Doctoral Programme in Integrative Life Science
Department of Medical and Clinical Genetics, Faculty of Medicine
Department of Veterinary Biosciences, Faculty of Veterinary Medicine
University of Helsinki
Folkhälsan Research Center

THE GENETIC BACKGROUND OF FIVE CANINE MODELS OF RARE HUMAN DISEASE

Julia Niskanen

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Medicine of the University of Helsinki, in lecture room PII, Porthania, Yliopistonkatu 3 on the 26th of August, 2022 at 12 o'clock.

Helsinki 2022

Supervisors

Professor Hannes Lohi, PhD
University of Helsinki
Folkhälsan Research Center
Helsinki, Finland

Docent Marjo Hytönen, PhD
University of Helsinki
Folkhälsan Research Center
Helsinki, Finland

Docent Maria Wiberg, DVM, PhD
University of Helsinki
Helsinki, Finland

Reviewers

Docent Sami Heikkinen, PhD
University of Eastern Finland
Kuopio, Finland

Docent Minna Ruddock, PhD
University of Oulu
Oulu, Finland

Opponent

Professor Johanna Schleutker, PhD
University of Turku
Turku, Finland

ISBN 978-951-51-8224-1 (softcover)
ISBN 978-951-51-8225-8 (PDF)

Unigrafia
Helsinki 2022

The Faculty of Medicine uses the Ouriginal system (plagiarism recognition) to examine all doctoral dissertations.

TIIVISTELMÄ

Tässä väitöskirjatyössä selvitettiin viiden koirilla esiintyvän perinnöllisen sairauden taustaa. Tutkimuksissa käytettiin genominlaajuista assosiaatioanalyysiä, seuraavan sukupolven sekvensointianalyysiä ja geneettisten varianttien validaatiota yhdessä yksityiskohtaisten kliinisten ja patologisten tutkimusten kanssa. Työn tuloksena kuvattiin uusia koirilla esiintyviä sairauksia ja tunnistettiin sairauksia aiheuttavia geenimuunnoksia ja niiden vaikutuksia koirien terveyteen ja ruumiinrakenteeseen.

Osatyössä I tunnistettiin seuraavan sukupolven sekvensointianalyysillä keskiaasiankoirien epidermolysis bullosa -ihosairauteen liittyvä ehdokasgeenivirhe *COL7A1*-geenissä. Geenivirheen yhteys sairauteen varmennettiin 190 koiran perimässä, ja sairauden alatyypiksi tarkennettiin resessiivinen dystrofinen epidermolysis bullosa. Immunohistokemialliset värjäykset ihonäytteissä paljastivat täyspitkän kollageeni VII -proteiinin puutoksen sairastuneen koiran ihossa.

Osatyössä II tunnistettiin genominlaajuista assosiaatioanalyysiä ja seuraavan sukupolven sekvensointianalyysiä käyttäen karjalankarhukoirien aivolisäkeperäiseen lyhytkasvuisuuteen liittyvä perimän alue ja ehdokasgeenivirhe. *POU1F1*-geenin silmukointialueella sijaitsevan geenivirheen yhteys sairauteen varmennettiin yli 8000 koiran perimässä. Laskennallinen ennuste viittasi geenivirheen sisältävän introni-eksoni -liitoksen silmukointialueen heikentymiseen.

Osatyössä III tunnistettiin homotsygotiakartoitusta ja seuraavan sukupolven sekvensointianalyysiä käyttäen rottweilerien synnynnäiseen kuurouteen liittyviä perimän ehdokasalueita ja -geenivirheitä. *LOXHD1*-geenissä sijaitsevan geenivirheen yhteys sairauteen varmennettiin yli 800 000 koiran perimässä. Varmennus paljasti myös geenivirheen esiintyvän lähes yksinomaan rottweiler-taustaisissa koirissa.

Osatyössä IV tarkennettiin varianttiseulontaa, kliinisiä tutkimuksia ja tilasto-analyysejä käyttäen aiemmin kuvatun *DVL2*-geenivirheen vaikutuksia koirien ruumiinrakenteeseen. Tulokset osoittivat geenivirheen yhteyden erilaisiin kaudaali-nikamien epämuodostumiin ja kuonon pituuden lyhenemiseen. Vaihteleva frekvenssi osoitti geenivirheeseen liittyvien terveysongelmien kuorman eroavan rotujen välillä.

Osatyössä V tunnistettiin kliinisiä ja patologisia tutkimuksia käyttäen pahanlaatuisia kammiorytmihäiriöitä ja äkillisiä sydänperäisiä kuolemia nuorissa leonberginkoirissa. Sairauden korkea esiintyvyys pentueittain ja suvuittain viittasi perinnölliseen taustaan, mutta genominlaajuista analyysiä käyttäen ei tunnistettu ehdokasalueita perimästä. Tämä johtui todennäköisesti geneettisestä tai fenotyyppisestä heterogeniasta, mikä pyritään tulevissa analyyseissä huomioimaan suuremmalla otoskoolla ja tarkemmilla kliinisillä ja molekyylitason tutkimuksilla.

Väitöstyön tuloksilla on useita tieteellisiä ja käytännön vaikutuksia. Löydökset edistävät diagnostiikkaa ja hoidon kehittämistä paljastamalla sairauksien taustalla olevat molekyylitason patofysiologiset mekanismit. Sairastuneet koirat voivat toimia mallieläiminä prekliinisissä tutkimuksissa, mikä hyödyttää sekä ihmis- että eläinlääketiedettä. Löydöksiin perustuvat geenitestit tukevat lisäksi koiranomistajia ja -kasvattajia jalostussuunnitelmien kehittämisessä terveempään suuntaan.

ABSTRACT

This thesis addresses the genetic background of five spontaneous canine models of rare human disease. By utilizing genome-wide mapping methods, next-generation sequencing analyses and variant validation combined with detailed clinical and post-mortem examinations, we characterized new canine models, identified novel disease-associated variants and dissected their effects on health and morphology.

In Study I, next-generation sequencing analysis in a Central Asian Shepherd dog affected by epidermolysis bullosa revealed a nonsense variant in *COL7A1*. Validation of the variant in 190 dogs confirmed the disease type as recessive dystrophic epidermolysis bullosa. Immunohistochemical stainings in skin samples illustrated the lack of full-length type VII collagen protein in the affected dog.

In Study II, genome-wide association analysis combined with next-generation sequence analysis identified a locus and candidate variant associated with recessive pituitary dwarfism in Karelian Bear Dogs. A splice site variant in *POU1F1* was confirmed with validation in over 8000 dogs. Computational predictions indicated weakening of the splice acceptor site at the affected intron-exon junction.

In Study III, homozygosity mapping combined with next-generation sequence analysis identified candidate regions and variants associated with recessive congenital hearing loss in Rottweilers. A missense variant in *LOXHD1* was confirmed with validation in over 800 000 dogs, revealing a link between the variant and Rottweiler breed background.

In Study IV, variant screening combined with clinical examinations and statistical analyses revealed novel morphological consequences of a previously described variant in *DVL2*. The findings showed that the variant is involved in variable caudal vertebral anomalies and contributes to a brachycephalic phenotype. Varying allele frequencies were identified across populations, indicating the differential impact of the variant on the genetic health of dog breeds.

In Study V, clinical and post-mortem examinations revealed malignant polymorphic ventricular arrhythmia and sudden cardiac death in young Leonberger dogs. The high prevalence of the disorder in litters and families strongly indicated a genetic aetiology. However, genome-wide association analyses failed to reveal associated loci, likely due to genetic or phenotypic heterogeneity. Future studies aim to overcome these obstacles with expanded cohorts and improved clinical and molecular phenotypes.

Our findings have multiple scientific and practical implications. The discoveries facilitate diagnostics and treatment by revealing the molecular and pathophysiological mechanisms of the disorders. Affected dogs also provide novel large animal models for preclinical studies, benefitting both human and veterinary medicine. Finally, the development of gene tests support dog owners and breeders in revising breeding programmes to improve the health of dog breeds.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications. In addition, unpublished data regarding Study V are presented.

I **Niskanen, J.**, Dillard, K., Arumilli, M., Salmela, E., Anttila, M., Lohi, H., & Hytönen, M. K. (2017). Nonsense variant in *COL7A1* causes recessive dystrophic epidermolysis bullosa in Central Asian Shepherd dogs. *PLOS ONE*, 12(5), e0177527. https://doi.org/10.1371/journal.pone.0177527

II Kyöstilä, K., **Niskanen, J. E.**, Arumilli, M., Donner, J., Hytönen, M. K., & Lohi, H. (2021). Intronic variant in *POU1F1* associated with canine pituitary dwarfism. *Human Genetics*. https://doi.org/10.1007/s00439-021-02259-2

III Hytönen, M. K., **Niskanen, J. E.**, Arumilli, M., Brookhart-Knox, C. A., Donner, J., & Lohi, H. (2021). Missense variant in *LOXHD1* is associated with canine nonsyndromic hearing loss. *Human Genetics*. https://doi.org/10.1007/s00439-021-02286-z

IV **Niskanen, J. E.**, Reunanen, V., Salonen, M., Bannasch, D., Lappalainen, A. K., Lohi, H., & Hytönen, M. K. (2021). Canine *DVL2* variant contributes to brachycephalic phenotype and caudal vertebral anomalies. *Human Genetics*. https://doi.org/10.1007/s00439-021-02261-8

V Wiberg, M., **Niskanen, J. E.**, Hytönen, M., Dillard, K., Hagner, K., Anttila, M., & Lohi, H. (2020). Ventricular arrhythmia and sudden cardiac death in young Leonbergers. *Journal of Veterinary Cardiology*, 27, 10–22. https://doi.org/10.1016/j.jvc.2019.11.006

The publications are referred to in the text by their Roman numerals.

AUTHOR'S CONTRIBUTION

The author contributed to each publication as follows:

I **Nonsense variant in *COL7A1* causes recessive dystrophic epidermolysis bullosa in Central Asian Shepherd dogs**

The author performed the next-generation sequencing analyses and variant screening, and drafted sections of the article text and figures.

II **Intronic variant in *POU1F1* associated with canine pituitary dwarfism**

The author contributed to the genome-wide association analyses and next-generation sequencing analyses and drafted sections of the article text and figures.

III **Missense variant in *LOXHD1* is associated with canine nonsyndromic hearing loss**

The author performed the homozygosity mapping, contributed to the next-generation sequencing analyses and drafted sections of the article text and all of the article figures.

IV **Canine *DVL2* variant contributes to brachycephalic phenotype and caudal vertebral anomalies**

The author participated in study design, recruited study participants, performed the variant screening and drafted most sections of the article text and figures.

V **Ventricular arrhythmia and sudden cardiac death in young Leonbergers**

The author recruited study participants, performed pedigree analyses and drafted sections of the article text and figures. Additionally, the author analysed the unpublished genetic data.

ABBREVIATIONS

AST	American Staffordshire Terrier
BAER	brainstem auditory evoked response
BOAS	brachycephalic obstructive airway syndrome
bp	base pairs
BT	Boston Terrier
CASD	Central Asian Shepherd Dog
ECG	electrocardiography
EDTA	ethylenediaminetetraacetic acid
DNA	deoxyribonucleic acid
EB	epidermolysis bullosa
EBD	English Bulldog
FBD	French Bulldog
FCI	Fédération Cynologique Internationale
FECD	Fuchs endothelial corneal dystrophy
GWAS	genome-wide association study
HWE	Hardy-Weinberg equilibrium
indel	small insertion or deletion
kb	kilobase
KBD	Karelian Bear Dog
LD	linkage disequilibrium
LEO	Leonberger
MAF	minor allele frequency
Mb	megabase
MDS	multidimensional scaling
MEI	mobile element insertion
MGI	Mouse Genome Informatics
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
QC	quality control
QQ	quantile-quantile
RNA	ribonucleic acid
ROH	runs of homozygosity
RW	Rottweiler
SBT	Staffordshire Bull Terrier
SCD	sudden cardiac death
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SV	structural variant
VA	ventricular arrhythmia
VPC	ventricular premature complex
VT	ventricular tachycardia
WES	whole-exome sequencing
WGS	whole-genome sequencing

1 REVIEW OF THE LITERATURE

1.1 Population genetics and genome of the domestic dog

1.1.1 Population history and structure

Dog (*Canis lupus familiaris*) is the first domesticated animal species and human's oldest companion.[1] While the fine details of where, when and how many times dogs were domesticated from wolves are still disputed, it is generally accepted that the event took place at latest around 16 000 years ago.[2] Used for several millennia in tasks aiding survival, such as hunting, herding and guarding, dogs have since been systematically bred to fulfil various nuanced purposes. There are presently over 350 breeds recognized by the worldwide canine organization, Fédération Cynologique Internationale (FCI).[3] Breed standards are typically focused on physical aspects, such as body proportions, coat type and colour, and shape of ears and tail, as well as behavioural traits, including the propensity for utilitarian and recreational activities like retrieving, tracking and obedience.

To achieve the desired, very specific physical and behavioural goals in each breed, dog breeders have carried out strict inbreeding. Modern dog breeding, often emphasizing appearance and pedigrees over function, is believed to have originated in Victorian England during the 19th century as a consequence of socioeconomic factors.[4] Thus, the current breeds have been established in less than 200 years. With very few exceptions, purebred dog populations are maintained so that only the offspring of same-breed parents are eligible for registration into that breed. This has effectively created a "breed barrier", an artificial interruption of gene flow resulting in closed gene pools.[5] The number of dogs in Europe further declined during the conflicts of the first half of 20th century, leading to very limited population sizes in many breeds.[1]

The bottlenecks within breeds are further escalated by the popular sire effect, where a small subpopulation of male dogs is used to contribute to a disproportionately large number of puppies. In a study covering dogs registered between 1980-2014 from all 215 breeds recognized by the UK Kennel Club, Lewis et al.[6] showed that every breed was affected by popular sires, with highest rates of inbreeding at 1980-1990. White Swiss Shepherd Dogs, a white-coated breed founded from German Shepherd Dogs, can be mentioned as an extreme example: in the breed standard, it is stated that the current population is largely based on the offspring of a single male dog.[7] Finally, gene flow is further restricted in many breeds due to geographical isolation or breeding strategies, such as in Labrador Retrievers, where the population is stratified by both country of origin (Europe/US) and function (working line/show line).[8]

1.1.2 Genomic architecture of modern breeds

The history of dogs has left distinct patterns in their genome. The two important bottlenecks, early domestication from wolves and recent breed creation, is reflected in the linkage disequilibrium (LD) structure: intrabreed blocks are up to several megabases long, but in contrast only extend over tens of kilobases between breeds.[9] Haplotype diversity is low in most breeds; the first study to measure LD in dogs in detail inspected intervals in five chromosomes and 156 breeds, and demonstrated that two to four haplotypes comprised 80 % of the haplotype variation in each breed.[10] Consequently, in the absence of other evidence, an initial hypothesis of a homogeneous and monogenic disease background in any given breed is often justified. Similarly, in disorders with suspected complex background, it is reasonable to assume that within a specific breed, the number of genes involved is smaller and their effect sizes larger than in the human disease orthologue. Furthermore, the extensive LD results in high numbers of SNPs tagging disease-causing variants, which supports mapping with arrays of relatively low density.[9]

Phylogenetic investigation of dog breeds has shown that they cluster into multi-breed clades representing dog types that existed before modern breeding started.[11] Typically, breeds that form a clade share long haplotypes largely only within that clade, excluding breeds that are newly created and therefore exhibit extensive recent admixture.[11] By capitalizing on the phenomenon that closely related breeds often share the haplotypes in which disease-causing variants reside, loci initially mapped in a single breed can be fine-mapped in related breeds.[12,13] Finally, cross-breed cohorts can also be utilized in variant analysis to exclude common, non-causal variation.[14]

1.1.3 Reference genome and annotations

The dog genome was first assembled in 2005 as an effort of the Dog Genome Sequencing Project led by the Broad Institute.[9] The initial assembly canFam1.0 and the updated high-quality draft canFam2.0 were based on a female Boxer, Tasha, and spanned 2.4 gigabases, covering 99 % of the genome at an average coverage of 7.5X.[9] The first compendium of SNPs was also introduced, containing over 2.5 million polymorphisms identified by comparing the Boxer against dogs from ten other breeds, wolves, and a coyote.[9] This set was complemented in 2013 by Axelsson et al.,[15] who detected over 3.6 million SNPs in wolves and dogs from 14 breeds. However, the presence of gaps, apparent assembly errors and sequencing errors highlighted the limited quality of canFam2.0 and emphasized the need for a more accurate reference genome; thus, an improved version, canFam3.1 was published in 2014.[16]

CanFam3.1, currently in wide use in the canine genetics community, introduced 85 megabases of novel reads and filled thousands of gaps present in canFam2.0. Furthermore, it provided new annotations for protein-coding genes, long non-coding RNAs and antisense transcripts through RNA sequencing of ten different tissues (Broad Improved Canine Annotation v1). The RNA annotations were again expanded in 2017 with the RNAseq-based FEELnc method, which yielded additional mRNA and long non-coding RNA models.[17] New polymorphisms were similarly catalogued, including sets of both germline and somatic variants identified from dogs with

lymphoma.[18] Additionally, computationally predicted gene models from the National Center for Biotechnology Information (NCBI)[19] and Ensembl[20] supplement the other annotations and are regularly maintained. Nonetheless, canFam3.1 is known to still contain gaps, and the reference has inherent limitations due to being based on a single Boxer dog. As a demonstration, pseudo-*de novo* assembly of unmapped reads from three breeds revealed several megabases of high-quality contigs with novel genes and thousands of variants, highlighting the need for more diverse high-quality canine reference genomes.[21]

New assemblies have started to emerge as successors to canFam3.1 in recent years (Table 1). They feature dogs from different breeds and both long-read and short-read sequencing technologies, providing various choices for reference genomes in future studies. Additionally, the Dog Genome Annotation project[22] is a new initiative that will provide new functional annotation, including enhancers and novel transcription start sites, for both current and emerging genomes and substantially facilitate the interpretation of variants.

Table 1. NCBI assembly identifiers, individual information and sequencing platforms of dog reference genomes generated after canFam3.1.

Assembly	**Breed**	**Sex**	**Name**	**Platform**
Dog10K_Boxer_Tasha[23]	Boxer	female	Tasha	PacBio Sequel
ROS_Cfam_1.0[24]	Labrador Retriever	male		PacBio Sequel, Illumina
UMICH_Zoey_3.1[25]	Great Dane	female	Zoey	PacBio RSII
UU_Cfam_GSD_1.0[26]	German Shepherd	female	Mischka	PacBio Sequel, Illumina HiSeq X
ASM864105v3[27]	German Shepherd	female	Nala	PacBio Sequel, PromethION, Illumina
UNSW_CanFamBas_1.2, Basenji_breed-1.1[28]	Basenji	female, male	China, Wags	PromethION, BGISEQ-500, Illumina NovaSeq

1.2 Spontaneous canine models of inherited disorders

1.2.1 Enrichment of hereditary diseases in dogs

In pursuit of desired traits, humans have inadvertently caused the enrichment of disease-causing variants in dogs.[29] Genetic disorders are quite common across breeds; in 2018, Donner et al.[30] screened 152 Mendelian disease-causing variants and showed that 40 % of over 100 000 purebred and mixed-breed dogs carried at least one of the variants in heterozygous or homozygous state. Specifically, 3.9 % of purebred and 1.3 % of mixed-breed dogs were homozygous for at least one recessive disorder.[30] Within breeds, disease prevalence can be very high: examples from the data of Donner et al.[30] include ichthyosis in Golden Retrievers (homozygous N=66, total N=330), multitype drug resistance 1 in Collies (homozygous N=42, total N=118) and exercise-induced collapse in Curly-Coated Retrievers (homozygous N=30, total N=156). Non-Mendelian disorders can be similarly prevalent, such as dilated cardiomyopathy in Dobermanns, which has a cumulative prevalence of 58 % in the European population.[31] In extreme cases, wild-type alleles can be lost from breeds entirely: for example, Dalmatians have become fixed for a variant causing hyperuricosuria.[32] This is due to apparent physical linkage of the causal gene, *SLC2A9* (solute carrier family 2 member 9), with a locus that has been under selection in the breed due to its involvement in the spotted pattern of the typical coat.[33]

In addition to unwanted, enriched disorders, some traits that are considered desired in specific breed standards and thus occur with high prevalence confer direct risk of health issues. These include disorders related to body conformation and traits with pleiotropic background. For example, the hairless phenotype of Chinese Crested Dogs and Mexican and Peruvian Hairless Dogs is the result of a 7-bp duplication and subsequent haploinsufficiency of *FOXI3* (forkhead box I3); however, due to its expression in not only hair follicles but also teeth during embryogenesis, the variant causes notable dentition abnormalities.[12] A similar trait-related disorder is seen in Rhodesian and Thai Ridgebacks, where a 133-kb duplication involving three fibroblast growth factor genes results in laterally oriented hair follicles, visible as a dorsal hair ridge.[34] The phenotype of heterozygotes is typically limited to the ridge, but homozygotes are predisposed to dermoid sinus, a tubular skin disease caused by defects in neural tube closure.[34] Body conformation in general is related to various other conditions, including but not limited to brachycephalic obstructive airway syndrome (BOAS) in brachycephalic dogs, musculoskeletal disorders in very small or large breeds, dermatitis in dogs with excessive skin folds and ulceration or irritation of eyes in dogs with protruding or sunken eyes.[35]

1.2.2 Physiological and genetic basis of canine models

The catalogue of Online Mendelian Inheritance in Animals currently lists 841 traits or diseases described in dogs.[36] Of these, 781 are classified as a defect, and 510 are considered a model of human disease, i.e. a similar phenotype in humans is described

in the catalogue of Online Mendelian Inheritance in Man (OMIM).[37] As large animals, dogs with spontaneously occurring disorders are invaluable models for human disease: compared to rodent models, dogs generally resemble humans more in terms of body conformation, longevity and living environment, which improves the potential for successful clinical translation.[38,39] In cardiac disorders, for example, large animal models feature several important anatomical similarities to humans, such as left ventricular wall tension, heart rate and vascular wall-to-lumen ratio.[40] In some phenotypes, successful therapy has already been developed and applied in dog models; one such demonstration is the rescue of rod function and preservation of retinal structure using adeno-associated virus vectors in dogs with *CNGB1* (cyclic nucleotide gated channel subunit beta 1) -related retinitis pigmentosa.[41]

Complementing the physiological similarities of dogs and humans, the high genetic homology between the two species further supports genetic research in canine models. The evolutionary distance between humans and dogs is shorter than that of humans and mice, likely influenced by the shorter generation time of murines.[42] From a bioinformatic perspective, the comparison of dogs and humans is fairly straightforward: in the last Ensembl release[20] (104) for canFam3.1 build, 85 % of the 20 199 protein coding genes have an orthologue mapped to GRCh38.p13. Conversely, the substantial functional information annotated to the human genome can be applied to syntenic canine regions, supplementing the scarcer dog-specific annotations.

While well-established disease-associated genes have often been revealed as causal in dogs, there are also studies where the first disease-causing variant has been described in dogs. One such case is a missense variant in *ATG4D* (autophagy related 4D cysteine pepsidase) in the Lagotto Romagnolo breed: the defect resulted in progressive cerebellar ataxia via altered autophagic flux and subsequent cellular vacuolization, and the gene was established as a candidate for neurodegenerative disorders for the first time.[43] The finding was recently followed up as the first human patient with similar symptoms and suspected pathogenic, compound heterozygous variants in *ATG4D* was described.[44] Similarly, another study in dogs investigated Lapponian Herders with progressive retinal atrophy and found an associated missense variant in *IFT122* (intraflagellar transport 122), a gene that had not previously been linked to retinal degeneration in mammals and was thus suggested as a new candidate for retinitis pigmentosa in humans.[45] A brief list of other examples includes various disorders caused by defects in the genes *FOXI3*, *LGI1* (leucine rich glioma inactivated 1), *DIRAS1* (DIRAS family GTPase 1), *SEL1L* (SEL1L adaptor subunit of ERAD E3 ubiquitin ligase), *HIVEP3* (HIVEP zinc finger 3), *ANLN* (anillin actin binding protein), and *SLC37A2* (solute carrier family 37 member 2).[12,46–51] However, unexpected species-specific cellular mechanisms of disease-associated genes may also be indicated, such as in Boxers and other breeds with a homozygous truncating *MPV17* (mitochondrial inner membrane protein MPV17) variant.[52] In humans and mice, *MPV17* defects cause hepatocerebral mitochondrial DNA depletion syndrome characterized primarily by hepatic and neurological manifestations.[53] In contrast, the genetically affected dogs were apparently healthy despite decreased *MPV17* mRNA expression in lymphocytes and complete absence of translated protein in transfected cells.[52] These differences illustrate the intricate molecular systems across species.

Collectively, studies that have identified causal genetic variants in dogs demonstrate the role of canine discoveries in highlighting novel candidate genes and confirming molecular diagnoses of human patients. Understanding the exact genetic background also enables precise and in-depth clinical examinations in dogs, increasing the informativity of preclinical and follow-up studies.

1.2.3 Key resources in canine research

In the field of canine genetics, there are several important factors that facilitate research markedly. First, resources and expertise available in state-of-the-art veterinary and pathology facilities provide a solid foundation for meaningful care and clinical characterization of canine patients.[38,54] Moreover, the equipment necessary for clinical examinations in dogs and other animals of equal size is generally more similar to that of humans than what is needed on small rodents, which may reduce the amount of specialized equipment and therefore additional costs involved.[38] In purebred dogs, routine examinations and health assessments are obligatory in some breeding programmes, such as the Finnish PEVISA programme for combating hereditary diseases and defects.[55] The results of these are often recorded in databases together with detailed pedigrees, which results in easier availability of essential phenotypic information.

