant partial epilepsy with auditory features.* Neurology, 2004. **62**(7): p. 1120-6.
299. Hsia, H.E., J. Tushaus, T. Brummer, Y. Zheng, S.D. Scilabra, and S.F. Lichtenthaler, *Functions of 'A disintegrin and metalloproteases (ADAMs)' in the mammalian nervous system.* Cell Mol Life Sci, 2019. **76**(16): p. 3055-3081.

300. Lancaster, E., E. Burnor, J. Zhang, and E. Lancaster, *ADAM23 is a negative regulator of Kv1.1/Kv1.4 potassium currents.* Neurosci Lett, 2019. **704**: p. 159-163.
301. Hivert, B., L. Marien, K.N. Agbam, and C. Faivre-Sarrailh, *ADAM22 and ADAM23 modulate the targeting of the Kv1 channel-associated protein LGI1 to the axon initial segment.* J Cell Sci, 2019. **132**(2).
302. Winden, K.D., A. Bragin, J. Engel, and D.H. Geschwind, *Molecular alterations in areas generating fast ripples in an animal model of temporal lobe epilepsy.* Neurobiol Dis, 2015. **78**: p. 35-44.
303. Karlskov-Mortensen, P., *Risk loci for idiopathic epilepsy in Belgian Shepherd dogs.* Anim Genet, 2021. **52**(1): p. 139-140.
304. Belanger, J.M., T. Heinonen, T.R. Famula, P.J.J. Mandigers, P.A. Leegwater, M.K. Hytonen, et al., *Validation of a Chromosome 14 Risk Haplotype for Idiopathic Epilepsy in the Belgian Shepherd Dog Found to Be Associated with an Insertion in the RAPGEF5 Gene.* Genes (Basel), 2022. **13**(7).
305. Yeh, C.W. and L.S. Hsu, *Zebrafish diras1 Promoted Neurite Outgrowth in Neuro-2a Cells and Maintained Trigeminal Ganglion Neurons In Vivo via Rac1-Dependent Pathway.* Mol Neurobiol, 2016. **53**(10): p. 6594-6607.
306. Tada, M., K. Gengyo-Ando, T. Kobayashi, M. Fukuyama, S. Mitani, K. Kontani, et al., *Neuronally expressed Ras-family GTPase Di-Ras modulates synaptic activity in Caenorhabditis elegans.* Genes Cells, 2012. **17**(9): p. 778-89.
307. Ekenstedt, K.J., E.E. Patterson, K.M. Minor, and J.R. Mickelson, *Candidate genes for idiopathic epilepsy in four dog breeds.* BMC Genet, 2011. **12**: p. 38.
308. Jones, G.M.C., H.A. Volk, and R.M.A. Packer, *Research priorities for idiopathic epilepsy in dogs: Viewpoints of owners, general practice veterinarians, and neurology specialists.* J Vet Intern Med, 2021. **35**(3): p. 1466-1479.
309. Alves, L., V. Hulsmeyer, A. Jaggy, A. Fischer, T. Leeb, and M. Drogemuller, *Polymorphisms in the ABCB1 gene in phenobarbital responsive and resistant idiopathic epileptic Border Collies.* J Vet Intern Med, 2011. **25**(3): p. 484-9.
310. Muñana, K.R., J.A. Nettifee-Osborne, R.L. Bergman, Jr., and K.L. Mealey, *Association between ABCB1 genotype and seizure outcome in Collies with epilepsy.* J Vet Intern Med, 2012. **26**(6): p. 1358-64.
311. Mizukami, K., A. Yabuki, H.S. Chang, M.M. Uddin, M.M. Rahman, K. Kushida, et al., *High frequency of a single nucleotide substitution (c.-6-180T>G) of the canine MDR1/ABCB1 gene associated with phenobarbital-resistant idiopathic epilepsy in Border Collie dogs.* Dis Markers, 2013. **35**(6): p. 669-72.
312. Gagliardo, T., G. Gandini, A. Gallucci, M. Menchetti, E. Bianchi, M.E. Turba, et al., *ABCB1 c.-6-180T>G polymorphism and clinical risk factors in a multi-breed cohort of dogs with refractory idiopathic epilepsy.* Vet J, 2019. **253**: p. 105378.
313. Schrickx, J.A. and J. Fink-Gremmels, *Implications of ABC transporters on the disposition of typical veterinary medicinal products.* Eur J Pharmacol, 2008. **585**(2-3): p. 510-9.
314. Mealey, K.L., S.A. Bentjen, J.M. Gay, and G.H. Cantor, *Ivermectin sensitivity in collies is associated with a deletion mutation of the mdr1 gene.* Pharmacogenetics, 2001. **11**(8): p. 727-33.
315. Mari, L., K. Matiasek, C.A. Jenkins, A. De Stefani, S.L. Ricketts, O. Forman, et al., *Hereditary ataxia in four related Norwegian Buhunds.* J Am Vet Med Assoc, 2018. **253**(6): p. 774-780.
316. The Kennel Club. *Comparative tables of registrations for the years 2012 - 2021 inclusive - Pastoral*. Breed registration statistics 2023 [cited 2023 16/01/2023];

Available from: https://www.thekennelclub.org.uk/media-centre/breed-registration-statistics/.
317. Matilla-Dueñas, A., M. Corral-Juan, V. Volpini, and I. Sanchez, *The Spinocerebellar Ataxias: Clinical Aspects And Molecular Genetics*, in *Neurodegenerative Diseases*, S.I. Ahmad, Editor. 2012, Springer US: New York, NY. p. 351-374.
318. Mhlanga-Mutangadura, T., G.S. Johnson, A. Ashwini, G.D. Shelton, S.A. Wennogle, G.C. Johnson, et al., *A Homozygous RAB3GAP1:c.743delC Mutation in Rottweilers with Neuronal Vacuolation and Spinocerebellar Degeneration.* J Vet Intern Med, 2016. **30**(3): p. 813-8.
319. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.
320. Adzhubei, I.A., S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
321. Schwarz, J.M., D.N. Cooper, M. Schuelke, and D. Seelow, *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods, 2014. **11**(4): p. 361-2.
322. Pruunsild, P. and T. Timmusk, *Structure, alternative splicing, and expression of the human and mouse KCNIP gene family.* Genomics, 2005. **86**(5): p. 581-93.
323. Yin, S., F. Ding, and N.V. Dokholyan, *Eris: an automated estimator of protein stability.* Nat Methods, 2007. **4**(6): p. 466-7.
324. An, W.F., M.R. Bowlby, M. Betty, J. Cao, H.P. Ling, G. Mendoza, et al., *Modulation of A-type potassium channels by a family of calcium sensors.* Nature, 2000. **403**(6769): p. 553-6.
325. Morohashi, Y., N. Hatano, S. Ohya, R. Takikawa, T. Watabiki, N. Takasugi, et al., *Molecular cloning and characterization of CALP/KChIP4, a novel EF-hand protein interacting with presenilin 2 and voltage-gated potassium channel subunit Kv4.* J Biol Chem, 2002. **277**(17): p. 14965-75.
326. Duarri, A., J. Jezierska, M. Fokkens, M. Meijer, H.J. Schelhaas, W.F. den Dunnen, et al., *Mutations in potassium channel kcnd3 cause spinocerebellar ataxia type 19.* Ann Neurol, 2012. **72**(6): p. 870-80.
327. Lee, Y.C., A. Durr, K. Majczenko, Y.H. Huang, Y.C. Liu, C.C. Lien, et al., *Mutations in KCND3 cause spinocerebellar ataxia type 22.* Ann Neurol, 2012. **72**(6): p. 859-69.
328. Smets, K., A. Duarri, T. Deconinck, B. Ceulemans, B.P. van de Warrenburg, S. Züchner, et al., *First de novo KCND3 mutation causes severe Kv4.3 channel dysfunction leading to early onset cerebellar ataxia, intellectual disability, oral apraxia and epilepsy.* BMC Med Genet, 2015. **16**: p. 51.
329. Coutelier, M., G. Coarelli, M.L. Monin, J. Konop, C.S. Davoine, C. Tesson, et al., *A panel study on patients with dominant cerebellar ataxia highlights the frequency of channelopathies.* Brain, 2017. **140**(6): p. 1579-1594.
330. Kurihara, M., H. Ishiura, T. Sasaki, J. Otsuka, T. Hayashi, Y. Terao, et al., *Novel De Novo KCND3 Mutation in a Japanese Patient with Intellectual Disability, Cerebellar Ataxia, Myoclonus, and Dystonia.* Cerebellum, 2018. **17**(2): p. 237-242.
331. Duarri, A., M.C. Lin, M.R. Fokkens, M. Meijer, C.J. Smeets, E.A. Nibbeling, et al., *Spinocerebellar ataxia type 19/22 mutations alter heterocomplex Kv4.3 channel function and gating in a dominant manner.* Cell Mol Life Sci, 2015. **72**(17): p. 3387-99.

332. Holmqvist, M.H., J. Cao, R. Hernandez-Pineda, M.D. Jacobson, K.I. Carroll, M.A. Sung, et al., *Elimination of fast inactivation in Kv4 A-type potassium channels by an auxiliary subunit domain.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 1035-40.
333. Liang, P., H. Wang, H. Chen, Y. Cui, L. Gu, J. Chai, et al., *Structural Insights into KChIP4a Modulation of Kv4.3 Inactivation.* J Biol Chem, 2009. **284**(8): p. 4960-7.
334. Norris, A.J., N.C. Foeger, and J.M. Nerbonne, *Interdependent roles for accessory KChIP2, KChIP3, and KChIP4 subunits in the generation of Kv4-encoded IA channels in cortical pyramidal neurons.* J Neurosci, 2010. **30**(41): p. 13644-55.
335. Zhou, J., Y. Tang, Q. Zheng, M. Li, T. Yuan, L. Chen, et al., *Different KChIPs compete for heteromultimeric assembly with pore-forming Kv4 subunits.* Biophys J, 2015. **108**(11): p. 2658-69.
336. Strassle, B.W., M. Menegola, K.J. Rhodes, and J.S. Trimmer, *Light and electron microscopic analysis of KChIP and Kv4 localization in rat cerebellar granule cells.* J Comp Neurol, 2005. **484**(2): p. 144-55.
337. Robinson, J.T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.
338. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
339. McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
340. McLaren, W., L. Gil, S.E. Hunt, H.S. Riat, G.R. Ritchie, A. Thormann, et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.
341. Tozaki, T., S. Mashima, K. Hirota, N. Miura, N.H. Choi-Miura, and M. Tomita, *Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods.* DNA Res, 2001. **8**(1): p. 33-45.
342. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.
343. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
344. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER.* Curr Protoc Bioinformatics, 2014. **47**: p. 5 6 1-32.
345. Shapovalov, M.V. and R.L. Dunbrack, Jr., *A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions.* Structure, 2011. **19**(6): p. 844-58.
346. Browning, S.R., *Missing data imputation and haplotype phase inference for genome-wide association studies.* Hum Genet, 2008. **124**(5): p. 439-50.
347. The International HapMap Consortium, K.A. Frazer, D.G. Ballinger, D.R. Cox, D.A. Hinds, L.L. Stuve, et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.
348. The 1000 Genomes Project Consortium, G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
349. McCarthy, S., S. Das, W. Kretzschmar, O. Delaneau, A.R. Wood, A. Teumer, et al., *A reference panel of 64,976 haplotypes for genotype imputation.* Nat Genet, 2016. **48**(10): p. 1279-83.

350. McCoy, A.M. and M.E. McCue, *Validation of imputation between equine genotyping arrays.* Anim Genet, 2014. **45**(1): p. 153.
351. Corbin, L.J., A. Kranis, S.C. Blott, J.E. Swinburne, M. Vaudin, S.C. Bishop, et al., *The utility of low-density genotyping for imputation in the Thoroughbred horse.* Genet Sel Evol, 2014. **46**: p. 9.
352. Schaefer, R.J., M. Schubert, E. Bailey, D.L. Bannasch, E. Barrey, G.K. Bar-Gal, et al., *Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds.* BMC Genomics, 2017. **18**(1): p. 565.
353. Chassier, M., E. Barrey, C. Robert, A. Duluard, S. Danvy, and A. Ricard, *Genotype imputation accuracy in multiple equine breeds from medium- to high-density genotypes.* J Anim Breed Genet, 2018. **135**(6): p. 420-431.
354. Pausch, H., B. Aigner, R. Emmerling, C. Edel, K.U. Gotz, and R. Fries, *Imputation of high-density genotypes in the Fleckvieh cattle population.* Genet Sel Evol, 2013. **45**: p. 3.
355. Hozé, C., M.N. Fouilloux, E. Venot, F. Guillaume, R. Dassonneville, S. Fritz, et al., *High-density marker imputation accuracy in sixteen French cattle breeds.* Genet Sel Evol, 2013. **45**: p. 33.
356. Korkuć, P., D. Arends, and G.A. Brockmann, *Finding the Optimal Imputation Strategy for Small Cattle Populations.* Front Genet, 2019. **10**: p. 52.
357. Gualdrón Duarte, J.L., R.O. Bates, C.W. Ernst, N.E. Raney, R.J. Cantet, and J.P. Steibel, *Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels.* BMC Genet, 2013. **14**: p. 38.
358. van den Berg, S., J. Vandenplas, F.A. van Eeuwijk, A.C. Bouwman, M.S. Lopes, and R.F. Veerkamp, *Imputation to whole-genome sequence using multiple pig populations and its use in genome-wide association studies.* Genet Sel Evol, 2019. **51**(1): p. 2.
359. Hayes, B.J., P.J. Bowman, H.D. Daetwyler, J.W. Kijas, and J.H. van der Werf, *Accuracy of genotype imputation in sheep breeds.* Anim Genet, 2012. **43**(1): p. 72-80.
360. Bolormaa, S., A.J. Chamberlain, M. Khansefid, P. Stothard, A.A. Swan, B. Mason, et al., *Accuracy of imputation to whole-genome sequence in sheep.* Genet Sel Evol, 2019. **51**(1): p. 1.
361. Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-8.
362. Sung, Y.J., K. Schwander, D.K. Arnett, S.L. Kardia, T. Rankinen, C. Bouchard, et al., *An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions.* Genet Epidemiol, 2014. **38**(4): p. 369-78.
363. Delaneau, O., J. Marchini, and J.F. Zagury, *A linear complexity phasing method for thousands of genomes.* Nat Methods, 2012. **9**(2): p. 179-81.
364. Browning, S.R. and B.L. Browning, *Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.* Am J Hum Genet, 2007. **81**(5): p. 1084-97.
365. Qanbari, S., *On the Extent of Linkage Disequilibrium in the Genome of Farm Animals.* Front Genet, 2020. **10**: p. 1304.
366. Arendt, M.L., M. Melin, N. Tonomura, M. Koltookian, C. Courtay-Cahen, N. Flindall, et al., *Genome-Wide Association Study of Golden Retrievers Identifies Germ-Line Risk Factors Predisposing to Mast Cell Tumours.* PLoS Genet, 2015. **11**(11): p. e1005647.

367. Biasoli, D., L. Compston-Garnett, S.L. Ricketts, Z. Birand, C. Courtay-Cahen, E. Fineberg, 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 Genet, 2019. **15**(3): p. e1007967.
368. Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee, *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.
369. Marchant, T.W., E. Dietschi, U. Rytz, P. Schawalder, V. Jagannathan, S. Hadji Rasouliha, et al., *An ADAMTS3 missense variant is associated with Norwich Terrier upper airway syndrome.* PLoS Genet, 2019. **15**(5): p. e1008102.
370. Jenkins, C.A., Dog Biomedical Variant Database Consortium, E.C. Schofield, C.S. Mellersh, L. De Risio, and S.L. Ricketts, *Improving the resolution of canine genome-wide association studies using genotype imputation: A study of two breeds.* Anim Genet, 2021. **52**(5): p. 703-713.
371. Katoh, M. and M. Katoh, *Identification and characterization of human FCHSD1 and FCHSD2 genes in silico.* Int J Mol Med, 2004. **13**(5): p. 749-54.
372. Abouelezz, A. and L. Almeida-Souza, *The mammalian endocytic cytoskeleton.* Eur J Cell Biol, 2022. **101**(2): p. 151222.
373. Almeida-Souza, L., R.A.W. Frank, J. Garcia-Nafria, A. Colussi, N. Gunawardana, C.M. Johnson, et al., *A Flat BAR Protein Promotes Actin Polymerization at the Base of Clathrin-Coated Pits.* Cell, 2018. **174**(2): p. 325-337 e14.
374. Xu, Y., J. Xia, S. Liu, S. Stein, C. Ramon, H. Xi, et al., *Endocytosis and membrane receptor internalization: implication of F-BAR protein Carom.* Front Biosci (Landmark Ed), 2017. **22**(9): p. 1439-1457.
375. Xiao, G.Y., A. Mohanakrishnan, and S.L. Schmid, *Role for ERK1/2-dependent activation of FCHSD2 in cancer cell-selective regulation of clathrin-mediated endocytosis.* Proc Natl Acad Sci U S A, 2018. **115**(41): p. E9570-E9579.
376. Xiao, G.Y. and S.L. Schmid, *FCHSD2 controls oncogenic ERK1/2 signaling outcome by regulating endocytic trafficking.* PLoS Biol, 2020. **18**(7): p. e3000778.
377. Zhai, X., Y. Shen, X. Zhang, T. Li, Q. Lu, and Z. Xu, *FCHSD2 cooperates with CDC42 and N-WASP to regulate cell protrusion formation.* Biochim Biophys Acta Mol Cell Res, 2022. **1869**(1): p. 119134.
378. Cao, H., X. Yin, Y. Cao, Y. Jin, S. Wang, Y. Kong, et al., *FCHSD1 and FCHSD2 are expressed in hair cell stereocilia and cuticular plate and regulate actin polymerization in vitro.* PLoS One, 2013. **8**(2): p. e56516.
379. Zhai, X., H. Du, Y. Shen, X. Zhang, Z. Chen, Y. Wang, et al., *FCHSD2 is required for stereocilia maintenance in mouse cochlear hair cells.* J Cell Sci, 2022.
380. Wang, H., D. Zhao, H. Du, X. Zhai, S. Wu, L. Lin, et al., *Deafness-related protein PDZD7 forms complex with the C-terminal tail of FCHSD2.* Biochem J, 2022. **479**(12): p. 1393-1405.
381. Coyle, I.P., Y.H. Koh, W.C. Lee, J. Slind, T. Fergestad, J.T. Littleton, et al., *Nervous wreck, an SH3 adaptor protein that interacts with Wsp, regulates synaptic growth in Drosophila.* Neuron, 2004. **41**(4): p. 521-34.
382. Del Signore, S.J., C.F. Kelley, E.M. Messelaar, T. Lemos, M.F. Marchan, B. Ermanoska, et al., *An autoinhibitory clamp of actin assembly constrains and directs synaptic endocytosis.* Elife, 2021. **10**.
383. Darling, N.J. and P. Cohen, *Nuts and bolts of the salt-inducible kinases (SIKs).* Biochem J, 2021. **478**(7): p. 1377-1397.

