ANNUAL REVIEWS

Annual Review of Animal Biosciences

Identification of Genetic Risk Factors for Monogenic and Complex Canine Diseases

Tosso Leeb,[1] Danika Bannasch,[2] and Jeffrey J. Schoenebeck[3]

[1]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland; email: tosso.leeb@vetsuisse.unibe.ch

[2]Department of Population Health and Reproduction, University of California, Davis, California, USA; email: dlbannasch@ucdavis.edu

[3]The Roslin Institute and Royal (Dick) School for Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom; email: jeff.schoenebeck@roslin.ed.ac.uk

Annu. Rev. Anim. Biosci. 2023. 11:183–205

First published as a Review in Advance on November 2, 2022

The *Annual Review of Animal Biosciences* is online at animal.annualreviews.org

https://doi.org/10.1146/annurev-animal-050622-055534

Keywords

Canis lupus familiaris, dog, veterinary medicine, One Health, precision medicine, animal model

Abstract

Advances in DNA sequencing and other technologies have greatly facilitated the identification of genetic risk factors for inherited diseases in dogs. We review recent technological developments based on selected examples from canine disease genetics. The identification of disease-causing variants in dogs with monogenic diseases may become a widely employed diagnostic approach in clinical veterinary medicine in the not-too-distant future. Diseases with complex modes of inheritance continue to pose challenges to researchers but have also become much more tangible than in the past. In addition to strategies for identifying genetic risk factors, we provide some thoughts on the interpretation of sequence variants that are largely inspired by developments in human clinical genetics.

1. INTRODUCTION

Linkage disequilibrium (LD): nonrandom association of alleles caused by limited or absent recombination between two neighboring loci

Heterogeneity: the same phenotype is caused by different variants in different individuals

Dogs provide an unparalleled model for human diseases that has garnered significant research interest. They share their environment intimately with people, including exposures to pesticides and toxins, water supply, and foods. They also receive veterinary care very similar to that of humans as far as both diagnostics and treatment modalities (1). Gene therapy, cancer treatment clinical trials, and studies of aging and behavior are just a few areas in which the canine model has been and will be used to advance human and veterinary medicine (2–8). Many examples of disease gene discoveries in dogs preceded the identification of homologous disease gene variants in human patients (9–15).

2. DOG POPULATION STRUCTURE

The many parallels between medical genetics applied to human patients and domestic animals are one of the reasons that domestic animals in general, and dogs in particular, are increasingly recognized as valuable models for biomedical research. Dogs are particularly attractive for forward genetic studies given their specific population structures (16).

2.1. Purebred Dogs

In the past 200 years, approximately 300–400 dog breeds have been formed from a limited number of founder animals (1, 17). Purebred dogs have been maintained in strictly closed populations for ~10–100 generations, requiring a considerable amount of inbreeding and resulting in a very special genetic makeup. Collectively, purebred dogs retain an amount of genetic variation comparable to that in humans, with several genetic variants per kilobase of genomic DNA on average. However, due to inbreeding, many purebred dogs also have long runs of homozygosity that may constitute a significant proportion of their genomes. Selection and breeding for breed-specific phenotypes have led to genetic fixation at a small number of loci in purebred dogs, but the vast majority of their genomes are still as variable and heterozygous as that of an average human (18). Nonetheless, breed foundation represented a severe genetic bottleneck and was based on very few founder chromosomes, resulting in relatively low haplotype diversity. Given the relatively short time span of fewer than 100 generations since breed foundation, intrabreed linkage disequilibrium (LD) is much longer than in average human populations and can still extend for more than 1 Mb in modern purebred dogs (16).

Closed populations with significant inbreeding promote the expression of simple and complex recessive diseases. If a new recessive disease arises within one breed, all affected dogs very often trace back to one founder animal and carry the same causal variant. The low level of heterogeneity within breeds greatly facilitates the identification of disease-causing variants. The long LD on the one hand helps to identify disease loci but on the other hand severely limits the ability for precise fine mapping. However, analyses using purebred animals need to consider confounding effects of artificial selection and genetic drift, which can manifest as false positive disease associations.

2.2. Random-Bred Dogs

Fewer attempts to identify disease-causing variants in random-bred dogs have been reported in the literature compared to the wealth of published studies on purebred dogs. Random-bred dogs have very heterogeneous population structures. As their pedigrees normally are undocumented, researchers must be aware of the breadth of possibilities. Although often not expected, random-bred dogs can be equally or more strongly inbred than purebred dogs. Brother–sister or parent–offspring matings are not uncommon. On the other hand, many random-bred dogs are

not closely inbred and will then more closely resemble the population structures of non-isolated human populations. Interest in mixed breeds or hybrid breeds (crosses between two purebreds) has increased recently. We distinguish these from random-bred dogs by the presence of a pedigree, which indicates the breeds and ancestry of the relatives. In the F1 generation, these animals should be highly heterozygous, depending on the relationship of the parent breeds. In one study, the Australian labradoodle showed low levels of inbreeding compared to the parent breeds; however, strong selection for traits from the poodle was also apparent (19). Mixed-breed dogs are not accepted or documented by kennel clubs, so their pedigrees are largely unknown; however, this appears to have led to significantly greater genetic diversity on average (20).