In countries where pet dogs are abundant, canine genetic research can be conducted with an approach that employs citizen science. For example, the number of dogs in Finland during the year 2016 was estimated to be 700 000, with almost 20 % of households having at least one dog.[56] By collaborating with breed clubs and voluntary dog owners that have donated samples from their dogs, the canine biobank in the University of Helsinki, Finland, has grown over a decade into world's largest canine biobank that currently houses almost 70 000 samples from more than 360 breeds and breed varieties. These continuous efforts, combined with interdisciplinary expertise, enable rapid utilization of large intra- and interbreed study cohorts in identifying loci and variants associated with monogenic[45,57–61] and complex diseases.[62–67] Collaboration with dog gene testing companies, which typically handle a higher volume of samples than academic research groups, can further increase sample sizes: by including candidate variants in panel testing, their segregation and prevalence can in only a few years be estimated in up to hundreds of thousands of dogs, providing critical evidence for interpretation of pathogenicity.[45,57,59,68,69] Indeed, the accumulating genetic and phenotypic data in biobanks across species accelerate discoveries in an unparalleled manner: as an example, the UK Biobank includes a wide range of genetic, lifestyle and environmental data on 500 000 human participants; between years 2012-2021, the resource has been utilized in more than 2300 publications.[70]

From a broader perspective, the recruitment of voluntary dog owners and their pet dogs in studies of inherited disorders improves animal welfare by enforcing the Three R's principles.[71] The number of animals is reduced, as no research colonies are generated, and their treatment is refined, since the dogs lead normal lives and undergo standard veterinary care. Typically, genetic analyses only require a blood sample or a

buccal swab for DNA, and necessary clinical examinations are conducted according to routine veterinary protocols. This minimal approach may also help in making the research more socially accepted in public opinion, as the use of species perceived to have 'higher' mental abilities, such as companion animals, in traditional animal experiments is generally not supported.[72]

1.2.4 Five genetic conditions addressed in the present thesis

This thesis study is based on the biological and genomic resources available in the canine biobank of the University of Helsinki and Folkhälsan Research Center. The study addresses the clinical and genetic features of five different canine models of human rare disorders: 1) epidermolysis bullosa, a dermatological disease characterized by blistering in Central Asian Shepherd Dogs (CASD) (Figure 1A) 2) pituitary dwarfism in Karelian Bear Dogs (KBD) (Figure 1B) 3) congenital non-syndromic hearing loss in Rottweilers (RW) (Figure 1C); Robinow-like syndrome in bulldog and pit bull type breeds (Figure 1D); and malignant ventricular arrhythmias and sudden cardiac death in Leonberger Dogs (LEO) (Figure 1E). The diversity of the phenotypes studied demonstrates the remarkable potential of spontaneous canine models in gene discovery across disease types.

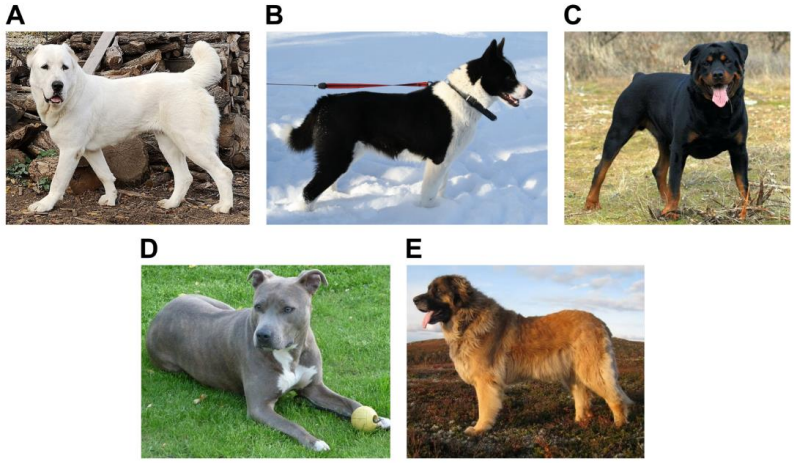

Figure 1. Dog breeds studied in this thesis. A) Central Asian Shepherd Dog (also known as Central Asian Ovcharka and Alabay). In some countries where it is not prohibited, their ears are cropped and tails docked. B) Karelian Bear Dog. C) Rottweiler. D) American Staffordshire Terrier, one of the breeds considered a pit bull type dog. E) Leonberger Dog. Images A, C, D and E were reproduced from Wikimedia Commons under the CC-BY license, and image B from Pixabay under the Pixabay license.

Epidermolysis bullosa

Epidermolysis bullosa is a heterogeneous group of dermatological disorders characterized by fragility of skin and mucous membranes that manifests as formation of erosions and blisters.[73] The underlying pathophysiological mechanism is the separation of skin layers at the dermo-epidermal junction in response to minimal mechanical stress.[74] Based on the zone of separation, the disorders are classified into four subtypes: epidermolysis bullosa simplex, junctional epidermolysis bullosa, dystrophic epidermolysis bullosa and Kindler's epidermolysis bullosa.[75] According to the most recent classification criteria,[75] a total of 16 genes have been linked to different forms of the disorder in humans: *CD151* (CD151 molecule, Raph blood group), *COL7A1* (collagen type VII alpha 1 chain), *COL17A1* (collagen type XVII alpha 1 chain), *DST* (dystonin), *EXPH5* (exophilin 5), *FERMT1* (FERM domain containing kindlin 1), *ITGA6* (integrin subunit alpha 6), *ITGB3* (integrin subunit beta 3), *ITGB4* (integrin subunit beta 4), *KLHL24* (kelch like family member 24), *KRT5* (keratin 5), *KRT14* (keratin 14), *LAMA3* (laminin subunit alpha 3), *LAMB3* (laminin subunit beta 3), *LAMC2* (laminin subunit gamma 2) and *PLEC* (plectin).

In dogs, the symptoms and severity of different forms of epidermolysis bullosa are similarly heterogeneous as in humans. Excluding the Central Asian Shepherd dogs described in this thesis, the disorder has been reported in eight breeds with known variants in four genes (Table 2).

Table 2. Canine epidermolysis bullosa and the respective causal genes and disease subtypes in non-CASD breeds.

breed	**subtype**	**gene**	**variant type**	**severity**
Eurasier[76]	simplex	*PLEC*	nonsense	severe
Collie[77]	simplex	unknown	unknown	not specified
German Pointer[78]	junctional	*LAMA3*	large insertion	variable
Australian Cattle dog cross[79]	junctional	*LAMA3*	missense	severe
Australian Shepherd dog[80]	junctional	*LAMB3*	missense	intermediate
Golden Retriever[81]	dystrophic	*COL7A1*	missense	intermediate
Basset Hound[82]	dystrophic	*COL7A1*	complex rearrangement	severe
Akita Inu[83]	dystrophic	unknown	unknown	mild

Pituitary dwarfism

Pituitary dwarfism, also known as growth hormone deficiency, is an endocrine disorder characterized by short stature and slowness or absence of growth.[84] Anterior pituitary gland is an important regulator of growth and many other physiological processes that acts by secreting hormones, including growth hormone, thyroid-stimulating hormone, follicle-stimulating hormone, luteinising hormone, prolactin and adrenocorticotropin.[85] The deficiency of growth hormone due abnormal development

or function of the pituitary gland, caused by genetic or non-genetic aetiologies, is the underlying cause of pituitary dwarfism.[86] The symptoms of the condition are heterogeneous and depend on whether the affected hormones include only growth hormone or additional pituitary hormones; the disorder is respectively classified into isolated growth hormone deficiency and combined pituitary hormone deficiency.[87]

In humans, the pathogenesis of combined pituitary hormone deficiency is known to involve 30 genes.[87] In dogs, proportionate pituitary dwarfism has originally been described in two breeds: German Shepherd Dogs and Karelian Bear Dogs. The causal variant in German Shepherds, also later found in Saarloos wolfdogs, Czechoslovakian wolfdogs and Tibetan Terriers, is caused by an intronic deletion and consequent aberrant splicing of *LHX3* (LIM homeobox 3).[88–90] In Karelian Bear Dogs, the condition was first characterized in 1978,[91] and the causal variant remained undiscovered for four decades until described in this thesis.

Hearing loss

Hearing loss is defined as a partial or total inability to hear. Caused by both genetic and non-genetic aetiologies, the prevalence of the disorder in humans is 0.1-0.2 % in newborns and over 50 % in 80-year-olds.[92] Hearing loss is a very heterogeneous disorder in terms of onset (congenital and early-onset to late-onset), progression (stable or progressive), symmetry (unilateral or bilateral), and symptoms (syndromic or non-syndromic).[93,94] Depending on the affected anatomical site, the disorder is categorized into conductive, sensorineural and mixed hearing loss.[94] Reflecting the complex biological mechanisms of hearing, both common and rare forms of genetic hearing impairment have been described, and over 100 genes have been associated with the disorder to date.[92]

In dogs, hearing loss has been reported in both syndromic and non-syndromic congenital forms as well as with adult onset. First, pigment-associated congenital sensorineural deafness occurs in various breeds with partial lack of pigmentation and involves *MITF* (melanocyte inducing transcription factor), a master regulator of melanocyte development; the underlying mechanism is the disturbance of melanoblast differentiation or migration from the neural crest to target sites, including hair follicles in the skin and stria vascularis in the inner ear.[95] Second, congenital deafness and vestibular dysfunction in Dobermanns is caused by variants in *MYO7A* (myosin VIIA) and *PTPRQ* (protein tyrosine phosphatase receptor type Q).[96,97] Third, a recent report linked a *KLF7* (Kruppel like factor 7) variant with incomplete penetrance to congenital sensorineural deafness in Australian Stumpy Tail Cattle Dogs.[98] Another recent study linked a mottled coat pattern, also called roaning, to an intronic tandem duplication in *USH2A* (usherin) and hypothesized a potential pleiotropic effect on both hearing and pigmentation as *USH2A* is known to cause congenital hearing loss and retinitis pigmentosa in humans.[99,100] In contrast, a locus on chromosome 6 has been associated with adult-onset deafness in Border Collies, but the causal gene remains uncertain.[101] Finally, in this thesis, congenital nonsyndromic hearing loss corresponding to a rare form of deafness in humans and an associated full-penetrance variant was reported for the first time in dogs, specifically in the Rottweiler breed.

Robinow-like syndrome

The canine Robinow-like syndrome is an orthologue of human Robinow syndrome, a rare developmental disorder. Inherited in both dominant and recessive manner and caused by signalling abnormalities in the non-canonical Wnt/planar cell polarity pathway,[102] Robinow syndrome in humans is characterized by variable symptoms, the most distinct of which are mesomelic dwarfism, craniofacial and orodental abnormalities, vertebral defects and genital hypoplasia.[103] To date, seven genes are known to be involved: *DVL1* (dishevelled segment polarity protein 1), *DVL2* (dishevelled segment polarity protein 2), *DVL3* (dishevelled segment polarity protein 3), *WNT5A* (Wnt family member 5A), *ROR2* (receptor tyrosine kinase like orphan receptor 2), *FZD2* (frizzled class receptor 2) and *NXN* (nucleoredoxin);[104–106] the first likely pathogenic variant in *DVL2* was identified very recently. The clinical variability of the disorder is considerable, stemming not only from allelic and locus heterogeneity but also from the complexity of how perturbations affect the Wnt pathway.[105]

In dogs, Robinow-like syndrome was first described in 2018 upon the discovery of a frameshift deletion in the *DVL2* gene.[107] The variant was found to be associated with caudal vertebral anomalies, also known as "screw tail", a trait that has been selected for in English Bulldogs, French Bulldogs and Boston Terriers. In addition to the caudal vertebral anomalies, these breeds are characterized by short and broad stature, wide head, brachycephalic skull shape and widely set eyes.[108–110] Importantly, *DVL2* had not been linked to Robinow syndrome in humans at the time. Therefore, due to the marked phenotypic similarities and homology of *DVL2* with the known causal genes *DVL1* and *DVL3*, Mansour et al. proposed that the canine phenotype be an equivalent of human Robinow syndrome, and that the *DVL2* variant would thus contribute to the entire bulldog type morphology.[107] However, due to allele fixation in the three affected breeds, the exact morphological consequences of the variant could not be studied further. In this thesis, the genotype-phenotype correlation of the canine *DVL2* variant with skeletal abnormalities is characterized in detail.

Ventricular arrhythmia and sudden cardiac death

Cardiac arrhythmia is defined as a variation from physiologically normal heart rate and rhythm.[111] Arrhythmias can occur in morphologically normal hearts or in conjunction with a structural heart disease,[112] and the underlying mechanisms are generally categorized into two groups: enhanced or abnormal impulse formation and conduction disturbances.[111] Depending on the type of arrhythmia and the anatomical site affected, the consequences range from benign to life-threatening.[113] Sudden cardiac death, defined as a sudden unexpected death due to a cardiovascular cause, is the most severe outcome, and it often results from ventricular tachyarrhythmias.[112]

The role of genetic factors in sudden cardiac death is most prominent in young individuals, whereas the impact of environmental factors, such as nutrition or exercise, is more evident after a longer period of life.[112] The genetic background of inherited arrhythmias is heterogeneous, and at least 39 genes encoding or interacting with components of ion channels are known to be involved in primary electrical diseases in humans.[112] In dogs, malignant ventricular arrhythmia in the absence of structural cardiac anomalies has previously been reported in two breeds: Rhodesian Ridgebacks

with an associated variant in *QIL1* (also known as *MICOS13*, mitochondrial contact site and cristae organizing system subunit 13),[114] and German Shepherds, where the causal gene remains unknown.[115] The disorder in Rhodesian Ridgebacks is inherited in a recessive manner and typically develops between 7-12 months of age, with a suggested molecular pathogenesis due to mitochondrial cristae abnormalities.[114] In German Shepherd Dogs, the incidence and severity of arrhythmias increased between 2-7 months and decreased between 7-11 months, demonstrating an age-dependent mechanism of the disease. Finally, in this thesis, Leonberger Dogs with juvenile ventricular arrhythmia and subsequent sudden cardiac death are described for the first time.

1.3 Methods in locus and variant discovery

An established approach to variant identification in dogs is a dual strategy combining locus mapping methods with NGS methods. By utilizing genome-wide association analysis, homozygosity mapping and/or genetic linkage analysis, candidate regions for subsequent variant analysis can be greatly narrowed. This procedure capitalizes on both the intrabreed and interbreed population structures of dog breeds, typically resulting in very efficient prioritization of associated loci and variants with relatively small discovery cohorts. The most widely used genetic analysis methods are presented in this section.

1.3.1 Genome-wide association study

Genome-wide association study (GWAS) is one of the most commonly utilized method to reveal disease-associated genomic regions. Since its introduction in 2005, over 55 000 loci have been identified for more than 5 000 phenotypes,[116] demonstrating the wide flexibility and applicability of the method to various study designs. The most simple approach is entirely based on comparing marker allele frequencies between cases and controls, and a myriad of more sophisticated GWAS methods have been developed to account for population structure, quantitative traits, covariates, and comorbid traits, for example.[117] As GWASes typically include large datasets which, depending on the species, consist of tens of thousands to millions of markers and up to hundreds of thousands of individuals, rigorous quality control and stringent multiple testing correction are essential to avoid spurious associations.[118]

In conjunction with the evolution of analysis methods and computational resources, SNP arrays have also expanded. The first experimental human array genotyping chip was developed by Affymetrix in 1998, and it could be used for simultaneous genotyping of 500 SNPs;[119] currently, the Affymetrix Genome-Wide Human SNP Array 6.0 (Thermo Fisher Scientific, CA, USA) contains more than 1.8 million markers, including autosomal, X and Y chromosomal and mitochondrial SNPs as well as copy number variants. Canine platforms have evolved similarly: first array designs included 22 000 (CanineSNP20 BeadChip, Illumina, CA, USA) and 27 000 markers (Affymetrix Canine array v1 #520170, Thermo Fisher Scientific), followed

by 50 000 markers (Affymetrix Canine array v2 #520431, Thermo Fisher Scientific) and the current, most widely used designs of 173 000 and 220 000 markers (CanineHD BeadChip, Illumina). Latest developments include the experimental chip designs Axiom Canine Genotyping Array Sets A and B (Thermo Fisher Scientific), containing 1.1 million markers in total, and the further refined Axiom Canine HD Array with 710 000 markers (Thermo Fisher Scientific). In addition, imputation and the advent of whole-genome sequences as a data source for GWAS have also brought versatility to study designs across species.[120]

The core principles in canine association studies are somewhat divergent from the classic common disease/common variant hypothesis.[121] Due to bottlenecks and founder effects in purebred dogs, many disorders considered rare in humans have a high-prevalence canine counterpart in specific breeds. The low intra-breed variation and extensive length of LD blocks enable high-power association studies on these disorders with relatively low sample size while simultaneously avoiding the obstacles caused by genetic heterogeneity and admixture encountered in human populations.[122] However, while admixture between established breeds is rare in dogs, substructures within breeds can be present due to diminished gene flow between e.g. working and show lines or between different countries.[8,123] Nevertheless, by utilizing a dual strategy of GWAS and NGS analysis, causal variants in canine models of human rare disease have repeatedly been discovered with association study cohorts of 50 or less individuals.[45,47,51,57,124] The challenge in combined GWAS and NGS is that candidate loci in canine populations typically span long regions, up to several megabases and tens of genes;[125,126] therefore, pinpointing the causal genes and variants can be difficult.

1.3.2 Homozygosity mapping

Homozygosity mapping, also known as autozygosity mapping, is an analysis method used to map genomic regions linked to recessive phenotypes in consanguineous families.[127] The underlying assumption is that the disease-causing allele responsible for the recessive phenotype resides in a haplotype that has been inherited from both parents by affected individuals. Therefore, due to linkage, not only the disease-causing allele but also the flanking alleles should be present in homozygous state in the affected population.[127] Consequently, by identifying homozygous identity-by-descent segments, also known as runs of homozygosity (ROH), that are shared by all cases, the genomic location of the disease allele can be narrowed down to candidate regions.[128] Of note, this method is not suitable for phenotypes with a compound heterozygous aetiology, although their pattern of inheritance may resemble autosomal recessive phenotypes in pedigrees.[129]

As an analysis method, homozygosity mapping is greatly dependent on the software of choice and its parameters. Currently, there are several options available, including programs such as PLINK, GERMLINE, AgileGenotyper, H^3M^2, HomSI, AutoMap, KinSNP, HomozygosityMapper and bcftools.[130–138] Among the software, there is great variation regarding the input data source (SNP array data, WGS data, or WES data), user interface (graphical or command line), mapping algorithm, and

output format and visualization. Furthermore, it has been shown that the accuracy in detection of ROH of different lengths varies considerably between programs,[139,140] and the parameters of each individual program may further have a major impact on the results, as demonstrated by Meyermans et al.[141] with PLINK's algorithm and SNP array data from multiple livestock species. Therefore, the selection of an appropriate mapping software is crucial during study design.

Homozygosity mapping is a method especially suitable for small study cohorts.[142] The advantage over GWAS for locus discovery is that homozygosity mapping can be performed with very few or even without controls; furthermore, multi-generational phenotypic and genetic information are not required, unlike in genetic linkage analysis.[143] However, without controls, the amount of resulting ROH can be very high in inbred populations with extensive homozygosity, such as in dogs.[144] In contrast, if controls are utilized to filter out candidate regions, there is a risk of discarding the causal region should a haplotype occur in the population with and without the disease allele, which is possible with variants of recent origin.[50]

1.3.3 Genetic linkage analysis

Genetic linkage analysis is a technique to map disease-associated loci by utilizing multigenerational phenotype information.[145] The analysis is based on the premises that the hypothetical causal allele segregates in a family according to the assumed mode of inheritance, and that the allele is in linkage with nearby markers. Therefore, by identifying alleles that cosegregate with the phenotype, candidate loci can be detected. Conducted first using microsatellite genotypes[146] and, with the advent of more high-throughput platforms, SNP arrays and WGS data, linkage analysis is an especially useful approach to pedigrees with well-defined phenotypes and mode of inheritance. The key difference of linkage analysis and GWAS in structured populations is that the former infers information directly from recombination events observed in a pedigree,[147] while the latter can take relatedness into account indirectly via kinship matrices in linear mixed models.[148]

In canine studies, linkage analysis has been successfully applied to discover loci for very different types of diseases and traits. The analysis is occasionally conducted in conjunction with or followed by homozygosity mapping or GWAS to obtain further supportive evidence or to fine-map the associated regions.[44,149–154] The strength of the approach in canine populations is that discoveries may be made with very small study cohorts, consisting of even less than ten individuals.[43,155–157] Another advantage is that compared to humans, dogs have a shorter generation time, which enables the collection of multigenerational samples more easily. However, analysis of large families with excessively complex pedigree loops, common in inbred dog populations, can result in extensive computation time or internal errors unless specialized software is utilized,[158] which is a potential limitation to the applicability of linkage analysis in canine studies.

1.3.4 Next-generation sequencing analysis and variant validation

Next-generation sequencing (NGS) is a widely utilized group of different high-throughput sequencing approaches and the core of current genomic methodology. Comprised of whole genome sequencing (WGS), whole exome sequencing (WES) and targeted sequencing, the data can be utilized to discover disease-associated variants by inspecting entire genomes, coding sequences or other, custom-defined regions. The generation and processing of NGS data involves high-throughput sequencing with short-read or long-read technologies,[159] base calling, alignment of the reads against the reference genome, and variant calling (Figure 2).[160] Rigorous quality control is warranted in each step of the process to ensure high-quality data and avoid biased calls.[160] Once the variants have been called, the essential effort in NGS analysis is to differentiate between likely causal variants and other, unrelated variation, as any given human or canine genome typically contains millions of variants.[14,161] Towards that, variants are evaluated in light of population-level information and functional annotation to retain or discard them from further analyses.

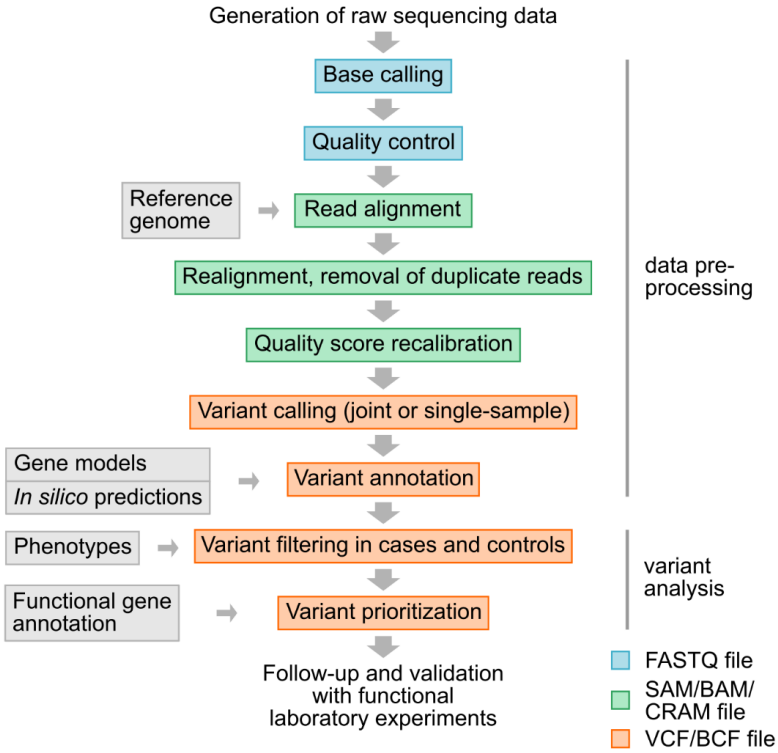

Figure 2. A flowchart depicting key steps in NGS data processing. The raw data generated with a DNA sequencer is processed using several types of specialized software and file formats to perform base calling (FASTQ format), read alignment (SAM/BAM/CRAM files) and variant calling (VCF files). External data is used throughout the process to annotate the variants in the context of genes and their impact

on coding sequences and other elements of interest. The figure was partially adapted from Altmann et al.[162] and Nielsen et al.[160]

To overcome the challenges caused by the abundance of variation, filtering variants based on their pattern of inheritance and frequencies in populations of affected and unaffected individuals can greatly narrow down the pool of candidate variants.[163] Common inclusion criteria include consistency with the segregation pattern of the assumed mode of inheritance, appropriate allele frequencies in specific populations, and compliance of genotypes with known or assumed disease penetrance. However, since filtering relies completely on underlying assumptions about the properties of the hypothetical causal variant, it is sensitive to the parameters employed and can result in both false negatives and false positives should the assumptions not hold true.

Filtering in canine populations is moderately different from human studies. The low intrabreed and high interbreed heterogeneity typically manifest as moderate or high frequency of disease-causing alleles in affected breeds, and low frequency or absence in other breeds.[30] While the utilization of multi-breed cohorts is a powerful strategy to identify variants specific to affected populations, the approach typically also yields numerous breed-specific, noncausal variants,[126] which can cause challenges in identifying the causal variant.

Further inspection of filtered variants is typically carried out with functional annotations. To guide the interpretation, predictions about the variant type (exonic, intronic, splice site or other region), consequence (missense, nonsense, frameshift, etc.), pathogenicity and conservation as well as knowledge about the expression, pathways and previous gene-disease associations of the affected gene are essential.[163] The resources for such information are abundant, and annotation data need to be carefully selected according to the design and aims of each study. Finally, many popular datasets, such as those in the catalogues of Human Protein Atlas,[164] GTEx[165] and Mouse Genome Informatics (MGI),[166] originate from human samples or common model organisms, including mice or zebrafish, which necessitates additional consideration on whether the data should and can be adapted to other species, such as dog.