384. Sasagawa, S., H. Takemori, T. Uebi, D. Ikegami, K. Hiramatsu, S. Ikegawa, et al., *SIK3 is essential for chondrocyte hypertrophy during skeletal development in mice.* Development, 2012. **139**(6): p. 1153-63.
385. Csukasi, F., I. Duran, M. Barad, T. Barta, I. Gudernova, L. Trantirek, et al., *The PTH/PTHrP-SIK3 pathway affects skeletogenesis through altered mTOR signaling.* Sci Transl Med, 2018. **10**(459).
386. Funato, H., C. Miyoshi, T. Fujiyama, T. Kanda, M. Sato, Z. Wang, et al., *Forward-genetics analysis of sleep in randomly mutagenized mice.* Nature, 2016. **539**(7629): p. 378-383.
387. Honda, T., T. Fujiyama, C. Miyoshi, A. Ikkyu, N. Hotta-Hirashima, S. Kanno, et al., *A single phosphorylation site of SIK3 regulates daily sleep amounts and sleep need in mice.* Proc Natl Acad Sci U S A, 2018. **115**(41): p. 10458-10463.
388. Wang, Z., J. Ma, C. Miyoshi, Y. Li, M. Sato, Y. Ogawa, et al., *Quantitative phosphoproteomic analysis of the molecular substrates of sleep need.* Nature, 2018. **558**(7710): p. 435-439.
389. Li, H., A. Russo, and A. DiAntonio, *SIK3 suppresses neuronal hyperexcitability by regulating the glial capacity to buffer K(+) and water.* J Cell Biol, 2019. **218**(12): p. 4017-4029.
390. Hong, G.S., B. Lee, J. Wee, H. Chun, H. Kim, J. Jung, et al., *Tentonin 3/TMEM150c Confers Distinct Mechanosensitive Currents in Dorsal-Root Ganglion Neurons with Proprioceptive Function.* Neuron, 2016. **91**(1): p. 107-18.
391. Hong, G.S., B. Lee, and U. Oh, *Evidence for Mechanosensitive Channel Activity of Tentonin 3/TMEM150C.* Neuron, 2017. **94**(2): p. 271-273 e2.
392. Dubin, A.E., S. Murthy, A.H. Lewis, L. Brosse, S.M. Cahalan, J. Grandl, et al., *Endogenous Piezo1 Can Confound Mechanically Activated Channel Identification and Characterization.* Neuron, 2017. **94**(2): p. 266-270 e3.
393. Anderson, E.O., E.R. Schneider, J.D. Matson, E.O. Gracheva, and S.N. Bagriantsev, *TMEM150C/Tentonin3 Is a Regulator of Mechano-gated Ion Channels.* Cell Rep, 2018. **23**(3): p. 701-708.
394. Lu, H.J., T.L. Nguyen, G.S. Hong, S. Pak, H. Kim, H. Kim, et al., *Tentonin 3/TMEM150C senses blood pressure changes in the aortic arch.* J Clin Invest, 2020. **130**(7): p. 3671-3683.
395. Parpaite, T., L. Brosse, N. Sejourne, A. Laur, Y. Mechioukhi, P. Delmas, et al., *Patch-seq of mouse DRG neurons reveals candidate genes for specific mechanosensory functions.* Cell Rep, 2021. **37**(5): p. 109914.
396. Katoh, H. and M. Negishi, *RhoG activates Rac1 by direct interaction with the Dock180-binding protein Elmo.* Nature, 2003. **424**(6947): p. 461-4.
397. Grimsley, C.M., J.M. Kinchen, A.C. Tosello-Trampont, E. Brugnera, L.B. Haney, M. Lu, et al., *Dock180 and ELMO1 proteins cooperate to promote evolutionarily conserved Rac-dependent cell migration.* J Biol Chem, 2004. **279**(7): p. 6087-97.
398. Etienne-Manneville, S. and A. Hall, *Rho GTPases in cell biology.* Nature, 2002. **420**(6916): p. 629-35.
399. Sato, Y., A. Sato, S. Mizuno, J.N. Hirota, S. Fujima, C. Ishii, et al., *Comparative gene expression analysis of the engulfment and cell motility (ELMO) protein family in the mouse brain.* Gene Expr Patterns, 2019. **34**: p. 119070.
400. Stevenson, C., G. de la Rosa, C.S. Anderson, P.S. Murphy, T. Capece, M. Kim, et al., *Essential role of Elmo1 in Dock2-dependent lymphocyte migration.* J Immunol, 2014. **192**(12): p. 6062-70.

401. Lu, Z., M.R. Elliott, Y. Chen, J.T. Walsh, A.L. Klibanov, K.S. Ravichandran, et al., *Phagocytic activity of neuronal progenitors regulates adult neurogenesis.* Nat Cell Biol, 2011. **13**(9): p. 1076-83.
402. Li, W., K.M.V. Tam, W.W.R. Chan, A.C. Koon, J.C.K. Ngo, H.Y.E. Chan, et al., *Neuronal adaptor FE65 stimulates Rac1-mediated neurite outgrowth by recruiting and activating ELMO1.* J Biol Chem, 2018. **293**(20): p. 7674-7688.
403. Chan, W.W.R., W. Li, R.C.C. Chang, and K.F. Lau, *ARF6-Rac1 signaling-mediated neurite outgrowth is potentiated by the neuronal adaptor FE65 through orchestrating ARF6 and ELMO1.* FASEB J, 2020. **34**(12): p. 16397-16413.
404. Kaler, P. and R. Prasad, *Molecular cloning and functional characterization of novel zinc transporter rZip10 (Slc39a10) involved in zinc uptake across rat renal brush-border membrane.* Am J Physiol Renal Physiol, 2007. **292**(1): p. F217-29.
405. Pawan, K., S. Neeraj, K. Sandeep, R. Kanta Ratho, and P. Rajendra, *Upregulation of Slc39a10 gene expression in response to thyroid hormones in intestine and kidney.* Biochim Biophys Acta, 2007. **1769**(2): p. 117-23.
406. Landry, G.M., E. Furrow, H.L. Holmes, T. Hirata, A. Kato, P. Williams, et al., *Cloning, function, and localization of human, canine, and Drosophila ZIP10 (SLC39A10), a Zn(2+) transporter.* Am J Physiol Renal Physiol, 2019. **316**(2): p. F263-F273.
407. Miyai, T., S. Hojyo, T. Ikawa, M. Kawamura, T. Irie, H. Ogura, et al., *Zinc transporter SLC39A10/ZIP10 facilitates antiapoptotic signaling during early B-cell development.* Proc Natl Acad Sci U S A, 2014. **111**(32): p. 11780-5.
408. Gao, H., L. Zhao, H. Wang, E. Xie, X. Wang, Q. Wu, et al., *Metal transporter Slc39a10 regulates susceptibility to inflammatory stimuli by controlling macrophage survival.* Proc Natl Acad Sci U S A, 2017. **114**(49): p. 12940-12945.
409. Hojyo, S., T. Miyai, H. Fujishiro, M. Kawamura, T. Yasuda, A. Hijikata, et al., *Zinc transporter SLC39A10/ZIP10 controls humoral immunity by modulating B-cell receptor signal strength.* Proc Natl Acad Sci U S A, 2014. **111**(32): p. 11786-91.
410. Berendt, M., H. Gredal, L.G. Pedersen, L. Alban, and J. Alving, *A cross-sectional study of epilepsy in Danish Labrador Retrievers: prevalence and selected risk factors.* J Vet Intern Med, 2002. **16**(3): p. 262-8.
411. Sauer-Delhees, S., F. Steffen, I. Reichler, and K. Beckmann, *[Clinical characteristics of Idiopathic Epilepsy in Greater Swiss Mountain Dogs in Switzerland].* Schweiz Arch Tierheilkd, 2020. **162**(11): p. 697-706.
412. Torkamani, A., N.E. Wineinger, and E.J. Topol, *The personal and clinical utility of polygenic risk scores.* Nat Rev Genet, 2018. **19**(9): p. 581-590.
413. Lewis, C.M. and E. Vassos, *Polygenic risk scores: from research tools to clinical instruments.* Genome Med, 2020. **12**(1): p. 44.
414. Ripatti, S., E. Tikkanen, M. Orho-Melander, A.S. Havulinna, K. Silander, A. Sharma, et al., *A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses.* Lancet, 2010. **376**(9750): p. 1393-400.
415. Mega, J.L., N.O. Stitziel, J.G. Smith, D.I. Chasman, M. Caulfield, J.J. Devlin, et al., *Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials.* Lancet, 2015. **385**(9984): p. 2264-2271.
416. Tada, H., O. Melander, J.Z. Louie, J.J. Catanese, C.M. Rowland, J.J. Devlin, et al., *Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history.* Eur Heart J, 2016. **37**(6): p. 561-7.
417. Hsu, L., J. Jeon, H. Brenner, S.B. Gruber, R.E. Schoen, S.I. Berndt, et al., *A model to determine colorectal cancer risk using common genetic susceptibility loci.* Gastroenterology, 2015. **148**(7): p. 1330-9 e14.

418. Seibert, T.M., C.C. Fan, Y. Wang, V. Zuber, R. Karunamuni, J.K. Parsons, et al., *Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts.* BMJ, 2018. **360**: p. j5757.
419. Maas, P., M. Barrdahl, A.D. Joshi, P.L. Auer, M.M. Gaudet, R.L. Milne, et al., *Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States.* JAMA Oncol, 2016. **2**(10): p. 1295-1302.
420. The Kennel Club. *Comparative tables of registrations for the years 2012 - 2021 inclusive - Gundog*. Breed registration statistics 2023 [cited 2023 06/01/2023]; Available from: https://www.thekennelclub.org.uk/media-centre/breed-registration-statistics/.
421. Lohmueller, K.E., C.L. Pearce, M. Pike, E.S. Lander, and J.N. Hirschhorn, *Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease.* Nat Genet, 2003. **33**(2): p. 177-82.
422. Garland, P., S. Quraishe, P. French, and V. O'Connor, *Expression of the MAST family of serine/threonine kinases.* Brain Res, 2008. **1195**: p. 12-9.
423. Landoulsi, Z., F. Laatar, E. Noe, S. Mrabet, M. Ben Djebara, G. Achaz, et al., *Clinical and genetic study of Tunisian families with genetic generalized epilepsy: contribution of CACNA1H and MAST4 genes.* Neurogenetics, 2018. **19**(3): p. 165-178.
424. Consortium, E., E.M. Consortium, M. Steffens, C. Leu, A.K. Ruppert, F. Zara, et al., *Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32.* Hum Mol Genet, 2012. **21**(24): p. 5359-72.
425. Hibar, D.P., H.H.H. Adams, N. Jahanshad, G. Chauhan, J.L. Stein, E. Hofer, et al., *Novel genetic loci associated with hippocampal volume.* Nat Commun, 2017. **8**: p. 13624.
426. Packer, R.M., N.K. Shihab, B.B. Torres, and H.A. Volk, *Clinical risk factors associated with anti-epileptic drug responsiveness in canine epilepsy.* PLoS One, 2014. **9**(8): p. e106026.
427. Clements, D.N., S. Carter, J. Innes, L. Kennedy, D. Bennett, and M. Day, *DNA archive for companion animals.* Vet Rec, 2007. **160**(22): p. 776.
428. He, C., J. Holme, and J. Anthony, *SNP genotyping: the KASP assay.* Methods Mol Biol, 2014. **1145**: p. 75-86.
429. Dang, D.T., J. Pevsner, and V.W. Yang, *The biology of the mammalian Kruppel-like family of transcription factors.* Int J Biochem Cell Biol, 2000. **32**(11-12): p. 1103-21.
430. Ghaleb, A.M. and V.W. Yang, *Kruppel-like factor 4 (KLF4): What we currently know.* Gene, 2017. **611**: p. 27-37.
431. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
432. Wernig, M., A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, et al., *In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state.* Nature, 2007. **448**(7151): p. 318-24.
433. Park, I.H., R. Zhao, J.A. West, A. Yabuuchi, H. Huo, T.A. Ince, et al., *Reprogramming of human somatic cells to pluripotency with defined factors.* Nature, 2008. **451**(7175): p. 141-6.
434. Moore, D.L., M.G. Blackmore, Y. Hu, K.H. Kaestner, J.L. Bixby, V.P. Lemmon, et al., *KLF family members regulate intrinsic axon regeneration ability.* Science, 2009. **326**(5950): p. 298-301.

435. Qin, S. and C.L. Zhang, *Role of Kruppel-like factor 4 in neurogenesis and radial neuronal migration in the developing cerebral cortex.* Mol Cell Biol, 2012. **32**(21): p. 4297-305.
436. Moon, B.S., J. Bai, M. Cai, C. Liu, J. Shi, and W. Lu, *Kruppel-like factor 4-dependent Staufen1-mediated mRNA decay regulates cortical neurogenesis.* Nat Commun, 2018. **9**(1): p. 401.
437. Sahin, G.S., M. Dhar, C. Dillon, M. Zhu, H. Shiina, B.D. Winters, et al., *Leptin stimulates synaptogenesis in hippocampal neurons via KLF4 and SOCS3 inhibition of STAT3 signaling.* Mol Cell Neurosci, 2020. **106**: p. 103500.
438. Kaushik, D.K., M. Gupta, S. Das, and A. Basu, *Kruppel-like factor 4, a novel transcription factor regulates microglial activation and subsequent neuroinflammation.* J Neuroinflammation, 2010. **7**: p. 68.
439. Li, L., X. Zi, D. Hou, and Q. Tu, *Kruppel-like factor 4 regulates amyloid-beta (Abeta)-induced neuroinflammation in Alzheimer's disease.* Neurosci Lett, 2017. **643**: p. 131-137.
440. Cheng, Z., W. Yang, B. Li, and R. Cui, *KLF4 Exerts Sedative Effects in Pentobarbital-Treated Mice.* J Mol Neurosci, 2021. **71**(3): p. 596-606.
441. Chen, Y., J. Chen, Y. Chen, and Y. Li, *miR-146a/KLF4 axis in epileptic mice: A novel regulator of synaptic plasticity involving STAT3 signaling.* Brain Res, 2022. **1790**: p. 147988.
442. Shi, Y., *Mechanistic insights into precursor messenger RNA splicing by the spliceosome.* Nat Rev Mol Cell Biol, 2017. **18**(11): p. 655-670.
443. Didychuk, A.L., S.E. Butcher, and D.A. Brow, *The life of U6 small nuclear RNA, from cradle to grave.* RNA, 2018. **24**(4): p. 437-460.
444. Doucet, A.J., G. Droc, O. Siol, J. Audoux, and N. Gilbert, *U6 snRNA Pseudogenes: Markers of Retrotransposition Dynamics in Mammals.* Mol Biol Evol, 2015. **32**(7): p. 1815-32.
445. Kraus, D.M., G.S. Elliott, H. Chute, T. Horan, K.H. Pfenninger, S.D. Sanford, et al., *CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues.* J Immunol, 2006. **176**(7): p. 4419-30.
446. Escudero-Esparza, A., N. Kalchishkova, E. Kurbasic, W.G. Jiang, and A.M. Blom, *The novel complement inhibitor human CUB and Sushi multiple domains 1 (CSMD1) protein promotes factor I-mediated degradation of C4b and C3b and inhibits the membrane attack complex assembly.* FASEB J, 2013. **27**(12): p. 5083-93.
447. Soltani Khaboushan, A., N. Yazdanpanah, and N. Rezaei, *Neuroinflammation and Proinflammatory Cytokines in Epileptogenesis.* Mol Neurobiol, 2022. **59**(3): p. 1724-1743.
448. Kopczynska, M., W.M. Zelek, S. Vespa, S. Touchard, M. Wardle, S. Loveless, et al., *Complement system biomarkers in epilepsy.* Seizure, 2018. **60**: p. 1-7.
449. Woo, J.J., J.G. Pouget, C.C. Zai, and J.L. Kennedy, *The complement system in schizophrenia: where are we now and what's next?* Mol Psychiatry, 2020. **25**(1): p. 114-130.
450. Xu, W., S. Cohen-Woods, Q. Chen, A. Noor, J. Knight, G. Hosang, et al., *Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1.* BMC Med Genet, 2014. **15**: p. 2.
451. Cukier, H.N., N.D. Dueker, S.H. Slifer, J.M. Lee, P.L. Whitehead, E. Lalanne, et al., *Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders.* Mol Autism, 2014. **5**(1): p. 1.

452. Ruiz-Martinez, J., L.J. Azcona, A. Bergareche, J.F. Marti-Masso, and C. Paisan-Ruiz, *Whole-exome sequencing associates novel CSMD1 gene mutations with familial Parkinson disease.* Neurol Genet, 2017. **3**(5): p. e177.
453. Protasova, M.S., F.E. Gusev, T.V. Andreeva, S.A. Klyushnikov, S.N. Illarioshkin, and E.I. Rogaev, *Novel genes bearing mutations in rare cases of early-onset ataxia with cerebellar hypoplasia.* Eur J Hum Genet, 2022. **30**(6): p. 703-711.
454. Costanzo, F., G. Zanni, E. Fuca, M. Di Paola, S. Barresi, L. Travaglini, et al., *Cerebellar Agenesis and Bilateral Polimicrogyria Associated with Rare Variants of CUB and Sushi Multiple Domains 1 Gene (CSMD1): A Longitudinal Neuropsychological and Neuroradiological Case Study.* Int J Environ Res Public Health, 2022. **19**(3).
455. Crippa, M., P. Malatesta, M.T. Bonati, F. Trapasso, F. Fortunato, G. Annesi, et al., *A familial t(4;8) translocation segregates with epilepsy and migraine with aura.* Ann Clin Transl Neurol, 2020. **7**(5): p. 855-859.
456. Naseer, M.I., A.G. Chaudhary, M. Rasool, G. Kalamegam, F.T. Ashgan, M. Assidi, et al., *Copy number variations in Saudi family with intellectual disability and epilepsy.* BMC Genomics, 2016. **17**(Suppl 9): p. 757.
457. Hong, D., X. Zhang, R. Li, J. Yu, Y. Lou, Q. He, et al., *Deletion of TMEM268 inhibits growth of gastric cancer cells by downregulating the ITGB4 signaling pathway.* Cell Death Differ, 2019. **26**(8): p. 1453-1466.
458. Qaiser, F., T. Sadoway, Y. Yin, Q. Zulfiqar Ali, C.M. Nguyen, N. Shum, et al., *Genome sequencing identifies rare tandem repeat expansions and copy number variants in Lennox-Gastaut syndrome.* Brain Commun, 2021. **3**(3): p. fcab207.
459. Doyle, G.A., B.C. Reiner, R.C. Crist, A.M. Rao, N.S. Ojeah, G. Arauco-Shapiro, et al., *Investigation of long interspersed element-1 retrotransposons as potential risk factors for idiopathic temporal lobe epilepsy.* Epilepsia, 2021. **62**(6): p. 1329-1342.
460. Leitner, D.F., S. Devore, J. Laze, D. Friedman, J.D. Mills, Y. Liu, et al., *Serotonin receptor expression in hippocampus and temporal cortex of temporal lobe epilepsy patients by postictal generalized electroencephalographic suppression duration.* Epilepsia, 2022. **63**(11): p. 2925-2936.
461. Jenkins, C.A., L. Kalmar, K. Matiasek, L. Mari, K. Kyostila, H. Lohi, et al., *Characterisation of canine KCNIP4: A novel gene for cerebellar ataxia identified by whole-genome sequencing two affected Norwegian Buhund dogs.* PLoS Genet, 2020. **16**(1): p. e1008527.
462. Cercos, P., D.A. Peraza, A. Benito-Bueno, P.G. Socuellamos, A. Aziz-Nignan, D. Arrechaga-Estevez, et al., *Pharmacological Approaches for the Modulation of the Potassium Channel K(V)4.x and KChIPs.* Int J Mol Sci, 2021. **22**(3).
463. Pacheco-Rojas, D.O., M. Delgado-Ramirez, K. Villatoro-Gomez, E.G. Moreno-Galindo, A.A. Rodriguez-Menchaca, J.A. Sanchez-Chapula, et al., *Riluzole inhibits Kv4.2 channels acting on the closed and closed inactivated states.* Eur J Pharmacol, 2021. **899**: p. 174026.
464. Manz, E., B. Tellhelm, and M. Krawczak, *Prospective evaluation of a patented DNA test for canine hip dysplasia (CHD).* PLoS One, 2017. **12**(8): p. e0182093.
465. Chatterjee, N., J. Shi, and M. Garcia-Closas, *Developing and evaluating polygenic risk prediction models for stratified disease prevention.* Nat Rev Genet, 2016. **17**(7): p. 392-406.
466. Leu, C., R. Stevelink, A.W. Smith, S.B. Goleva, M. Kanai, L. Ferguson, et al., *Polygenic burden in focal and generalized epilepsies.* Brain, 2019. **142**(11): p. 3473-3481.

467. Gramm, M., C. Leu, E. Perez-Palma, L. Ferguson, L. Jehi, M.J. Daly, et al., *Polygenic risk heterogeneity among focal epilepsies.* Epilepsia, 2020. **61**(11): p. e179-e185.
468. Moreau, C., R.M. Rebillard, S. Wolking, J. Michaud, F. Tremblay, A. Girard, et al., *Polygenic risk scores of several subtypes of epilepsies in a founder population.* Neurol Genet, 2020. **6**(3): p. e416.
469. Oliver, K.L., C.A. Ellis, I.E. Scheffer, S. Ganesan, C. Leu, L.G. Sadleir, et al., *Common risk variants for epilepsy are enriched in families previously targeted for rare monogenic variant discovery.* EBioMedicine, 2022. **81**: p. 104079.
470. Martin, A.R., M.J. Daly, E.B. Robinson, S.E. Hyman, and B.M. Neale, *Predicting Polygenic Risk of Psychiatric Disorders.* Biol Psychiatry, 2019. **86**(2): p. 97-109.
471. Brainstorm Consortium, V. Anttila, B. Bulik-Sullivan, H.K. Finucane, R.K. Walters, J. Bras, et al., *Analysis of shared heritability in common disorders of the brain.* Science, 2018. **360**(6395).
472. Mahajan, A., D. Taliun, M. Thurner, N.R. Robertson, J.M. Torres, N.W. Rayner, et al., *Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps.* Nat Genet, 2018. **50**(11): p. 1505-1513.
473. Gettler, K., R. Levantovsky, A. Moscati, M. Giri, Y. Wu, N.Y. Hsu, et al., *Common and Rare Variant Prediction and Penetrance of IBD in a Large, Multi-ethnic, Health System-based Biobank Cohort.* Gastroenterology, 2021. **160**(5): p. 1546-1557.
474. Marquez-Luna, C., P.R. Loh, C. South Asian Type 2 Diabetes, S.T.D. Consortium, and A.L. Price, *Multiethnic polygenic risk scores improve risk prediction in diverse populations.* Genet Epidemiol, 2017. **41**(8): p. 811-823.
475. Whiteley, M.H., J.S. Bell, and D.A. Rothman, *Novel allelic variants in the canine cyclooxgenase-2 (Cox-2) promoter are associated with renal dysplasia in dogs.* PLoS One, 2011. **6**(2): p. e16684.
476. Safra, N., L.J. Hayward, M. Aguilar, B.N. Sacks, J.L. Westropp, F.C. Mohr, et al., *DNA Sequence Variants in the Five Prime Untranslated Region of the Cyclooxygenase-2 Gene Are Commonly Found in Healthy Dogs and Gray Wolves.* PLoS One, 2015. **10**(8): p. e0133127.
477. McNamara, J.W., M. Schuckman, R.C. Becker, and S. Sadayappan, *A Novel Homozygous Intronic Variant in TNNT2 Associates With Feline Cardiomyopathy.* Front Physiol, 2020. **11**: p. 608473.
478. Schipper, T., A. Ohlsson, M. Longeri, J.J. Hayward, L. Mouttham, P. Ferrari, et al., *The TNNT2:c.95-108G>A variant is common in Maine Coons and shows no association with hypertrophic cardiomyopathy.* Anim Genet, 2022. **53**(4): p. 526-529.
479. Panagiotou, O.A., C.J. Willer, J.N. Hirschhorn, and J.P. Ioannidis, *The power of meta-analysis in genome-wide association studies.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 441-65.
480. Gershony, L.C., J.M. Belanger, M.K. Hytonen, H. Lohi, and A.M. Oberbauer, *Whole Genome Sequencing Reveals Multiple Linked Genetic Variants on Canine Chromosome 12 Associated with Risk for Symmetrical Lupoid Onychodystrophy (SLO) in the Bearded Collie.* Genes (Basel), 2021. **12**(8).
481. Letko, A., K.M. Minor, E.M. Norton, V.D. Marinescu, M. Drogemuller, E. Ivansson, et al., *Genome-Wide Analyses for Osteosarcoma in Leonberger Dogs Reveal the CDKN2A/B Gene Locus as a Major Risk Locus.* Genes (Basel), 2021. **12**(12).
482. Jiang, L., Z. Li, J.J. Hayward, K. Hayashi, U. Krotscheck, R.J. Todhunter, et al., *Genomic Prediction of Two Complex Orthopedic Traits Across Multiple Pure and Mixed Breed Dogs.* Front Genet, 2021. **12**: p. 666740.

483. Srikanth, K., D.J.F. von Pfeil, B.J. Stanley, C. Griffitts, and H.J. Huson, *Genome Wide Association Study with Imputed Whole Genome Sequence Data Identifies a 431 kb Risk Haplotype on CFA18 for Congenital Laryngeal Paralysis in Alaskan Sled Dogs.* Genes (Basel), 2022. **13**(10).
484. DoGA consortium. *Dog Genome Annotation Project*. 2023 [cited 2023 22/01/23]; Available from: www.doggenomeannotation.org.
485. Kaukonen, M., I.B. Quintero, A.K. Mukarram, M.K. Hytonen, S. Holopainen, K. Wickstrom, et al., *A putative silencer variant in a spontaneous canine model of retinitis pigmentosa.* PLoS Genet, 2020. **16**(3): p. e1008659.
486. Koboldt, D.C., *Best practices for variant calling in clinical sequencing.* Genome Med, 2020. **12**(1): p. 91.
487. Ho, S.S., A.E. Urban, and R.E. Mills, *Structural variation in the sequencing era.* Nat Rev Genet, 2020. **21**(3): p. 171-189.
488. Ye, K., M.H. Schulz, Q. Long, R. Apweiler, and Z. Ning, *Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.* Bioinformatics, 2009. **25**(21): p. 2865-71.
489. Wong, K., T.M. Keane, J. Stalker, and D.J. Adams, *Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly.* Genome Biol, 2010. **11**(12): p. R128.
490. Rausch, T., T. Zichner, A. Schlattl, A.M. Stutz, V. Benes, and J.O. Korbel, *DELLY: structural variant discovery by integrated paired-end and split-read analysis.* Bioinformatics, 2012. **28**(18): p. i333-i339.
491. Layer, R.M., C. Chiang, A.R. Quinlan, and I.M. Hall, *LUMPY: a probabilistic framework for structural variant discovery.* Genome Biol, 2014. **15**(6): p. R84.
492. Chen, X., O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Kallberg, et al., *Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications.* Bioinformatics, 2016. **32**(8): p. 1220-2.
493. Alkan, C., B.P. Coe, and E.E. Eichler, *Genome structural variation discovery and genotyping.* Nat Rev Genet, 2011. **12**(5): p. 363-76.
494. Murgiano, L., D. Becker, D. Torjman, J.K. Niggel, A. Milano, C. Cullen, et al., *Complex Structural PPT1 Variant Associated with Non-syndromic Canine Retinal Degeneration.* G3 (Bethesda), 2019. **9**(2): p. 425-437.
495. Sanchez, L., E. Beltran, A. de Stefani, L.T. Guo, A. Shea, G.D. Shelton, et al., *Clinical and genetic characterisation of dystrophin-deficient muscular dystrophy in a family of Miniature Poodle dogs.* PLoS One, 2018. **13**(2): p. e0193372.
496. Salmela, E., J. Niskanen, M. Arumilli, J. Donner, H. Lohi, and M.K. Hytonen, *A novel KRT71 variant in curly-coated dogs.* Anim Genet, 2019. **50**(1): p. 101-104.
497. Gentilini, F., M.E. Turba, F. Giancola, R. Chiocchetti, C. Bernardini, M. Dajbychova, et al., *A large deletion in the GP9 gene in Cocker Spaniel dogs with Bernard-Soulier syndrome.* PLoS One, 2019. **14**(9): p. e0220625.
498. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.
499. Karyadi, D.M., E. Karlins, B. Decker, B.M. vonHoldt, G. Carpintero-Ramirez, H.G. Parker, et al., *A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit.* PLoS Genet, 2013. **9**(3): p. e1003409.
500. Serres-Armero, A., B.W. Davis, I.S. Povolotskaya, C. Morcillo-Suarez, J. Plassais, D. Juan, et al., *Copy number variation underlies complex phenotypes in domestic dog breeds and other canids.* Genome Res, 2021. **31**(5): p. 762-774.

501. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Curr Protoc Hum Genet, 2013. **Chapter 7**: p. Unit7 20.
502. Gauderman, W.J., *Sample size requirements for association studies of gene-gene interaction.* Am J Epidemiol, 2002. **155**(5): p. 478-84.
503. Johnson, J.L. and G.R. Abecasis, *GAS Power Calculator: web-based power calculator for genetic association studies.* bioRxiv, 2017: p. 164343.
504. Skol, A.D., L.J. Scott, G.R. Abecasis, and M. Boehnke, *Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies.* Nat Genet, 2006. **38**(2): p. 209-13.
505. Kise, Y., G. Kasuya, H.H. Okamoto, D. Yamanouchi, K. Kobayashi, T. Kusakizako, et al., *Structural basis of gating modulation of Kv4 channel complexes.* Nature, 2021. **599**(7883): p. 158-164.
506. Meuwissen, T.H.E. and Z. Luo, *Computing inbreeding coefficients in large populations.* Genetics Selection Evolution, 1992. **24**(4): p. 305.

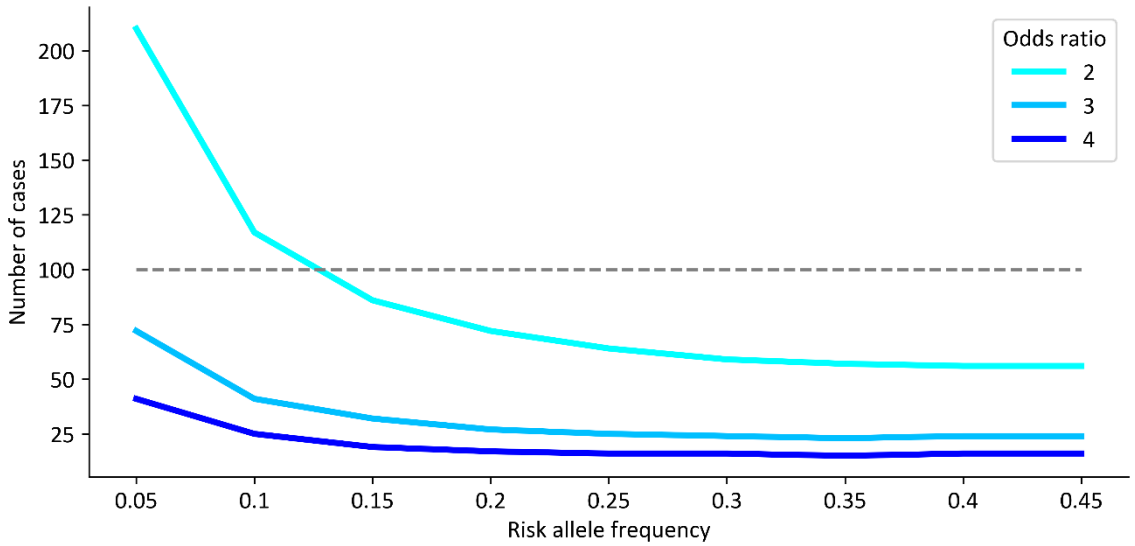
10 Appendices

i. Power calculations

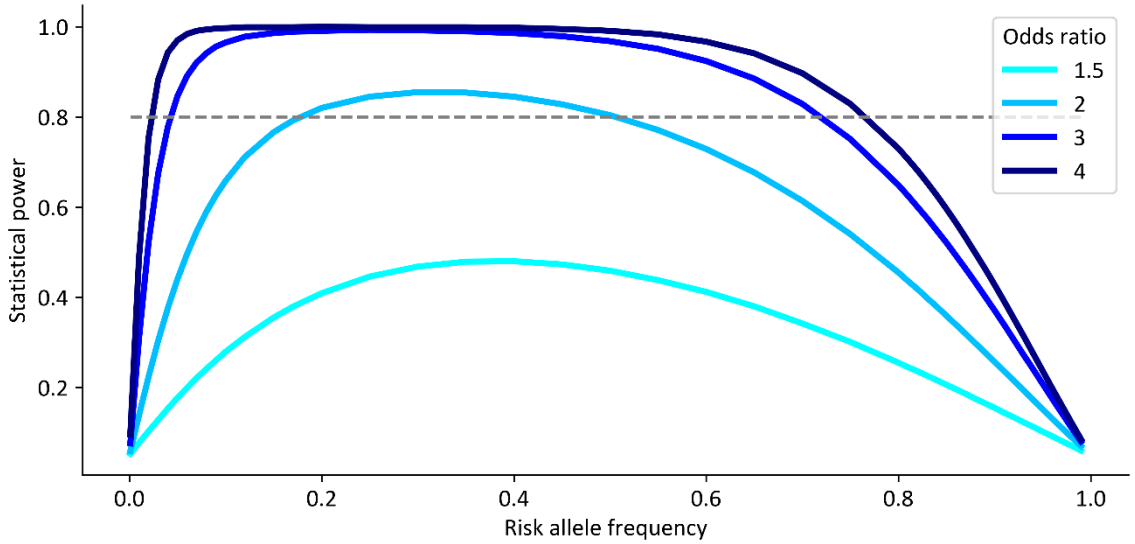
Appendix i.i. Border Collie IE study power calculations

A power calculation was performed for the Border Collie IE study using Quanto [502] to determine the number of cases required, with controls outnumbering cases at a ratio of 1.5:1, to detect associations (desired study statistical power: 0.8) for a range of risk allele frequencies and moderate to high effect sizes. Calculations were based on a significance threshold of 0.05, a log-additive mode of inheritance, and a disease prevalence of 0.85% based on published data [203]. The power calculations suggested that a study including 100 cases and 150 controls would have sufficient power to detect associations of variants with risk allele frequency 0.15 – 0.45 and odds ratio ≥ 2(**Appendix i.ii**).

A second power calculation was performed using the online genetic association study (GAS) power calculator [503], which is based on the power calculator for two stage association studies (CaTS) [504]. This power calculation was used to examine the power of a study of 100 cases and 150 controls to detect associations with a range of risk allele frequencies and effect sizes. Calculations were based on an additive disease model, a significance threshold of 0.05, and a disease prevalence of 0.85% [203]. The power calculation confirmed that a study of 100 cases and 150 controls should have sufficient power to detect associations with odds ratios > 3 and low to high risk allele frequencies (**Appendix i.iii**). The results suggested that such a study would also be expected to detect variants with a moderate effect (odds ratio ~2) and risk allele frequencies between 0.2 and 0.5.

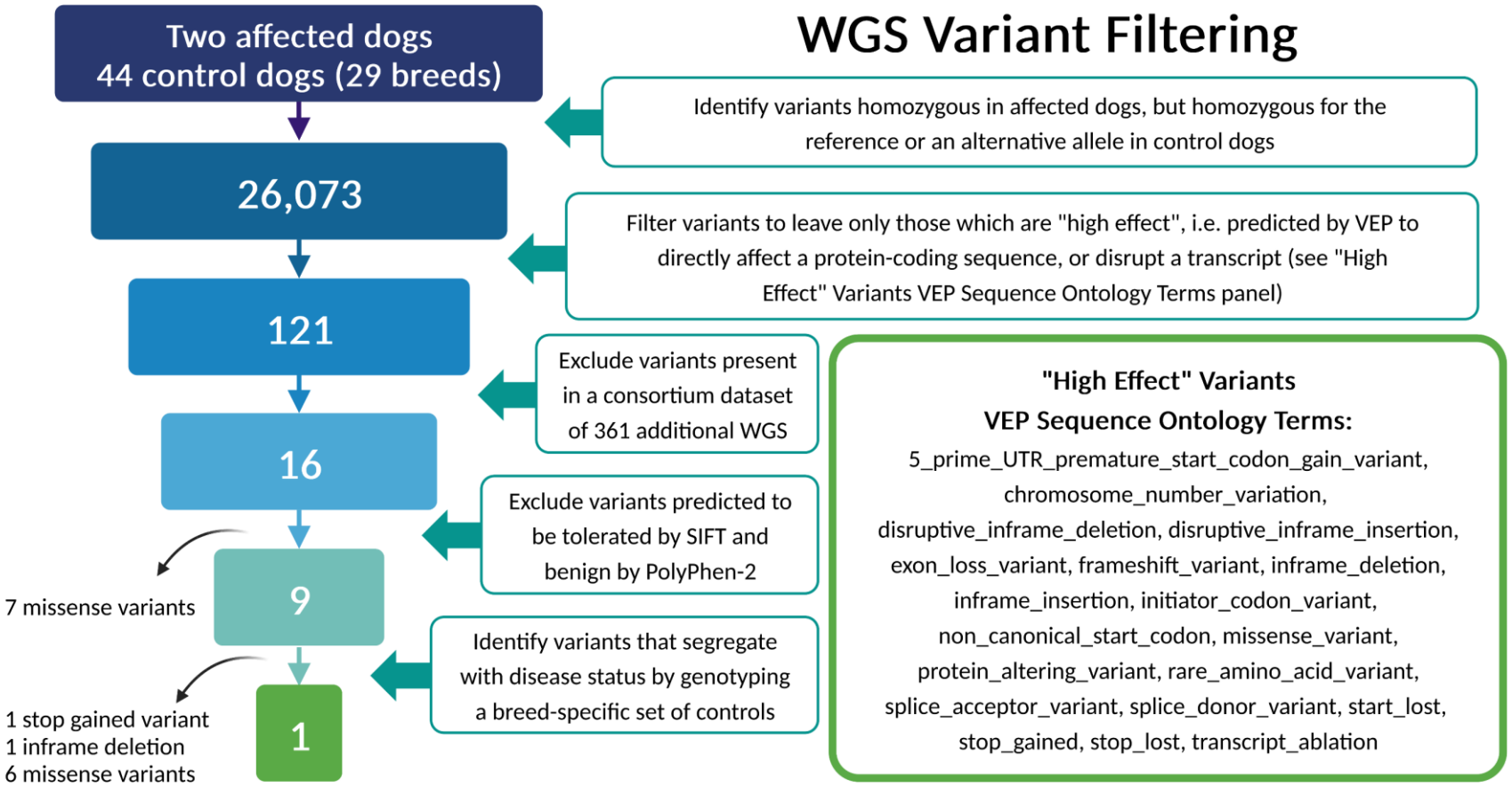


Appendix i.ii. Border Collie idiopathic epilepsy power calculation indicating case numbers required to detect variants with moderate to high effect on risk and a range of risk allele frequencies. Based on a control to case ratio of 1.5:1. Dashed line indicates a study of 100 cases and 150 controls.

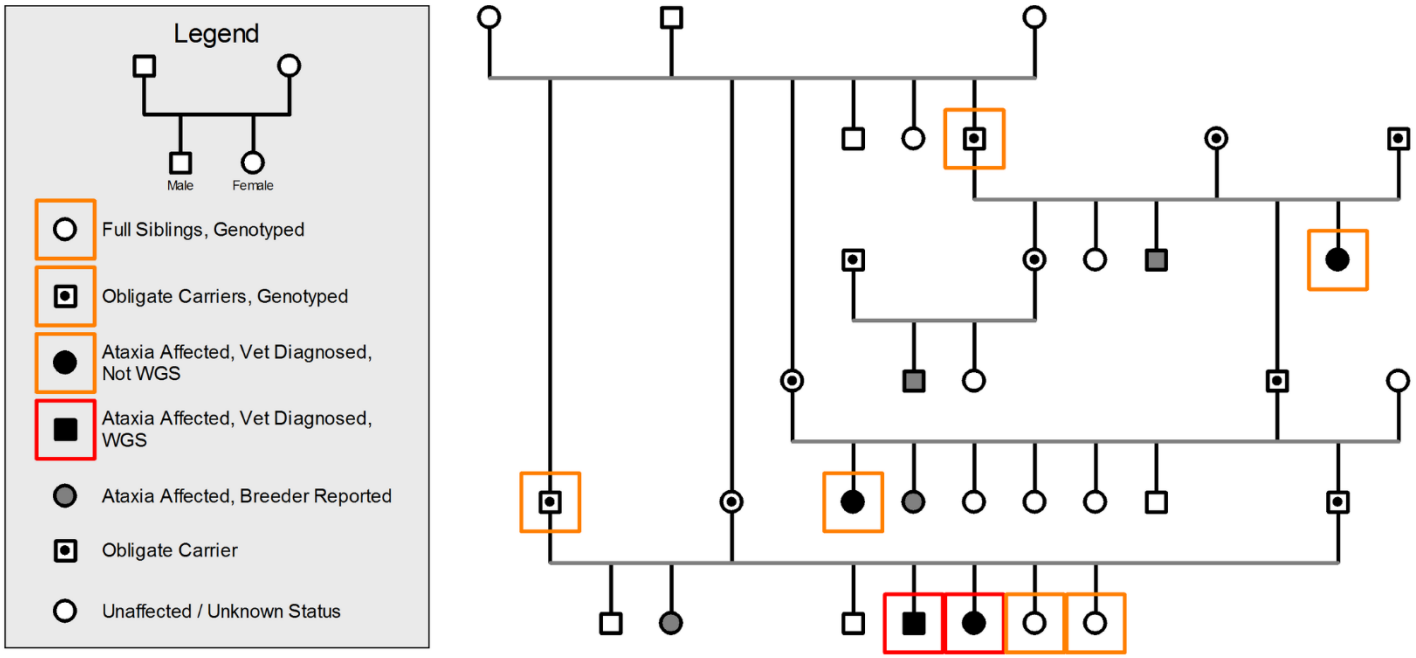


Appendix i.iii. Border Collie idiopathic epilepsy power calculation indicating the statistical power of a study of 100 cases and 150 controls to detect associations with a moderate to high effect on risk and a range of risk allele frequencies. Dashed line indicates the desired statistical power (0.8).

ii. Manuscript 1 appendices



Appendix ii.i. Figure demonstrating the process used for filtering whole genome sequence (WGS) variants. The number of variants remaining after each step are shown in each box. The number and predicted effect of variants excluded using *in silico* tools and by genotyping controls are given. The Ensembl Variant Effect Predictor (VEP) sequence ontology terms used to identify variants with a predicted "high effect" are shown. Created with BioRender.com.



Appendix ii.ii. Pedigree of affected dogs with obligate carriers and related dogs used for genotyping highlighted.

Appendix ii.iii. Summary of the genotypes of 14 Buhunds for nine candidate variants.