To confirm the phenotype-genotype correlation and predicted impact of candidate variants, *in silico* analyses are typically followed by laboratory experiments. Genetic validation is performed to further confirm the segregation of candidate causal variants. Excluding large, consortium-based resources, NGS can usually be performed to a relatively limited number of samples due to high cost and time required, which limits the informativity of the analysis. To complement this, candidate causal variants can be screened in larger populations with known phenotypes by Sanger sequencing, TaqMan genotyping or similar lower cost approaches,[167] which can either strengthen the correlation or discard the variant from further consideration. Pedigree information can bring an additional dimension to the screening,[168] but is not strictly necessary. Second, functional validation is ideally performed to determine the effects of the variant on RNA, protein or tissue level in patient tissue or in a model, such as cell cultures (possibly derived from induced pluripotent stem cells), organoids or a knock-down system.[169] The approaches for this are numerous and depend on the study design,

and they include e.g. quantification of RNA and protein, RNA sequencing, different stainings to determine localization and consequence on cellular morphology, mutagenesis, bait and prey systems to detect protein-protein interactions, and many more.[169] To summarize, while NGS analyses form the basis on which modern disease genomics is built, laboratory experiments remain essential in validating the molecular consequences of candidate variants identified through *in silico* methods.

2 AIMS OF THE STUDY

The primary goal of this PhD thesis was to characterize the clinical and genetic background of five canine models of rare disease: epidermolysis bullosa (Study I), pituitary dwarfism (Study II), congenital hearing loss (Study III), Robinow-like syndrome (Study IV) and malignant ventricular arrhythmia and sudden cardiac death (Study V). The genetic causes of four conditions (Studies I-III, V) were previously undescribed, and they were hypothesized to be hereditary based on their high prevalence within dog families and breeds. While the genetic cause of one of the studied phenotypes (Study IV) was previously established, the precise genotype-phenotype correlation remained elusive and was hypothesized to be clarified with additional breed screenings combined with detailed clinical measurements. To test the hypotheses and to reach the goal of the study, the following specific aims were set:

1. To collect cohorts of affected and unaffected dogs (Studies I-III, V) or dogs with known *DVL2* genotypes (Study IV) and confirm their phenotypes with clinical (Studies I-V) or post-mortem examinations (Studies I and V)

2. To map disease-associated loci (Studies I-III, V) and identify candidate genes and variants (Studies I-III) by utilizing appropriate gene mapping and next generation sequencing analysis methods

3. To confirm and investigate the consequences of candidate causal variants (Studies I-III) and the previously known *DVL2* variant (Study IV) with genetic validation (Studies I-IV), functional experiments (Study I) and statistical analyses (Study IV)

3 MATERIALS AND METHODS

All studies were primarily conducted in Professor Hannes Lohi's research group in the University of Helsinki and Folkhälsan Research Center in collaboration with veterinarians in the University of Helsinki Small Animal Hospital. Experiments, analyses and protocols not conducted by the author of this thesis are specifically noted in each section.

3.1 Ethical statement

All samples in Studies I-V were collected from privately owned pet dogs with the dog owners' informed consent. Participation in the studies was voluntary and the dog owners had an opportunity to withdraw from the studies at any time without penalty. Ethylenediaminetetraacetic acid (EDTA) whole blood and tissue sample collection was ethically approved by the Animal Ethics Committee of State Provincial Office of Southern Finland under permits ESAVI/7482/04.10.07/2015, ESAVI/343/04.10.07/2016 and ESAVI/25696/2020. Clinical examinations in Study V were performed under permit ESAVI/8176/04.10.07. All dogs on whom a post-mortem examination was performed in Studies I and V had been sent for the necropsy voluntarily by the dog owners.

3.2 Study cohorts and samples

Previously collected EDTA blood samples in the canine biobank of the University of Helsinki were extensively utilized in Studies I-IV. Additional blood or tissue sample collection was targeted to affected dogs (Studies I-V) and their close relatives (Studies I-III, V). First, collection of blood samples was conducted at veterinary clinics, or by qualified researchers and laboratory technicians of the Lohi research group. Second, collection of tissue samples during post-mortem examinations was conducted by veterinary pathologists. Finally, extraction and quantification of DNA from the blood and tissue samples was performed by laboratory technicians of the Lohi research group at the University of Helsinki. The extraction was performed with the semi-automatic Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany) and Chemagic 360 (Chemagen) extraction robots. DNA concentration as well as 260:280 and 260:230 ratios were determined with either NanoDrop ND-1000 UV/Vis Spectrophotometer (Thermo Fisher Scientific, WA, USA), Qubit 3.0 Fluorometer (Thermo Fisher Scientific) or DeNovix DS-11 Spectrophotometer (DeNovix Inc., DE, USA).

The cohorts in Studies I-III consisted of affected pedigrees, biobank samples from the affected breeds, and additional biobank cohorts from closely related breeds (Studies I and II). First, the cohort in Study I comprised two CASD puppies suspected

to be affected by epidermolysis bullosa, their six unaffected littermates and parents, and 37 other CASD, as well as 39 Caucasian Shepherd Dogs, three South Russian Ovcharkas, six Kuvasz Dogs, 19 Slovakian Chuvach Dogs and 76 Tibetan Mastiffs (total N=190). Second, Study II included eight KBD with suspected pituitary dwarfism and their close relatives in a cohort of altogether 642 KBD, 263 Lapponian Herders and 87 Laika Dogs (total N=992). Finally, Study III involved four RW puppies affected by hearing loss, one unaffected littermate, their unaffected parents and 578 other RW (total N=585).

In addition to the pedigree and biobank samples, Studies II and III included large multi-breed cohorts submitted to commercial genetic testing. During the studies, sets of 7025 dogs from 206 breeds (Study II) and 28 116 dogs from 374 breeds (Study III) were tested for candidate causal variants at Genoscoper Laboratories (Wisdom Health, Finland), and a sample of 771 864 mixed-breed dogs (Study III) was screened with the Wisdom Panel test (Wisdom Health, WA, USA).

In contrast to studies I-III, Study IV did not consist of pedigree cohorts; it was instead based solely on breed samples from the canine biobank or collected during the study. The cohort included 1954 dogs from the following 15 breeds: 165 Boston Terriers (BT), 211 English Bulldogs (EBD), 297 French Bulldogs (FBD), 4 Olde English Bulldogges, 11 American Bulldogs, 285 American Staffordshire Terriers (AST), 714 Staffordshire Bull Terriers (SBT), 73 Dogues de Bordeaux, 16 Bull Terriers, 13 Miniature Bull Terriers, 40 Lhasa Apsos, 20 Shih Tzus, 46 Tibetan Spaniels, 10 Pekingese, 47 King Charles Spaniels and two mixed breed dogs (FBD × German Shepherd Dog).

Finally, the cohort of Study V consisted of 65 LEO that underwent either a clinical examination or a post-mortem examination. The recruitment of the clinical study cohort was focused on pedigrees previously known to be affected by ventricular arrhythmia or sudden cardiac death.

3.3 Pedigree information

Pedigree information was utilized in Studies I-III and Study V. The data was obtained from the public databases of the Finnish Kennel Club (https://jalostus.kennelliitto.fi) and the Estonian Kennel Club (https://register.kennelliit.ee/) (Study I), and from individual dog owners. In Studies I and V, pedigrees were built with the AncesTrim software.[170] Finally, GenoPro 2011 (Genopro Limited, Hong Kong) was used for visualization.

3.4 Phenotyping

3.4.1 Clinical examinations

Clinical examinations were conducted in all studies to establish case-control cohorts (Studies I-III, Study V) or to dissect the qualitative and quantitative effects of a genetic variant (Study IV). The examinations were performed by experienced veterinarians at the Small Animal Hospital of University of Helsinki, Finland or in other veterinary facilities in Finland or Sweden.

In Study I, a clinical examination was performed to two CASD puppies soon after birth. The dogs, born at full term, presented severe lesions, blisters and ulcers on their feet, ears, muzzle and oral mucosa. Based on the symptoms, the puppies were suspected to be affected by EB and, due to poor prognosis, were humanely euthanized by the veterinarian due to the owner's request.

In Studies II and III, clinical examinations had been conducted prior to our study and the reports were collected from the dog owners. In Study II, one KBD suspected to be affected by pituitary dwarfism underwent blood urea nitrogen testing, renal ultrasound and thyroid hormone testing. Another suspected case was clinically examined due to skin anomalies. A third suspected case in Sweden was tested for serum insulin-like growth factor. Finally, in Study III, four RW puppies with impaired hearing underwent brainstem auditory evoked response (BAER) testing.

In Study IV, phenotyping was conducted with computer tomography (CT) imaging and by retrospectively assessing spinal radiographs. First, 19 AST with known *DVL2* genotypes (eight wild types, eight heterozygotes and three homozygotes) and one wild-type SBT were recruited to full-body computed tomography (CT) imaging and physical examination. The resulting CT images were interpreted by a clinical instruction unaware of the dogs' *DVL2* genotypes. The images were classified for hemivertebrae and other vertebral malformations according to previous guidelines,[171] and all thoracic, lumbar and sacral vertebrae were measured. The length of the bony tail and left and right radii were also recorded. Craniometric measurements included facial length and width, cranial length and width, mandibular length, skull length and width, skull base length, distance between eyes, nasal bone length, hard palate length, and soft palate length and thickness. Finally, skull, cranial and facial indices were calculated as previously described.[172] From the radiographs, the vertebrae of 23 ASTs (21 wild types, one heterozygote and one homozygote) were similarly measured and classified for abnormalities. Images of the thoracic and lumbar spine and pelvis were available from all dogs, and caudal vertebral images from the homozygous dog. Finally, spinal and tail radiographs from three FBD × German Shepherd Dog crosses were classified for vertebral anomalies.

In Study V, clinical examinations, echocardiography, five-minute ECG and 24-hour Holter monitoring was performed to 46 LEO less than four years old. The echocardiography was conducted according to standard, and 20 echocardiographic parameters were recorded: left ventricular end-diastolic and end-systolic volume and volume indices; ejection fraction and fractional shortening; left ventricular end-diastolic and end-systolic internal diameter; interventricular septum end-diastolic and

end-systolic thickness; left ventricular posterior wall end-diastolic and end-systolic thickness; E-point-to-septal separation; sphericity index; left atrial and aortic root diameters and their ratio; aortic hinge diameter; and aortic and pulmonic peak velocity. The Holter monitoring was performed for 24 hours while the dogs wore the monitor in their home environment, and owners were requested to record the dogs' activities in a diary. The recordings were analysed both automatically and manually for heart rate, number of VPCs, ventricular couplets and triplets, VT, sinus pauses, supraventricular premature complexes, ventricular escape complexes and second-degree atrioventricular blocks. Finally, blood samples were analysed for complete blood cell count, serum biochemistry profile, high-sensitivity troponin I and for vector-borne diseases with a Snap4Dx test.

3.4.2 Post-mortem examinations

Post-mortem examinations were conducted in Studies I and V. The dogs on whom the examination was done had been voluntarily sent for necropsy by the dog owners. The examinations were performed by experienced veterinary pathologists at the Veterinary Bacteriology and Pathology Research Unit, Finnish Food Authority, Helsinki, Finland and at the Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, Finland.

In Study I, two CASD puppies suspected to be affected by EB were sent for a full necropsy, and in Study V, necropsy reports of 21 LEO with suspected SCD examined prior to our study were reviewed retrospectively. During necropsy, tissue samples of all major organs, including brain, respiratory and gastrointestinal tract, heart, liver, spleen, pancreas, kidneys, thyroid and adrenal glands, gonads and skin (Studies I and V) as well as oral mucosa (Study I) were collected, formalin fixed, routinely processed and paraffin embedded, followed by sectioning at 4 µm and haematoxylin and eosin staining. Additionally, periodic acid-Schiff staining was performed in Study I on skin samples to determine the zone of separation in the dermo-epidermal junction. Finally, liver samples were collected and stored at -20°C for later DNA extraction.

3.4.3 Owner reports

In Studies II-IV, where phenotypes could be partially assessed without specific expertise or protocols, reports of the participating dogs' health were collected from the owners. In Study II, information of the weight, height and coat quality of KBD suspected to be affected by pituitary dwarfism was collected, along with photographs from selected individuals. In Study III, RW used as controls in the homozygosity mapping were confirmed to have normal hearing by the dog owners. Finally, in Study IV, descriptions and images of participating dogs' tails were obtained when available to evaluate the presence of tail kinks.

3.5 Sequences

The gene models used in all studies were annotated to the canFam3.1 reference sequence. The NCBI Nucleotide collection[173] geneID, mRNA and protein accession numbers of the sequences used in downstream analyses and primer design are as follows: 403467, NM_001002980.1 and NP_001002980.1 (Study I); 403753, NM_001006949.1 and NP_001006950.1 (Study II); and 490467, XM_022421426.1 and XP_022277134.1 (Study III).

3.6 SNP array genotyping and quality control

DNA samples were genotyped on a commercial SNP array for Studies II, III and V (unpublished data). The array genotyping was performed in several batches with the CanineHD Whole-Genome Genotyping BeadChip (Lincoln, NE, USA) containing 173,662 markers along the canFam3.1 reference genome. The cohorts included five cases and 139 controls (Study II), four cases and three controls (Study III) and 28 cases and 19 controls (Study V unpublished data). Pre-analytical quality control was conducted with PLINK version 1.9[131] and included pruning for marker call rate of > 95 % (all studies), sample call rate of > 95 % (Study II and Study V unpublished data) or > 93 % (Study III), HWE p-value > 1×10^{-8} (all studies) and MAF > 5 % (Study II and Study V unpublished data). MAF pruning was not performed in Study III, where homozygosity mapping was utilized as an analysis method.

3.7 SNP data analysis

3.7.1 Genome-wide association study

GWAS was performed in Study II and Study V (unpublished data). The analysis in Study II was performed jointly with a co-author. In Study II, the cohort included five dogs suspected to be affected by pituitary dwarfism and 139 unaffected controls. After pre-analytical quality control (QC), the data contained 96 274 markers. The analysis was conducted with PLINK 1.9[131] with basic case/control association (--assoc), and resulting p-values corrected for inflation with genomic control correction and for multiple testing with the Bonferroni method. Post-analytical QC included evaluation of stratification from QQ and MDS plots. Multidimensional scaling was performed with PLINK 1.9 with options --cluster and --mds-plot 2. The results were visualized with in-house Python 3 scripts using the matplotlib module.[174,175]

In Study V (unpublished data), the cohort included 28 affected and 19 unaffected LEO. Of the affected dogs, 9/28 had a diagnosis of ventricular arrhythmia based on 24-hour Holter monitoring. The other 19/28 affected dogs had died a sudden cardiac death and undergone a post-mortem examination, with necropsy findings indicating acute heart failure. Finally, the 19 control dogs were determined to be unaffected based on 24-hour Holter monitoring.

The pre-analytical QC and association analyses in Study V (unpublished data) were performed with two study populations: by using only the Holter-diagnosed cases and 98 508 QC-passing markers; and second, by using both Holter-diagnosed and necropsy cases, and 98 137 QC-passing markers. Both analyses were performed with a univariate linear mixed model approach in GEMMA version 0.98.1[176] and corrected for inflation with genomic control correction and for multiple testing with the Bonferroni and M_{eff}[177] methods. Number of markers used for M_{eff} correction was 8183 for Holter-diagnosed cases and 9654 for Holter-diagnosed and necropsy cases. Post-analytical QC was performed as described in Study II. Additionally, marker distributions in each chromosome and across bins of 1 Mb were inspected in pruned and unpruned datasets with an in-house Python 3 script and visualized using matplotlib.[174,175]

3.7.2 Homozygosity mapping

Homozygosity mapping was conducted in Study III. After pre-analytical QC, the data included three cases, three controls and 154 235 markers. The mapping was performed with PLINK 1.9,[131] and the parameters of the mapping algorithm were optimized before the analysis. Genome-coverage based approach, outlined by Meyermans et al.,[141] was conducted by running ROH detection on simulated data.

The maximum genome coverage approach, presented here with detailed parameters due to the complexity of the algorithm, was as follows: First, minimum marker size for ROH (--homozyg-snp) was set to 70 based on the formula presented by Purfield et al.,[178] with $\alpha = 0.05$, $N_s = 154\,235$, $N_i = 6$ and mean SNP heterozygosity = 0.21 as estimated with an in-house Python 3 script.[174] Second, scanning window parameters were set to 20 for window size (--homozyg-window-snp); 1 and 0 for missing (--homozyg-window-missing) and heterozygous (--homozyg-window-het) markers, respectively; 0.05 for ROH inclusion threshold (--homozyg-window-threshold); and 1000 kb for minimum ROH length (--homozyg-kb). Finally, ROH density (--homozyg-density) was set to 200 kb/snp and maximum SNP gap (--homozyg-gap) to 2000. With these settings, ROH detection was performed on a simulated dataset of one individual homozygous at every marker, and the total length of the resulting ROH was recorded as maximum genome coverage. With the genome coverage established, ROH detection was repeated by varying ROH density between 10-125 kb/snp in increments of 5 kb and maximum SNP gap between 20-1000 kb in increments of 20 kb; the ratio of resulting ROH to the maximum genome coverage was then recorded. Based on the coverage ratios, optimal ROH density was determined to be 30 kb/snp and maximum SNP gap 200 kb, as coverage reached 100 % for density at 30 kb/snp and increased only negligibly for maximum SNP gap after 200 kb.

With ROH density and maximum SNP gap established, the homozygosity mapping was performed with the following parameters: scanning window size 50, missing and heterozygous marker thresholds at 1 marker, ROH inclusion threshold 0.05, ROH marker size 70, ROH length 1000 kb, ROH density 30 kb/snp and maximum SNP gap 200 kb. The resulting ROH were pooled (--homozyg group), and required to be allelically shared by the three cases and allelically different or absent in

the three controls. The results were visualized with an in-house Python 3 script using matplotlib.[174,175]

3.8 Next-generation sequencing

Next-generation sequencing was conducted in Studies I-III at specialised facilities. Processing of the raw genomic data from FASTQ to VCF files, including alignment, variant calling and annotation, was performed by bioinformaticians of the Lohi research group at the University of Helsinki. The data were stored and processed on the servers of CSC - IT Center for Science, Finland.

3.8.1 Whole-genome sequencing

Whole-genome sequencing was conducted in Studies I-III for a total of five samples. The sequencing was performed in two facilities: the University of Bern (Bern, Switzerland) (Studies I-III) and the Novogene Bioinformatics Institute (Beijing, China) (Study II).

The sequencing at the University of Bern was performed to three fragment libraries prepared from DNA samples. The libraries were sequenced on an Illumina HiSeq2000 platform with paired-end read length of 2×125 bp and average coverages of 12.7-15.0X. Similarly, the sequencing at the Novogene Bioinformatics Institute was performed for two samples on an Illumina HiSeqX platform with paired-end read length of 2×150 bp and average coverages of 25.9X and 30.5X.

The reads generated from the sequencing were aligned to the canFam3.1 reference sequence with the Burrows-Wheelers Aligner (BWA)[179] version 0.5.9-r16 (Studies I and II) or version 0.7.15 (Study III). Next, the reads were sorted by position and duplicates were marked with Picard tools[180] (Studies I and II) or SAMBLASTER (Study III).[181] Finally, the Genome Analysis Tool Kit (GATK)[182] version 2.3-6 (Studies I and II) or version 4.1 (Study III) was used for local realignment, followed by single nucleotide variant (SNV) and small insertion and deletion (indel) calling with the UnifiedGenotyper (Studies I and II) or the HaplotypeCaller (Study III) module. The resulting SNVs and indels were annotated with gene models from Ensembl[20] and NCBI's *Canis lupus familiaris* annotation,[19] and snpEff[183] (Study I) or ANNOVAR[184] (Studies II and III) were used to predict their functional effects.

In addition to SNVs and indels, structural variants (SV) and mobile element insertions (MEI) were called in Studies II and III. SVs, including insertions, deletions, duplications and inversions, were called with DELLY2[185] and MEIs with MELT.[186]

3.8.2 Whole-exome sequencing

Whole-exome sequencing was performed in Study III. An exome library was prepared from one DNA sample with the Roche NimbleGen SeqCap EZ target enrichment design 140702_canFam3_exomeplus_BB_EZ_HX1 with a capture size of 152 Mb.[187] The library was sequenced at the Biomedicum Functional Genomics Unit (University

of Helsinki, Finland) using an Illumina NextSeq500 platform with paired-end read length of 2×75 bp and an average coverage of 38.4X. Alignment, SNV and indel calling and annotation to the canFam3.1 reference genome were performed as described with whole-genome sequencing data using BWA[179] version 0.7.12, Picard tools,[180] GATK version 3.5.3[182] with the HaplotypeCaller module and ANNOVAR.[184]

3.9 Next-generation sequence data analysis

Next-generation sequencing data analyses, including variant filtering and downstream *in silico* analyses, were conducted in Studies I-III. The analyses were conducted jointly with co-authors in Studies II and III.

3.9.1 Variant filtering

Variant analyses were conducted in Studies I-III. The analyses were performed with an in-house Python script (Study I) and with Genotype Query Tools (GQT) command line[188] and the webGQT interface[189] (Studies II-III). DNA from the affected individuals were sequenced for the studies, and control genomes were available both in-house, sequenced for previous studies, and from the database of the DBVDC.[14]

In Study I, SNVs and indels of one CASD affected by epidermolysis bullosa were filtered against 31 unaffected controls from nine other breeds. As the disease was suspected to be inherited in a recessive manner, variants homozygous in the case and heterozygous, wild-type or absent in the controls were retained for further analysis.

In Study II, SNVs and indels of three KBD affected by pituitary dwarfism were filtered against 804 control genomes and 231 control exomes from 144 breeds or breed varieties, including eight unaffected KBD. Based on genotypes observed in the candidate regions of GWAS, one affected dog was not regarded as an obligate homozygote in the analysis; therefore, assuming autosomal recessive inheritance, variants homozygous in two cases and heterozygous, wild-type or absent in the controls were retained for further analysis. SVs and MEIs were filtered similarly, with the exception that only WGS samples were used; therefore, 256 WGS samples were utilized as controls and heterozygous genotypes were allowed in a maximum of ten control samples.

In Study III, SNVs and indels of two RW affected by congenital hearing loss (one WGS and one WES) were filtered against 537 control genomes from 105 breeds or breed varieties, including three unaffected RW. Assuming autosomal recessive inheritance, variants homozygous in the two affected dogs and heterozygous, wild-type or absent in the controls were retained for further analysis. In SV and MEI filtering, only WGS samples were used; thus, one affected genome was filtered against 290 control genomes. The affected genome was allowed both homozygous and heterozygous calls to account for inaccuracies in calling.

3.9.2 Downstream *in silico* analysis

In Studies I-III, candidate causal variants from the variant filtering were prioritised according to their functional effect and known gene function. In all analyses, exonic and splice site variants were emphasized over intronic, UTR, upstream, downstream and intergenic variants. In Study I, the analysis was focused on the 19 genes classified as EB candidate genes at the time of the study: *COL7A1*, *COL17A1*, *DSP* (desmoplakin), *DST*, *EXPH5*, *FERMT1*, *ITGA3*, *ITGA6*, *ITGB4*, *JUP* (junctional plakoglobin), *KLHL24*, *KRT5*, *KRT14*, *LAMA3*, *LAMB3*, *LAMC2*, *PLEC*, *PKP1* (plakophilin 1) and *TGM5* (transglutaminase 5). In Studies II and III, the function of all remaining genes were inspected from literature and databases, including PubMed Central[190], OMIM,[37] MGI[166], UniProt[191] and HUGO Gene Nomenclature Committee.[192]

After gene prioritization, the predicted functional impact and conservation of selected candidate variants were further inspected (Studies II and III). Predicted impact was assessed with PredictSNP[193] (Study II) or PROVEAN[194] and PolyPhen-2[195] (Study III). NNSPLICE version 0.9[196] was additionally utilized in Study II to evaluate the effect of a candidate variant on a splice acceptor site. Finally, conservation was estimated by retrieving orthologues for canine sequences either from the Entrez protein database[197] (Study II) or by utilizing NCBI's blastp web interface[198] (Study III) and alignment of the sequences with Clustal Omega[199] (Study II) or COBALT[200] (Study III).

3.10 Variant validation

Sanger sequencing was conducted in Studies I-IV, and TaqMan genotyping in Study II. The Sanger sequencing was performed by the author in Studies I and IV, and the Sanger sequencing in Studies II and III as well as TaqMan genotyping by co-authors or laboratory technicians.

3.10.1 Sanger sequencing

Sanger sequencing was performed in Studies I-IV to assess genotype-phenotype correlation of variants in the genes *COL7A1*, *POU1F1*, *LOXHD1* and *DVL2*. Standard PCR with Taq polymerase (Biotools B&M labs, Spain) was followed by treatment with exonuclease I (New England Biolabs, MA, USA) and rapid alkaline phosphatase (Roche Diagnostics, Switzerland). The following primer pairs designed with Primer3[201] were used: 5'-CTCTGCTCTTCTTCCCCAGG-3' and 5'-CCTCCCTCTGCTTACCACTG-3' (Study I); 5′-CCAGGAAAAGTGTGATCGGG-3′ and 5′-TCCATCTCCTCTGTACGTTTTG-3′ (Study II); and 5′-GCTGTGTGTTGGAGAAGCAA-3′ and 5′-TAGTTGCCTGACACCCTGAG-3′ (Study III). Primers designed by Mansour et al.[107] were used in Study IV: 5′-CGGCTAGCTGTCAGTTCTGG-3′ and 5′-CAGTGAGTCTGAGCCCTCCA-3′. Sanger sequencing was performed with an ABI 3730 capillary sequencer (Life Technologies) at the Finnish Institute for Molecular Medicine using one primer from

each pair. The resulting sequences were aligned and analysed with Sequencher version 5.3 (GeneCodes, MI, USA) or UGENE version 1.29.0.[202]

3.10.2 TaqMan genotyping

TaqMan genotyping was performed in Study II to screen a candidate causal variant in the *POU1F1* gene. The following primer and probe pairs were used: 5′-TTTGCATTGTTTTAGAAAGAAAATTTGAAACTCAAA-3′ (forward) and 5′-CCTTTTTCTTTCATTTGCTCCCACTT-3′ (reverse); and 5′-VIC-ATTCCCCATTACAGCTTT-3′ (reference allele) and 5′-FAM-CTCACCGTAGTCCCCCAT-3′ (variant allele). The reaction was carried out with a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Inc., CA, USA).