			Obligate Carriers			Unaffected Dogs			
Chromosome	Position	Gene	Homozygous Reference	Heterozygous	Homozygous Variant	Homozygous Reference	Heterozygous	Homozygous Variant	Total
3	88890674	*KCNIP4*	0	2	0	8	4	0	14
12	3016408	*ITPR3*	0	0	1	0	3	9	13
12	1357280	*ZBTB12*	0	0	2	0	3	9	14
15	16999815	*ENSCAFG00000030728*	2	0	0	8	3	1	14
15	19175300	*ENSCAFG00000030891*	2	0	0	8	2	1	13
17	61194104	*ENSCAFG00000030938*	1	0	1	0	4	8	14
18	39244371	*ENSCAFG00000031871*	0	0	2	0	3	9	14
20	1178402	*PLXNA1*	0	2	0	2	5	5	14
32	37511942	*PRSS12*	1	1	0	3	7	2	14

For two of the variants (ITPR3 and ENSCAFG00000030891) one dog failed genotyping.

Appendix ii.iv. Genotypes of multi-breed panel by breed.

Breed	Total	T/T	T/C	C/C
Affenpinscher	3	3	0	0
Afghan Hound	3	3	0	0
Airedale Terrier	3	3	0	0
Alaskan Malamute	3	3	0	0
American Bulldog	3	3	0	0
American Cocker Spaniel	3	3	0	0
Australian Cattle Dog	3	3	0	0
Australian Kelpie	3	3	0	0
Australian Shepherd	3	3	0	0
Basset Fauve De Bretagne	3	3	0	0
Basset Hound	3	3	0	0
Beagle	3	3	0	0
Bearded collie	3	3	0	0
Belgian Shepherd Dog (Groenendael)	1	1	0	0
Belgian Shepherd Dog (Tervueren)	2	2	0	0
Berger Picard	3	3	0	0
Bloodhound	3	3	0	0
Border Collie	3	3	0	0
Border Terrier	3	3	0	0
Borzoi	3	3	0	0
Boston Terrier	3	3	0	0
Boxer	3	3	0	0
Brittany Spaniel	3	3	0	0
Bull Terrier	3	3	0	0
Bulldog	3	3	0	0
Catalan Sheepdog	3	3	0	0
Cavalier King Charles Spaniel	3	3	0	0
Cesky Terrier	2	2	0	0
Chesapeake Bay Retriever	3	3	0	0
Chow Chow	2	2	0	0
Cocker Spaniel	3	3	0	0
Curly Coated Retriever	3	3	0	0
Dalmatian	3	3	0	0
Dandie Dinmont	3	3	0	0
Doberman	3	3	0	0
English Setter	3	3	0	0
English Springer Spaniel	3	3	0	0
Field Spaniel	3	3	0	0
Finnish Lapphund	3	3	0	0
Flat Coated Retriever	3	3	0	0
French Bull Dog	3	3	0	0
German Pinscher	3	3	0	0
German Shepherd Dog	3	3	0	0
German Spitz (Klein)	2	2	0	0
German Wire-haired Pointer	3	3	0	0
Giant Schnauzer	3	3	0	0
Glen of Imaal Terrier	3	3	0	0

Breed	Total	T/T	T/C	C/C
Golden Retriever	3	3	0	0
Gordon Setter	3	3	0	0
Grand Basset Griffon Vendeen	3	3	0	0
Great Dane	3	3	0	0
Greater Swiss Mountain Dog	3	3	0	0
Greyhound	3	3	0	0
Havanese	3	3	0	0
Hovawart	3	3	0	0
Icelandic Sheepdog	3	3	0	0
Irish Red and White Setter	3	3	0	0
Irish Setter	3	3	0	0
Irish Water Spaniel	3	3	0	0
Italian Greyhound	3	3	0	0
Italian Spinone	3	3	0	0
Jack Russell Terrier	3	3	0	0
Japanese Akita	3	3	0	0
Japanese Shiba Inu	3	3	0	0
Keeshond	3	3	0	0
Labrador Retriever	3	3	0	0
Lagotto Romagnolo	3	3	0	0
Lakeland Terrier	3	3	0	0
Lancashire Heeler	3	3	0	0
Large Münsterländer	3	3	0	0
Leonberger	3	3	0	0
Lhasa Apso	3	3	0	0
Long-Haired Dachshund (Standard)	3	3	0	0
Manchester Terrier	3	3	0	0
Miniature Smooth Haired Dachshund	3	3	0	0
Miniature Bull Terrier	3	3	0	0
Miniature Long-haired Dachshund	3	3	0	0
Miniature Schnauzer	3	3	0	0
Miniature Wire Haired Dachshund	3	3	0	0
Norfolk Terrier	3	3	0	0
Northern Inuit	3	3	0	0
Norwegian Elkhound	3	3	0	0
Norwich terrier	3	3	0	0
Nova Scotia Duck Tolling Retrievers	3	3	0	0
Old English Mastiff	3	3	0	0
Old English Sheepdog	3	3	0	0
Otterhound	3	3	0	0
Parson Russell Terrier	3	3	0	0
Pembroke Welsh Corgi	3	3	0	0
Petit Basset Griffon Vendeen	3	3	0	0
Pharaoh Hound	3	3	0	0
Polish Lowland Sheepdog	3	3	0	0
Pug	2	2	0	0
Rhodesian Ridgeback	3	3	0	0
Rough Collie	3	3	0	0
Saint Bernard	3	3	0	0

Breed	Total	T/T	T/C	C/C
Saluki	3	3	0	0
Samoyed	3	3	0	0
Scottish Deerhounds	3	3	0	0
Scottish Terrier	3	3	0	0
Shar Pei	3	3	0	0
Shetland Sheepdogs	3	3	0	0
Shih Tzu	3	3	0	0
Siberian Husky	3	3	0	0
Skye Terrier	3	3	0	0
Smooth Haired Dachshund (Standard)	3	3	0	0
Soft-coated Wheaten Terrier	3	3	0	0
Staffordshire Bull Terrier	3	3	0	0
Standard Poodle	3	3	0	0
Standard Schnauzer	3	3	0	0
Swedish Vallhund	3	3	0	0
Tibetan Spaniel	3	3	0	0
Tibetan Terrier	3	3	0	0
Utonagan	3	3	0	0
Vizsla (smooth coat)	3	3	0	0
Vizsla (Wire-haired)	3	3	0	0
Weimaraner	3	3	0	0
Welsh Springer Spaniel	3	3	0	0
West Highland White Terrier	3	3	0	0
Whippet	3	3	0	0
Wire-haired Dachshund	3	3	0	0
Working Sheepdog	3	3	0	0

Appendix ii.v. The number of individuals for each breed represented in the 802 whole genome sequences used.

Breed	Number
Affenpinscher	2
Airedale Terrier	5
Alaskan Husky	2
Alaskan Malamute	4
Alpine Dachsbracke	2
American Bulldog	2
American Cocker Spaniel	1
American Staffordshire Terrier	2
Australian Cattle Dog	5
Australian Shepherd	3
Australian Terrier	1
Basenji	7
Grand Basset Griffon Vendeen	1
Basset Hound	9
Bavarian Hound	1
Beagle	8
Bearded Collie	13
Bedlington Terrier	1
Berger Blanc Suisse	1
Berger Picard	2
Bichon Frise	6
Black Russian Terrier	1
Bloodhound	1
Border Collie	44
Border Collie Cross	1
Border Terrier	9
Boston Terrier	1
Boxer	2
Briard	2
Brussels Griffon	3
Bull Terrier	3
Bulldog	3
Bullmastiff	2
Cairn Terrier	4
Cane Corso	5
Cardigan Welsh Corgi	1
Cavalier King Charles Spaniel	6
Central Asian Shepherd dog	1
Cesky Terrier	1
Chesapeake Bay Retriever	1
Chihuahua	6
Chinese Crested Dog	3
Chinese Indigenous Dog	28
Chow Chow	2
Cocker Spaniel	1
Collie	1

Breed	Number
Cross Breed	4
Curly Coated Retriever	5
Dachshund	6
Dalmatian	4
Dandie Dinmont Terrier	5
Doberman Pinscher	5
Dogue de Bordeaux	6
Dutch Shepherd	1
Elo	1
English Bulldog	1
English Cocker Spaniel	3
English Mastiff	2
English Setter	1
English Springer Spaniel	3
Entlebucher Sennenhund	8
Eurasier	2
Field Spaniel	1
Finnish Lapphund	2
Flat Coated Retriever	4
French Bulldog	7
Friesian Stabyhoun	2
German Shepherd Dog	20
German Wirehaired	1
Giant Schnauzer	5
Glen of Imaal Terrier	1
Golden Retriever	13
Gordon Setter	1
Great Dane	2
Great Pyrenees	1
Greater Swiss Mountain Dog	6
Greyhound	7
Griffon Bruxellois	1
Havanese	3
Heideterrier	1
Hovawart	3
Irish Red and White Setter	4
Irish Setter	2
Irish Soft Coated Wheaten Terrier	1
Irish Terrier	3
Irish Water Spaniel	1
Irish Wolfhound	8
Italian Greyhound	1
Italian Spinone	2
Jack Russell Terrier	8
Jagdterrier	2
Japanese Akita	1
Japanese Chin	1
Karelian Bear dog	1
Keeshond	3

Breed	Number
Kerry Blue Terrier	2
Kromfohrlander	1
Kunming Dog	10
Labrador Retriever	14
Lagotto Romagnolo	10
Lakeland Terrier	1
Lancashire Heeler	2
Landseer	2
Large Munsterlander	1
Leonberger	55
Lhasa Apso	1
Malinois	7
Maltese	1
Miniature Bull terrier	2
Miniature Long Haired Dachshund	2
Miniature Poodle	2
Miniature Schnauzer	25
Miniature Wire Haired Dachshund	1
Mixed Breed	8
Newfoundland	3
Northern Inuit	2
Norwegian Buhund	3
Norwich Terrier	5
Nova Scotia Duck Tolling Retriever	2
Old English Sheepdog	2
Otterhound	2
Papillon	4
Pembroke Welsh Corgi	4
Perro de Agua Espanol	1
Petit Basset Griffon Vendeen	5
Pomeranian	4
Poodle	19
Portuguese Podengo	1
Portuguese Water Dog	3
Pug	22
Rhodesian Ridgeback	4
Rottweiler	4
Rough Collie	1
Saluki	2
Scottish Deerhound	3
Scottish Terrier	6
Shar Pei	2
Shetland Sheepdog	4
Shih Tzu	5
Siberian Husky	5
Skye Terrier	2
Sloughi	3
Smooth Collie	1
Soft Coated Wheaten Terrier	2

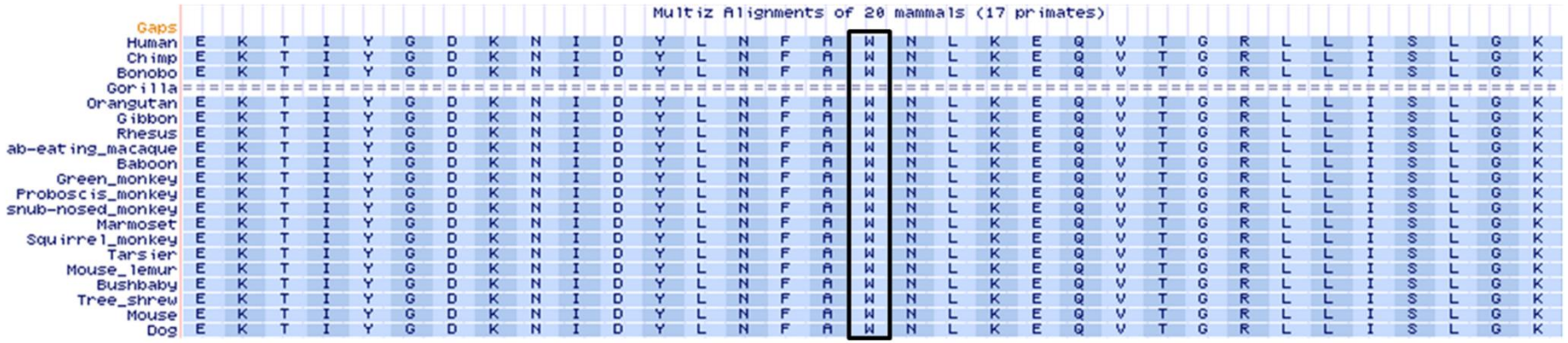
Breed	Number
Spitz (Grossspitz)	1
St. Bernard	2
Staffordshire Bull Terrier	1
Standard Poodle	2
Standard Schnauzer	1
Swedish Vallhund	3
Tibetan Mastiff	10
Tibetan Spaniel	1
Tibetan Terrier	4
Vizsla (smooth coat)	4
Vizsla (wire-haired)	1
Weimaraner	3
Welsh Springer Spaniel	9
West Highland White Terrier	20
Whippet	3
White Shepherd	1
Wolf	8
Yorkshire Terrier	69
Total	802

Appendix ii.vi. UCSC Multiz Alignments of 100 Vertebrates Human GRCh38/hg38 Assembly 4:20,734,676-20,734,678. There are no alignment data for this region in the Gorilla.

Species	Sequence
Human	CCA
Chimp	CCA
Orangutan	CCA
Gibbon	CCA
Rhesus	CCA
Crab-eating macaque	CCA
Baboon	CCA
Green monkey	CCA
Marmoset	CCA
Squirrel monkey	CCA
Bushbaby	CCA
Chinese tree shrew	CCA
Squirrel	CCA
Lesser Egyptian jerboa	CCA
Prairie vole	CCA
Chinese hamster	CCA
Golden hamster	CCA
Mouse	CCA
Rat	CCA
Naked mole-rat	CCA
Guinea pig	CCA
Chinchilla	CCA
Brush-tailed rat	CCA
Rabbit	CCA
Pika	CCA
Pig	CCA
Alpaca	CCA
Bactrian camel	CCA
Dolphin	CCA
Killer whale	CCA
Tibetan antelope	CCA
Cow	CCA
Sheep	CCA
Domestic goat	CCA
Horse	CCA
White rhinoceros	CCA
Cat	CCA
Dog	CCA
Ferret	CCA
Panda	CCA
Pacific walrus	CCA
Weddell seal	CCA
Black flying-fox	CCA
Megabat	CCA
Big brown bat	CCA
David's myotis (bat)	CCA

Species	Sequence
Microbat	CCA
Hedgehog	CCA
Shrew	CCA
Star-nosed mole	CCA
Elephant	CCA
Cape elephant shrew	CCA
Manatee	CCA
Cape golden mole	CCA
Tenrec	CCA
Aardvark	CCA
Armadillo	CCA
Opossum	CCA
Tasmanian devil	CCA
Wallaby	CCA
Platypus	CCA
Rock pigeon	CCA
Saker falcon	CCA
Peregrine falcon	CCA
Collared flycatcher	CCA
White-throated sparrow	CCA
Medium ground finch	CCA
Zebra finch	CCA
Tibetan ground jay	CCA
Budgerigar	CCA
Parrot	CCA
Scarlet macaw	CCA
Mallard duck	CCA
Chicken	CCA
Turkey	CCA
American alligator	CCA
Green seaturtle	CCA
Painted turtle	CCA
Chinese softshell turtle	CCA
Spiny softshell turtle	CCA
Lizard	CCA
X. tropicalis	CCA
Coelacanth	CCA
Tetraodon	CCA
Fugu	CCA
Yellowbelly pufferfish	CCA
Nile tilapia	CCA
Princess of Burundi	CCA
Burton's mouthbreeder	CCA
Zebra mbuna	CCA
Pundamilia nyererei	CCA
Medaka	CCA
Southern platyfish	CCA
Stickleback	CCA
Atlantic cod	CCA

Species	Sequence
Zebrafish	CCA
Mexican tetra (cavefish)	CCA
Spotted gar	CCA
Lamprey	CCA
Gorilla	===



Appendix ii.vii. UCSC Multiz Alignments of 20 species of mammal (17 primates) Human GRCh38/hg38 Assembly 4: 20,734,628-20,734,726. The amino acid at the location of the mutation is boxed, and the 16 flanking amino acids in each direction are shown. There are no alignment data for the Gorilla in this region.

Appendix ii.viii. Canine KCNIP4 transcripts and protein isoform annotations, and the details of those demonstrated to be expressed in canine cerebellum.

Split into three parts (A, B, and C) to fit on page. Original Excel file available at: https://doi.org/10.1371/journal.pgen.1008527 ('S5 Table')

Appendix ii.viii (A)

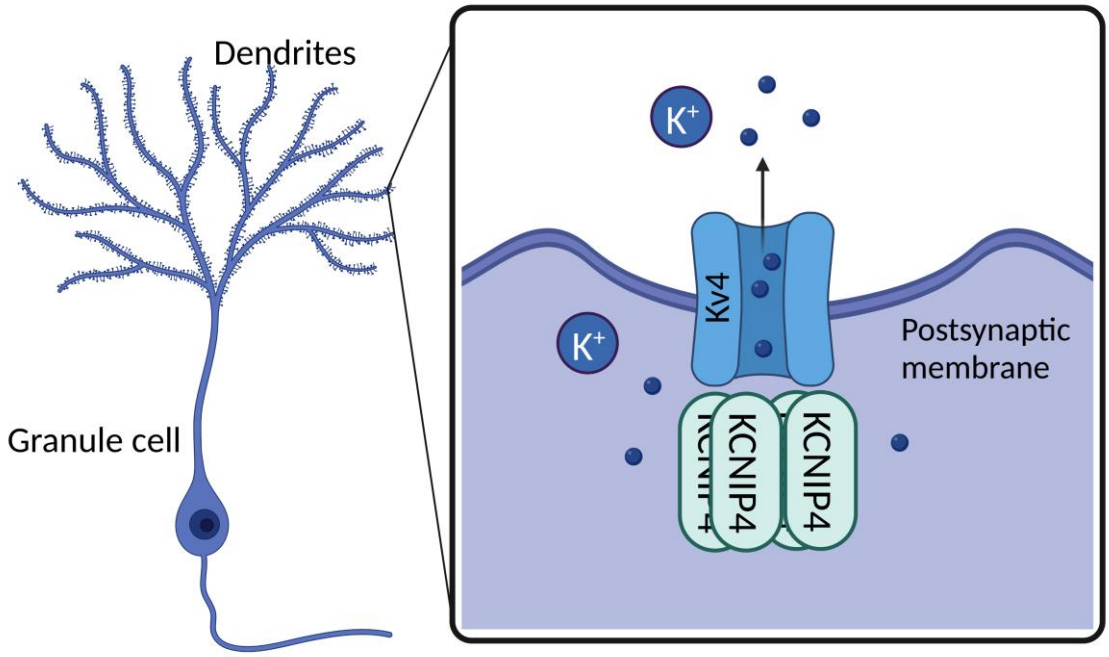
Canine RefSeq Transcript	Canine Ensembl Transcript	CanFam3.1 Genomic Position (Annotated)	Canine Protein RefSeq	Confirmed in Canine Cerebellum	Length (aa)
N/A	ENSCAFT00000090603.1 (Partial)	3: 87,771,875-88,894,884 (Ensembl)	N/A	Yes	250
N/A	ENSCAFT00000090861.1 (Partial)	3: 87,771,875-88,894,884 (Ensembl)	N/A	Yes	216
XM_014112663.2	ENSCAFT00000060142.1	3: 88,013,563-88,894,884 (Ensembl)	XP_013968138.1	Yes	225
XM_003434400.4	ENSCAFT00000083618.1	3: 88,370,168-88,894,884 (Ensembl)	XP_003434448.1	Yes	229
XM_005618660.3	N/A	3: 88,665,414-88,896,093 (RefSeq)	XP_005618717.1	Yes	233
XM_536275.6	N/A	3: 88,754,085..88,894,884 (RefSeq)	XP_536275.4	No	213
N/A	ENSCAFT00000026195.4	3: 88,665,407-88,894,884 (Ensembl)	N/A	No	267
N/A	ENSCAFT00000079461.1	3: 88,665,407-88,894,884 (Ensembl)	N/A	No	280
N/A	ENSCAFT00000061238.1	3: 88,754,036-88,894,884 (Ensembl)	N/A	No	297

Appendix ii.viii (B)

Canine RefSeq Transcript	Canine Ensembl Transcript	Pruunsild et al (2005)	Human Ensembl Transcript ID	Human UniProt Identifier	Mass (Da)
N/A	ENSCAFT00000090603.1 (Partial)	KCNIP4-1a	ENST00000382152.7	Q6PIL6-1	28,729
N/A	ENSCAFT00000090861.1 (Partial)	KCNIP4-1aΔ2	ENST00000447367.6	Q6PIL6-2	25,008
XM_014112663.2	ENSCAFT00000060142.1	KCNIP4-1bΔ2	ENST00000382148.7	Q6PIL6-5	26,267
XM_003434400.4	ENSCAFT00000083618.1	KCNIP4-1dΔ2	ENST00000382150.8	Q6PIL6-4	26,502
XM_005618660.3	N/A	KCNIP4-1eΔ2	ENST00000382149.9	Q3YAB7	26,929
XM_536275.6	N/A	N/A	N/A	N/A	N/A
N/A	ENSCAFT00000026195.4	N/A	N/A	N/A	N/A
N/A	ENSCAFT00000079461.1	N/A	N/A	N/A	N/A
N/A	ENSCAFT00000061238.1	N/A	N/A	N/A	N/A

Appendix ii.viii (C)

Canine RefSeq Transcript	Canine Ensembl Transcript	Confirmed Canine First Exon Coordinates (translated region)(CanFam 3.1)	Confirmed Canine Second Exon Coordinates (CanFam 3.1)
N/A	ENSCAFT00000090603.1 (Partial)	3:87771818-87771878 (Gap in genome until 87,771,874)	3:88754097-88754198
N/A	ENSCAFT00000090861.1 (Partial)	3:87771818-87771878 (Gap in genome until 87,771,874)	3:88780584-88780708
XM_014112663.2	ENSCAFT00000060142.1	3:88013563-88013650	3:88780584-88780708
XM_003434400.4	ENSCAFT00000083618.1	3:88370168-88370267	3:88780584-88780708
XM_005618660.3	N/A	3:88665441-88665552	3:88780584-88780708
XM_536275.6	N/A	N/A	N/A
N/A	ENSCAFT00000026195.4	N/A	N/A
N/A	ENSCAFT00000079461.1	N/A	N/A
N/A	ENSCAFT00000061238.1	N/A	N/A



Appendix ii.ix. Simplified illustration of the cellular localisation of *KCNIP4* in its suggested role in granule cell dendrites. Informed by Kise et all (2021) [505]. Created with BioRender.com.

Appendix ii.x. Candidate causal variants identified from whole genome sequencing, and primer sequences used for Sanger sequencing and fragment length analysis.

Variant Location	Variant cDNA	Variant Protein	Gene	Forward Primer Sequence **[Tail]**	Reverse Primer Sequence
3:88890674	c.538T>C	p.Trp180Arg	*KCNIP4*	AAATCCAGAGGGCTGTAACC	TTGTGTGGGTGATGGTGAG
18:39244371	c.410G>A	p. Asp137Asn	ENSCAFG00000031871	TCCAATGTGAATGTGTTACTGG	CCTTTGCTGGCTGTGTATGT
20:1178402	c.4697G>A	p.Arg1566His	*PLXNA1*	CCTTGTCACTCCCTTGTGCT	ATGTGTGCCCTGTCATCCA
12:3016408	c.1021C>T	p.Arg341Tryp	*ITPR3*	CAGCAGGAATGGTGAGGAC	GAGACAAGGGACAGGACAGC
17:61194104	c.7088G>A	p.Arg2363Lys	*FLG2* (ENSCAFG00000030938)	CCTGAGCCAAAGCCATGT	CTGGACAGAACGAATCTGGA
32:37511942	c.2023G>T	p.Glu675*	*PRSS12*	TTGGAGGTGGTTTTCTGTGG	TAGGACTGAGGGTGGTGAGG
12:1357280	c.899_922del	p.Val300_ Ala307del	*ZBTB12*	**[TGACCGGCAGCAAAATTG]** GAGGTTGCTGCTGTGGTTG	GGGAGGGTTTGCTGTTGAT
15:16999815	c.274C>G	p.Pro92Ala	ENSCAFG00000030728	CGAGGTAACAGGGAACAGGA	GCTGGCGTGGGACATAATAA
15:19175300	c.226C>T	p.Ala76Val	ENSCAFG00000030891	CCCCAATCGCCTTAGAAACA	CTGGAGGAGGTCCCTGGAG

Appendix ii.xi. Primer sequences used for allelic discrimination assay of *KCNIP4* variant.

Primer / Probe	Primer / Probe Sequence
KCNIP4 Forward	CATCAAGGGTCTTTCCATT
KCNIP4 Reverse	TGGTGAGGCATTTCTTACTT
KCNIP4 HEX (Variant Probe)	AACTCAACTGGGCATTTAACTTGTAT
KCNIP4 FAM (Reference Probe)	AACTCAACCGGGCATTTAACTT

Appendix ii.xii. Transcript-specific primers for *KCNIP4*.

Transcript	Forward Primer	Expected PCR Product Size (bp)
KCNIP4-1a	CCACAGGCGGTTTCCTCTAC	678
KCNIP4-1aΔ2	GAGCTCCACAGGCGACAG	581
KCNIP4-1bΔ2	TCATACAGACAGCGTGGAAGA	575
KCNIP4-1dΔ2	TTGAAGCAGGTTTAGAAGACAGC	585
KCNIP4-1eΔ2	TCGACTTCTCGGAAGACAGC	582

Shared reverse primer: GCATGGAGCGCATTATGTTT

iii. Manuscript 2 appendices

Appendix iii.i. Affiliations and funding information for DBVDC members

Dog Biomedical Variant Database Consortium (DBVDC) Members:

Gustavo Aguirre

Department of Clinical Sciences & Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, United States of America

Catherine André

University of Rennes, CNRS, IGDR – UMR 6290, France

Danika Bannasch

School of Veterinary Medicine, University of California, Davis, United States of America

Doreen Becker

Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany

Brian Davis

College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, United States of America

Cord Drögemüller

Institute of Genetics, Vetsuisse Faculty, University of Bern, Switzerland

Kari Ekenstedt

Department of Basic Medical Sciences, College of Veterinary Medicine, Purdue University, West Lafayette, Indiana, United States of America

Kiterie Faller

Royal (Dick) School of Veterinary Studies, University of Edinburgh, United Kingdom

Oliver Forman

Wisdom Health, Waltham Centre for Pet Nutrition, Leicestershire, United Kingdom

Steven Friedenberg

Department of Veterinary Clinical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, United States of America

Eva Furrow

Department of Veterinary Clinical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, United States of America

Urs Giger

Section of Medical Genetics (PennGen), University of Pennsylvania, Philadelphia, United States of America

Christophe Hitte

University of Rennes, CNRS, IGDR – UMR 6290, France

Marjo K. Hytönen

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Vidhya Jagannathan

Institute of Genetics, Vetsuisse Faculty, University of Bern, Switzerland

Tosso Leeb

Institute of Genetics, Vetsuisse Faculty, University of Bern, Switzerland

Hannes Lohi

Department of Medical and Clinical Genetics, and Department of Veterinary Biosciences, University of Helsinki; and Folkhälsan Research Center, Helsinki, Finland

Cathryn S. Mellersh

Kennel Club Genetics Centre, Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

James R. Mickelson

College of Veterinary Medicine, University of Minnesota, St Paul, United States of America

Leonardo Murgiano

Department of Clinical Sciences & Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, United States of America

Anita Oberbauer

Department of Animal Science, University of California, Davis, United States of America

Sheila Schmutz

Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon, Canada

Jeffrey J. Schoenebeck

Royal (Dick) School of Veterinary Studies, University of Edinburgh, United Kingdom

Kim M. Summers

Mater Research Institute-University of Queensland, Brisbane, Australia

Frank G. van Steenbeek

Faculty of Veterinary Medicine, Department of Clinical Sciences of Companion Animals, Utrecht University, the Netherlands

Claire Wade

School of Life and Environmental Sciences, University of Sydney, Australia

Funding information:

Catherine André and Christophe Hitte were supported by the French Cani-DNA CRB (http://dog-genetics.genouest.org), which is part of the CRB-Anim infrastructure, ANR-11-INBS-0003. Kari J. Ekenstedt was supported by the Office of the Director, National Institutes of Health (NIH) under award number K01-OD027051. Eva Furrow was supported by the Office of the Director, National Institutes of Health (NIH) under award

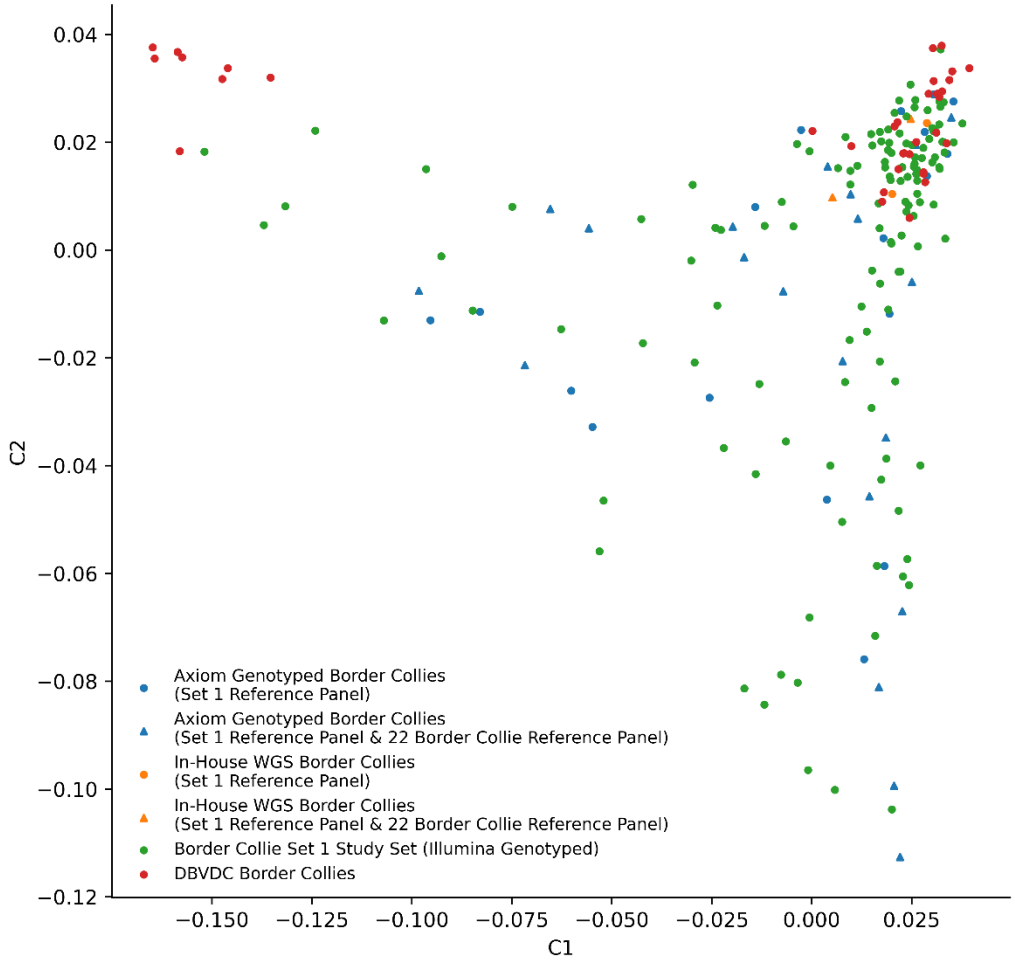
number K01-OD019912. Tosso Leeb was supported by the Albert-Heim Foundation (project no. 105). Kim M. Summers receives core support from the Mater Foundation, Brisbane, Australia.

Appendix iii.ii. The number of individuals for each of the 93 breeds, and mixed breeds, included in the dataset of 186 in-house WGS.

Breed	Count
Affenpinscher	1
Airedale Terrier	2
Alaskan Malamute	1
American Cocker Spaniel	1
Australian Shepherd	1
Basset Hound	4
Beagle	3
Bearded collie	2
Bedlington terrier	1
Berger Picard (Picardy sheepdog)	1
Bloodhound	1
Border Collie	7
Border Terrier	9
Boxer	1
Briard	2
Bull Terrier	1
Bulldog	3
Cairn Terrier	1
Cavalier King Charles Spaniel	3
Cesky Terrier	1
Chesapeake Bay Retriever	1
Chihuahua	1
Chinese Crested	2
Chow Chow	1
Corgi	1
Mixed Breed	5
Dalmatian	1
Dandie Dinmont	4
Dobermann	2
English Setter	1
English Springer Spaniel	2
Field Spaniel	1
Finnish Lapphund	1
Flat Coated Retriever	4
French Bull Dog	4
German Shepherd Dog	1
Giant Schnauzer	5
Glen of Imaal Terrier	1
Golden Retriever	2
Gordon Setter	1
Grand Basset Griffon Vendeen	1
Great Dane	1
Greyhound	2
Griffon Bruxellois	1

Breed	Count
Irish Red and White Setter	4
Irish Setter	2
Irish Terrier	1
Irish Water Spaniel	1
Irish Wolfhound	6
Italian Spinone	2
Japanese Akita	1
Keeshond	3
Labrador Retriever	3
Lagotto Romagnolo	1
Lakeland Terrier	1
Lancashire Heeler	2
Large Munsterlander	1
Leonberger	2
Lhasa Apso	1
Maltese	1
Miniature Long-haired Dachshund	2
Miniature Poodle	1
Miniature Schnauzer	2
Miniature Wire Haired Dachshund	1
Newfoundland	1
Northern Inuit	2
Norwegian Buhund	3
Norwich terrier	1
Nova Scotia Duck Tolling Retriever	1
Old English Sheepdog	1
Otterhound	2
Papillon	1
Petit Basset Griffon Vendeen	2
Pug	3
Rottweiler	1
Rough Collie	1
Scottish Terrier	5
Shar Pei	2
Shetland Sheepdog	1
Shih Tzu	1
Siberian Husky	3
Skye Terrier	2
Smooth Collie	1
Soft-coated Wheaten Terrier	2
Staffordshire Bull Terrier	1
Swedish Vallhund	1
Tibetan Spaniel	1
Tibetan Terrier	2
Vizsla (smooth coat)	4
Vizsla (Wire-haired)	1
Weimaraner	1

Breed	Count
Welsh Springer Spaniel	7
West Highland White Terrier	1
Whippet	1



Appendix iii.iii. Multidimensional scaling (MDS) plot of 39 Axiom genotyped Border Collies, five in-house WGS Border Collies, 35 DBVDC WGS Border Collies, and 130 Border Collie Set 1 individuals genotyped using the Illumina array. MDS data are based on 100,535 SNPs and were generated using PLINK (v1.90). Triangles indicate individuals that were included in the '22 Border Collie Reference Panel'.

Appendix iii.iv. Comparison of the expected frequency of the allele coded as '1' (provided by IMPUTE2) for imputed SNPs across grouped Info scores for the three datasets.

	Border Collie Set 1			**Border Collie Set 2**			**Italian Spinone**		
Info group	Mean expected allele frequency	Standard deviation	SNPs (n)	Mean expected allele frequency	Standard deviation	SNPs (n)	Mean expected allele frequency	Standard deviation	SNPs (n)
0-0.1	0.0006	0.02	27,743	0.0015	0.04	34,608	0.0031	0.06	66,315
>0.1-0.2	0.0031	0.04	2,424	0.0070	0.07	2,337	0.0050	0.05	2,522
>0.2-0.3	0.0053	0.05	1,665	0.0056	0.05	1,548	0.0087	0.07	1,644
>0.3-0.4	0.0065	0.05	1,480	0.0105	0.08	1,251	0.0074	0.04	1,267
>0.4-0.5	0.0131	0.08	1,524	0.0154	0.09	1,172	0.0149	0.08	1,173
>0.5-0.6	0.0275	0.12	1,575	0.0290	0.12	1,173	0.0237	0.10	1,167
>0.6-0.7	0.0466	0.15	2,524	0.0534	0.17	2,023	0.0317	0.11	1,533
>0.7-0.8	0.0598	0.17	5,225	0.0744	0.19	3,960	0.0571	0.16	2,603
>0.8-0.9	0.0877	0.20	13,238	0.1078	0.23	10,855	0.1098	0.22	6,882
>0.9-1	0.3228	0.28	268,393	0.3257	0.27	261,962	0.3539	0.28	286,775

Appendix iii.v. Comparison of the number of imputed SNPs with an expected frequency of the allele coded as '1' (provided by IMPUTE2) lower than 0.05 across grouped Info scores for the three datasets. The percentage of the total number of imputed SNPs with an expected allele frequency lower than 0.05 that are within each Info group is shown.

	Border Collie Set 1		**Border Collie Set 2**		**Italian Spinone**	
Info group	SNPs with expected allele frequency <0.05 (n)	Percent of all SNPs with expected allele frequency <0.05 (%)	SNPs with expected allele frequency <0.05 (n)	Percent of all SNPs with expected allele frequency <0.05 (%)	SNPs with expected allele frequency <0.05 (n)	Percent of all SNPs with expected allele frequency <0.05 (%)
0-0.1	27,727	27.16	34,558	34.45	66,111	53.54
>0.1-0.2	2,420	2.37	2,325	2.32	2,515	2.04
>0.2-0.3	1,660	1.63	1,544	1.54	1,635	1.32
>0.3-0.4	1,474	1.44	1,241	1.24	1,262	1.02
>0.4-0.5	1,502	1.47	1,151	1.15	1,149	0.93
>0.5-0.6	1,512	1.48	1,123	1.12	1,121	0.91
>0.6-0.7	2,314	2.27	1,841	1.84	1,412	1.14
>0.7-0.8	4,603	4.51	3,342	3.33	2,168	1.76
>0.8-0.9	10,008	9.8	7,539	7.52	4,622	3.74
>0.9-1	48,855	47.86	45,648	45.51	41,496	33.6

iv. Manuscript 3 appendices

Appendix iv.i. GWAS data and Chi-squared analysis of the 44 SNPs with a GEMMA-adjusted GWAS P-value < 1×10^{-3}

			GWAS (GEMMA-adjusted)					Unadjusted	Replication			
SNP ID	**Genomic Pos.** [a]	**Nearest known gene**	**ca/co** [b] **(n)**	**P-value**	**Alleles (risk/non-risk)**	**Risk allele freq. (ca/co** [b]**)**	**Odds ratio (95% CI)**	**χ^2 P-value (GWAS dogs)**	**Pop. set risk allele freq.** [c]	**χ^2 P-value** [d]	**Additional ca/co** [b] **(n)**	**Risk allele freq. (ca/co** [b]**)** [e]
AX-167241819	5:16414896	*PFAH1B2*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.18	0.01	10/45	0.15/0.20
AX-167180215	5:16428240	*PFAH1B2*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.18	0.01	10/45	0.15/0.20
AX-167198131	5:16498068	*SIK3*	24/24	8.5×10^{-4}	C/T	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167217929	5:16524042	*SIK3*	24/24	8.5×10^{-4}	C/T	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167216958	5:16547914	*SIK3*	24/24	8.5×10^{-4}	T/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167401878	5:16564735	*SIK3*	23/24	7.9×10^{-4}	A/G	0.59/0.27	1.48 (1.21-1.82)	3.5×10^{-3}	0.16	3.6×10^{-4}	10/45	0.15/0.13
AX-167254147	5:16572747	*SIK3*	24/24	8.5×10^{-4}	T/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.85/0.87
AX-167763190	5:16581759	*SIK3*	24/24	8.5×10^{-4}	A/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-168270348	5:16591140	*SIK3*	24/24	8.5×10^{-4}	T/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167160608	5:16613986	*SIK3*	24/24	8.5×10^{-4}	C/A	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167757392	5:16616413	*SIK3*	24/24	8.5×10^{-4}	T/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-168258838	5:16625279	*SIK3*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13

			GWAS (GEMMA-adjusted)					Unadjusted	Replication			
SNP ID	**Genomic Pos. [a]**	**Nearest known gene**	**ca/co [b] (n)**	**P-value**	**Alleles (risk/non-risk)**	**Risk allele freq. (ca/co [b])**	**Odds ratio (95% CI)**	**χ^2 P-value (GWAS dogs)**	**Pop. set risk allele freq. [c]**	**χ^2 P-value [d]**	**Additional ca/co [b] (n)**	**Risk allele freq. (ca/co [b]) [e]**
AX-167737405	5:16638591	*SIK3*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167894416	5:16647269	*SIK3*	24/24	8.5×10^{-4}	G/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167197869	5:16654822	*SIK3*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167164472	5:16667564	*SIK3*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167231875	5:16677678	*SIK3*	24/24	8.5×10^{-4}	C/T	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167822748	5:16692418	*SIK3*	24/24	8.5×10^{-4}	G/A	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167198539	5:16729119	*SIK3*	24/24	8.5×10^{-4}	T/C	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167562834	5:16729648	*SIK3*	24/24	8.5×10^{-4}	A/G	0.54/0.27	1.50 (1.19-1.88)	0.01	0.16	1.1×10^{-3}	10/45	0.15/0.13
AX-167923327	10:54532815	*ASB3*	24/24	4.2×10^{-4}	A/C	0.23/0.04	1.74 (1.31 - 2.31)	3.5×10^{-3}	0.23	0.21	10/45	0.50/0.30
AX-167671433	12:22674130	*GFRAL*	24/23	9.4×10^{-4}	A/G	0.35/0.15	1.64 (1.25 - 2.15)	5.6×10^{-3}	0.21	0.14	10/41	0.25/0.24
AX-167677441	14:48725029	*ELMO1*	23/24	8.9×10^{-4}	T/C	0.67/0.44	1.49 (1.17 - 1.89)	0.02	0.33	3.5×10^{-3}	10/45	0.05/0.24
AX-167244070	15:50013246	*FBXW7*	24/23	4.8×10^{-4}	A/C	0.92/0.74	1.71 (1.30-2.26)	0.01	0.80	0.14	10/45	0.85/0.82
AX-167162511	15:50036671	*FBXW7*	24/24	5.6×10^{-4}	C/T	0.96/0.77	1.75 (1.29-2.36)	3.5×10^{-3}	0.89	0.14	10/45	0.95/0.97
AX-167885106	15:50109542	*FBXW7*	24/24	5.6×10^{-4}	C/T	0.96/0.77	1.75 (1.29-2.36)	3.5×10^{-3}	0.88	0.10	10/45	0.95/0.96