3.11 Immunohistochemistry

An immunohistochemical staining was conducted in Study I by a co-author at the Veterinary Bacteriology and Pathology Research Unit, Finnish Food Authority, Helsinki, Finland to examine the tissue-level impact of a candidate variant in the *COL7A1* gene. The staining was performed with a primary collagen VII antibody (Monoclonal Anti-Collagen Type VII Antibody, sab4200686, Sigma-Aldrich, USA). In short, skin sections of two EB-affected CASD puppies were deparaffinised and treated with proteinase K, followed by incubation with the antibody and staining with the UltraVision LP Detection System HRP Polymer/AEC Chromogen kit (Thermo Fischer Scientific Inc.). The staining was performed according to the kit protocol, excluding the hydrogen peroxide block step. Skin samples of two 0 to 2-day-old puppies with normal skin, sent for necropsy for non-EB causes, were used as controls.

3.12 Statistical analyses

Statistical analyses were conducted in Studies IV and V. All analyses were performed by co-authors or statisticians in the University of Helsinki, Finland.

In Study IV, the association of the previously described canine *DVL2* variant[107] to skeletal morphology in 19 AST was assessed. Briefly, linear models and generalized linear models were used to examine the following measurements, ratios and indices: body height and height at withers; radius length; length and height of the soft palate; length of the hard palate and skull base; ratio of hard palate length to skull base length; and facial, cranial and skull indices. These variables were explained with the dog's *DVL2* genotype, sex, and body weight if it improved the model fit. After selection of distributions and link functions, examination of outliers, evaluation of multicollinearity and assessment of linearity, an analysis of variance was performed to estimate the overall effects of the variables. Mean estimates and pairwise comparisons were obtained, and all p-values controlled for false discovery rate. The

significance cut-off value was set to $p < 0.05$. All analyses were performed using R version 3.6.2.[203]

In Study V, echocardiographic variables were compared between three clinical LEO study groups: VA-affected, VA-borderline and VA-unaffected. In short, the relationship of the 20 echocardiographic variables described earlier to background variables age, sex and body weight in 46 LEO were first analysed with linear regression and independent samples t-test. Depending on the effects of the background variables, either analysis of variance or analysis of covariance was conducted to examine the differences of the variables between the study groups. The significance cut-off value was set to $p < 0.05$. All analyses were performed using SAS version 9.4 (SAS Institute, NC, USA).

4 RESULTS

4.1 A nonsense variant in the *COL7A1* gene causes epidermolysis bullosa in Central Asian Shepherd dogs (Study I)

4.1.1 Epidermolysis bullosa in Central Asian Shepherd dogs

In Study I, two Central Asian Shepherd Dog puppies presenting severe lesions, blisters and ulcers on their skin and oral mucosa were brought to a veterinary clinic by the dog owner soon after birth (Figure 3). The puppies were diagnosed with epidermolysis bullosa, and due to the poor prognosis of the severe form of the disease, they were humanely euthanized by the owner's request and underwent a post-mortem examination with histological analyses. In periodic acid-Schiff stainings, separation of the dermis and epidermis was evident, with vacuoles of different size and cleft formation at the dermo-epidermal junction (Study I/Figure 2). The stainings also indicated that the basement membrane was present on both the roof and floor of the vacuoles, with strands crossing larger clefts. Signs of secondary inflammation, including mononuclear cells, neutrophils, necrotic debris and erythrocytes were observed, but no specific macroscopic or histological changes were present in other organs.

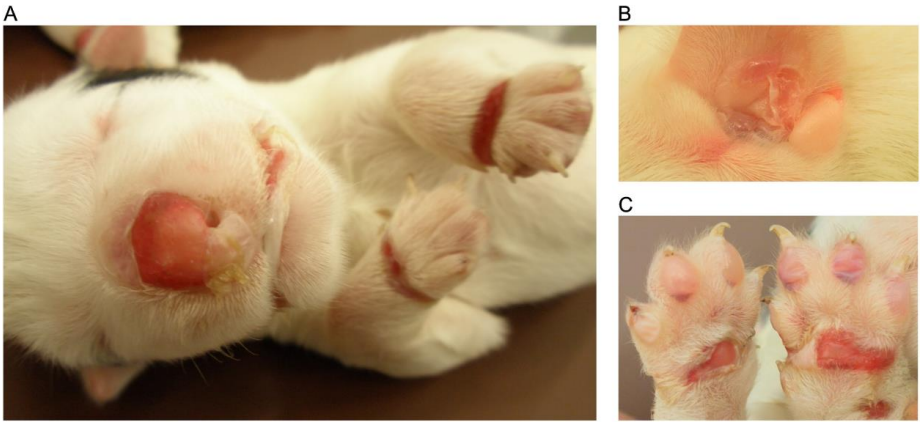

Figure 3. A Central Asian Shepherd Dog puppy affected by epidermolysis bullosa. The photographs demonstrate numerous large erupted and intact blisters A) on the nasal plane, lips and dorsal skin of the front paws B) in the inner ear lobe C) on the digital, metacarpal and carpal pads. The figure was reproduced under the CC-BY license from Niskanen et al.[204]

4.1.2 WGS analysis identifies a candidate variant in *COL7A1*

Whole-genome sequencing was performed on one affected CASD, and variants from the sequencing were filtered against 31 dogs from nine other breeds with an assumed recessive mode of inheritance. From the 6 880 938 called variants, 67 213 were specific to and homozygous in the affected dog, and 401 of them were in coding regions. The variants were further narrowed down by retaining those in known EB candidate genes, and as a result, a single candidate variant emerged. A C>T substitution was identified in *COL7A1* at chr20:40,532,043, predicted to cause a premature stop codon (c.4579C>T, p.(R1527*)) in the triple-helical domain. The genotype-phenotype correlation of the variant was initially confirmed by Sanger sequencing in the immediate pedigree of the affected puppies, followed by screening in 37 other CASD and 143 dogs from five related breeds. The variant was homozygous in the two affected dogs, heterozygous in the littermates and parents, and heterozygous or wild-type in other CASD. The five other breeds were completely wild-type for the variant. The validation therefore indicated complete segregation, suggesting that EB in the CASD breed is caused by the variant.

4.1.3 Functional experiments indicate impairment of type VII collagen in affected dogs

The effect of the *COL7A1* p.(R1527*) variant on the type VII collagen protein was confirmed by immunohistochemical staining in the skin samples of one EB-affected homozygous CASD and two unaffected non-CASD puppies. In the control skin, collagen VII staining was positive in the basement membrane; in contrast, the staining was negative in the affected puppy (Study I/Figure 5). As the antibody specifically recognizes an epitope in the C-terminal region, the lack of staining indicated that there was no full-length intact collagen VII in the skin of homozygous dogs. The C-terminal domain is essential for the dimerization of collagen VII homotrimers into anchoring fibrils, which are critical for the structural integrity of the dermo-epidermal junction.[205] Therefore, the results of the immunohistochemical experiments established a link between the disease and the variant.

4.1.4 Results confirm recessive dystrophic form of epidermolysis bullosa specific to Central Asian Shepherd Dogs

Based on the mode of inheritance and the affected gene, the results defined the disease subtype in CASD as recessive dystrophic epidermolysis bullosa. The validation further indicated that the *COL7A1* variant is specific to the CASD breed, despite their close relation to e.g. Caucasian Shepherd Dogs, and that the variant is of recent origin. Specifically, the heterozygotes and homozygotes shared a common ancestor three to four generations ago (Figure 4). Due to the family bias and small size of our cohort, the allele frequency in the CASD breed could not be accurately estimated. The allele and carrier frequencies in the entire CASD sample were 13.8 % and 27.7 %,

respectively; excluding the affected litter and their parents, the allele frequency was 6.8 % and carrier frequency 13.5 %.

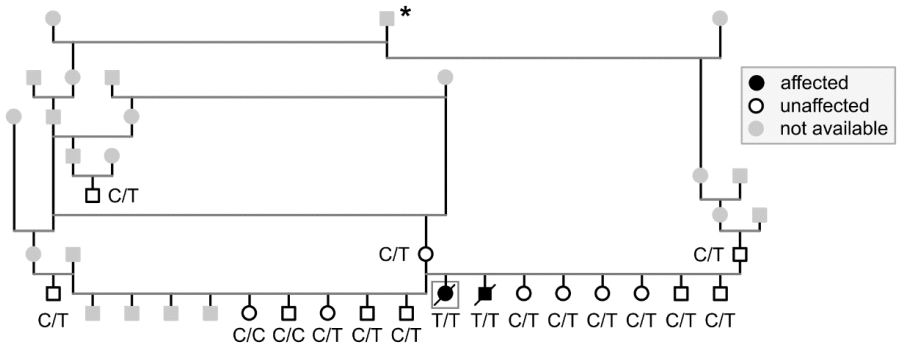

Figure 4. An extended pedigree of the CASD heterozygous or homozygous for the chr20:40,532,043C>T variant. A common ancestor of the dogs with the variant is marked with an asterisk. Females are denoted with a circle and males with a square symbol, dogs confirmed to be deceased with a diagonal line and the whole-genome sequenced sample with a grey box. Dogs from whom a sample was not available are marked with a light grey symbol.

Since our original publication in 2017, the *COL7A1* variant has been tested in 34817 purebred dogs from over 350 breeds or breed variants and over one million mixed-breed dogs on a commercial gene panel, Wisdom Panel Premium (personal communication with Jonas Donner). All dogs in both cohorts were wild-type, including three CASD, which further supports the breed specificity and recent origin of the variant.

4.2 A splice site variant in the *POU1F1* gene is associated with pituitary dwarfism in Karelian Bear Dogs (Study II)

4.2.1 Pituitary dwarfism in Karelian Bear Dogs

In Study II, clinical information of eight Karelian Bear Dogs suspected to be affected by pituitary dwarfism were collected from the dog owners. Five of these were identified before our study due to their symptoms, and three after screening of a candidate causal variant. The major characteristics of the clinical presentation were proportionate small stature and coat abnormalities (Figure 5). The typical height and weight of adult KBD are 49-55 cm (females) to 54-60 cm (males), and 17-20 kg (females) to 25-28 kg (males).[206] The affected KBD were measured at different timepoints, and the results included 0.5 kg lighter weight compared to littermates at five weeks (one dog), 37-40 cm and 7-10 kg compared to littermates of 20 kg at six months (three dogs), and 34 cm and 4 kg at 9.5 months (one dog). One dog was

measured for serum insulin-like growth factor 1, which was below reference range (83 µg/l, reference 130-180 µg/l), thus confirming the diagnosis of pituitary dwarfism.

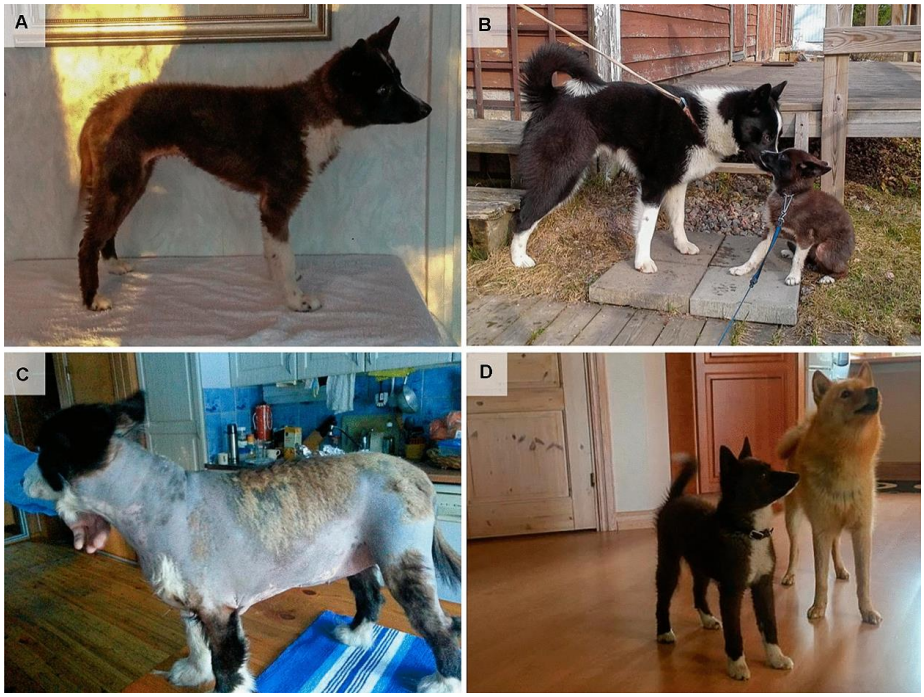

Figure 5. Photographs of four Karelian Bear Dogs suspected to be affected by pituitary dwarfism. All of the dogs exhibited small stature and coat abnormalities. A) Female, 9 months old. B) Female, 10 months (right) and an unaffected male littermate (left). C) Female, adult. D) Female, 6 months (left) and a dog from another breed (right). The figure was reproduced under the CC-BY license from Kyöstilä et al.[68]

In addition to the major findings, other symptoms were variable and not shared between all affected dogs. They included increased blood urea and normal thyroid hormone levels (one dog), hypothyroidism (one dog), early-onset blindness or sight problems (two dogs) and seizures and internal hydrocephalus (one dog).

4.2.2 GWAS reveals associated loci

To map loci associated with pituitary dwarfism in KBD, a GWAS was conducted. After pre-analytical QC, the cohort included five cases and 139 controls as well as 96 274 markers. A basic case-control association with PLINK 1.9[131] revealed two statistically significant loci on chromosomes 9 and 31 (Study II/Figure 3). First, the chromosome 9 signal spanned 3.5 Mb with eleven markers at 9:32,542,006-36,040,587 and three top SNPs at 9:34,197,651, 9:34,236,659 and 9:34,275,570 ($P_{GC} = 1.59\times10^{-10}$). Similarly, the chromosome 31 signal extended over a 6.0 Mb region at 31:2,799,588-8,844,745 with fifteen markers, including the top SNP at 31:4,769,865

($P_{GC} = 6.83\times10^{-15}$). Inspection of genotypes at the loci revealed a homozygous block shared by all cases on chromosome 31 but not on chromosome 9, where one of the affected dogs presented a different haplotype. Therefore, the signal at chromosome 9 was likely influenced by the notable stratification of the study population as indicated by a high lambda value (1.88 before genomic control correction).

4.2.3 WGS analysis suggests a candidate variant in *POU1F1*

The associated loci discovered with GWAS were investigated in detail with a WGS analysis. Whole-genome variants shared by two affected KBD were filtered against 804 whole genomes and 231 whole exomes (SNVs/indels) or 256 whole genomes (SVs/MEIs), which resulted in the discovery of 321 SNVs and indels, two SVs and three MEIs. Of these, 14 SNVs/indels resided on chromosome 9 and 281 SNVs/indels, two MEIs and the two SVs on chromosome 31. The variants were prioritized by predicted functional effect, which resulted in the inclusion of five exonic variants on chromosome 9 and one splice site variant on chromosome 31 in further analyses.

The impact of the exonic and splice site variants were further assessed. The splice site variant resided in *POU1F1* (POU class 1 homeobox 1, also known as pituitary transcript factor 1) at chr31:784,534 (c.605-3C>A). The variant was predicted to weaken the intron 4 splice acceptor from a score of 0.97 to 0.67, and its conservation was evaluated in the dog and 18 other mammalians, which revealed either C or T in all sequences and suggested that A might not be tolerated at this position. Finally, as *POU1F1* is known to cause CPHD in humans[87] and proportional dwarfism in mice,[207,208] the variant was retained for further analysis.

Of the five exonic variants, two were synonymous and predicted not to affect splicing, two were missense and predicted to be neutral, and one in-frame deletion/insertion variant, called as four separate variants by the pipeline, did not affect any conserved amino acids. The affected genes, *ANKFN1* (ankyrin repeat and fibronectin type III domain containing 1), *LPO* (lactoperoxidase) and *TSPOAP1* (TSPO associated protein 1) are not known to be associated with pituitary phenotypes or function. Furthermore, the variants were wild-type in the affected dog that did not present the homozygous block on chromosome 9; thus, they were classified as likely not disease-causing and discarded from further analyses.

4.2.4 Cohort screening supports association of *POU1F1* with pituitary dwarfism

The segregation of the splice site variant in *POU1F1* was investigated in a larger cohort with Sanger sequencing and TaqMan SNP genotyping. The variant was genotyped in 642 KBD, of which 572 were wild-type, 62 heterozygous and eight homozygous. All of the homozygotes presented symptoms, indicating a complete segregation of the variant with the disease. Excluding the affected dogs and their close relatives (N=23), the allele and carrier frequencies in our KBD cohort were 4 % and 8 %, respectively.

In addition to KBD, closely related breeds were also screened. These included 263 Lapponian Herders, 50 East Siberian Laikas, 17 West Siberian Laikas and 20 Russo-European Laikas. Heterozygotes were only discovered in Lapponian Herders, with seven identified in this screening and one in the WGS control genomes, and the other breeds were completely wild-type for the allele.

Finally, to supplement the Sanger and TaqMan screening, we inspected the variant in a cohort of 7925 dogs from 206 breeds submitted for commercial genetic testing. In the entire cohort, only one heterozygous KBD was found, and the other dogs were wild-type, indicating that the allele is absent or very rare in other breeds than KBD and Lapponian Herders.

4.3 A missense variant in *LOXHD1* is associated with congenital nonsyndromic hearing loss in Rottweilers (Study III)

4.3.1 Nonsyndromic congenital hearing loss in Rottweilers

In Study III, four RW in a litter of ten were examined with a BAER test due to suspected hearing impairment. The results of the BAER confirmed that the dogs, tested at the age of four (N=2), five or 19 months, had no auditory response, which indicated profound hearing loss. The dog owners reported signs of hearing impairment already when the dogs were a few weeks old, which suggests that the hearing loss was congenital. No other clinical signs were observed. The parents of the litter had normal hearing, suggesting a recessive mode of inheritance.

4.3.2 Homozygosity mapping identifies candidate regions

To map regions associated with hearing loss in RW, the four affected puppies and three unaffected RW were SNP array genotyped to conduct homozygosity mapping. Due to poor call rate, one case was discarded from the analysis, and after pre-analytical QC, three cases, three controls and 154 235 markers remained. Case-specific allelically shared ROH were detected with PLINK 1.9[131], resulting in 22 regions on 12 different chromosomes with a total length of 62.3 Mb (Study III/Figure 1).

4.3.3 NGS analysis suggests a candidate variant in *LOXHD1*

Candidate causal variants for the hearing loss in RW were identified with NGS analysis. The SNVs and indels shared in homozygous state by two affected RW (one WGS and one WES) were filtered against 537 control genomes from 105 breeds or breed varieties. SVs and MEIs of the WGS case were similarly filtered against 290 control genomes. This resulted in the discovery of 32 SNVs/indels, 63 SVs and 32 MEIs, of which six SNVs/indels and one MEI resided in case-specific ROH. These

variants were further prioritized for predicted functional effect; therefore, two exonic SNVs were retained for further analysis.

First, a C>T missense variant at chr7:44,806,821 was predicted to result in a glycine-to-alanine substitution (p.G1914A) in the *LOXHD1* (lipoxygenase homology domains 1) gene, which is known to cause hearing loss in humans and mice.[209] Pathogenicity assessment of the variant classified it as "deleterious" (PROVEAN, -4.517) and "probably damaging" (PolyPhen-2, 0.992). Second, a C>T missense variant at chr24:785,932 in the *MROH8* (maestro heat like repeat family member 8) gene was predicted to cause a glycine-to-serine substitution. The pathogenicity predictions for the variant were "deleterious" (PROVEAN, -3.336) and "possibly damaging" (PolyPhen-2, 0.550). As *MROH8* has been linked to red blood cell volume, body mass index, telomere length and hippocampal atrophy in humans[210] but not any hearing-related function or phenotype, and the predicted pathogenic impact of the variant was slightly lower, it was discarded from further analysis.

Finally, we evaluated the conservation of the *LOXHD1* variant in 99 Eutherian protein orthologues, including dog. Multiple alignment of the sequences revealed complete conservation of the affected residue (G1914) and its flanking amino acids, indicating that a G-to-A substitution might impair the normal function of the LOXHD1 protein.

4.3.4 Cohort screening confirms *LOXHD1* segregation and indicates a Rottweiler-specific rare variant

To investigate the segregation of the *LOXHD1* variant with hearing loss, we genotyped the variant in 585 RW with Sanger sequencing. In the screening, 548 dogs were wild-type, 33 dogs heterozygous and the four affected dogs homozygous, indicating complete segregation. Excluding the affected family, the allele and carrier frequencies in the study population were 2.6 % and 5.3 %, respectively.

The variant was further screened in large multi-breed cohorts submitted for commercial genetic testing to estimate its prevalence across breeds. First, the alternate allele was not found in 28 116 purebred dogs or designer dog mixes from 374 breeds or breed varieties. In contrast, allele and carrier frequencies of 0.04 % and 0.08 %, respectively, were discovered in a cohort of 771 864 dogs. Specifically, six homozygous dogs were found in the screening, and owners of 4/6 dogs could be contacted for further information. The four dogs were all reported to have hearing loss or deafness; furthermore, one of them was a purebred RW and two had RW ancestry. In general, 63.4 % of the dogs with at least one copy of the alternate allele showed direct evidence of RW ancestry in their three-generation pedigrees. Altogether, these results indicate that the *LOXHD1* variant is very rare and connected to the RW breed.

Finally, in addition to the hearing phenotype, *LOXHD1* variants in heterozygous state specifically have been suggested to contribute to late-onset Fuchs endothelial corneal dystrophy (FECD) in humans.[211] To inspect our cohort for any signs of FECD, we retrospectively reviewed the results of routine eye examinations available from 2/4 homozygotes and 22/33 heterozygotes. The two homozygotes had been found healthy at 2 and 8 years, and of the heterozygotes, examined between 1 and 7 years old, three

had been diagnosed with different forms of cataract, a fairly common eye disease in the breed. Corneal dystrophy was not found in any of the examined dogs, which suggests that *LOXHD1* heterozygosity is not linked to FECD in RW.

4.4 *DVL2* variant causes caudal vertebral anomalies and contributes to brachycephalic phenotype in dogs (Study IV)

4.4.1 Cohort screening uncovers variable allele frequencies in Bulldog and Pit Bull type breeds

In Study IV, we screened the previously identified *DVL2* variant[107] in a cohort of 1954 dogs from 15 breeds to investigate its prevalence (Figure 6, Study IV/Table 1). The results revealed complete fixation of the alternate allele in English Bulldogs (N=297), French Bulldogs (N=211) and Boston Terriers (N=165), supporting previous findings. Both homozygotes and heterozygotes were discovered in ASTs, with five homozygotes, 79 heterozygotes and 201 wild types, and in SBTs, with one homozygote, 135 heterozygotes and 578 wild types. Of note, the genotypes observed in SBTs were not in Hardy-Weinberg equilibrium ($P < 0.015$). Additionally, the allele was also present in Dogues de Bordeaux, American Bulldogs and Olde English Bulldogges. In the five breeds with allele variation, carrier frequencies and cohort sizes varied considerably, from 18.2 to 50.0 % carriers (N=4-714). Finally, only wild types were found in Bull Terriers and Miniature Bull Terriers, Lhasa Apsos, Shih Tzus, Tibetan Spaniels, Pekingese and King Charles Spaniels (N=13-47).

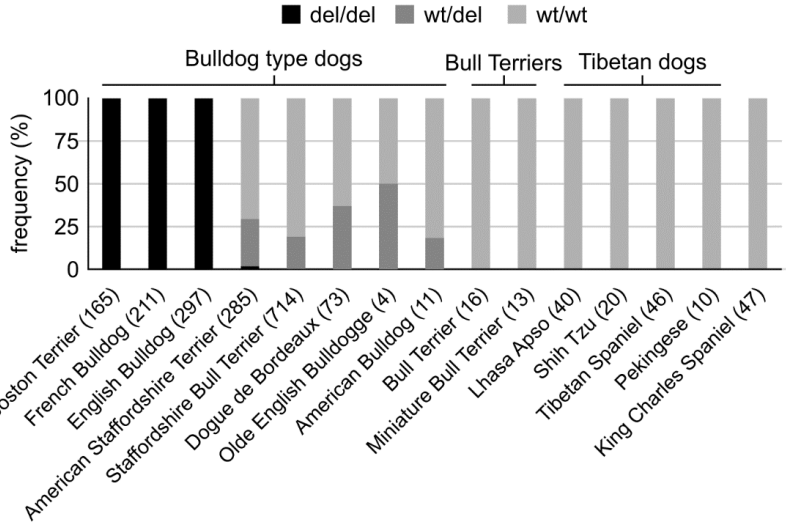

Figure 6. Relative frequencies of *DVL2* genotypes in 15 dog breeds. Sample sizes in each breed are indicated in parentheses.

4.4.2 Clinical examinations reveal fully penetrant caudal vertebral anomalies in *DVL2* homozygotes

The discovery of allelic variation enabled us to examine the qualitative and quantitative phenotypic consequences of the *DVL2* allele on skeletal morphology. Through full-body CT examinations of 19 AST (eight wild types, eight heterozygotes and three homozygotes) and retrospective review of radiographs of 23 AST (21 wild types, one heterozygote and one homozygote), we documented vertebral abnormalities. In all four homozygotes, we observed variable caudal vertebral anomalies, including block vertebra, butterfly vertebra, dorso-lateral hemivertebra, lateral hemivertebra, ventral wedge shape, unclassified congenital malformation and abnormally short vertebra. The number of malformed caudal vertebrae per dog was 1, 8, 10 and 14. Some of these anomalies were also seen in 3/8 heterozygotes and 2/8 wild types. Finally, one homozygote had malformations in the sacrum, and thoracic or lumbar vertebral anomalies were not found any dogs in the CT and radiograph study populations.