			GWAS (GEMMA-adjusted)					Unadjusted	Replication			
SNP ID	**Genomic Pos. [a]**	**Nearest known gene**	**ca/co [b] (n)**	**P-value**	**Alleles (risk/non-risk)**	**Risk allele freq. (ca/co [b])**	**Odds ratio (95% CI)**	**χ^2 P-value (GWAS dogs)**	**Pop. set risk allele freq. [c]**	**χ^2 P-value [d]**	**Additional ca/co [b] (n)**	**Risk allele freq. (ca/co [b]) [e]**
AX-167791796	15:50165261	*FBXW7*	24/24	5.6×10^{-4}	A/G	0.96/0.77	1.75 (1.29-2.36)	3.5×10^{-3}	0.90	0.13	10/44	0.95/0.97
AX-167578307	15:50168839	*FBXW7*	24/24	5.6×10^{-4}	A/G	0.96/0.77	1.75 (1.29-2.36)	3.5×10^{-3}	0.90	0.14	10/45	0.95/0.97
AX-167631839	15:50180761	*FBXW7*	24/24	7.7×10^{-4}	G/C	0.94/0.77	1.70 (1.25-2.29)	0.01	0.90	0.28	10/45	0.95/0.97
AX-167169849	15:50444714	*FBXW7*	24/23	4.5×10^{-4}	T/C	0.96/0.76	1.77 (1.32-2.39)	2.5×10^{-3}	0.90	0.13	10/45	0.95/0.97
AX-167780040	15:50486447	*TMEM154*	24/23	4.3×10^{-4}	T/C	0.96/0.76	1.79 (1.33 - 2.42)	2.5×10^{-3}	0.90	0.13	10/45	0.95/0.97
AX-168017403	15:50494888	*TMEM154*	24/24	5.6×10^{-4}	A/G	0.96/0.77	1.75 (1.29-2.36)	3.5×10^{-3}	0.81	0.04	10/44	0.85/0.81
AX-167233313	21:25108139	*FCHSD2*	24/24	9.5×10^{-4}	C/G	0.31/0.06	1.72 (1.33 - 2.24)	3.5×10^{-4}	0.007	5.8×10^{-7}	10/45	0.00/0.00
AX-167664915	32:5897058	*RASGEF1B*	23/24	6.3×10^{-4}	G/T	0.76/0.52	1.56 (1.22 – 1.99)	0.02	0.64	0.06	10/45	0.65/0.60
AX-168031858	32:6522283	*TMEM150C*	24/23	9.2×10^{-4}	T/C	0.69/0.37	1.43 (1.17 - 1.76)	6.8×10^{-3}	0.30	1.2×10^{-4}	10/45	0.30/0.21
AX-167541099	32:6531625	*TMEM150C*	24/24	9.7×10^{-4}	G/A	0.69/0.40	1.46 (1.18 - 1.81)	6.8×10^{-3}	0.32	3.2×10^{-4}	8/39	0.31/0.24
AX-167757670	32:6537118	*TMEM150C*	24/23	6.4×10^{-4}	C/G	0.71/0.39	1.46 (1.19-1.80)	5.1×10^{-3}	0.53	3.3×10^{-3}	4/20	0.50/0.45
AX-167854555	32:6572078	*SCD5*	24/24	7.6×10^{-4}	A/G	0.73/0.50	1.54 (1.20-1.98)	0.02	0.69	0.17	10/43	0.65/0.64
AX-167206452	32:6656290	*SCD5*	24/24	7.6×10^{-4}	T/C	0.73/0.50	1.54 (1.20-1.98)	0.02	0.69	0.19	10/45	0.65/0.64
AX-167858314	32:6753215	*SCD5*	24/24	7.6×10^{-4}	T/C	0.73/0.50	1.54 (1.20-1.98)	0.02	0.69	0.19	10/45	0.65/0.64

			GWAS (GEMMA-adjusted)					Unadjusted	Replication			
SNP ID	**Genomic Pos. [a]**	**Nearest known gene**	**ca/co [b] (n)**	**P-value**	**Alleles (risk/non-risk)**	**Risk allele freq. (ca/co [b])**	**Odds ratio (95% CI)**	**χ^2 P-value (GWAS dogs)**	**Pop. set risk allele freq. [c]**	**χ^2 P-value [d]**	**Additional ca/co [b] (n)**	**Risk allele freq. (ca/co [b]) [e]**
AX-167828838	32:6759073	*SEC31A*	24/23	6.1×10^{-4}	T/C	0.73/0.50	1.56 (1.22 – 2.00)	0.02	0.69	0.20	10/45	0.65/0.64
AX-167803841	37:2515946	*TMEFF2*	24/23	6.5×10^{-4}	G/A	0.71/0.46	1.56 (1.23-1.97)	0.01	0.64	0.03	10/45	0.80/0.72
AX-167666536	37:5073636	*SLC39A10*	24/23	9.2×10^{-4}	T/C	0.46/0.17	1.55 (1.25 - 1.93)	4.9×10^{-3}	0.18	1.4×10^{-3}	10/44	0.15/0.13
AX-167857485	37:5161255	*SLC39A10*	23/24	4.3×10^{-4}	G/A	0.61/0.27	1.59 (1.30 - 1.94)	2.5×10^{-3}	0.30	0.07	10/44	0.15/0.31

The SNPs are ordered by chromosome and genomic position. [a] CanFam3.1 genomic location of the SNP in the format chromosome: bp position. [b] ca/co = case/control. [c] The risk allele frequency in the full population dataset including up to 232 dogs. [d] χ2 P-value for a combined analysis of the dogs included in the GWAS and up to 10 cases and 45 controls. [e] Risk allele frequencies in the additional cases and controls from the population dataset.

Appendix iv.ii. Chi-squared analysis of chromosome 32 top SNPs using only dogs with non-missing genotype data for all SNPs.

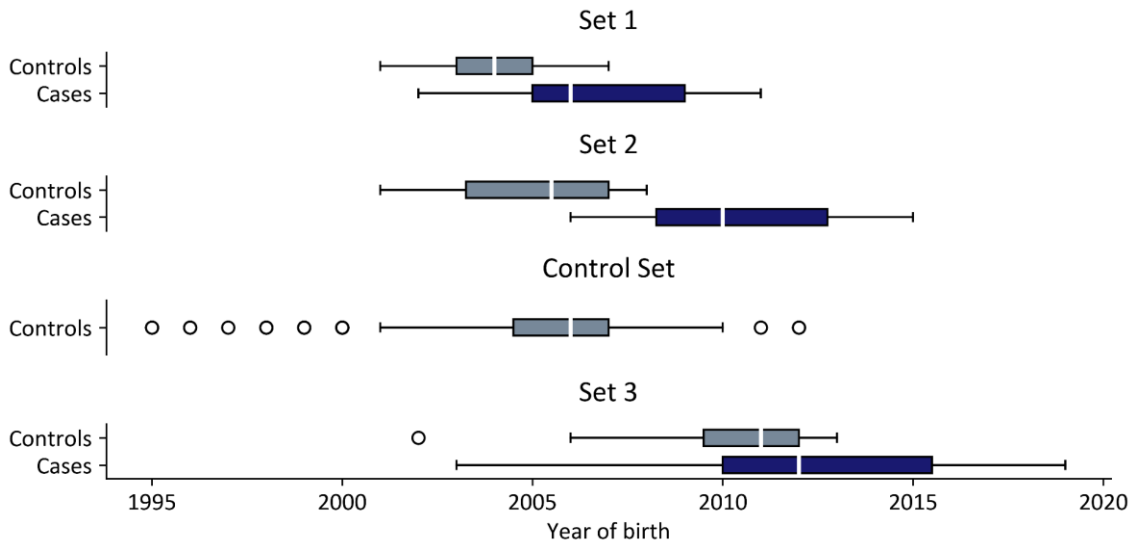
				GWAS Dogs			Replication Set		
SNP ID	**Genomic Pos. [a]**	**Nearest gene**	**Alleles (risk/non-risk)**	**ca/co [b] (n)**	**Risk allele freq. (ca/co [b])**	**χ^2 P-value (GWAS dogs)**	**Additional ca/co [b] (n)**	**Risk allele freq. (ca/co [b]) [c]**	**χ^2 P-value**
AX-167828838	32:6759073	SEC31A	T/C	24/22	0.73 / 0.48	9.3×10^{-3}	8/39	0.25/0.32	0.15
AX-168031858	32:6522283	TMEM150C	T/C	24/22	0.69 / 0.34	2.4×10^{-3}	8/39	0.31/0.24	9.8×10^{-5}
AX-167541099	32:6531625	TMEM150C	G/A	24/22	0.69 / 0.36	2.7×10^{-3}	8/39	0.31/0.24	1.2×10^{-4}

[a] The CanFam3.1 genomic location of the SNP in the format chromosome: bp position. [b] ca/co = case/control. [c] Risk allele frequencies in the additional cases and controls from the population dataset.

v. Manuscript 4 appendices

Appendix v.i. Sample details including collection years, years of birth, country of origin, and method of case diagnosis

		Set 1	Set 2	Control Set	Set 3
Number of cases and controls		29 cases, 29 controls	23 cases, 22 controls	175 controls	23 cases, 23 controls
Samples collected (year)		2007-2014	2014-2017	2007-2019	2007-2021
Years of birth	**Cases**	2002-2011	2006-2015 (1 unknown)	-	2003-2019 (4 unknown)
	Controls	2001-2007	2001-2008	1995-2012	2002-2013
Country of origin (n)		UK (49), USA (7), Australia (2)	UK (32), USA (12), Germany (1)	UK (118), USA (54), Italy (4), Netherlands (2), Belgium (1), Denmark (1), Romania (1)	UK (38), USA (4), Canada (2), Finland (2)
Case diagnosis methodology		Breed-wide survey. Animal Health Trust neurology unit.	Owner reported questionnaire and veterinary records, where available. Animal Health Trust neurology unit, UK.	-	Owner reported questionnaire and veterinary records, where available. Animal Health Trust neurology unit, UK. Linnaeus referral veterinary hospitals, UK. Two samples from University of Helsinki, Finland



Appendix v.ii. Year of birth for cases and controls by study set. Grey: Controls. Blue: Cases. Boxes indicate the lower to upper quartiles, and the white lines the median. Whiskers extend to the first datum beyond 1.5 times the interquartile range from the lower and upper quartiles. Circles indicate outliers. The plots share an x-axis.

Appendix v.iii. Additional methodology and results: kinship calculations and selection of controls for Set 2

Due to the selection criteria requiring control dogs be over the age of seven years, the cases and controls in Set 2 had only partially overlapping years of birth. Therefore, two separate random sample sets were used in the analysis of the Set 2 case-control set, one comprising n=21 dogs born between 2008 and 2016 mirroring the 'cases', and the other comprising 25 dogs born between 2001 and 2009 as per the controls.

Relationships between each pair of dogs in a particular cohort were determined using kinship coefficients (the kinship between two individuals is equal to the inbreeding coefficient of their offspring). Ancestors of the dogs in each cohort were isolated from the Kennel Club pedigree, and 'dummy' progeny were created for each pair of individuals in the cohort. Inbreeding coefficients for the 'dummy' progeny (i.e. kinship coefficients of each pair) were calculated utilising the algorithm of Meuwissen and Luo (1992) [506] using a script written in MATLAB.

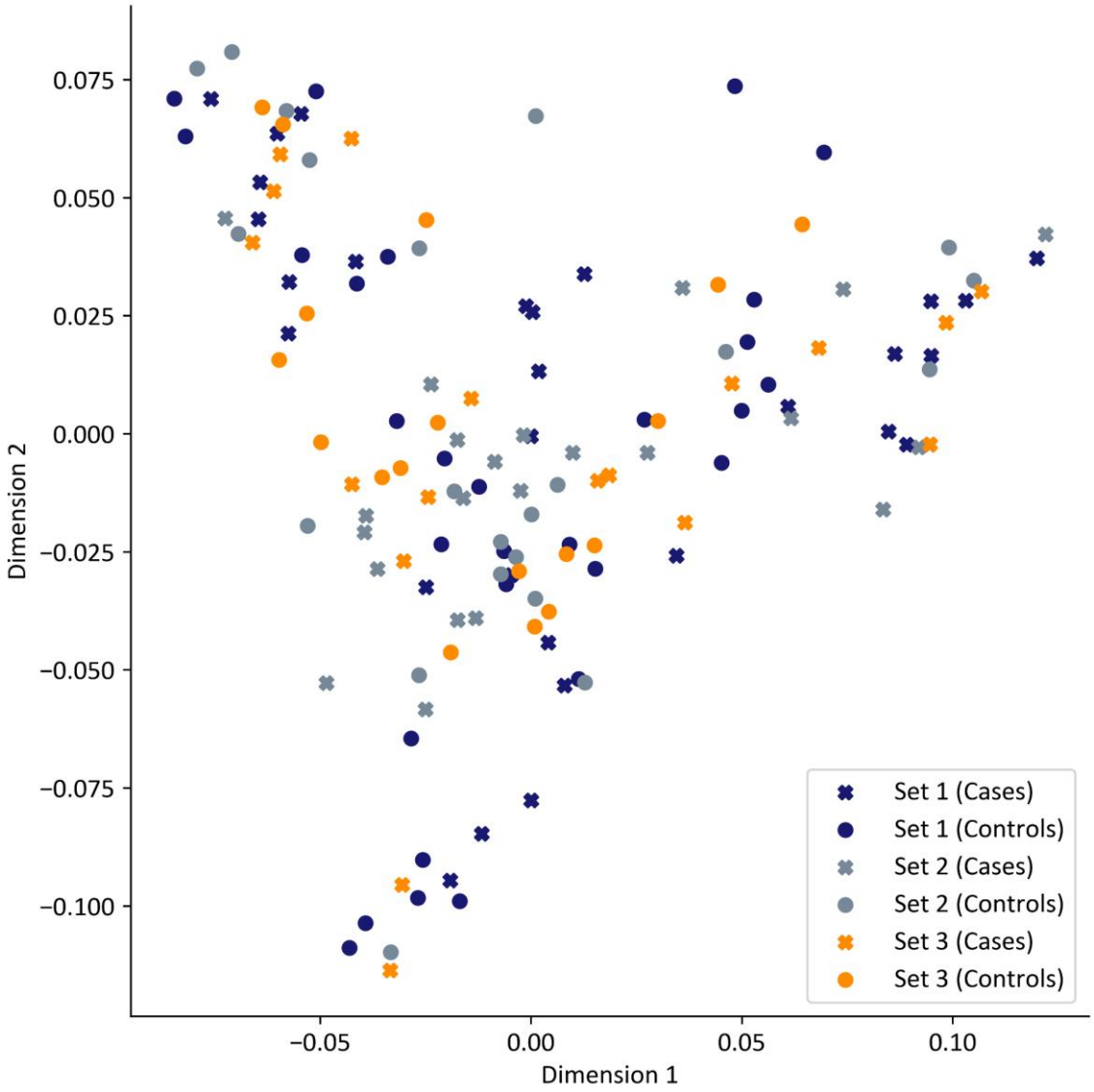
The mean kinships among and between the 'cases' and 'controls' cohorts used in this study were representative of random samples of dogs from similar birth years (**Appendix v.iv**).

Standard deviations of kinship coefficients among the cases were within but at the higher end of the 95% confidence intervals from the 1,000 random samples in each case: 0.0233 – 0.0515 for samples representing 'cases', 0.0339 – 0.0593 in samples representing 'controls, and 0.0266 – 0.0486 between. A possible reason for this comparatively higher variance in pair-wise kinships among actual case and control cohorts is the close relationships between a small number of dogs therein (full-siblings and parent/progeny).

Appendix v.iv. Mean kinship coefficients among cases, controls, and random samples of the Kennel Club registered population

Mean kinship	**Actual cases**	**Actual controls**	**Random sample ('cases')**	**Random sample ('controls')**
Actual cases	0.0490	-	-	-
Actual controls	0.0464	0.0579	-	-
Random sample ('cases')	0.0404	-	0.0400	-
Random sample ('controls')	-	0.0569	0.0429	0.0548

Mean kinship coefficients among and between dogs within actual case and control cohorts, and the mean of 1,000 average kinship coefficients among and between samples of n=21 ('cases') or n=25 ('control') randomly selected dogs, and between these samples and actual case and control cohorts.

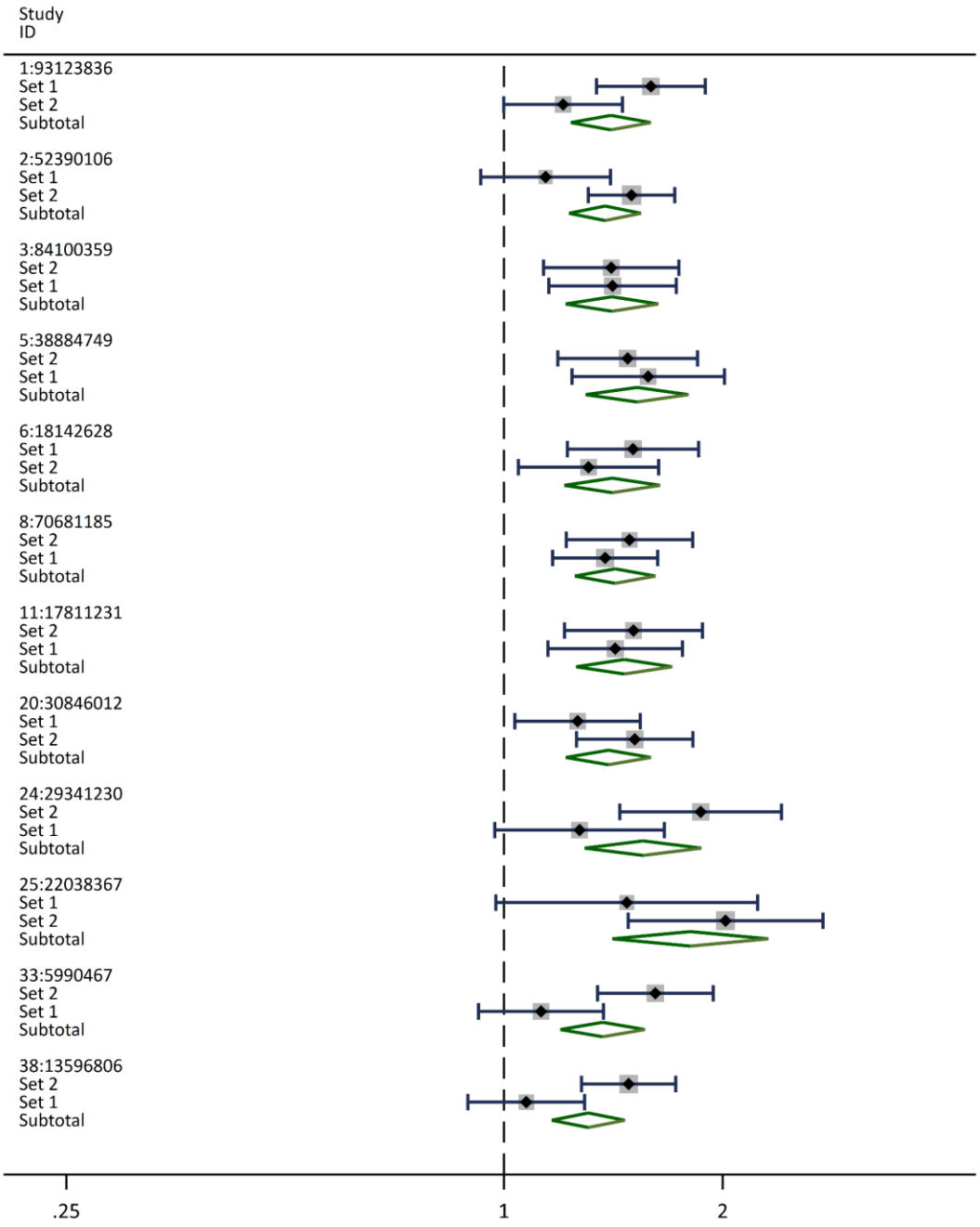


Appendix v.v. MDS plot of the three Italian Spinone IE study sets.

Appendix v.vi. Allelic discrimination assay primer and reporter probe sequences used to genotype the 10 SNPs identified from the GWAS meta-analysis of IE

Genomic position*	Forward Primer	Reverse Primer	Reporter 1 (VIC)	Reporter 2 (FAM)
1:93123836	AGCACAAAGGAACAAGCTAAACCT	TGCAATGCTAGTGACACTGTGT	CAGAGGCTTTAGACGTGCT	CAGAGGCTTTAAACGTGCT
2:52390106	CTCACCTTACCCATTTGCTTTTGTG	GGGAGCCATAGTTAAATGAAATGTGATCT	CAAACTTATCCTGTTCTTTGAGAT	AACTTATCCTGTTCCTTGAGAT
3:84100359	TGTAGTACTGTAGCAGATGTATCATGAGTTATAA	CCAAGGCTGGGAGATAAATACCAAT	CTGGGAACAGTCAGTTTA	TGGGAACAGTCACTTTA
5:38884749	GCTGCGCGTTCACTTTGG	TCCAGGCTGATTCATCATTGTTACC	CCAAATTTACAGATAACAATAT	AAATTTACAGATGACAATAT
6:18142628	GCTCAGCAATGCCAGAGACATAAA	ACTAGCAGCAGATAAGCTCTGTCT	TTAGATTACACGACAGATTT	ATTACACGGCAGATTT
8:70681185	TCCATAGAGTTAGGACCCCTTGTG	CAGCTTCTGTCCATATGTGAAGTCAA	TGCTTTTCTACAAGCGTTTG	TGCTTTTCTACAATCGTTTG
11:17811231	TGATACATAAAGTAAGCAAGGGAGATCCA	ACCACACCCTTTTTCATTGCAAAAT	CAAAGTGTGTCTTTTAACCAG	AAGTGTGTCTTCTAACCAG
20:30846012	CGATTGTGTCCAAAGAAGGAATGTC	GACCAAGATCTGTCCTTGAAAATGTC	CAACTTATGGTTTGCTTTAA	ACTTATGGTTCGCTTTAA
24:29341230	AATTGGGTGTCATTTATTTATTTATTTTTGCTAAA	GTCTGGCTCCCTCTAGTATGGA	CATCAACGGTCTGACCC	TCAACGGCCTGACCC
25:22038367	CAAGGCACACTAAGCAGACCAT	GGATCCCTCATAGTCCATAATAGAAAAGT	TAGCCTCACTTTATGGTTC	TAGCCTCACTTTTTGGTTC

*CanFam3.1 in the format: chromosome: base pair



Appendix v.vii. Forest plot showing the consistency of associations between study sets for the 12 most significantly associated SNPs from the meta-analysis. The forest plot was generated using Stata's 'metan' command. SNP IDs are the CanFam3.1 genomic position in the format chromosome: base pair. Black diamonds are the odds ratio point estimates for each study. Grey box size indicates study weighting. Whiskers indicate lower and upper 95% confidence intervals of the odds ratio. 'Subtotal' green diamonds represent the odds ratio point estimate (centre) and the lower and upper 95% confidence intervals of the odds ratio (left and right points respectively) for the meta-analysis.

Appendix v.viii. Results from analysis of the five-SNP genetic risk score SNPs individually and combined as a genetic risk score, and risk allele frequencies in cases and controls.

	Set 1		Set 2		GWAS		Set 3			
Genomic Pos. [a]	**ca/co** [b]	**Risk Allele Freq. (ca/co** [b]**)**	**ca/co** [b]	**Risk Allele Freq. (ca/co** [b]**)**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**ca/co** [b]	**Risk Allele Freq. (ca/co** [b]**)**
1:93123836	29/29	0.84/0.47	22/22	0.68/0.43	4.61 (2.28 - 9.32)	1.5×10^{-6}	1.45 (0.62 - 3.42)	0.39	23/23	0.59/0.50
2:52390106	29/29	0.50/0.34	22/22	0.70/0.16	3.38 (1.86 - 6.16)	1.3×10^{-5}	1.00 (0.45 - 2.24)	1.00	23/23	0.46/0.46
11:17811231	29/29	0.84/0.62	22/22	0.91/0.61	5.64 (2.47 - 12.86)	5.3×10^{-6}	1.59 (0.61 - 4.17)	0.34	23/23	0.78/0.70
20:30846012	29/29	0.69/0.50	22/22	0.75/0.36	4.57 (2.13 - 9.80)	9.9×10^{-6}	1.27 (0.58 - 2.81)	0.55	23/23	0.52/0.46
25:22038367	29/29	0.10/0.05	22/22	0.20/0.00	6.72 (1.80 - 25.05)	1.2×10^{-3}	0.45 (0.07 - 2.76)	0.38	23/23	0.04/0.09
Five-SNP genetic risk score	29/29	-	22/22	-	5.69 (2.97 - 10.90)	5.2×10^{-17}	1.23 (0.75 - 2.00)	0.40	23/23	-

[a] CanFam 3.1 genomic location of the SNP in the format chromosome: bp position. Genomic positions shown in bold indicate the most significantly associated SNPs in the GWAS meta-analysis. [b] case / control. Only individuals with genotype data for all five SNPs were included in the analysis.

Appendix v.ix. Results of the meta-analysis with and without inclusion of a subset of 18 cases and 18 controls of the replication subset (Set 3), and association analysis statistics of the Set 3 subset alone.

	GWAS Sets 1 and 2			**GWAS Sets 1, 2, and Replication Set 3**			**Set 3 only**	
Genomic pos. *	**Odds ratio (95% CI)**	**P-value**	**P-value for het.**	**Odds ratio (95% CI)**	**P-value**	**P-value for het.**	**Odds ratio (95% CI)**	**P-value**
1:93092920	1.34 (1.18-1.52)	7.2×10^{-6}	0.08	1.29 (1.15-1.44)	1.9×10^{-5}	0.08	1.07 (0.82-1.41)	0.46
1:93123836	**1.40 (1.24-1.59)**	**1.8×10^{-7}**	**0.03**	**1.36 (1.21-1.52)**	**1.4×10^{-7} ****	**0.06**	**1.19 (0.92-1.54)**	**0.18**
1:93134637	1.34 (1.18-1.53)	6.6×10^{-6}	0.09	1.29 (1.15-1.45)	1.2×10^{-5}	0.10	1.11 (0.85-1.44)	0.33
1:93143512	1.35 (1.19-1.53)	4.6×10^{-6}	0.10	1.30 (1.16-1.45)	8.5×10^{-6}	0.11	1.11 (0.85-1.44)	0.33
1:93151289	1.34 (1.19-1.52)	3.1×10^{-6}	0.10	1.30 (1.16-1.45)	5.7×10^{-6}	0.11	1.11 (0.85-1.44)	0.33
1:93166693	1.34 (1.19-1.52)	3.1×10^{-6}	0.10	1.30 (1.16-1.45)	5.7×10^{-6}	0.11	1.11 (0.85-1.44)	0.33
1:93168992	1.34 (1.19-1.52)	3.1×10^{-6}	0.10	1.30 (1.16-1.45)	5.5×10^{-6}	0.11	1.11 (0.86-1.44)	0.32
1:93177398	1.34 (1.19-1.52)	3.1×10^{-6}	0.10	1.30 (1.16-1.45)	5.7×10^{-6}	0.11	1.11 (0.85-1.44)	0.33
1:93179652	1.34 (1.19-1.52)	3.1×10^{-6}	0.10	1.30 (1.16-1.45)	5.7×10^{-6}	0.11	1.11 (0.85-1.44)	0.33
1:93250465	1.37 (1.20-1.56)	3.2×10^{-6}	0.22	1.31 (1.16-1.47)	6.9×10^{-6}	0.17	1.11 (0.85-1.44)	0.43
1:93262646	1.38 (1.21-1.57)	1.7×10^{-6}	0.26	1.32(1.17-1.48)	3.7×10^{-6}	0.18	1.11 (0.85-1.44)	0.43
1:93269238	1.38 (1.21-1.57)	1.7×10^{-6}	0.26	1.32(1.17-1.48)	3.7×10^{-6}	0.18	1.11 (0.85-1.44)	0.43
1:93395919	1.39 (1.22-1.58)	7.1×10^{-7}	0.20	1.33 (1.19-1.49)	1.2×10^{-6}	0.16	1.13 (0.88-1.46)	0.31
2:45972651	1.40 (1.20-1.62)	9.1×10^{-6}	0.24	1.25 (1.11-1.41)	2.0×10^{-4}	0.03	1.02 (0.83-1.25)	0.94
2:52390106	**1.38 (1.23-1.54)**	**3.6×10^{-8}**	**0.03**	**1.30 (1.17-1.44)**	**9.6×10^{-7}**	**4.0×10^{-3}**	**1.03 (0.81-1.32)**	**1.00**

	GWAS Sets 1 and 2			GWAS Sets 1, 2, and Replication Set 3			Set 3 only	
Genomic pos. *	**Odds ratio (95% CI)**	**P-value**	**P-value for het.**	**Odds ratio (95% CI)**	**P-value**	**P-value for het.**	**Odds ratio (95% CI)**	**P-value**
2:52396456	1.36 (1.21-1.53)	1.6×10^{-7}	0.04	1.28 (1.15-1.42)	3.4×10^{-6}	5.9×10^{-3}	1.03 (0.81-1.32)	1.00
2:54074065	1.49 (1.25-1.77)	8.4×10^{-6}	0.09	1.48 (1.26-1.74)	2.9×10^{-6} **	0.24	1.42 (0.88-2.27)	0.21
2:56417302	1.33 (1.17-1.51)	9.2×10^{-6}	0.19	1.21 (1.09-1.35)	5.9×10^{-4}	5.7×10^{-3}	1.11 (0.88-1.38)	0.46
2:56423516	1.32 (1.17-1.50)	9.2×10^{-6}	0.21	1.21 (1.09-1.35)	5.7×10^{-4}	6.2×10^{-3}	1.11 (0.88-1.38)	0.46
2:73837742	1.41 (1.21-1.64)	9.7×10^{-6}	0.22	1.39 (1.21-1.58)	1.3×10^{-6} **	0.43	1.32 (1.01-1.72)	0.15
3:84100359	**1.41 (1.22-1.63)**	**5.0×10^{-6}**	**0.98**	**1.39 (1.22-1.59)**	**1.2×10^{-6} ****	**0.94**	**1.32 (0.96-1.83)**	**0.10**
5:38871847	1.54 (1.29-1.85)	2.6×10^{-6}	0.55	1.51 (1.28-1.77)	5.2×10^{-7} **	0.73	1.39 (0.98-1.96)	0.12
5:38878225	1.54 (1.29-1.85)	2.6×10^{-6}	0.55	1.51 (1.28-1.77)	5.2×10^{-7} **	0.73	1.39 (0.98-1.96)	0.12
5:38884749	**1.52 (1.29-1.79)**	**4.1×10^{-7}**	**0.70**	**1.46 (1.26-1.68)**	**2.3×10^{-7} ****	**0.51**	**1.26 (0.94-1.70)**	**0.13**
5:38900229	1.51 (1.28-1.77)	5.5×10^{-7}	0.88	1.45 (1.26-1.67)	2.8×10^{-7} **	0.58	1.26 (0.94-1.70)	0.13
5:57150319	1.43 (1.24-1.66)	2.1×10^{-6}	0.02	1.29 (1.13-1.47)	1.4×10^{-4}	7.7×10^{-4}	1.13 (0.85-1.50)	0.22
6:18142628	**1.41 (1.21-1.64)**	**9.4×10^{-6}**	**0.36**	**1.35 (1.18-1.54)**	**7.7×10^{-6} ****	**0.34**	**1.18 (0.92-1.53)**	**0.20**
8:70681185	**1.42 (1.25-1.62)**	**6.9×10^{-8}**	**0.56**	**1.36 (1.21-1.52)**	**1.6×10^{-7}**	**0.23**	**1.13 (0.88-1.45)**	**0.39**
8:70685342	1.43 (1.25-1.62)	8.2×10^{-8}	0.55	1.36 (1.21-1.52)	2.0×10^{-7}	0.22	1.13 (0.88-1.45)	0.39
8:70913880	1.38 (1.22-1.56)	2.8×10^{-7}	0.74	1.31 (1.18-1.46)	8.1×10^{-7}	0.23	1.11 (0.88-1.39)	0.46
11:17811231	**1.46 (1.26-1.70)**	**9.8×10^{-7}**	**0.71**	**1.40 (1.23-1.60)**	**7.4×10^{-7} ****	**0.46**	**1.21 (0.92-1.59)**	**0.13**
20:18573917	1.34 (1.18-1.52)	7.3×10^{-6}	0.05	1.24 (1.11-1.39)	1.6×10^{-4}	7.3×10^{-3}	1.06 (0.83-1.37)	0.45

	GWAS Sets 1 and 2			GWAS Sets 1, 2, and Replication Set 3			Set 3 only	
Genomic pos. *	**Odds ratio (95% CI)**	**P-value**	**P-value for het.**	**Odds ratio (95% CI)**	**P-value**	**P-value for het.**	**Odds ratio (95% CI)**	**P-value**
20:18728616	1.34 (1.18-1.52)	5.4×10^{-6}	0.03	1.24 (1.11-1.39)	1.8×10^{-4}	2.8×10^{-3}	1.10 (0.85-1.42)	0.28
20:24952889	1.41 (1.22-1.62)	3.3×10^{-6}	0.02	1.37 (1.21-1.56)	9.5×10^{-7} **	0.06	1.26 (0.96-1.65)	0.20
20:26103063	1.45 (1.24-1.69)	3.5×10^{-6}	0.64	1.37 (1.19-1.58)	8.4×10^{-6}	0.31	1.12 (0.82-1.52)	0.88
20:28920317	1.39 (1.20-1.60)	9.95×10^{-6}	1.00	1.36 (1.19-1.55)	4.1×10^{-6} **	0.83	1.25 (0.92-1.69)	0.48
20:30846012	**1.39 (1.22-1.59)**	**1.6×10^{-6}**	**0.19**	**1.34 (1.19-1.50)**	**1.2×10^{-6} ****	**0.21**	**1.18 (0.94-1.49)**	**0.24**
24:29341230	**1.55 (1.29-1.87)**	**3.2×10^{-6}**	**0.04**	**1.36 (1.16-1.59)**	**1.0×10^{-4}**	**4.7×10^{-3}**	**1.00 (0.76-1.32)**	**0.56**
25:2967994	1.55 (1.29-1.86)	3.3×10^{-6}	0.27	1.43 (1.22-1.69)	1.6×10^{-5}	0.11	1.07 (0.74-1.54)	0.54
25:2992154	1.58 (1.30-1.92)	3.7×10^{-6}	0.36	1.45 (1.22-1.72)	2.1×10^{-5}	0.11	1.07 (0.74-1.54)	0.54
25:3084236	1.55 (1.29-1.86)	3.3×10^{-6}	0.27	1.43 (1.22-1.69)	1.6×10^{-5}	0.11	1.07 (0.74-1.54)	0.54
25:3466059	1.55 (1.29-1.86)	3.3×10^{-6}	0.27	1.43 (1.22-1.69)	1.6×10^{-5}	0.11	1.07 (0.74-1.54)	0.54
25:17906526	1.50 (1.26-1.80)	6.8×10^{-6}	0.01	1.37 (1.16-1.61)	1.5×10^{-4}	1.5×10^{-3}	1.17 (0.79-1.75)	0.34
25:21307564	1.77 (1.38-2.26)	5.2×10^{-6}	0.09	1.48 (1.19-1.83)	3.6×10^{-4}	2.9×10^{-3}	1.22 (0.78-1.91)	0.32
25:22038367	**1.81 (1.41-2.31)**	**2.9×10^{-6}**	**0.24**	**1.50 (1.21-1.86)**	**2.5×10^{-4}**	**4.9×10^{-3}**	**1.22 (0.78-1.91)**	**0.32**
25:24411272	1.70 (1.35-2.14)	5.5×10^{-6}	0.05	1.46 (1.19-1.79)	2.7×10^{-4}	2.3×10^{-3}	1.22 (0.78-1.91)	0.32
25:24411325	1.70 (1.35-2.14)	5.5×10^{-6}	0.05	1.46 (1.19-1.79)	2.7×10^{-4}	2.3×10^{-3}	1.22 (0.78-1.91)	0.32
25:33935066	1.64 (1.32-2.02)	5.4×10^{-6}	1.4×10^{-3}	1.46 (1.21-1.76)	6.4×10^{-5}	5.6×10^{-4}	1.01 (0.69-1.48)	0.59

* CanFam 3.1 genomic location of the SNP in the format chromosome: bp position. The SNPs are ordered by chromosome and GWAS P-value. Bold text indicates the most significantly associated SNPs in the GWAS meta-analysis. ** P-values lower after the addition of Set 3 to the meta-analysis.

vi. Manuscript 5 appendices

Appendix vi.i. Countries of origin and case and control definitions for dogs included in seven sample sets.

	GWAS			Replication			
	UH GWAS	**KCGC GWAS 1**	**KCGC GWAS 2**	**KCGC**	**UM**	**RVC**	**UU**
Countries of origin	Germany (69.1%), Switzerland (27.3%), Netherlands	UK (KCGC GWAS 1: 96.9%, KCGC GWAS 2: 89.2%), Australia, Czech Republic, Germany, Channel Islands, Netherlands, Norway, and USA		UK (60.9%); Italy and USA; Australia, Canada, Channel Islands; and nine other European countries (Belgium, France, Germany, Luxembourg, Netherlands, Portugal, Sweden, and Switzerland).	UK	UK	Predominantly from the Netherlands, but also dogs from neighbouring countries including Germany and Belgium
Case definition	Diagnosis confirmed by a board-certified veterinary neurologist	Diagnosis in the Animal Health Trust Centre for Small Animal Studies neurology unit, Newmarket, UK; and Linnaeus referral veterinary hospitals in the UK; or assessment of owner reported questionnaires, veterinary records, and epileptic seizure video footage where available, by LDR (co-author). The diagnosis of IE was based on IVETF criteria.			Dogs with epilepsy undergoing treatment and routine screening of serum phenobarbital concentrations	Big Brainy Border Collie study (BBBCS) and the Idiopathic Epilepsy and Anxiety Study (IDEAS). Cases met, as a minimum, the IVETF Tier I criteria for IE diagnosis.	IE-affected as part of a study characterising the phenotype of IE in the Border Collie (Santifort et al., 2022). A minimum of a IVETF Tier I diagnosis.
Control definition	Not reported to have ever had a seizure (55.9% over the age of seven years)	Over the age of eight years not reported to have ever had a seizure.			-	BBBCS; owner reported free of epilepsy. Ages ranging from one to nine years (34.5% over the age of six years).	Owner reported free from epilepsy over six years of age. 80% were over nine years of age.

UH: University of Helsinki. KCGC: Kennel Club Genetics Centre. UM: University of Manchester. RVC: Royal Veterinary College. UU: University of Utrecht. IVETF: International Veterinary Epilepsy Task Force

Appendix vi.ii. The number of individuals for each of the 97 breeds, and mixed breeds, included in the dataset of 219 in-house WGS.