To complement the CT and radiographic studies in ASTs and evaluate tail anomalies in *DVL2* heterozygotes and wild types, we assessed owner reports of ASTs, SBTs, Dogues de Bordeaux, Olde English Bulldogges and American Bulldogs. We also performed a CT examination on one wild-type SBT with a short and kinked tail and analysed radiographs of three French Bulldog × German Shepherd Dog crosses, two of which were available for DNA sampling and heterozygous for *DVL2*. Combining the CT studies, radiographic studies and owner reports, a total of 12 heterozygotes and 13 wild types from five breeds presented evidence of caudal vertebral anomalies (Study V/Table 2). All three of the French Bulldog × German Shepherd Dog crosses had normal spines and full-length tails without any vertebral anomalies.

In addition to vertebral anomalies, two homozygotes in the CT study presented asymmetrical frontal sinuses and a mild heart murmur. Both had undergone echocardiography prior to our study: one of them had been diagnosed with aortic stenosis and the other, though findings within the normal range, had been recommended a follow-up study by a cardiologist. Finally, the owner of one homozygote unavailable for a CT examination reported a mild, asymptomatic heart murmur.

4.4.3 Statistical analyses confirm quantitative effect of the *DVL2* variant on proportions of the skull

Body and craniometric measurements recorded from the CT examined dogs were analysed with linear models and generalized linear models to dissect the quantitative effects of the *DVL2* variant on skeletal morphology. After accounting for background factors, *DVL2* genotype was found to significantly affect hard palate length, skull base length, ratio of hard palate length to skull base length, facial index and soft palate height (Figure 7). The most significant effects were observed in hard palate length, where differences were found between all genotypes: heterozygotes had a shorter hard palate than wild types, and homozygotes similarly had a shorter hard palate than

heterozygotes. In skull base length and facial index, homozygotes differed from heterozygotes and wild types, and the ratio of hard palate length to skull base length showed differences between wild types and other genotypes. Finally, soft palate height differed between homozygotes and the other genotypes, but no difference was found in soft palate length.

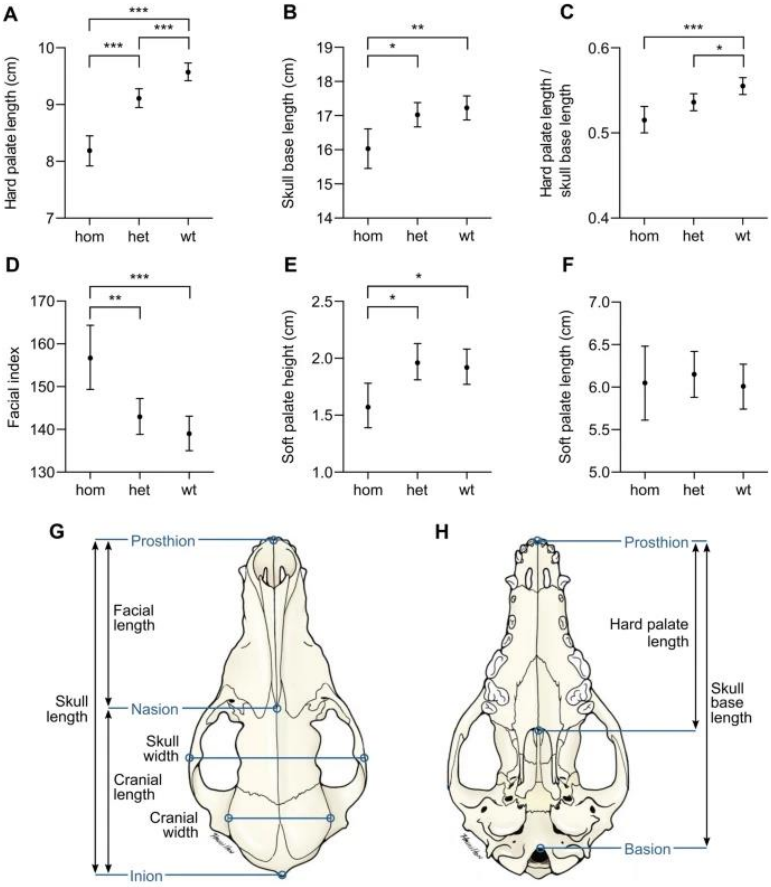

Figure 7. Association of *DVL2* genotype with craniometric measurements, ratios, and indices, and schematic representations of a dog skull. The figures A-E demonstrate the association of the genotype with A) hard palate length, B) skull base length, C) ratio of hard palate length to skull base length, D) facial index and E) soft palate height. F) Soft palate length was not associated with *DVL2* genotype. The measurements are represented in the diagrams of G) dorsal and H) ventral views of a dog skull. The figure was reproduced under the CC-BY license from Niskanen et al.[212]

4.4.4 Results suggest differential effects of *DVL2* alternate allele copy number on caudal vertebral malformations and skull proportions

Based on the results of the clinical and statistical analyses, the *DVL2* variant seems to act on the caudal vertebral and craniometric phenotypes differentially depending on alternative allele copy number. Caudal vertebral anomalies were fully penetrant in homozygotes, but the frequency of the anomalies in heterozygotes was similar to that of wild types, indicating that *DVL2* heterozygosity does not confer major risk to abnormal caudal vertebral development. In contrast, proportions of the bony skull were variably different between the three genotypes, with the strongest effect on hard palate length, where all genotypes differed significantly. Taken together, the effect of the allele may be tissue-dependent. Finally, it is notable that soft palate length did not differ between genotypes, indicating that the soft palate is proportionally longer in heterozygous and homozygous dogs.

4.5 Ventricular arrhythmia and sudden cardiac death in juvenile Leonberger dogs (Study V and unpublished data)

4.5.1 Clinical and post-mortem examinations confirm ventricular arrhythmia and sudden cardiac death without prior symptoms

Study V was initiated upon contact with a dog breeder that reported sudden unexplained deaths of several young Leonberger dogs. To inspect the deaths in the breed, we retrospectively reviewed necropsy reports of LEO examined during years 2008-2018 and discovered 21 dogs (8 females, 13 males) that had died at less than three years of age with findings indicating sudden cardiac death. In the affected dogs, acute congestion and oedema of the lungs and other organs, mainly the liver, was observed. The hearts were macroscopically normal, with no specific histopathological changes. Other significant organ changes were not present. Therefore, arrhythmia was determined as the likely cause of death.

To investigate whether cardiac abnormalities occur in the breed, we recruited 46 LEO less than four years old to prospective clinical examinations. Recruitment was prioritized on families in which SCD had previously occurred. Through 24-hour Holter monitoring, ventricular arrhythmia was discovered in 14 dogs. The arrhythmias were variable and consisted of VPCs, couplets, triplets and/or ventricular tachycardia; furthermore, there were differences both between affected dogs and within the first examinations and follow-ups of single dogs. Based on the findings, the dogs were classified into three Holter groups: VA-affected (N=7, >100 VPCs/24 hours and/or multiple couplets, triplets, or >1 non-sustained VT period), VA-borderline (N=7, <100 VPCs/24 hours and one to two couplets/triplets but no VT) and VA-unaffected (N=32, <50 VPCs/24 hours and no couplets, triplets or VT). Notably, two dogs in the VA-affected group with VT died during follow-up: one was euthanized at 21 months due

to acute collapse with signs of shock an arrhythmia after strenuous exercise, and the other died suddenly at 32 months.

The cohort was further assessed for differences between the Holter groups. No significant findings were revealed in heart rate, RR, PQ or QT intervals, echocardiographic variables, physical examination and auscultation, blood cell count or biochemistry profile. Females and males were equally affected.

Together, the clinical and post-mortem examinations confirm the phenotype as an electrical abnormality in the absence of structural cardiac disease. Thus, the disorder in Leonberger dogs was defined as idiopathic ventricular arrhythmia and subsequent sudden cardiac death.

4.5.2 Pedigree analysis and high prevalence indicate a genetic background

To determine the mode of inheritance for the arrhythmias and sudden cardiac death, we inspected the pedigrees of the clinical and post-mortem study cohorts (Figure 8). The pedigrees comprised 65 dogs and 39 litters, of which 2/39 included both SCD and VA cases, 17/39 included SCD cases only, and 4/39 included VA cases only; finally, 16/39 litters were unaffected. Full litters were not available for clinical studies, and the mean number of phenotyped siblings per litter was 1.67 (minimum: 1, maximum: 6). However, four litters included multiple affected siblings, and many of the cases were otherwise close relatives. Unfortunately, the parents of the affected litters were not available for clinical examinations at the typical age of onset, and multigenerational phenotypes could not be acquired. Therefore, it was not possible to establish a likely mode of inheritance. Nevertheless, the high breed-specific prevalence together with multiple affected siblings per litter indicate that the arrhythmia and SCD are of genetic origin.

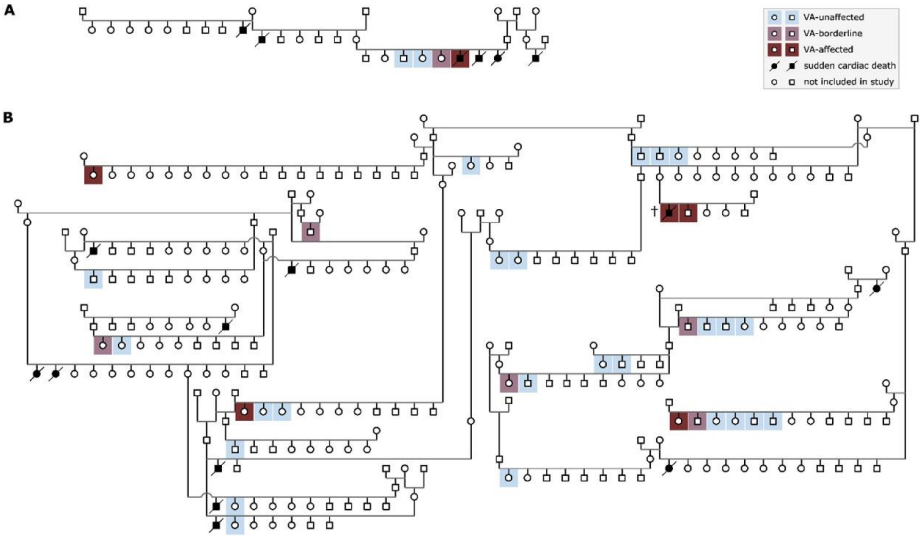

Figure 8. Two extended pedigrees (A and B) of the clinical and post-mortem Leonberger cohorts. The pedigrees illustrated here do not include the entire cohorts. Dogs unaffected by ventricular arrhythmia are indicated in light blue, those in the borderline group in pale red and those affected by ventricular arrhythmia in dark red. Dogs that died a sudden cardiac death are indicated in black and with a diagonal line. One affected dog marked with a dagger (†) was euthanized. The figure was reproduced under author's rights from Wiberg et al.[213]

4.5.3 GWAS fails to reveal associated loci

As the likely mode of inheritance could not be established, we sought to map disease-associated loci with GWAS (unpublished data). The analyses were performed with two cohorts of affected dogs: first, with VA-affected cases only; and second, with both VA-affected and SCD cases. With the VA-affected dogs, a significant association was observed on chromosome 37, with two significant SNPs at chr37:13,758,816 and chr37:13,769,994 (p_{Meff}=0.026, Figure 9A,C). However, addition of the SCD cases failed to replicate the locus at chr37:13.7 Mb (p_{Meff}=0.80, Figure 9C); instead, three SNPs at chr2:1,086,929, chr5:78,301,761 and chrX:43,900,465 reached statistical significance (p_{Meff}=0.043, Figure 9B). Other significant SNPs were not observed in the immediate surroundings of these three SNPs, indicating that the significance was due to genotyping artefacts caused by batch effect rather than a true associated haplotype (Figure 9D). Finally, considering that the Holter and post-mortem phenotypes are of the same phenotypic continuum and that the chromosome 37 locus should be expected to be replicated if based on a true biological association, it is likely that the statistical significance in the first analysis was largely influenced by stratification.

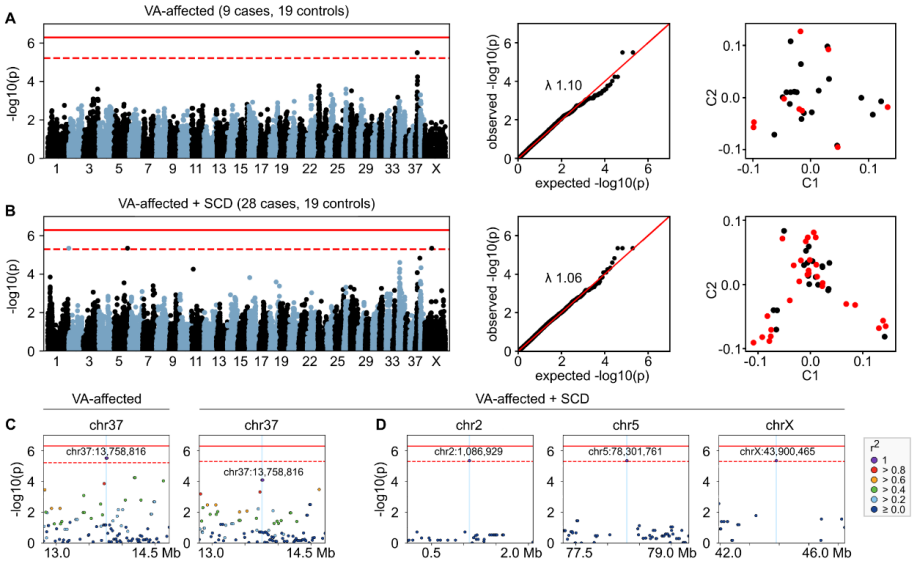

Figure 9. Results of a univariate linear mixed model GWAS in 47 LEO. In the Manhattan and locus plots, solid and dotted red lines indicate Bonferroni and M_{eff} corrected thresholds, respectively. In the MDS plots, cases are indicated in red and controls in black. SNP correlation (r^2) in images C and D was calculated with PLINK1.9. A) Manhattan, QQ and MDS plots of an analysis with nine VA cases and 19 unaffected controls. B) Manhattan, QQ and MDS plots of an analysis with nine VA cases, 19 SCD cases and 19 controls. C) Locus plots of chr37:13.0-14.5 Mb in both analyses. D) Locus plots of chr2:0.0-2.0 Mb, chr5:77.5-79.0Mb and chrX:42.0-46.0 Mb in the VA + SCD analysis.

Due to the failure of the association analyses, we inspected the distribution of the markers in the unpruned and pruned datasets in detail, focusing specifically on the analysis with both the VA-affected and SCD cases. During QC, 43.3 % of the markers were removed, mostly due to the minimum MAF threshold of 5 %, which predictably resulted in decreased marker density throughout the genome (Figure 10). Mean density across all chromosomes was 13.5 kb/SNP in the unpruned data and 23.7 kb/SNP in the pruned data. Conversely, by dividing the genome into bins of 1 Mb, we observed an average of 73.8 and 41.9 SNPs per bin in the unpruned and pruned datasets, respectively. In intermarker distances, the pruned dataset also showed higher median distance and more variation, with a median of 15.19 kb in the unpruned data and 11.99 kb in the pruned data. Additionally, the longest distances were much larger in the pruned data (Figure 10B). For example, a region spanning over 2.7 Mb was present in chromosome X. These expected differences illustrate the large regions completely devoid of markers that can be generated during standard QC.

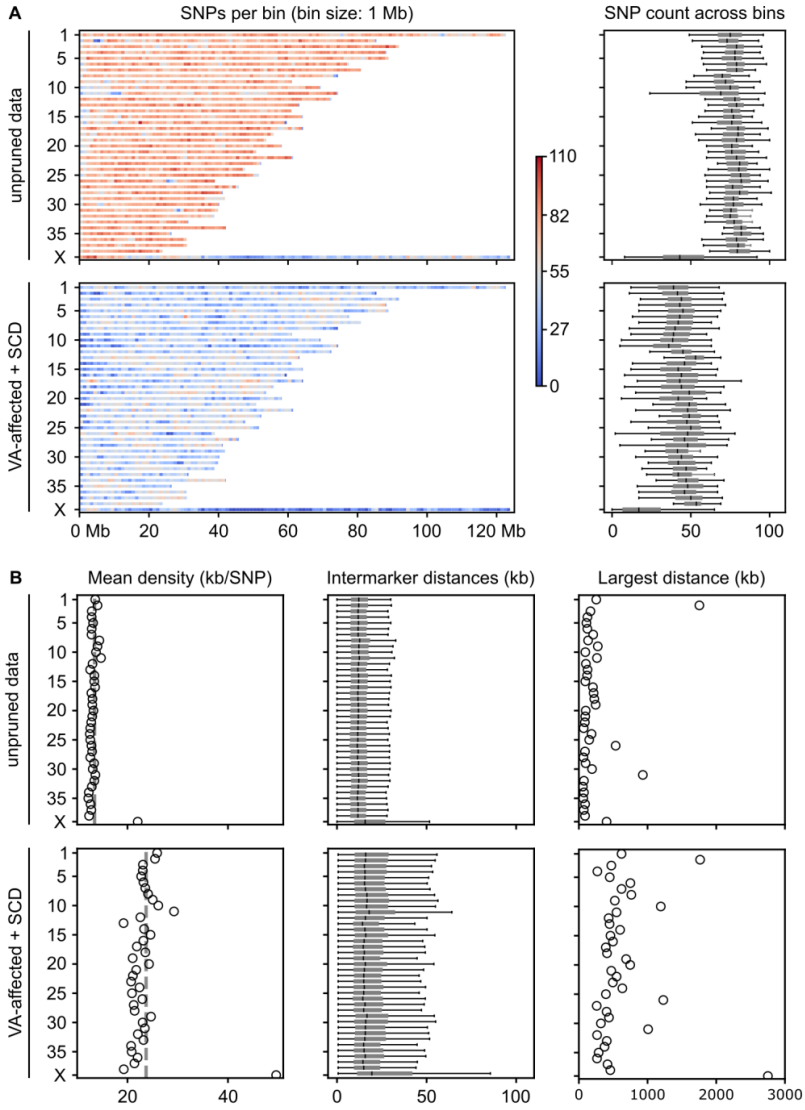

Figure 10. Marker counts and densities in each chromosome in the unpruned (172 963 markers) and pruned (98 137 markers) datasets. The analysis with VA-affected and SCD cases was utilized as the pruned dataset. A) Count of markers in bins of 1 Mb in relation to chromosomal coordinates (left), and distributions of SNP counts per bin (right). The colours indicate the number of markers in each bin. B) Mean densities (left), distributions of intermarker distances (middle) and largest intermarker distances (right) in each chromosome. The dashed lines in the leftmost image indicate the mean density across the entire genome.

5 DISCUSSION

In this thesis, novel canine models of rare human disease and causal genetic variants were characterized: *COL7A1*-related dystrophic epidermolysis bullosa (Study I), *POU1F1*-related pituitary dwarfism (Study II), *LOXHD1*-related congenital hearing loss (Study III), *DVL2*-related Robinow-like syndrome (Study IV) and malignant ventricular arrhythmias and sudden cardiac death (Study V). The aetiologies of the disorders in studies I-III and V were previously unknown, and we hypothesized a genetic background based on their high prevalence within dog breeds and families. In Study IV, the underlying genetic variant in *DLV2* was previously described but the genotype-phenotype correlations of the disease allele remained unclear, and we hypothesized that the morphological consequences of the allele could be dissected with additional variant screening conducted in conjunction with detailed clinical examinations. Our hypotheses were fully confirmed in Studies I-III, where we identified causal variants, and in Study IV, where we revealed the differential cross-breed distributions of the variant and illustrated its phenotypic consequences on skeletal morphology. In Study V, our hypothesis was partially confirmed as we showed strong evidence for a genetic aetiology.

The identification of novel causal genetic variants in this thesis is a major breakthrough for the diagnostics of the disorders in the affected breeds. Several new gene tests have been developed based on the findings, helping veterinarians to unambiguously diagnose the conditions and breeders to improve breeding programmes to eradicate the conditions from the breeds. We demonstrate the impact of industrial collaboration with unprecedented massive screening of candidate variants in over 800 000 dogs to validate our findings and understand their true significance. The affected breeds may also serve as important complementary animal models for the corresponding human conditions by enabling additional functional or preclinical studies. For example, the arrhythmia-affected Leonbergers, clinically characterized for the first time in this thesis, may highlight novel arrhythmia candidate genes in future genetic studies and provide a large animal model for ventricular arrhythmias and tachycardia.

Together, these studies highlight the remarkable potential of canine models in dissecting the genetic background of rare human disorders and emphasize the power of canine populations in gene mapping and NGS approaches. They also demonstrate the importance of essential infrastructure, including the canine biobank and computational pipelines, and interdisciplinary collaboration that brings together genetics, veterinary medicine and bioinformatics. Our results have implications on both human and veterinary medicine by providing new knowledge about the molecular mechanisms of these disorders, shedding light on their pathogenesis and enabling the improvement of diagnostics and treatment. Finally, we contribute critical new information on the genetic and clinical health of various dog breeds to support ethical discussion on breeding programmes and their implications to dogs' welfare.

Mapping and sequencing methods pinpoint causal variants in three canine models of rare disease

Epidermolysis bullosa

In Study I, we utilized NGS analysis in conjunction with clinical and pathological examinations to characterize severe dystrophic epidermolysis bullosa in dogs for the first time, specifically in Central Asian Shepherd Dogs.[204] Epidermolysis bullosa in humans is a well characterized group of genetically and clinically heterogeneous, rare genetic dermatoses.[214] The disorder is similarly heterogeneous in dogs, and three forms of EB had been described prior to our study: EB simplex in Eurasiers and Collies,[76,77] junctional EB in German Pointers[78] and intermediate and mild dystrophic EB in Golden Retrievers and Akita Inu, respectively.[81,83] In CASD, the disorder is caused by a nonsense variant, p.(1527*), in the *COL7A1* gene (Figure 11).

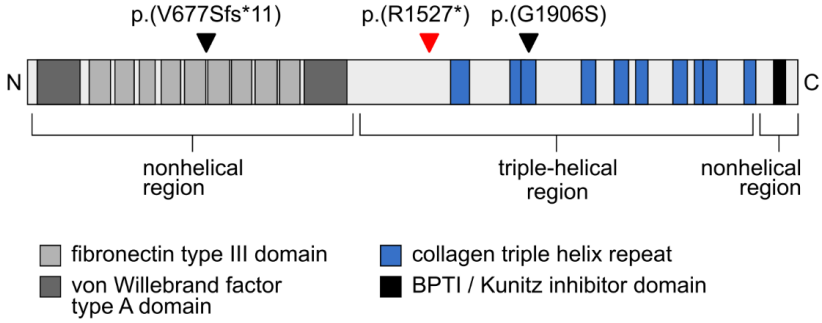

Figure 11. A schematic representation of the domain structure of the canine COL7A1 protein (NCBI ID: NP_001002980.1) and variants associated with dystrophic epidermolysis bullosa. The CASD variant, p.(R1527*), is indicated with a red triangle and the Basset Hound and Golden Retriever variants, p.(V677Sfs*11) and p.(G1906S), with black triangles. Domains for NP_001002980.1 were retrieved from the Conserved Domain Database.[215]

The tissue-level consequences of the predicted truncating p.(R1527*) variant were confirmed with an immunohistochemical staining, which showed the absence of full-length type VII collagen in the skin of an affected puppy (Study I/Figure 5). Together with the extensive and severe early-onset phenotype, this indicates that the disorder in CASD most appropriately corresponds to the Hallopeau-Siemens recessive dystrophic EB subtype described in human patients with *COL7A1* variants.[216] This subtype is most commonly caused by premature stop codons,[216] which is in in line with our discovery in the CASD. Similarly, a frameshift variant leading to a premature stop codon, p.(V677Sfs*11), underlies severe recessive dystrophic EB in Basset Hounds, although its predicted truncating effect on the collagen VII protein was not confirmed with functional experiments.[82] In contrast, the p.(G1906S) missense variant in Golden Retrievers causes an intermediate phenotype, corresponding to that between Hallopeau-Siemens and mild forms of EB.[217] These clinical manifestations reflect the

differential contribution of nonsense and missense variants to phenotypes of varying severities, as observed in human patients.[216]

The poor prognosis of the affected puppies significantly limits the potential of the CASD model in preclinical studies. However, there is great interest in studying potential EB therapies in large animal models. While Golden Retrievers with EB have been used to investigate the efficacy of retroviral-corrected epidermal autographs and intravenous recombinant collagen VII injections,[218,219] the missense variant that underlies the pathophysiological mechanism does not fully correspond to the premature stop codons that are more commonly seen in the severe Hallopeau-Siemens form in humans and in CASD. Therapies specific to EB caused by a premature stop codon in *COL7A1* are being developed, such as the drug amlexanox that has been found to induce read-through of the stop codon in affected fibroblast and keratinocyte cultures;[220] however, the extent of the blistering and erosions in the CASD puppies at birth challenges potential intervention studies. Finally, the topic is accompanied by substantial ethical and moral questions on animal welfare: in addition to providing new understanding on pathophysiology of EB, the fundamental aim of the study was to enable the complete and permanent eradication of this crippling disease from the CASD breed.

Pituitary dwarfism

In Study II, we used GWAS and NGS analysis to identify a putative splice variant in *POU1F1* associated with pituitary dwarfism in Karelian Bear Dogs,[68] a condition that was first described 44 years prior to our study.[91] An association analysis with five affected and 139 unaffected KBD highlighted two genome-wide significant loci on chromosomes 9 and 31 (Study II/Figure 3), and case-specific variants were subsequently analysed in these regions. Though the GWAS revealed two associated regions, one of the affected dogs did not share the same haplotype and filtered variants as the four other cases on chromosome 9 (Study II/Table 1). Therefore, the locus on chromosome 31 and a singular case-specific splice site variant, c.605-3C>A, in *POU1F1* were prioritized.

POU1F1, also known as *PIT-1*, codes for a protein that acts as a transcription factor and is involved in the development of the anterior pituitary gland.[221] In human patients with combined pituitary hormone deficiency, more than 25 variants of significance have been catalogued in *POU1F1*, including some that disrupt splicing.[222,223] These patients are primarily characterized by severe growth retardation and distinctive facial features, including mid-facial hypoplasia, a prominent forehead and a depressed nasal bridge.[222] The affected KBD similarly exhibited proportionate small stature but also an abnormal, "puppy-like" coat with or without hair loss.

The effect of the canine c.605-3C>A variant was predicted to weaken the splice acceptor site at the junction of intron 4 and exon 5 (Figure 12A). Human studies indicate that *POU1F1* is exclusively expressed in the anterior pituitary,[224] which was not available from the affected KBD to confirm the prediction with functional experiments. Nevertheless, the evidence from the association and NGS analyses as well as the phenotypical parallels between human patients and the KBD support the pathogenicity of the variant.

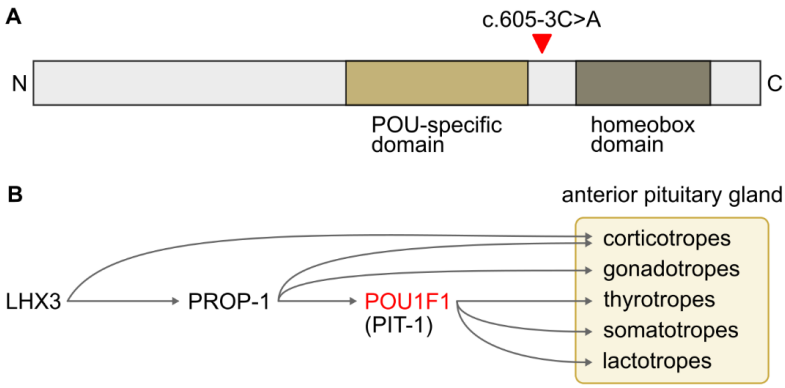

Figure 12. A) A schematic representation of the domain structure of the canine POU1F1 protein (NCBI ID: NP_001006950.1). The putative splice variant c.605-3C>A observed in KBD is indicated with a red triangle. The domains for NP_001006950.1 were retrieved from the Conserved Domain Database.[215] B) A simplified schematic representation of the involvement of POU1F1 and LHX3 in the development of anterior pituitary gland. The genes regulate the differentiation of specialized, hormone-secreting cell populations by acting as transcription factors. Intermediate regulators in the pathway were omitted from the image for clarity. Image B was adapted from Prince et al.[225] and Pavlović et al.[226]

The KBD model marks the second molecular characterization of pituitary dwarfism in dogs. The disorder has also been described in German Shepherd Dogs, Saarloos wolfdogs, Czechoslovakian wolfdogs and Tibetan Terriers, where the causal variant is a 7-bp deletion in an intronic repeat region in the *LHX3* gene.[88–90] LHX3 and POU1F1 are both involved in the pathway that regulates the development of pituitary gland and differentiation of the specialized hormone-secreting cells, including somatotropes that produce growth hormone (Figure 12B).[225] This is reflected in the similar symptoms of the affected dogs: like the KBD, the German Shepherd Dogs, Saarloos wolfdogs, Czechoslovakian wolfdogs and Tibetan Terriers are also characterized by growth retardation, failure to transition into adult coat and alopecia.[89,90,227] The levels of growth hormone and thyroid-stimulating hormone are low in affected German Shepherd Dogs and Saarloos and Czechoslovakian wolfdogs.[89,228] A full phenotypic comparison with the affected KBD was not possible due to lack of hormone measurements for most of the affected dogs; however, one KBD was measured for and showed low levels of serum insulin-like growth factor 1, a hormone whose excretion is induced by growth hormone.[229] Finally, thyroid-stimulating hormone showed variation in the KBD: one case had been diagnosed with hypothyroidism prior to our study whereas another case showed a normal level of thyroid hormone at the age of 13 weeks.

Together, our results demonstrate the molecular mechanisms of pituitary dwarfism in KBD and establish a basis for genetic testing in the breed. Ultimately, the revision of breeding programmes with the help of genetic testing will enable the eradication of

the disease from the affected breed. In the meanwhile, the affected dogs provide a new preclinical model to improve treatment of combined pituitary hormonal deficiency across species.

Hearing loss

In Study III, we performed homozygosity mapping and NGS analysis to reveal the first full-penetrance causal variant for nonsyndromic congenital hearing loss in dogs.[69] Homozygous regions and variants specific to affected Rottweilers were investigated, and one of the analysed variants was a missense variant in *LOXHD1*. In humans, *LOXHD1* is one of the 90 genes with moderate or strong evidence in nonsyndromic hearing loss;[230] to date, close to 70 different variants in this gene have been described in patients with autosomal recessive nonsyndromic hearing loss type 77 (OMIM ID: DFNB77).[231]

LOXHD1 was linked to hearing loss for the first time in the *samba* mouse line, where it was demonstrated that murine *Loxhd1* was specifically expressed in the inner ear hair cells and a missense variant in the gene lead to perturbed hair cell function and degeneration.[209] A recent study further elaborated this by showing that a *Loxhd1*$^{T1308X/T1308X}$ mouse model expressing an isoform with a deletion in the 10th PLAT domain had drastically reduced mechanotransductive activity in inner hair cells.[232] The onset of the loss of mechanotransduction was determined to happen between P7 and P11, but the kinetic properties of the channels and hair bundle morphology were preserved at both timepoints; this suggested a critical phase of hair cell maturation in the postnatal period.[232]

The p.(G1914A) variant present in the affected RW was predicted to affect the 14th PLAT domain of canine LOXHD1 (UniProt ID: J9PAE4). Several variants in the corresponding domain of the human LOXHD1 protein sequence have been described (Figure 13); one of them is a glycine-to-arginine substitution in a position orthologous to the canine p.(G1914A) variant. The patient, compound heterozygous for the p.(G1849R) and a p.(Y1541*) variant, was diagnosed with severe hearing loss at the age of two years.[233] In general, there is great variation in the genetic variant type, age of onset, severity and progression of *LOXHD1*-related hearing loss in human patients.[231] Therefore, an in-depth comparison of age of onset between the affected RW, human patients and *Loxhd1*$^{T1308X/T1308X}$ mice is not very informative. It is known that brainstem auditory evoked potential thresholds are mature in 20-day-old puppies;[234] observations from the owners of the RW suggest that the puppies were affected by hearing impairment already at a few weeks of age, but they did not undergo BAER testing until 4 to 19 months of age. Therefore, the exact onset could not be accurately established, but the findings indicate that the hearing loss in the RW is congenital.

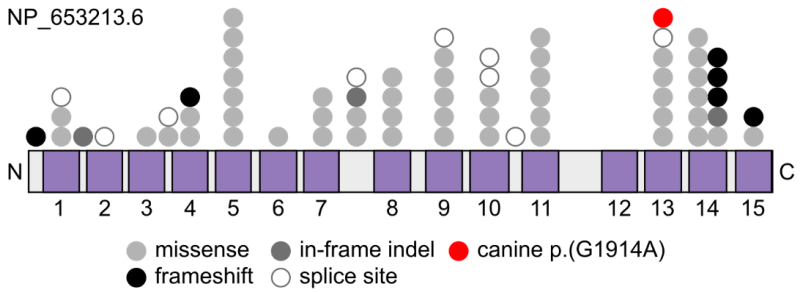

Figure 13. A schematic representation of the domain structure of human LOXHD1 (NCBI ID: NP_653213.6). The circular markers denote variants reported within and between the PLAT domains (indicated in purple and numbered sequentially from 1 to 15). The variants visualized here include those reported in human patients with autosomal recessive nonsyndromic hearing loss type 77[231] as well as the orthologous position of the canine p.(G1914A) variant.

Nonsyndromic congenital hearing loss has previously been reported in one other breed. In a recent study, a missense variant in *KLF7* was linked to congenital sensorineural deafness in Australian Stumpy Tail Cattle Dogs.[98] The variant, discovered with a cross-breed GWAS followed by NGS analysis in three affected and one unaffected dog, was estimated to have a penetrance of 0.75. In contrast to the bilateral autosomal recessive deafness in RW, hearing loss in Australian Stumpy Tail Cattle Dogs is additionally linked to coat colour and occurs in both unilateral and bilateral type, which suggests that the genetic background is heterogeneous.[235] Hence, our study highlights the affected RW as the first spontaneous canine model of full-penetrance nonsyndromic congenital hearing loss.

Screening of *DVL2* deletion reveals cross-breed distribution and consequences on skeletal morphology

In Study IV, we investigated the distribution of the canine *DVL2* deletion variant[107] by Sanger sequencing in 1954 dogs from 15 breeds, which showed differential frequencies of the variant across breeds. We also assessed computed tomography and radiographic images from American Staffordshire Terriers, which indicated qualitative and quantitative consequences on skeletal morphology. Together, our results shed light on the impact of the *DVL2* variant on the health of the affected breeds and strengthen its role as a candidate gene for Robinow syndrome in humans.

Screening of the deletion allele in 15 breeds revealed differential distributions across populations. Boston Terriers, French Bulldogs and English Bulldogs, which have been selected for short and kinked tails and an extreme brachycephalic skull shape[108–110] were predictably fixed for the deletion. Five other bulldog type breeds, including American Staffordshire Terrier, Staffordshire Bull Terrier, Dogue de Bordeaux, Olde English Bulldogge and American Bulldog, showed variable frequencies, likely reflecting that a short and kinked tail is considered undesired or

even a fault in these breeds.[236–239] Bull Terriers and Miniature Bull Terriers, despite their close relation to bulldog type dogs[11] did not carry the allele in our cohorts; while our sample sizes were small and definitive conclusions cannot thus be drawn, the finding is in accordance with the breed definitions that emphasize a long head shape and straight tail.[240,241] Finally, in contrast to previous findings,[107] we did not observe the allele in Shih Tzus or related small Tibetan type dogs.

Inspection of CT and radiographic images in ASTs representing all three *DVL2* genotypes revealed qualitative effects on vertebral morphology. Variable caudal vertebral malformations were present in all four *DVL2* homozygotes (Study IV/Figure 1), and sacral vertebral malformations in one of them. Other vertebrae were normal. This suggests that the *DVL2* deletion mainly affects caudal vertebrae; however, variation in both the type and number of malformations in the homozygous ASTs was considerable. Furthermore, there are notable differences between breeds: BTs, FBDs and EBDs exhibit a more extreme phenotype with truncated and tightly curled tails.[107] The difference to homozygous ASTs might be caused by different genetic backgrounds that could either interact with *DVL2* or act independently to cause the anomalies. However, these hypotheses will need to be tested in future studies.

Our findings in ASTs are in concordance with previous studies that indicate that caudal vertebral anomalies are caused by various different genes. In our cohort, we observed kinked tails in altogether 12 heterozygous and 13 wild-type bulldog type dogs. In the CT cohort, 3/8 heterozygotes and 2/8 wild types had caudal vertebral anomalies; while it seems that *DVL2* heterozygosity does not markedly increase the risk, our analysis was limited by our small sample size and the results should ideally be confirmed in a larger study. However, short tails with prominent kinks are common in King Charles Spaniels,[242] all of which were wild-type for *DVL2* in our cohort, supporting the hypothesis of additional causes. The complexity of caudal vertebral development is further demonstrated in Pembroke Welsh Corgis and other breeds with a short tail, caused by an autosomal dominant missense variant in *TBXT* (T-box transcription factor T).[242,243] Not all short-tailed breeds, however, carry the *TBXT* variant.[242] Therefore, accounting for both *DVL2* and *TBXT* still leaves a large number of various anomalies unexplained. While some of these might be spontaneous developmental defects, the myriad of genotypes associated with abnormal tail morphology (MGI ID: MP:0002111) in mice[166,244] suggest that additional genetic variants still remain to be discovered in dogs.

Besides caudal vertebral defects, the *DVL2* deletion has also been linked to thoracic vertebral malformations in BTs, FBDs and EBDs, with penetrance ranging from 45 to 100 %.[107] In contrast, defects in thoracic vertebrae were not observed in any ASTs, regardless of genotype. Similarly to the variation in the extent of caudal vertebral anomalies, the differences in the thoracic vertebrae could be related to the breed-specific genetic backgrounds as discussed above.

In addition to vertebral morphology, the *DVL2* deletion was also associated with proportions of the skull in ASTs (Study IV/Figure 4). The association was strongest with hard palate length, where all three genotypes significantly ($p<0.001$) differed from each other, such that deletion homozygotes had the shortest and wild types the longest hard palate. Significant differences were also seen in skull base length, ratio

of hard palate length to skull base length, facial index, soft palate height and soft palate length, but the groups of genotypes between which these differences were observed were not consistent. Skull base length, facial index and soft palate height were different between wild types and non-wild types, but the ratio of hard palate length to skull base length was different between homozygotes and non-homozygotes. This likely reflects our small CT study cohort (N=19), and in a larger sample the findings from a similar analysis could be more consistent. Nevertheless, our findings support the conclusion that each copy of the *DVL2* deletion results in a proportionally shorter skeletal muzzle.

In contrast to the skeletal measurements, length of the soft palate did not correlate with *DVL2* genotype. Therefore, it was of similar length in all ASTs regardless of hard palate length. This mismatch of proportions, i.e. an elongated soft palate, is a primary abnormality in BOAS.[245,246] In affected dogs, the disproportionate amount of soft tissue compared to the space available results in partial obstruction of the nasopharynx and larynx.[246] Combined with stenotic nares, this increases airflow resistance and impairs thermoregulation, which causes clinical signs such as respiratory distress, dyspnoea, and heat and exercise intolerance.[247] The symptoms can severely and permanently affect the dog's quality of life, even with medical or surgical management.[247] Our findings therefore indicate that the *DVL2* deletion may cause adverse health effects via increased risk of BOAS.

Brachycephaly in dogs is a polygenic trait, and two other associated genes have so far been described. A fragmented long interspersed nuclear element insertion in *SMOC2*[248] and a missense variant in *BMP3*[249] occur in brachycephalic breeds, but their distributions overlap that of the *DVL2* deletion only partially. To demonstrate, the *DVL2* allele has so far not been found in Pugs, but the *BMP3* variant occurs in the majority of the breed.[249] Conversely, the *BMP3* variant has not been observed Dogues de Bordeaux[249] whereas the *DVL2* deletion does segregate in the breed. In contrast, EBDs, FBDs and BTs are fixed for both variants.[107,249] Therefore, some breeds carry more than one brachycephaly-causing variant; however, it is currently unknown whether the combined effects of these variants are additive or epistatic.

When Mansour et al.[107] first described the *DVL2* variant in EBDs, FBDs and BTs, they noted the homology of *DVL2* with *DVL1* and *DVL3*, genes that are known to cause Robinow syndrome in humans, and the similarity of the bulldog type morphology with human Robinow syndrome patients. Therefore, they hypothesized that canine *DVL2* would contribute broadly to body conformation and proposed to use the name "Robinow-like syndrome" in dogs. Our findings verify that the canine phenotype is not limited to caudal vertebral anomalies only and therefore support this hypothesis. Indeed, the first likely pathogenic variant in *DVL2* in a human Robinow syndrome patient[106] was identified after Mansour's and our studies. The phenotype of the patient with a frameshift variant matches the morphology observed in dogs, with the most striking similarities including a low nasal bridge, hypertelorism, mesomelic shortening of the upper limb, and thoracic vertebral anomalies.[106] As part of our analyses, we also investigated the association of the canine *DVL2* variant with radius length and height at withers, which would correspond to mesomelic limb length and overall stature in humans. We observed a slight trend with *DVL2* genotype, but the

results were not statistically significant. To confirm whether these traits are part of the phenotype in dogs, the analyses will need to be repeated in a larger cohort.

In addition to the skeletal findings, we incidentally discovered mild cardiac anomalies in homozygous ASTs. Two of the three homozygotes in our CT study presented with a mild heart murmur, and they had undergone an echocardiographic examination prior to our study. As a result, one of them had been diagnosed with aortic stenosis, whereas the other had had echocardiographic measurements within normal range but had nonetheless been recommended a follow-up examination by a cardiologist. Additionally, the owner of one homozygous AST unavailable for a CT study reported a mild, asymptomatic heart murmur. Importantly, severe cardiac developmental anomalies are observed in $Dvl2^{-/-}$ mice, where a signaling defect in cardiac neural crest leads to malformations of the outflow tract, including transposition of the great arteries, double outlet right ventricle, and persistent truncus arteriosus.[250] In addition to being involved in the septation of the common outflow tract into the aorta and pulmonary trunk, cardiac neural crest cells also participate in the development of aortic and pulmonary valves from endocardial cushions.[251–256] Based on this common developmental origin, it can be speculated that *DVL2* might additionally affect cardiac development in homozygous dogs. Since aortic and pulmonary stenosis are among the most commonly documented congenital heart defects in EBDs, FBDs and ASTs,[257] this suggestion warrants further research.

Altogether, our findings demonstrate differential allele frequencies of the canine *DVL2* deletion across bulldog type and related dog breeds. The allele affects the development of caudal vertebrae and the skull, with an indication of adverse health effects linked to brachycephaly and potentially cardiac anomalies. Together with the study of Mansour et al.,[107] our results further establish *DVL2* as a candidate gene for human Robinow syndrome. The study also provides valuable information for communities of dog owners who wish to improve breeding programmes to avoid and eradicate health issues related to brachycephaly.

Lack of associated locus suggests genetic or phenotypic heterogeneity in Leonberger dogs

In Study V, we characterized malignant ventricular arrhythmias in juvenile Leonberger dogs and suggested it as a hereditary condition for the first time. By conducting a GWAS (unpublished data) with a cohort of clinically confirmed affected and unaffected dogs as well as a post-mortem cohort of sudden cardiac death cases, we aimed to map associated loci. However, the analyses did not yield significant results despite the appropriate use of gene mapping methods, which warrants discussion of factors that could reduce the power of the GWAS.

Mapping of loci associated with full-penetrance monogenic disorders in dogs is typically successful even with very small samples sizes due to the extensive LD present in purebred dogs. Recent examples, conducted with similar methods and single-breed cohorts, include studies regarding surfactant metabolism dysfunction in Airedale Terriers,[57] progressive retinal atrophy in Lapponian Herders,[45] succinic

semialdehyde dehydrogenase deficiency in Salukis[258] and exfoliative cutaneous lupus erythematosus in German Shorthaired Pointers.[259] The association analyses in the studies included 29, 20, 35 and 43 individuals, respectively, and the resulting genome-wide significant regions spanned several megabases. As the LEO breed is characterized by low genetic diversity and high levels of inbreeding,[260] it should be expected that the presumed disease allele is surrounded by an LD block of observable length. Furthermore, our GWAS cohort is of comparable size to the successful studies, consisting of altogether 47 dogs. Therefore, we would expect to observe a significant locus if the following premises were true: that the arrhythmias in LEO were of monogenic origin, homogeneous and fully penetrant; that the disease allele was surrounded by a unique LD block; and that the phenotyping was successful. However, the GWAS did not yield results; thus, our assumptions about either the genetic background of the disease or the accuracy of the phenotyping are not valid.

Breed-specific genetic disorders in dog breeds are often homogeneous, but not always. Despite the past bottlenecks, LEO are affected by several forms of polyneuropathy, caused by variants in *ARHGEF10*, *GJA9* and *CNTNAP1*.[261–263] Similarly, two genetically distinct forms of progressive retinal atrophy with different age of onset and disease progression occur in Miniature Schnauzers, a breed that has undergone significant founder effects.[49,264] These examples demonstrate that it may be disadvantageous to assume that a single risk variant is shared by all LEO affected by ventricular arrhythmias, as inadvertently treating a heterogeneous population as a single cohort will inevitably lead to unexpected loss of power and potential false negative findings in basic GWAS. The analysis is designed to test the association of single polymorphisms with the phenotype, and oligogenic or polygenic traits or disorders require substantially larger study cohorts for successful mapping than those with simple Mendelian inheritance. In a similar fashion, even if the disorder was homogeneous but the disease allele so recent that the surrounding haplotype was not unique, such as in Dalmatians with acute respiratory distress syndrome,[50] the SNP array might not contain markers tagging the disease allele exclusively. This could, depending on the number of unaffected dogs with a haplotype identical to that shared by the affected dogs, lead to lack of association in the region.

Another genetic mechanism that could affect association analyses is penetrance. If some of the genetically affected individuals do not exhibit the phenotype at the level it is measured, the cohort can include samples that are misclassified from the genetic perspective. Similarly to the heterogeneity discussed above, incomplete penetrance will cause loss of power in basic GWAS due to the discordant genetic and phenotypic information, preventing discoveries in small cohorts.

The level of phenotyping also influences the success of the analyses. Historically, GWAS in disease genetics has been used to study medical phenotypes at the level of an organism, such as disease state or clinical measurements. Instead of defaulting to "clinical GWAS", it is important to consider other levels of information: "molecular GWAS" focuses on molecular phenotypes, including RNA, proteins and metabolites; similarly, "cellular GWAS" highlights cellular processes.[265] This framework is highly relevant for ventricular arrhythmias, as cardiac excitation and contraction is a dynamic process involving multiple scales at molecular, tissue and organ levels; an essential

endeavor in understanding the mechanisms of arrhythmias is to be able to relate molecular behaviors to electrical wave dynamics in time and space.[266] Therefore, while the assumed disease allele in genetically affected LEO should cause anomalies at the molecular level, it may not be valid to assume that the process would consistently manifest as observable arrhythmias in their Holter recordings. However, the molecular approach for GWAS is significantly challenged by the scarce availability of fresh cardiac tissue from both affected and unaffected LEO in the appropriate age window. Therefore, we have initiated a collaboration to identify arrhythmia-related biomarkers from serum samples collected during the clinical examinations of our LEO cohort. Our preliminary results suggest potential autoantibodies similar to those seen in human Brugada syndrome patients (personal communication with Robert Hamilton); this may facilitate future genetic studies by highlighting specific candidate genes and by providing novel, molecular phenotypes for association studies.

In our study, we assessed the clinical phenotypes with 24-hour Holter monitoring. Dogs with frequent ventricular arrhythmias in the Holter recording can be assigned as affected with fairly high confidence, but borderline and negative findings include the risk of misclassification. We showed that the arrhythmias in affected LEO had great variation both between dogs and the examinations of single dogs. While borderline affected dogs were not included in the analyses, the possibility of our GWAS cohort including false negative phenotypes cannot be excluded. The role of 24-hour Holter monitoring compared to longer periods has been discussed: a recent study in dogs demonstrated that a 48-hour monitoring period, compared to 24 hours, increased the likelihood of identifying an arrhythmia only slightly.[267] The authors concluded that the additional diagnostic yield was minimal and noted that longer recording periods should be considered. Another study utilizing seven-day Holter monitoring reported a significant increase in diagnosing ventricular arrhythmias in Dobermanns by comparing the initial 24 hours to the entire seven-day recording period, indicating that the longer monitoring is markedly more sensitive with phenotypes exhibiting high day-to-day variation.[268] Even longer periods have been tested in humans, as it has been shown that a 14-day adhesive patch electrocardiographic monitor significantly increased detection of arrhythmia events compared to 24-hour Holter monitoring.[269] A similar pilot study in four Boxer dogs, either affected by arrhythmogenic right ventricular cardiomyopathy or unaffected, found that the use of adhesive patch monitor in dogs is feasible, although the dogs' activity caused intermittent motion artifacts.[270] Taken together, the method for clinical phenotyping of LEO could be reassessed, taking diagnostic and practical aspects and costs involved into account.

In our GWAS, we utilized cohorts with and without sudden cardiac death cases. Notably, two of the dogs that exhibited ventricular tachycardia in Holter monitoring died during the follow-up: one was euthanized due to life-threatening shock and arrhythmias, and the other died suddenly during rest. The post-mortem examinations of these dogs indicated acute heart failure, consistent with the clinical findings. Therefore, there is strong evidence that the ventricular arrhythmias and the sudden cardiac death in LEO fall along the same phenotypic continuum and that these cohorts can thus be grouped together in association analyses. However, the addition of the

SCD cases to the GWAS did not strengthen the tentative association obtained from the smaller cohort including only nine Holter cases, which suggests that the low sample size in the first analysis lead to stratification-caused artifacts.

In addition to genetic and phenotypic factors, the mapping attempt could be affected by the SNP array itself. The 173 000 markers in the Illumina SNP array were selected from the SNP set generated along with the canFam2.0 reference genome; this set containing 2.5 million SNPs was produced by comparing the Boxer reference to dogs from ten other breeds, wolves, and a coyote.[9] The diversity of the selected markers was confirmed with validation in 26 different breeds, with polymorphic sites per breed ranging from 85 000 (Greenland Sled Dog) to 126 000 (Jack Russel Terrier).[271] By design, the array therefore yields variable numbers of informative markers for different populations. In our LEO analysis with VA-affected and SCD cases, 98 137 markers remained after pre-analytical QC, resulting in an average of 41.9 SNPs/Mb and mean density of 23.7 kb/SNP. While relatively large gaps were present in the pruned data, the longest spanning up to several megabases in some chromosomes, the breed is characterized by low diversity and long runs of homozygosity with an average length of 5.88 Mb.[260] Therefore, even though there is a risk of not capturing the hypothetical disease haplotype with the relatively scarce marker set, it is simultaneously not unreasonable to assume that some markers tagging the disease allele would still be present in the dataset even after QC.

As a summary, our clinical and post-mortem study strongly indicates genetic background for the ventricular arrhythmias and sudden cardiac death in young Leonbergers. However, genetic analyses have so far not revealed the exact risk loci and variants. Various factors could cause heterogeneity in the genetic or phenotypic data, preventing discovery with GWAS. Efforts to expand our cohort and explore improved clinical as well as molecular phenotyping are ongoing, together with GWAS and NGS analyses.

6 CONCLUDING REMARKS

This thesis study is the outcome of several interdisciplinary and international collaborations combining basic and clinical research at the junction of molecular genetics, bioinformatics and veterinary medicine. The discoveries demonstrate the importance of broad expertise in collecting meaningful phenotypic information, processing and analysing genetic data, and validating *in silico* findings *in vitro*. Critical infrastructure utilized in this thesis include the comprehensive canine biobank and extensive network of dog owners, established bioinformatic pipelines, high-power computing environments, and veterinary and pathology collaborations. These resources were essential to form a sound scientific and methodological basis for gene mapping and variant identification.