Breed	Count
Affenpinscher	1
Airedale Terrier	2
Alaskan Malamute	1
American Cocker Spaniel	1
Australian Shepherd	1
Basset Hound	19
Bassett Griffon Vendeen	1
Beagle	3
Bearded Collie	2
Bedlington terrier	1
Berger Picard	1
Bloodhound	1
Border Collie	7
Border Terrier	10
Boxer	1
Briard	2
Bull Terrier	1
Bulldog	3
Cairn Terrier	1
Cavalier King Charles Spaniel	3
Cesky Terrier	1
Chesapeake Bay Retriever	1
Chihuahua	1
Chinese Crested	2
Chow Chow	1
Cocker Spaniel	4
Corgi	1
Dalmatian	1
Dandie Dinmont	4
Dobermann	2
English Setter	1
English Springer Spaniel	2
Field Spaniel	2
Finnish Lapphund	1
Flat Coated Retriever	4
French Bull Dog	5
German Shepherd	1
German Wirehaired	1
Giant Schnauzer	5
Glen of Imaal Terrier	1
Golden Retriever	2
Gordon Setter	1
Great Dane	1
Greyhound	2

Breed	Count
Griffon bruxellois	1
Irish Red and White Setter	4
Irish Setter	2
Irish Terrier	1
Irish Water Spaniel	1
Irish Wolfhound	6
Italian Spinone	4
Japanese Akita	1
Keeshond	3
Labrador Retriever	6
Lagotto Romagnolo	1
Lakeland Terrier	1
Lancashire Heeler	2
Large Munsterlander	1
Leonberger	2
Lhasa Apso	1
Maltese	1
Miniature Long-haired Dachshund	2
Miniature Poodle	1
Miniature Schnauzer	2
Miniature Wire Haired Dachshund	1
Mixed Breed	5
Newfoundland	1
Northern Inuit	2
Norwegian Buhund	3
Norwich Terrier	1
Nova Scotia Duck Tolling Retriever	1
Old English Sheepdog	1
Otterhound	2
Papillon	1
Petit Basset Griffon Vendeen	2
Pug	3
Pyrenean Mountain dog	1
Rottweiler	1
Rough Collie	1
Russian Black Terrier	1
Scottish Terrier	5
Shar Pei	3
Shetland Sheepdog	1
Shih Tzu	1
Siberian Husky	3
Skye Terrier	2
Smooth Collie	1
Soft-coated Wheaten Terrier	2
Staffordshire Bull Terrier	1
Swedish Vallhund	1
Tibetan Spaniel	1

Breed	Count
Tibetan Terrier	3
Vizsla (smooth coat)	4
Vizsla (wire haired)	1
Weimaraner	1
Welsh Springer Spaniel	7
West Highland White Terrier	1
Whippet	2

Appendix vi.iii. The numbers of dogs and SNPs included in the WGS reference panels used to impute three study sets

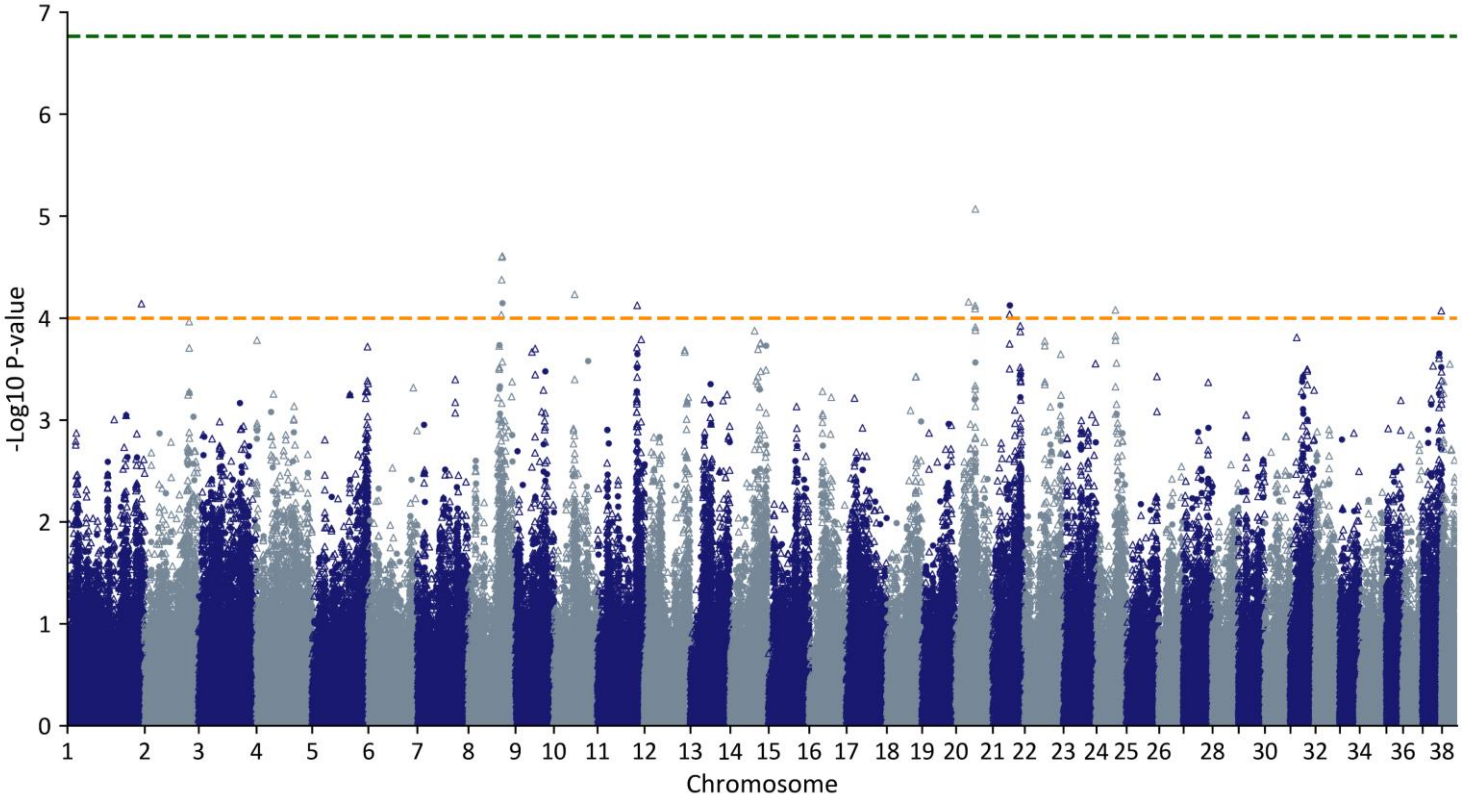
Study Set:	**KCGC GWAS 1 and 2**	**UH GWAS**
Total SNPs	12,110,068	12,114,126
Pool of individuals (n)	818	817
Minimum individuals (n)	786	785
Maximum individuals (n)	817	816
Border Collies (n)	40	39

Appendix vi.iv. Mean IMPUTE2 estimated concordance across chromosomes, and concordance estimates for the chromosomes with the highest and lowest concordance, for WGS-density imputed datasets.

	KCGC GWAS set 1	**KCGC GWAS set 2**	**UH GWAS**
Mean estimated concordance (%)	96.5	96.6	96.6
Lowest concordance (%)	95.0	95.3	95.0
Highest concordance (%)	97.6	97.4	97.8

Appendix vi.v. The number of cases, controls, and SNPs in each study set after quality control filtering

	Study set	Cases (n)	Controls (n)	SNPs (n)
Imputed to array SNP density	KCGC GWAS set 1	57	73	310,596
	KCGC GWAS set 2	26	60	310,282
	UH GWAS	21	34	307,561
Imputed to WGS SNP density	KCGC GWAS set 1	73	89	6,498,131
	KCGC GWAS set 2	29	63	6,531,278
	UH GWAS	21	34	6,478,397



Appendix vi.vi. Array SNP density imputed genome-wide association study meta-analysis of 104 Border Collie idiopathic epilepsy cases and 167 controls with 16 SNPs as covariates (291,431 SNPs). Plot of negative log (base 10) transformed P-values. X-axis is SNP location by chromosome (left to right, autosomes 1 to 38). Solid circles indicate array genotyped SNPs, hollow triangles denote SNPs that were imputed for any of the three datasets. Green (upper) line shows Bonferroni corrected threshold for statistical significance (1.7×10^{-7}). Orange (lower) line indicates the empirical threshold for suggestive association (1×10^{-4}).

Appendix vi.vii. The SNPs identified in the GWAS meta-analyses, at array (with and without 16 SNPs as covariates) and WGS SNP level.

	Genomic Pos. [a]	Odds ratio (95% CI)	P-value	Heterogeneity P-value	Alleles (risk/non-risk)	Nearest known gene [b]
Top SNPs from	1:87042337	1.21 (1.12 - 1.31)	1.2×10^{-6}	0.30	C/T	*TRPM3*
array-density	5:15989786	1.20 (1.10 - 1.30)	2.2×10^{-5}	0.78	G/A	*DSCAML1*
analysis	8:41418834	1.18 (1.09 - 1.28)	2.7×10^{-5}	0.14	T/C	*TMEM229B* / *PLEK2*
	9:51228405	1.32 (1.16 - 1.52)	5.1×10^{-5}	0.23	A/T	*MRPS2*
	11:68785445	1.19 (1.09 - 1.30)	9.6×10^{-5}	0.62	G/A	*C9orf91* (*TMEM268*)
	12:60209578	1.29 (1.14 - 1.46)	7.7×10^{-5}	0.15	T/C	*GRIK2*
	14:28706441	1.28 (1.14 - 1.44)	4.3×10^{-5}	0.24	T/A	*ETV1*
	16:55920429	1.21 (1.11 - 1.32)	2.0×10^{-5}	0.05	G/T	*CSMD1*
	18:55320649	1.19 (1.10 - 1.30)	1.8×10^{-5}	0.29	T/C	*CD6*
	22:43948947	1.18 (1.09 - 1.28)	5.6×10^{-5}	0.24	A/G	*GPC5*
	25:19226599	1.18 (1.09 - 1.27)	6.0×10^{-5}	0.32	G/A	*PALLD*
	27:32560742	1.18 (1.09 - 1.28)	8.3×10^{-5}	0.87	G/A	*NMDE2_CANFA* (*GRIN2B* in humans)
	28:40887640	1.28 (1.15 - 1.43)	1.3×10^{-5}	0.67	C/T	*ZNF511*
	35:23042846	1.22 (1.11 - 1.33)	2.6×10^{-5}	0.26	C/T	*FAM65B*
	36:20271612	1.21 (1.12 - 1.30)	8.1×10^{-7}	0.01	A/C	*MTX2*
	37:8098533	1.30 (1.14 - 1.48)	7.6×10^{-5}	0.08	C/T	*SATB2*

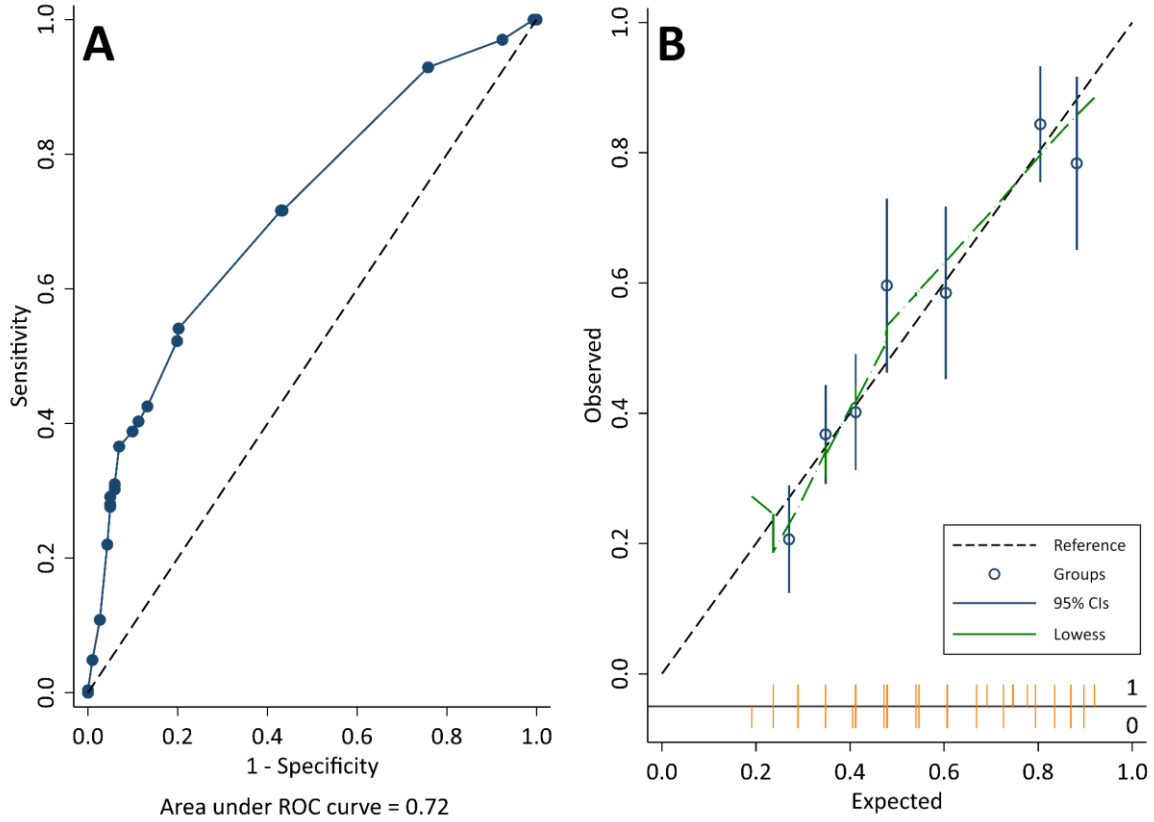
	Genomic Pos. [a]	Odds ratio (95% CI)	P-value	Heterogeneity P-value	Alleles (risk/non-risk)	Nearest known gene [b]
Array SNPs	1:117532714	1.13 (1.06 - 1.20)	7.2×10^{-5}	0.20	C/T	*GRAMD1A*
identified by	8:53942648	1.19 (1.10 - 1.28)	2.5×10^{-5}	0.31	T/A	*SEL1L*
conditional	11:62873725	1.15 (1.07 - 1.23)	7.5×10^{-5}	0.74	T/C	*KLF4*
covariate analysis	37:30026463	1.18 (1.09 - 1.28)	8.5×10^{-5}	0.53	T/C	*ERICH1*
Top SNPs from	1:89337587	1.19 (1.11 - 1.28)	1.6×10^{-6}	0.43	T/C	*DOCK8*
WGS SNP density	6:77085450	1.37 (1.19 - 1.57)	8.0×10^{-6}	0.007	T/G	WLS
analysis	11:68783067	1.20 (1.11 - 1.30)	4.1×10^{-6}	0.79	C/T	*C9orf91* (*TMEM268*)
	12:58476156	1.26 (1.14 - 1.40)	9.5×10^{-6}	0.29	G/T	*SIM1*
	16:54266694	1.29 (1.16 - 1.42)	7.9×10^{-7}	0.06	C/T	*ARHGEF10*
	17:45586797	1.20 (1.11 - 1.29)	7.8×10^{-6}	0.50	A/G	*LRRTM4*
	18:55240220	1.21 (1.12 - 1.30)	4.5×10^{-7}	0.16	G/A	*CD5*
	35:23052056	1.24 (1.14 - 1.35)	9.6×10^{-7}	0.02	A/C	*FAM65B*
	36:26774691	1.21 (1.11 - 1.31)	6.3×10^{-6}	0.52	G/A	*STK19*
	4:28189911*	1.26 (1.15 - 1.38)	7.7×10^{-7}	0.54	C/T	*RPS24*
	4:28195881*	1.26 (1.15 - 1.38)	8.3×10^{-7}	0.61	G/T	*RPS24*
	5:17717963	1.29 (1.17 - 1.43)	3.9×10^{-7}	0.54	G/A	*CADM1*
	8:41418371	1.20 (1.11 - 1.29)	1.3×10^{-6}	0.19	G/C	*TMEM229B* / *PLEK2*

SNPs shown are those that were most significantly associated on each chromosome in each analysis that passed the thresholds for suggestive association (array density analyses: $P < 1 \times 10^{-4}$. WGS SNP density analysis: $P < 1 \times 10^{-5}$). * CanFam3.1 4:28195881 was in linkage disequilibrium (R2 = 0.93) with 4:28189911 and was genotyped as a proxy. [a] The CanFam3.1 genomic location of the SNP in the format chromosome: bp position. [b] Nearest known gene annotated on CanFam3.1. CI: confidence interval.

Appendix vi.viii. Results of the Border Collie IE replication study of the 27 SNPs identified through the GWAS meta-analyses.

	Three GWAS sets			**Replication sets**		
Genomic pos. [a]	ca/co [b] (n)	Odds Ratio (95% CI)	P-value	ca/co [b] (n)	Odds Ratio (95% CI)	P-value
1:87042337	96/155	2.18 (1.53-3.11)	1.2×10^{-5}	270/306	1.23 (0.95-1.58)	0.12
1:89337587	121/186	1.96 (1.42-2.69)	3.3×10^{-5}	269/302	1.12 (0.86-1.46)	0.39
1:117532714	97/161	1.89 (1.34-2.67)	2.4×10^{-4}	269/299	0.86 (0.67-1.09)	0.21
4:28195881	122/183	2.63 (1.75-3.95)	1.0×10^{-6}	271/303	0.91 (0.67-1.23)	0.54
5:15989786	103/167	2.09 (1.44-3.04)	8.9×10^{-5}	270/303	0.90 (0.69-1.18)	0.45
5:17717963	114/183	2.59 (1.62-4.13)	6.4×10^{-5}	261/288	1.26 (0.92-1.74)	0.15
8:41418834	100/164	1.99 (1.40-2.84)	9.4×10^{-5}	269/307	0.93 (0.72-1.22)	0.61
8:53942648	102/160	1.76 (1.12-2.78)	0.01	269/306	0.92 (0.68-1.25)	0.59
9:51228405	93/156	3.42 (1.79-6.53)	2.2×10^{-4}	267/301	1.33 (0.96-1.85)	0.09
11:62873725*	102/160	1.74 (1.19-2.54)	3.7×10^{-3}	268/304	1.25 (0.97-1.61)	0.09
11:68783067	94/132	1.79 (1.23-2.62)	2.1×10^{-3}	270/306	1.28 (0.99-1.66)	0.05
11:68785445*	104/164	1.86 (1.28-2.71)	9.1×10^{-4}	271/306	1.29 (1.00-1.67)	0.05
12:58476156	118/181	2.61 (1.64-4.15)	5.7×10^{-5}	268/306	1.09 (0.78-1.52)	0.60
12:60209578	101/165	2.82 (1.57-5.08)	5.8×10^{-4}	268/299	1.27 (0.82-1.96)	0.28
14:28706441	104/167	2.68 (1.55-4.62)	4.1×10^{-4}	269/306	0.72 (0.50-1.03)	0.07
16:55920429*	104/167	1.97 (1.32-2.94)	8.1×10^{-4}	271/306	1.39 (1.04-1.84)	0.02
18:55240220	117/174	2.16 (1.55-3.03)	4.5×10^{-6}	270/303	1.20 (0.94-1.53)	0.15
18:55320649	94/155	2.07 (1.42-3.02)	1.1×10^{-4}	269/307	1.11 (0.87-1.42)	0.40
22:43948947	104/167	2.15 (1.49-3.10)	2.5×10^{-5}	269/305	1.07 (0.83-1.38)	0.60
25:19226599	104/166	1.98 (1.39-2.83)	1.4×10^{-4}	260/290	1.08 (0.83-1.40)	0.56
28:40887640	98/148	3.21 (1.90-5.43)	1.9×10^{-6}	268/297	1.42 (1.01-2.01)	0.04
35:23042846	95/153	2.26 (1.48-3.45)	1.7×10^{-4}	269/306	1.20 (0.87-1.64)	0.27
35:23052056	106/162	2.42 (1.62-3.62)	1.6×10^{-5}	268/302	1.02 (0.76-1.36)	0.91
36:20271612	99/159	2.34 (1.65-3.32)	7.6×10^{-7}	271/304	1.01 (0.78-1.29)	0.96
36:26774691	120/175	2.05 (1.42-2.95)	8.1×10^{-5}	265/304	0.99 (0.75-1.30)	0.93
37:8098533	102/163	2.66 (1.51-4.69)	7.0×10^{-4}	270/306	1.06 (0.72-1.58)	0.76
37:30026463	97/153	1.44 (0.89-2.33)	0.13	262/282	1.01 (0.73-1.41)	0.95

Results from logistic regression analysis of the three GWAS sets and the replication sets. [a] The CanFam3.1 genomic location of the SNP in the format chromosome: bp position. [b] case/control. * SNPs taken forward to be tested as a weighted risk score. CI: confidence interval.



Appendix vi.ix. Receiver operating characteristic (ROC) curve and calibration plot for an unweighted three-SNP risk score for idiopathic epilepsy based on risk allele counts not weighted by effect in Border Collie replication sets. Plot A is a ROC curve; points represent each potential risk score cut off for defining cases, from the highest (0,0) to the lowest (1,1). Sensitivity: fraction of cases correctly classified. Specificity: fraction of controls correctly classified (1 – (minus) specificity is the false-positive fraction). The area under the ROC curve (AUC) is given below the plot. An AUC of 0.5 (indicated by the dashed line) would represent a test unable to discriminate cases from controls. On the calibration plot (B) points represent ten equally sized groups of individuals divided by predicted risk. Observed: the proportion of cases in each group. Expected: the average (mean) of the predicted probabilities generated from the risk score logistic regression model. The 95% confidence intervals are shown for each group. The dashed reference line indicates perfect risk score calibration where predicted risk matches the observed proportion of affected dogs within each group. Locally weighted scatterplot smoothing (LOWESS) is displayed in green. The orange lines at the base of the graph are a spike plot indicating the distribution of IE cases (1) and controls (0).