The field of canine genetics has changed rapidly during this thesis. Next-generation sequencing has become a standard, and cohort sizes have expanded significantly. In Study I in 2017, one affected and 31 control genomes were used; in Study II in 2021, three affected genomes, 804 control genomes and 231 control exomes were utilized. This 30-fold increase illustrates the ongoing methodological transition from small case studies to large joint efforts. The Dog Biomedical Variant Database Consortium has been an important community facilitating this development, as large variant datasets have been released for use of the consortium members. The availability of genomes and related metadata has enabled and continues to enable a rich variety of different studies. Another development is anticipated with the Dog10K consortium, which aims to sequence 10 000 genomes in the upcoming years to capture the genetic diversity of canids across the world.[272] This unprecedented endeavour will require re-evaluation of study designs as we start to understand the intricacies of allele and haplotype sharing and ancestry across dogs and other canid species.

In addition to constantly increasing sample numbers, the horizon of traditional association analyses has been broadened by large-scale integration of molecular and cellular phenotypes. Perhaps the most distinguished multi-omic project, GTEx[165] is a fine example of interpretation of genetic variants through the context of transcriptome and tissues. With over 40 consortium publications, the project has shed light on the interplay of numerous phenomena, including but not limited to gene regulation, cell and tissue specific expression, genetic variation, allelic expression and pleiotropy. While such resources have previously not been systematically produced in the canine genetics community, the first comprehensive functional annotation of the dog genome is currently under work by the Dog Genome Annotation consortium.[22] The transcriptomic and regulatory information generated during the project, similar to the FANTOM expression atlas,[273] will be imperative to truly dissect the molecular mechanisms of diseases and traits in canine models, such as the arrhythmias in Leonbergers with yet unknown genetic background.

The findings of this thesis have contributed new knowledge to the current understanding of the health of dog breeds. The accumulating information on the distribution and prevalence of disease-causing alleles has revealed variable allele

burden across populations and pinpointed breeds that are highly predisposed to inherited disorders. One example of these are the severely brachycephalic breeds, which are known to suffer from several health issues caused by their body conformation.[274] Expanding evidence clearly shows that inbreeding and pursuit of extreme morphologies significantly increase morbidity.[275] Therefore, as the species that is fully responsible for the health of our companion animals, we humans are morally obligated to ask ourselves: are we providing the healthiest possible life to our best friends, and if not, what must change? Fortunately, there have been signs of a positive change during the past years as the increasing awareness of inherited disorders has driven breed clubs to act and revise their programmes, sometimes by introducing genetic diversity from other breeds. However, combating inherited disorders by crossbreeding is currently not the norm; therefore, genetic tests will continue to play an important role in dog breeders' decision-making. While the utility of gene tests for the *COL7A1*, *LOXHD1* and *POU1F1* variants identified in this thesis is mostly limited to a relatively small number of affected breeds, we showed that the *DVL2* deletion occurs in several highly popular breeds. The impact of the gene test is differential across breeds, however; some breeds are fixed for the deletion and *DVL2*-related health issues in these populations cannot be addressed without crossbreeding, while other breeder communities seeking to avoid brachycephaly will directly benefit from the gene test. As the volume of genetic tests performed on dogs increases, our discovery may have a profound impact on reducing brachycephaly-related health issues and improving the welfare of pet dogs worldwide.

The molecular characterization of the disorders in this thesis provides novel spontaneous animal models for preclinical studies. The study of treatment strategies and disease progression and prevention benefits both veterinary and human medicine. Some of the disorders had been identified in the affected dog breeds decades prior to our studies, with a definitive molecular diagnosis finally available. Of the newly characterized models, ventricular arrhythmia in Leonbergers is perhaps the most severe: the disorder is life-threatening, and clinical signs or symptoms may not occur until sudden cardiac death. Ventricular arrhythmias and subsequent cardiac death is similarly a major contributor to mortality in human populations,[112] and prospective identification of families and individuals at genetic risk would be vital in reducing the disease burden. Therefore, discovering the yet unknown genetic architecture underlying rare cardiac disorders is essential. Towards this, we will continue to investigate the affected Leonberger population to identify causal genes, with potential new avenues to explore in improving clinical and molecular phenotypic data.

Altogether, the resources that support the study of spontaneous genetic disorders in dogs form a solid basis for unique research opportunities. The upcoming methodological developments offer possibilities for canine studies of unprecedented scale, and it might be anticipated that phenome-wide studies as well as reverse genomics and other data mining-based approaches, already observed in human genetic research, may become established in dog genetics as well. The past five years during which this thesis was completed have shown the remarkable transformation of our field; it will be exciting and fascinating to observe the advancements during the next five years as well.

ACKNOWLEDGEMENTS

This thesis was carried out in the Doctoral Programme in Integrative Life Science in the Department of Medical and Clinical Genetics, Faculty of Medicine and the Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki as well as the Folkhälsan Research Center. The study was supervised by Professor Hannes Lohi, Docent Marjo Hytönen and Docent Maria Wiberg. The projects were financially supported by the Academy of Finland, Jane and Aatos Erkko Foundation, Biocentrum Helsinki, Orion Research Foundation, Morris Animal Foundation, Jenny and Antti Wihuri Foundation, the Canine Health Research Fund, Helsinki Institute of Life Science, Wisdom Health, Doctoral Programme in Integrative Life Science, American Kennel Club Canine Health Foundation, Finnish Foundation for Cardiovascular Research, Leonberger Health Foundation International, The Leonberger Club of Great Britain, Kennel Club Charitable Trust and Biomedicum Helsinki Foundation.

I would like to warmly thank Hannes for your supervision. I arrived in the group as a fresh Bachelor of Science seven and a half years ago, and I could never have imagined at that time that I would become a doctoral researcher. You saw the potential in me, and during the course of my studies I have been able to drastically accumulate new skills and knowledge for the better of not only myself, but also our entire group. It has been delightful to develop new tools and workflows to support our research. I am also glad that you have trusted me in the opportunities and challenges I've faced, even when I have not done so myself.

I would also like to thank Marjo for being an amazing supervisor, colleague and friend. Words can hardly capture the depth of my gratitude: never before in my life have I met someone with whom I share such a connection. Working with you has been exceptionally warm, easy and rewarding. You have always given me kind yet constructive feedback that has truly improved my work. We have shared joys and endured sorrows together, and with your support I've found courage to be the person I am. I'm so very proud of what we have achieved, and I gladly look forward to everything that awaits us in the future.

I'd like to express my thanks to Maria Wiberg for your supervision and excellent clinical expertise. Your contribution to our studies, both within and outside this thesis, has been essential. I hope that with our collaboration we can continue to try to solve the mysteries that still remain and bring new knowledge into our fields.

This thesis has also been supported by various other skilled researchers. I thank Professor Markus Perola and Docent Maija Wessman for being my thesis committee members and providing guidance on my work. I also thank Docent Minna Ruddock and Docent Sami Heikkinen for reviewing my thesis. All of my thesis projects are the outcome of combined efforts of multiple people, and for that I thank all our numerous collaborators during these years.

Our research is based on the contributions of voluntary dog owners, breeders and breeds clubs that have diligently collected and donated samples to our biobank. I

would like to thank all of our participants: without you, our projects would not exist. Your efforts are essential for our work.

I am grateful to all the colleagues that have made our research group into the warm, welcoming community it is. I'd like to thank all my past and present coworkers, especially the tight-knit group of my closest peers Lea, Tiina, Milla, Salla, Riika, Jenni, Emma, Sini and Inka. A very special thanks goes to Lea: in addition to your brilliant research skills and guidance, you have taught me a lot about this world by confidently being yourself. That is a lesson I will carry with me beyond the academia.

I would also like to mention key persons that have influenced and strengthened my commitment to this field. For cultivating my interest in natural sciences, critical thinking, and argumentation and communication, I want to thank Esa Sederholm, Aki Holm, Sirpa Lappalainen, Petra Luukko, Susanna Kauppi, Minna Oittinen and Vanessa Fuller. The skills you have taught me have served me handsomely and will surely continue to be essential in my future endeavours.

Finally, I want to thank my family and family-in-law for standing by me throughout this process. These years have included some of the most absurd and challenging events in my life, and you have offered a respite in your homes. Many things have become better by taking a break from the hectic tempo of the capital region. Lastly, my greatest thanks I want to say to Miro, my steadfast partner in life and science: with your companionship, my journey with all its ups and downs has been so much better. Your absolute and unwavering support has been the one constant I have always been able to rely on. I hope that I can be equally supportive to you as we continue to pursue our dreams and goals together.

REFERENCES

1. Larson, G. *et al.* Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci.* **109**, 8878–8883 (2012).
2. Perri, A. A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. *J. Archaeol. Sci.* **68**, 1–4 (2016).
3. Fédération Cynologique Internationale. Presentation of our organisation. http://www.fci.be/en/Presentation-of-our-organisation-4.html.
4. Ritvo, H. Pride and Pedigree: The Evolution of the Victorian Dog Fancy. *Vic. Stud.* **29**, 227–253 (1986).
5. Parker, H. G. *et al.* Genetic Structure of the Purebred Domestic Dog. *Science* **304**, 1160–1164 (2004).
6. Lewis, T. W., Abhayaratne, B. M. & Blott, S. C. Trends in genetic diversity for all Kennel Club registered pedigree dog breeds. *Canine Genet. Epidemiol.* **2**, 13 (2015).
7. Fédération Cynologique Internationale. FCI-Standard N° 347: Berger Blanc Suisse. http://www.fci.be/Nomenclature/Standards/347g01-en.pdf (2011).
8. Lampi, S., Donner, J., Anderson, H. & Pohjoismäki, J. Variation in breeding practices and geographic isolation drive subpopulation differentiation, contributing to the loss of genetic diversity within dog breed lineages. *Canine Med. Genet.* **7**, 5 (2020).
9. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
10. Sutter, N. B. *et al.* Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Res.* **14**, 2388–2396 (2004).
11. Parker, H. G. *et al.* Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep.* **19**, 697–708 (2017).
12. Drögemüller, C. *et al.* A Mutation in Hairless Dogs Implicates FOXI3 in Ectodermal Development. *Science* **321**, 1462–1462 (2008).
13. Parker, H. G. *et al.* Breed relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**, 1562–1571 (2007).
14. Jagannathan, V., Drögemüller, C. & Leeb, T. A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Anim. Genet.* (2019) doi:10.1111/age.12834.
15. Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–364 (2013).
16. Hoeppner, M. P. *et al.* An Improved Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding Transcripts. *PLOS ONE* **9**, e91172 (2014).

17. Wucher, V. *et al.* FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* gkw1306 (2017) doi:10.1093/nar/gkw1306.
18. Elvers, I. *et al.* Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* **25**, 1634–1645 (2015).
19. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
20. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
21. Holden, L. A. *et al.* Assembly and Analysis of Unmapped Genome Sequence Reads Reveal Novel Sequence and Variation in Dogs. *Sci. Rep.* **8**, 10862 (2018).
22. DoGA consortium. The Dog Genome Annotation (DoGA) Project. *The Dog Genome Annotation (DoGA) Project* https://www.doggenomeannotation.org.
23. Jagannathan, V. *et al.* Dog10K_Boxer_Tasha_1.0: A Long-Read Assembly of the Dog Reference Genome. *Genes* **12**, 847 (2021).
24. Eory, L. *et al.* ROS_Cfam_1.0: A high-quality, de novo assembly of a male Labrador retriever. in (International Society for Animal Genetics, 2021).
25. Halo, J. V. *et al.* Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes. *Proc. Natl. Acad. Sci.* **118**, e2016274118 (2021).
26. Wang, C. *et al.* A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Commun. Biol.* **4**, 185 (2021).
27. Field, M. A. *et al.* Canfam_GSD: De novo chromosome-length genome assembly of the German Shepherd Dog (Canis lupus familiaris) using a combination of long reads, optical mapping, and Hi-C. *GigaScience* **9**, giaa027 (2020).
28. Edwards, R. J. *et al.* Chromosome-length genome assembly and structural variations of the primal Basenji dog (Canis lupus familiaris) genome. *BMC Genomics* **22**, 188 (2021).
29. Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci.* **113**, 152–157 (2016).
30. Donner, J. *et al.* Frequency and distribution of 152 genetic disease variants in over 100,000 mixed breed and purebred dogs. *PLOS Genet.* **14**, e1007361 (2018).
31. Wess, G. *et al.* Prevalence of Dilated Cardiomyopathy in Doberman Pinschers in Various Age Groups: Prevalence of DCM in Dobermans. *J. Vet. Intern. Med.* **24**, 533–538 (2010).
32. Bannasch, D. *et al.* Mutations in the SLC2A9 Gene Cause Hyperuricosuria and Hyperuricemia in the Dog. *PLoS Genet.* **4**, e1000246 (2008).
33. Schaible, R. Linkage of a pigmentary trait with a high-level of uric-acid excretion in dalmation dog. in *Genetics* vol. 83 S68–S68 (428 EAST PRESTON ST, BALTIMORE, MD 21202, 1976).

34. Salmon Hillbertz, N. H. C. *et al.* Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* **39**, 1318–1320 (2007).
35. Asher, L., Diesel, G., Summers, J. F., McGreevy, P. D. & Collins, L. M. Inherited defects in pedigree dogs. Part 1: Disorders related to breed standards. *Vet. J.* **182**, 402–411 (2009).
36. Nicholas, F. W. Online Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic Acids Res.* **31**, 275–277 (2003).
37. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
38. Ribitsch, I. *et al.* Large Animal Models in Regenerative Medicine and Tissue Engineering: To Do or Not to Do. *Front. Bioeng. Biotechnol.* **8**, 972 (2020).
39. Hoffman, J. M., Creevy, K. E., Franks, A., O'Neill, D. G. & Promislow, D. E. L. The companion dog as a model for human aging and mortality. *Aging Cell* **17**, e12737 (2018).
40. Silva, K. A. S. & Emter, C. A. Large Animal Models of Heart Failure. *JACC Basic Transl. Sci.* **5**, 840–856 (2020).
41. Petersen-Jones, S. M. *et al.* Patients and animal models of CNGβ1-deficient retinitis pigmentosa support gene augmentation approach. *J. Clin. Invest.* **128**, 190–206 (2017).
42. Hoffman, M. M. & Birney, E. Estimating the Neutral Rate of Nucleotide Substitution Using Introns. *Mol. Biol. Evol.* **24**, 522–531 (2006).
43. Kyöstilä, K. *et al.* A Missense Change in the ATG4D Gene Links Aberrant Autophagy to a Neurodegenerative Vacuolar Storage Disease. *PLOS Genet.* **11**, e1005169 (2015).
44. Tamargo-Gómez, I. *et al.* ATG4D is the main ATG8 delipidating enzyme in mammalian cells and protects against cerebellar neurodegeneration. *Cell Death Differ.* **28**, 2651–2672 (2021).
45. Kaukonen, M. *et al.* A missense variant in IFT122 associated with a canine model of retinitis pigmentosa. *Hum. Genet.* (2021) doi:10.1007/s00439-021-02266-3.
46. Seppälä, E. H. *et al.* LGI2 Truncation Causes a Remitting Focal Epilepsy in Dogs. *PLoS Genet.* **7**, e1002194 (2011).
47. Wielaender, F. *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.* **114**, 2669–2674 (2017).
48. Kyöstilä, K. *et al.* A SEL1L Mutation Links a Canine Progressive Early-Onset Cerebellar Ataxia to the Endoplasmic Reticulum–Associated Protein Degradation (ERAD) Machinery. *PLoS Genet.* **8**, e1002759 (2012).
49. Kaukonen, M. *et al.* A putative silencer variant in a spontaneous canine model of retinitis pigmentosa. *PLOS Genet.* **16**, e1008659 (2020).

50. Holopainen, S. *et al.* ANLN truncation causes a familial fatal acute respiratory distress syndrome in Dalmatian dogs. *PLOS Genet.* **13**, e1006625 (2017).
51. Hytönen, M. K. *et al.* Molecular Characterization of Three Canine Models of Human Rare Bone Diseases: Caffey, van den Ende-Gupta, and Raine Syndromes. *PLoS Genet.* **12**, e1006037 (2016).
52. Hänninen, R. L. *et al.* Canine MPV17 truncation without clinical manifestations. *Biol. Open* **4**, 1253–1258 (2015).
53. El-Hattab, A. W. *et al.* MPV17-related mitochondrial DNA maintenance defect: New cases and review of clinical, biochemical, and molecular aspects. *Hum. Mutat.* **39**, 461–470 (2018).
54. Gurda, B. L., Bradbury, A. M. & Vite, C. H. Canine and Feline Models of Human Genetic Diseases and Their Contributions to Advancing Clinical Therapies. *Yale J. Biol. Med.* **90**, 417–431 (2017).
55. Suomen Kennelliitto. Perinnöllisten vikojen ja sairauksien vastustamisohjelman ohje (PEVISA-ohje). (2019).
56. Tilastokeskus. *Suomen virallinen tilasto (SVT): Kotitalouksien kulutus*. http://www.stat.fi/til/ktutk/index.html.
57. Dillard, K. J. *et al.* Recessive missense LAMP3 variant associated with defect in lamellar body biogenesis and fatal neonatal interstitial lung disease in dogs. *PLoS Genet.* **16**, e1008651 (2020).
58. Hytönen, M. & Lohi, H. A frameshift insertion in SGK3 leads to recessive hairlessness in Scottish Deerhounds: a candidate gene for human alopecia conditions. *Hum. Genet.* **138**, 535–539 (2019).
59. Hytönen, M. K. *et al.* In-frame deletion in canine PITRM1 is associated with a severe early-onset epilepsy, mitochondrial dysfunction and neurodegeneration. *Hum. Genet.* (2021) doi:10.1007/s00439-021-02279-y.
60. Hytönen, M. K., Arumilli, M., Sarkiala, E., Nieminen, P. & Lohi, H. Canine models of human amelogenesis imperfecta: identification of novel recessive ENAM and ACP4 variants. *Hum. Genet.* **138**, 525–533 (2019).
61. Kyöstilä, K. *et al.* A homozygous missense variant in the alkaline phosphatase gene ALPL is associated with a severe form of canine hypophosphatasia. *Sci. Rep.* **9**, 973 (2019).
62. Mikkola, L. *et al.* An across-breed validation study of 46 genetic markers in canine hip dysplasia. *BMC Genomics* **22**, 68 (2021).
63. Mikkola, L. *et al.* Genetic dissection of canine hip dysplasia phenotypes and osteoarthritis reveals three novel loci. *BMC Genomics* **20**, 1027 (2019).
64. Sarviaho, R. *et al.* A novel genomic region on chromosome 11 associated with fearfulness in dogs. *Transl. Psychiatry* **10**, 169 (2020).
65. Sarviaho, R. *et al.* Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Transl. Psychiatry* **9**, 18 (2019).
66. Koskinen, L. L. E. *et al.* ADAM23 is a common risk gene for canine idiopathic epilepsy. *BMC Genet.* **18**, 8 (2017).
67. Koskinen, L. L. E. *et al.* Identification of a common risk haplotype for canine idiopathic epilepsy in the ADAM23 gene. *BMC Genomics* **16**, 465 (2015).

68. Kyöstilä, K. *et al.* Intronic variant in POU1F1 associated with canine pituitary dwarfism. *Hum. Genet.* (2021) doi:10.1007/s00439-021-02259-2.
69. Hytönen, M. K. *et al.* Missense variant in LOXHD1 is associated with canine nonsyndromic hearing loss. *Hum. Genet.* (2021) doi:10.1007/s00439-021-02286-z.
70. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
71. Russell, W. M. S. & Burch, R. L. *The principles of humane experimental technique*. (Methuen, 1959).
72. Ormandy, E. & Schuppli, C. Public Attitudes toward Animal Research: A Review. *Animals* **4**, 391–408 (2014).
73. Has, C. & Fischer, J. Inherited epidermolysis bullosa: New diagnostics and new clinical phenotypes. *Exp. Dermatol.* **28**, 1146–1152 (2019).
74. Has, C., Bruckner-Tuderman, L. & Uitto, J. Epidermolysis Bullosa. in *Reference Module in Biomedical Sciences* B9780128012383056000 (Elsevier, 2014). doi:10.1016/B978-0-12-801238-3.05619-1.
75. Has, C. *et al.* Consensus reclassification of inherited epidermolysis bullosa and other disorders with skin fragility. *Br. J. Dermatol.* **183**, 614–627 (2020).
76. Mauldin, E. A., Wang, P., Olivry, T., Henthorn, P. S. & Casal, M. L. Epidermolysis bullosa simplex in sibling Eurasier dogs is caused by a PLEC non-sense variant. *Vet. Dermatol.* **28**, 10-e3 (2017).
77. Scott, D. W. & Schultz, R. D. Epidermolysis bullosa simplex in the Collie dog. *J. Am. Vet. Med. Assoc.* **171**, 721–727 (1977).
78. Capt, A. *et al.* Inherited Junctional Epidermolysis Bullosa in the German Pointer: Establishment of a Large Animal Model. *J. Invest. Dermatol.* **124**, 530–535 (2005).
79. Herrmann, I. *et al.* Canine junctional epidermolysis bullosa due to a novel mutation in LAMA3 with severe upper respiratory involvement. *Vet. Dermatol.* **32**, 379 (2021).
80. Kiener, S. *et al.* LAMB3 Missense Variant in Australian Shepherd Dogs with Junctional Epidermolysis Bullosa. *Genes* **11**, 1055 (2020).
81. Baldeschi, C. *et al.* Genetic correction of canine dystrophic epidermolysis bullosa mediated by retroviral vectors. *Hum. Mol. Genet.* **12**, 1897–1905 (2003).
82. Garcia, T. M., Kiener, S., Jagannathan, V., Russell, D. S. & Leeb, T. A COL7A1 Variant in a Litter of Neonatal Basset Hounds with Dystrophic Epidermolysis Bullosa. *Genes* **11**, 1458 (2020).
83. Nagata, M. *et al.* Dystrophic form of inherited epidermolysis bullosa in a dog (Akita Inu). *Br. J. Dermatol.* **133**, 1000–1003 (1995).
84. Dattani, M. T. & Malhotra, N. A review of growth hormone deficiency. *Paediatr. Child Health* **29**, 285–292 (2019).
85. Perez-Castro, C., Renner, U., Haedo, M. R., Stalla, G. K. & Arzt, E. Cellular and Molecular Specificity of Pituitary Gland Physiology. *Physiol. Rev.* **92**, 1–38 (2012).
86. Kim, S. Y. Diagnosis and Treatment of Hypopituitarism. *Endocrinol. Metab.* **30**, 443 (2015).

87. Fang, Q. *et al.* Genetics of Combined Pituitary Hormone Deficiency: Roadmap into the Genome Era. *Endocr. Rev.* **37**, 636–675 (2016).
88. Voorbij, A. M. W. Y. *et al.* A Contracted DNA Repeat in LHX3 Intron 5 Is Associated with Aberrant Splicing and Pituitary Dwarfism in German Shepherd Dogs. *PLoS ONE* **6**, e27940 (2011).
89. Voorbij, A. M. W. Y., Leegwater, P. A. & Kooistra, H. S. Pituitary Dwarfism in Saarloos and Czechoslovakian Wolfdogs is Associated with a Mutation in LHX3. *J. Vet. Intern. Med.* **28**, 1770–1774 (2014).
90. Thaiwong, T., Corner, S., Forge, S. L. & Kiupel, M. Dwarfism in Tibetan Terrier dogs with an LHX3 mutation. *J. Vet. Diagn. Invest.* **33**, 740–743 (2021).
91. Allan, G. S. *et al.* Pituitary dwarfism in German Shepherd Dogs. *J. Small Anim. Pract.* **19**, 711–727 (1978).
92. Raviv, D., Dror, A. A. & Avraham, K. B. Hearing loss: a common disorder caused by many rare alleles. *Ann. N. Y. Acad. Sci.* **1214**, 168–179 (2010).
93. Cunningham, L. L. & Tucci, D. L. Hearing Loss in Adults. *N. Engl. J. Med.* **377**, 2465–2473 (2017).
94. Lieu, J. E. C., Kenna, M., Anne, S. & Davidson, L. Hearing Loss in Children: A Review. *JAMA* **324**, 2195 (2020).
95. Strain, G. M. The Genetics of Deafness in Domestic Animals. *Front. Vet. Sci.* **2**, (2015).
96. Guevar, J., Olby, N. J., Meurs, K. M., Yost, O. & Friedenberg, S. G. Deafness and vestibular dysfunction in a Doberman Pinscher puppy associated with a mutation in the *PTPRQ* gene. *J. Vet. Intern. Med.* **32**, 665–669 (2018).
97. Webb, A. A., Ruhe, A. L. & Neff, M. W. A missense mutation in MYO7A is associated with bilateral deafness and vestibular dysfunction in the Doberman pinscher breed. *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire* **83**, 142–148 (2019).
98. Xu, F., Shan, S., Sommerlad, S., Seddon, J. M. & Brenig, B. A Missense Mutation in the KLF7 Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs. *Genes* **12**, 467 (2021).
99. Kawakami, T. *et al.* R-locus for roaned coat is associated with a tandem duplication in an intronic region of USH2A in dogs and also contributes to Dalmatian spotting. *PLOS ONE* **16**, e0248233 (2021).
100. Adato, A. *et al.* Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells. *Hum. Mol. Genet.* **14**, 3921–3932 (2005).
101. Yokoyama, J. S. *et al.* Variation in Genes Related to Cochlear Biology Is Strongly Associated with Adult-Onset Deafness in Border Collies. *PLoS Genet.* **8**, e1002898 (2012).
102. Mazzeu, J. F. & Brunner, H. G. 50 years of Robinow syndrome. *Am. J. Med. Genet. A.* **182**, 2005–2007 (2020).
103. Patton, M. A. Robinow syndrome. *J. Med. Genet.* **39**, 305–310 (2002).
104. Zhang, C. *et al.* Novel pathogenic genomic variants leading to autosomal dominant and recessive Robinow syndrome. *Am. J. Med. Genet. A.* **185**, 3593–3600 (2021).

105. White, J. J. *et al.* WNT Signaling Perturbations Underlie the Genetic Heterogeneity of Robinow Syndrome. *Am. J. Hum. Genet.* **102**, 27–43 (2018).
106. Zhang, C. *et al.* Novel pathogenic variants and quantitative phenotypic analyses of Robinow syndrome: WNT signaling perturbation and phenotypic variability. *Hum. Genet. Genomics Adv.* **3**, 100074 (2022).
107. Mansour, T. A. *et al.* Whole genome variant association across 100 dogs identifies a frame shift mutation in DISHEVELLED 2 which contributes to Robinow-like syndrome in Bulldogs and related screw tail dog breeds. *PLoS Genet.* **14**, e1007850 (2018).
108. Fédération Cynologique Internationale. FCI-Standard N° 101: French Bulldog. http://www.fci.be/Nomenclature/Standards/101g09-en.pdf (2015).
109. Fédération Cynologique Internationale. FCI-Standard N° 149: Bulldog. http://www.fci.be/Nomenclature/Standards/149g02-en.pdf (2011).
110. Fédération Cynologique Internationale. FCI-Standard N° 140: Boston Terrier. http://www.fci.be/Nomenclature/Standards/140g09-en.pdf (2014).
111. Antzelevitch, C. & Burashnikov, A. Overview of Basic Mechanisms of Cardiac Arrhythmia. *Card. Electrophysiol. Clin.* **3**, 23–45 (2011).
112. Marsman, R. F., Tan, H. L. & Bezzina, C. R. Genetics of sudden cardiac death caused by ventricular arrhythmias. *Nat. Rev. Cardiol.* **11**, 96–111 (2014).
113. Bigger, J. T. Definition of benign versus malignant ventricular arrhythmias: Targets for treatment. *Am. J. Cardiol.* **52**, C47–C54 (1983).
114. Meurs, K. *et al.* A QIL1 Variant Associated with Ventricular Arrhythmias and Sudden Cardiac Death in the Juvenile Rhodesian Ridgeback Dog. *Genes* **10**, 168 (2019).
115. Moise, N. S. *et al.* Inherited ventricular arrhythmias and sudden death in German shepherd dogs. *J. Am. Coll. Cardiol.* **24**, 233–243 (1994).
116. Loos, R. J. F. 15 years of genome-wide association studies and no signs of slowing down. *Nat. Commun.* **11**, 5900 (2020).
117. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Primer* **1**, 59 (2021).
118. Turner, S. *et al.* Quality Control Procedures for Genome-Wide Association Studies. *Curr. Protoc. Hum. Genet.* **68**, (2011).
119. Wang, D. G. *et al.* Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science* **280**, 1077–1082 (1998).
120. Quick, C. *et al.* Sequencing and imputation in GWAS: Cost-effective strategies to increase power and genomic coverage across diverse populations. *Genet. Epidemiol.* **44**, 537–549 (2020).
121. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
122. Liu, J., Lewinger, J. P., Gilliland, F. D., Gauderman, W. J. & Conti, D. V. Confounding and Heterogeneity in Genetic Association Studies with Admixed Populations. *Am. J. Epidemiol.* **177**, 351–360 (2013).

123. Quignon, P. *et al.* Canine Population Structure: Assessment and Impact of Intra-Breed Stratification on SNP-Based Association Studies. *PLoS ONE* **2**, e1324 (2007).
124. Kaukonen, M. *et al.* Maternal Inheritance of a Recessive RBP4 Defect in Canine Congenital Eye Disease. *Cell Rep.* **23**, 2643–2652 (2018).
125. Ke, X. *et al.* Assessment of the functionality of genome-wide canine SNP arrays and implications for canine disease association studies: Assessment of the functionality of genome-wide canine SNP arrays. *Anim. Genet.* **42**, 181–190 (2011).
126. Plassais, J. *et al.* Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019).
127. Vahidnezhad, H., Youssefian, L., Jazayeri, A. & Uitto, J. Research Techniques Made Simple: Genome-Wide Homozygosity/Autozygosity Mapping Is a Powerful Tool for Identifying Candidate Genes in Autosomal Recessive Genetic Diseases. *J. Invest. Dermatol.* **138**, 1893–1900 (2018).
128. Pemberton, T. J. *et al.* Genomic Patterns of Homozygosity in Worldwide Human Populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
129. Kamphans, T. *et al.* Filtering for Compound Heterozygous Sequence Variants in Non-Consanguineous Pedigrees. *PLoS ONE* **8**, e70151 (2013).
130. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
131. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).
132. Gusev, A. *et al.* Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* **19**, 318–326 (2009).
133. Carr, I. M. *et al.* Autozygosity Mapping with Exome Sequence Data. *Hum. Mutat.* **34**, 50–56 (2013).
134. Magi, A. *et al.* H3M2 : detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics* **30**, 2852–2859 (2014).
135. Görmez, Z., Bakir-Gungor, B. & Sağıroğlu, M. Ş. HomSI: a homozygous stretch identifier from next-generation sequencing data. *Bioinformatics* **30**, 445–447 (2014).
136. Quinodoz, M. *et al.* AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. *Nat. Commun.* **12**, 518 (2021).
137. Seelow, D., Schuelke, M., Hildebrandt, F. & Nurnberg, P. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Res.* **37**, W593–W599 (2009).
138. Narasimhan, V. *et al.* BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).
139. Howrigan, D. P., Simonson, M. A. & Keller, M. C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics* **12**, 460 (2011).

140. Pippucci, T., Magi, A., Gialluisi, A. & Romeo, G. Detection of Runs of Homozygosity from Whole Exome Sequencing Data: State of the Art and Perspectives for Clinical, Population and Epidemiological Studies. *Hum. Hered.* **77**, 63–72 (2014).
141. Meyermans, R., Gorssen, W., Buys, N. & Janssens, S. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics* **21**, (2020).
142. Alkuraya, F. S. Homozygosity mapping: One more tool in the clinical geneticist's toolbox. *Genet. Med.* **12**, 236–239 (2010).
143. Hildebrandt, F. *et al.* A Systematic Approach to Mapping Recessive Disease Genes in Individuals from Outbred Populations. *PLoS Genet.* **5**, e1000353 (2009).
144. Sams, A. J. & Boyko, A. R. Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. *G3 Bethesda Md* **9**, 117–123 (2019).
145. Ott, J., Wang, J. & Leal, S. M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat. Rev. Genet.* **16**, 275–284 (2015).
146. Hearne, C., Ghosh, S. & Todd, J. Microsatellites for linkage analysis of genetic traits. *Trends Genet.* **8**, 288–294 (1992).
147. Pulst, S. M. Genetic Linkage Analysis. *Arch. Neurol.* **56**, 667 (1999).
148. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
149. Letko, A. *et al.* A Missense Variant in SCN8A in Alpine Dachsbracke Dogs Affected by Spinocerebellar Ataxia. *Genes* **10**, 362 (2019).
150. Kiener, S. *et al.* ABHD5 frameshift deletion in Golden Retrievers with ichthyosis. *G3 GenesGenomesGenetics* jkab397 (2021) doi:10.1093/g3journal/jkab397.
151. Christen, M. *et al.* MIA3 Splice Defect in Cane Corso Dogs with Dental-Skeletal-Retinal Anomaly (DSRA). *Genes* **12**, 1497 (2021).
152. Störk, T. *et al.* TSEN54 missense variant in Standard Schnauzers with leukodystrophy. *PLOS Genet.* **15**, e1008411 (2019).
153. Murgiano, L. *et al.* CCDC66 frameshift variant associated with a new form of early-onset progressive retinal atrophy in Portuguese Water Dogs. *Sci. Rep.* **10**, 21162 (2020).
154. Ahonen, S. J., Arumilli, M. & Lohi, H. A CNGB1 Frameshift Mutation in Papillon and Phalène Dogs with Progressive Retinal Atrophy. *PLoS ONE* **8**, e72122 (2013).
155. Caduff, M., Bauer, A., Jagannathan, V. & Leeb, T. OCA2 splice site variant in German Spitz dogs with oculocutaneous albinism. *PLOS ONE* **12**, e0185944 (2017).
156. Steffen, F. *et al.* A Nonsense Variant in COL6A1 in Landseer Dogs with Muscular Dystrophy. *G3 GenesGenomesGenetics* **5**, 2611–2617 (2015).
157. Christen, M. *et al.* Deletion of the SELENOP gene leads to CNS atrophy with cerebellar ataxia in dogs. *PLOS Genet.* **17**, e1009716 (2021).

158. Becker, A., Geiger, D. & Schäffer, A. A. Automatic Selection of Loop Breakers for Genetic Linkage Analysis. *Hum. Hered.* **48**, 49–60 (1998).
159. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
160. Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **12**, 443–451 (2011).
161. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
162. Altmann, A. *et al.* A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Hum. Genet.* **131**, 1541–1554 (2012).
163. Sefid Dashti, M. J. & Gamieldien, J. A practical guide to filtering and prioritizing genetic variants. *BioTechniques* **62**, 18–30 (2017).
164. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
165. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
166. Smith, C. L. & Eppig, J. T. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **1**, 390–399 (2009).
167. Fitarelli-Kiehl, M. *et al.* Comparison of multiple genotyping methods for the identification of the cancer predisposing founder mutation p.R337H in TP53. *Genet. Mol. Biol.* **39**, 203–209 (2016).
168. Sosnay, P. R. & Cutting, G. R. Interpretation of genetic variants. *Thorax* **69**, 295–297 (2014).
169. Rodenburg, R. J. The functional genomics laboratory: functional validation of genetic variants. *J. Inherit. Metab. Dis.* **41**, 297–307 (2018).
170. Niskanen, J., Salmela, E. & Lohi, H. AncesTrim - a tool for trimming complex pedigrees. *J. Open Source Softw.* **2**, (2017).
171. Gutierrez-Quintana, R. *et al.* A proposed radiographic classification scheme for congenital thoracic vertebral malformations in brachycephalic 'screw-tailed' dog breeds. *Vet. Radiol. Ultrasound Off. J. Am. Coll. Vet. Radiol. Int. Vet. Radiol. Assoc.* **55**, 585–591 (2014).
172. Evans, H. & de Lahunta, A. *Miller's Anatomy of the Dog*. (Saunders, 2013).
173. National Library of Medicine (US), National Center for Biotechnology Information. Nucleotide [Internet]. https://www.ncbi.nlm.nih.gov/nucleotide/.
174. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual*. (CreateSpace, 2009).
175. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
176. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
177. Gao, X., Becker, L. C., Becker, D. M., Starmer, J. D. & Province, M. A. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* **34**, 100-n/a (2010).

178. Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. Runs of homozygosity and population history in cattle. *BMC Genet.* **13**, 70 (2012).
179. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
180. Picard toolkit. *Broad Institute, GitHub repository* (2019).
181. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
182. O'Connor, B. D. & Auwera, G. van der. *Genomics analysis with Spark, Docker, and clouds: a guide to big data tools for genomics research*. (2017).
183. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
184. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
185. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
186. Gardner, E. J. *et al.* The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res.* **27**, 1916–1929 (2017).
187. Broeckx, B. J. G. *et al.* Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Sci. Rep.* **5**, 12810 (2015).
188. Layer, R. M., Kindlon, N., Karczewski, K. J. & Quinlan, A. R. Efficient genotype compression and analysis of large genetic-variation data sets. *Nat. Methods* **13**, 63–65 (2016).
189. Arumilli, M., Layer, R. M., Hytönen, M. K. & Lohi, H. webGQT: A Shiny Server for Genotype Query Tools for Model-Based Variant Filtering. *Front. Genet.* **11**, 152 (2020).
190. NCBI Resource Coordinators *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
191. The UniProt Consortium *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
192. Tweedie, S. *et al.* Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* **49**, D939–D946 (2021).
193. Bendl, J. *et al.* PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Comput. Biol.* **10**, e1003440 (2014).
194. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
195. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
196. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).

197. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* **35**, D26–D31 (2007).
198. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
199. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013).
200. Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007).
201. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
202. Okonechnikov, K., Golosova, O. & Fursov, M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).
203. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020).
204. Niskanen, J. *et al.* Nonsense variant in COL7A1 causes recessive dystrophic epidermolysis bullosa in Central Asian Shepherd dogs. *PLOS ONE* **12**, e0177527 (2017).
205. Chung, H. J. & Uitto, J. Type VII Collagen: The Anchoring Fibril Protein at Fault in Dystrophic Epidermolysis Bullosa. *Dermatol. Clin.* **28**, 93–105 (2010).
206. Fédération Cynologique Internationale. FCI-Standard N° 48: Karjalankarhukoira (Karelian Bear Dog). http://www.fci.be/Nomenclature/Standards/048g05-en.pdf (2015).
207. Camper, S. A., Saunders, T. L., Katz, R. W. & Reeves, R. H. The Pit-1 transcription factor gene is a candidate for the murine Snell dwarf mutation. *Genomics* **8**, 586–590 (1990).
208. Li, S. *et al.* Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* **347**, 528–533 (1990).
209. Grillet, N. *et al.* Mutations in LOXHD1, an Evolutionarily Conserved Stereociliary Protein, Disrupt Hair Cell Function in Mice and Cause Progressive Hearing Loss in Humans. *Am. J. Hum. Genet.* **85**, 328–337 (2009).
210. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
211. Riazuddin, S. A. *et al.* Mutations in LOXHD1, a Recessive-Deafness Locus, Cause Dominant Late-Onset Fuchs Corneal Dystrophy. *Am. J. Hum. Genet.* **90**, 533–539 (2012).
212. Niskanen, J. E. *et al.* Canine DVL2 variant contributes to brachycephalic phenotype and caudal vertebral anomalies. *Hum. Genet.* **140**, 1535–1545 (2021).
213. Wiberg, M. *et al.* Ventricular arrhythmia and sudden cardiac death in young Leonbergers. *J. Vet. Cardiol.* **27**, 10–22 (2020).
214. Bardhan, A. *et al.* Epidermolysis bullosa. *Nat. Rev. Dis. Primer* **6**, 78 (2020).
215. Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).

216. Varki, R., Sadowski, S., Uitto, J. & Pfendner, E. Epidermolysis bullosa. II. Type VII collagen mutations and phenotype-genotype correlations in the dystrophic subtypes. *J. Med. Genet.* **44**, 181–192 (2006).
217. Magnol, J.-P. *et al.* Characterization of a canine model of dystrophic bullous epidermolysis (DBE). Development of a gene therapy protocol. *Bull. Acad. Natl. Med.* **189**, 107–121 (2005).
218. Gache, Y., Pin, D., Gagnoux-Palacios, L., Carozzo, C. & Meneguzzi, G. Correction of Dog Dystrophic Epidermolysis Bullosa by Transplantation of Genetically Modified Epidermal Autografts. *J. Invest. Dermatol.* **131**, 2069–2078 (2011).
219. Gretzmeier, C. *et al.* Systemic Collagen VII Replacement Therapy for Advanced Recessive Dystrophic Epidermolysis Bullosa. *J. Invest. Dermatol.* S0022202X21022739 (2021) doi:10.1016/j.jid.2021.09.008.
220. Atanasova, V. S. *et al.* Amlexanox Enhances Premature Termination Codon Read-Through in COL7A1 and Expression of Full Length Type VII Collagen: Potential Therapy for Recessive Dystrophic Epidermolysis Bullosa. *J. Invest. Dermatol.* **137**, 1842–1849 (2017).
221. Tatsumi, K. *et al.* Cretinism with combined hormone deficiency caused by a mutation in the PIT1 gene. *Nat. Genet.* **1**, 56–58 (1992).
222. Majdoub, H. *et al.* Extreme Short Stature and Severe Neurological Impairment in a 17-Year-Old Male With Untreated Combined Pituitary Hormone Deficiency Due to POU1F1 Mutation. *Front. Endocrinol.* **10**, 381 (2019).
223. Gergics, P. *et al.* High-throughput splicing assays identify missense and silent splice-disruptive POU1F1 variants underlying pituitary hormone deficiency. *Am. J. Hum. Genet.* **108**, 1526–1539 (2021).
224. Bodner, M. *et al.* The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* **55**, 505–518 (1988).
225. Prince, K. L., Walvoord, E. C. & Rhodes, S. J. The role of homeodomain transcription factors in heritable pituitary disease. *Nat. Rev. Endocrinol.* **7**, 727–737 (2011).
226. Pavlović, M. *et al.* Human anterior pituitary's ACTH cells during the aging process: immunohistochemic and morphometric study. *Anat. Sci. Int.* **96**, 250–257 (2021).
227. Voorbij, A. & Kooistra, H. S. Pituitary dwarfism in German shepherd dogs. *J Vet Clin Sci* **2**, 4–11 (2009).
228. Hamann, F. *et al.* Pituitary function and morphology in two German shepherd dogs with congenital dwarfism. *Vet. Rec.* **144**, 644–646 (1999).
229. Laron, Z. Insulin-like growth factor 1 (IGF-1): a growth hormone. *Mol. Pathol.* **54**, 311–316 (2001).
230. Abou Tayoun, A. N. *et al.* Improving hearing loss gene testing: a systematic review of gene evidence toward more efficient next-generation sequencing–based diagnostic testing and interpretation. *Genet. Med.* **18**, 545–553 (2016).
231. Bai, X. *et al.* Five Novel Mutations in LOXHD1 Gene Were Identified to Cause Autosomal Recessive Nonsyndromic Hearing Loss in Four Chinese Families. *BioMed Res. Int.* **2020**, 1–9 (2020).

232. Trouillet, A. *et al.* Loxhd1 Mutations Cause Mechanotransduction Defects in Cochlear Hair Cells. *J. Neurosci.* **41**, 3331–3343 (2021).
233. Plevova, P. *et al.* STRC Deletion is a Frequent Cause of Slight to Moderate Congenital Hearing Impairment in the Czech Republic. *Otol. Neurotol.* **38**, e393–e400 (2017).
234. Strain, G. M., Tedford, B. L. & Jackson, R. M. Postnatal development of the brain stem auditory-evoked potential in dogs. *Am. J. Vet. Res.* **52**, 410–415 (1991).
235. Sommerlad, S. *et al.* Congenital Sensorineural Deafness in Australian Stumpy-Tail Cattle Dogs Is an Autosomal Recessive Trait That Maps to CFA10. *PLoS ONE* **5**, e13364 (2010).
236. Fédération Cynologique Internationale. FCI-Standard N° 286: American Staffordshire Terrier. http://www.fci.be/Nomenclature/Standards/286g03-en.pdf (1997).
237. Fédération Cynologique Internationale. FCI-Standard N° 76: Staffordshire Bull Terrier. http://www.fci.be/Nomenclature/Standards/076g03-en.pdf (1998).
238. Fédération Cynologique Internationale. FCI-Standard N° 116: Dogue de Bordeaux. http://www.fci.be/Nomenclature/Standards/116g02-en.pdf (2009).
239. United Kennel Club. Olde English Bulldogge. Official UKC Breed Standard. *Olde English Bulldogge* https://www.ukcdogs.com/docs/breeds/olde-english-bulldogge.pdf (2013).
240. Fédération Cynologique Internationale. FCI-Standard N° 11: Bull Terrier. http://www.fci.be/Nomenclature/Standards/011g03-en.pdf (2011).
241. Fédération Cynologique Internationale. FCI-Standard N° 359: Miniature Bull Terrier. http://www.fci.be/Nomenclature/Standards/359g03-en.pdf (2011).
242. Hytönen, M. K. *et al.* Ancestral T-Box Mutation Is Present in Many, but Not All, Short-Tailed Dog Breeds. *J. Hered.* **100**, 236–240 (2009).
243. Haworth, K. *et al.* Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm. Genome* **12**, 212–218 (2001).
244. Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A. & Richardson, J. E. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
245. Stockard, C. R. *The genetic and endocrinic basis for differences in form and behavior : as elucidated by studies of contrasted pure-line dog breeds and their hybrids*. (The Wistar Institute of Anatomy and Biology, 1941).
246. Harvey, C. E. Inherited and congenital airway conditions. *J. Small Anim. Pract.* **30**, 184–187 (1989).
247. Packer, R. M. & Tivers, M. S. Strategies for the management and prevention of conformation-related respiratory disorders in brachycephalic dogs. *Vet. Med. Auckl.* **6**, 219–232 (2015).
248. Marchant, T. W. *et al.* Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. *Curr. Biol. CB* **27**, 1573-1584.e6 (2017).
249. Schoenebeck, J. J. *et al.* Variation of BMP3 Contributes to Dog Breed Skull Diversity. *PLoS Genet.* **8**, e1002849 (2012).

250. Hamblet, N. S. *et al.* Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. *Dev. Camb.* **129**, 5827–5838 (2002).
251. Waldo, K., Miyagawa-Tomita, S., Kumiski, D. & Kirby, M. L. Cardiac neural crest cells provide new insight into septation of the cardiac outflow tract: aortic sac to ventricular septal closure. *Dev. Biol.* **196**, 129–144 (1998).
252. Nakajima, Y., Yamagishi, T., Hokari, S. & Nakamura, H. Mechanisms involved in valvuloseptal endocardial cushion formation in early cardiogenesis: Roles of transforming growth factor (TGF)-β and bone morphogenetic protein (BMP). *Anat. Rec.* **258**, 119–127 (2000).
253. Jiang, X., Rowitch, D. H., Soriano, P., McMahon, A. P. & Sucov, H. M. Fate of the mammalian cardiac neural crest. *Dev. Camb. Engl.* **127**, 1607–1616 (2000).
254. Délot, E. C. Control of endocardial cushion and cardiac valve maturation by BMP signaling pathways. *Mol. Genet. Metab.* **80**, 27–35 (2003).
255. Brown, C. B. & Baldwin, H. S. Neural Crest Contribution to the Cardiovascular System. in *Neural Crest Induction and Differentiation* vol. 589 134–154 (Springer US, 2006).
256. Hutson, M. R. & Kirby, M. L. Model systems for the study of heart development and disease: Cardiac neural crest and conotruncal malformations. *Semin. Cell Dev. Biol.* **18**, 101–110 (2007).
257. Brambilla, P. G. *et al.* Epidemiological study of congenital heart diseases in dogs: Prevalence, popularity, and volatility throughout twenty years of clinical practice. *PloS One* **15**, e0230160 (2020).
258. Vernau, K. M. *et al.* A Missense Variant in ALDH5A1 Associated with Canine Succinic Semialdehyde Dehydrogenase Deficiency (SSADHD) in the Saluki Dog. *Genes* **11**, 1033 (2020).
259. Leeb, T. *et al.* A Missense Variant Affecting the C-Terminal Tail of UNC93B1 in Dogs with Exfoliative Cutaneous Lupus Erythematosus (ECLE). *Genes* **11**, 159 (2020).
260. Letko, A. *et al.* Genomic diversity and population structure of the Leonberger dog breed. *Genet. Sel. Evol.* **52**, 61 (2020).
261. Ekenstedt, K. J. *et al.* An ARHGEF10 Deletion Is Highly Associated with a Juvenile-Onset Inherited Polyneuropathy in Leonberger and Saint Bernard Dogs. *PLoS Genet.* **10**, e1004635 (2014).
262. Becker, D. *et al.* A GJA9 frameshift variant is associated with polyneuropathy in Leonberger dogs. *BMC Genomics* **18**, 662 (2017).
263. Letko, A. *et al.* A CNTNAP1 Missense Variant Is Associated with Canine Laryngeal Paralysis and Polyneuropathy. *Genes* **11**, 1426 (2020).
264. Deveyrac, K. & Miniature Schnauzer Club of Canada Board of Directors. Potential consequences of proposed bans on ear cropping and tail docking for the miniature schnauzer gene pool. *Can. Vet. J. Rev. Veterinaire Can.* **51**, 929 (2010).
265. Ko, D. C. & Urban, T. J. Understanding Human Variation in Infectious Disease Susceptibility through Clinical and Cellular GWAS. *PLoS Pathog.* **9**, e1003424 (2013).

266. Qu, Z. & Weiss, J. N. Mechanisms of Ventricular Arrhythmias: From Molecular Fluctuations to Electrical Turbulence. *Annu. Rev. Physiol.* **77**, 29–55 (2015).
267. Mavropoulou, A., Oliveira, P. & Willis, R. Holter monitoring in dogs: 24 h vs. 48 h. *Vet. J.* **272**, 105628 (2021).
268. Gunasekaran, T., Olivier, N. B. & Sanders, R. A. Comparison of single- versus seven-day Holter analysis for the identification of dilated cardiomyopathy predictive criteria in apparently healthy Doberman Pinscher dogs. *J. Vet. Cardiol.* **27**, 78–87 (2020).
269. Barrett, P. M. *et al.* Comparison of 24-hour Holter Monitoring with 14-day Novel Adhesive Patch Electrocardiographic Monitoring. *Am. J. Med.* **127**, 95.e11-95.e17 (2014).
270. Lichtenberger, J., Meurs, K. M. & Côté, E. Preliminary Assessment of a Novel 14-Day Electrocardiographic Adhesive Patch Monitor in Dogs. *J. Am. Anim. Hosp. Assoc.* **54**, 138–143 (2018).
271. Illumina, Inc. *Data Sheet: DNA Genotyping. #370-2009-018.* https://support.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_caninehd.pdf.
272. Ostrander, E. A. *et al.* Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. *Natl. Sci. Rev.* **6**, 810–824 (2019).
273. Lizio, M. *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res.* **47**, D752–D758 (2019).
274. Packer, R. M. A., O'Neill, D. G., Fletcher, F. & Farnworth, M. J. Great expectations, inconvenient truths, and the paradoxes of the dog-owner relationship for owners of brachycephalic dogs. *PLOS ONE* **14**, e0219918 (2019).
275. Bannasch, D. *et al.* The effect of inbreeding, body size and morphology on health in dog breeds. *Canine Med. Genet.* **8**, 12 (2021).