WGS: whole-genome sequencing

Neomorphic allele: an allele with a new function; may lead to new phenotypes that require revision of known genotype–phenotype correlations

3. EXPERIMENTAL STRATEGIES TO IDENTIFY DISEASE GENES

3.1. Functional Candidate Genes

Many inherited diseases present with highly specific phenotypes that enable the generation of hypotheses regarding possible underlying genes. The selection of appropriate functional candidate genes is facilitated by increasing knowledge from studies in other species, including humans, and the increasingly sophisticated diagnostic approaches available in veterinary clinics and pathology.

As an example, in German hunting terriers with an autosomal recessively inherited exercise-induced myopathy, an acylcarnitine profile screening in blood showed increased tetradecenoylcarnitine (C14:1) levels. This suggested a deficiency in acyl-CoA dehydrogenase very long chain (ACADVL). Subsequent genetic analysis confirmed that the myopathic dogs carried an *ACADVL* nonsense variant in homozygous state (21).

A precise and accurate characterization of the phenotype is key to the functional candidate gene approach. This approach relies on expert knowledge of published genotype–phenotype correlations and the molecular etiology of inherited diseases, including biochemistry and pathophysiology. Although the choice of a single or a clearly delimited group of functional candidate genes is relatively easy for some highly characteristic phenotypes, it may be very challenging for others. A notorious example are neurodegenerative diseases that often result from general defects in cellular metabolism. Neurons often do not divide, are unable to regenerate, and are consequently particularly sensitive to metabolic disturbances. Therefore, many genetic defects that affect basic cellular functions result in neurological phenotypes, as neurons are usually the first cell type to die from a general defect.

Due to advances in DNA sequencing, the functional candidate gene approach has become feasible even in situations involving a fairly large number of candidate genes. For example, there are at least 37 known candidate genes for epidermolysis bullosa (22). Analysis of these 37 genes in whole-genome sequencing (WGS) data helped to identify a frameshift variant in *COL7A1* as a cause for recessive dystrophic epidermolysis bullosa in basset hounds (23).

Finally, however, the hypothesis-driven functional candidate gene approach is bound to fail if an inherited disease is caused by a variant in an orphan gene whose function has not been characterized previously. There are still thousands of genes in the mammalian genome with partially or completely unknown functions. Other instances in which the functional candidate gene approach will not work include variants leading to neomorphic alleles (24) or diseases due to newly inserted active retrogenes (25, 26).

3.2. Hypothesis-Free Holistic Approaches

Hypothesis-free approaches for disease gene identification rely on statistical methods and have proven immensely powerful in the last decades. Their main advantage is that they do not require any mechanistic knowledge about disease etiology. For successful disease gene mapping, it is

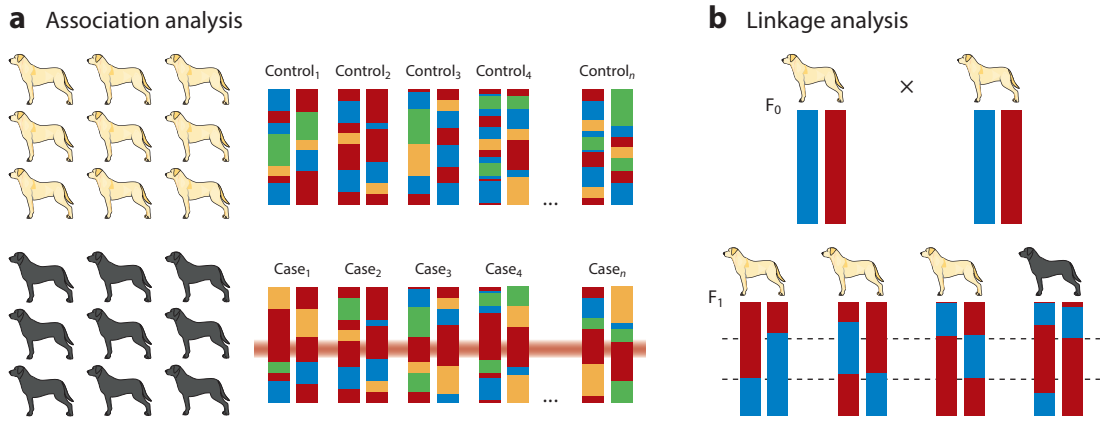

Figure 1

Approaches to map the position of disease-causing variants. The examples illustrate a monogenic autosomal recessive trait indicated by coat color. However, these methods work for any mode of inheritance, including complex traits and diseases. (*a*) An association study is a population-based method that requires cohorts of cases and controls. The relationships between the dogs are typically not known beforehand. (*b*) Linkage analysis requires complete families. The example shows a typical situation in which two heterozygous parents carry the same disease-associated chromosome (haplotype). However, linkage analysis would also work in a situation of allelic heterogeneity, i.e., if the parents carried different pathogenic alleles and two different disease-associated haplotypes.

sufficient to classify dogs as affected (cases) or nonaffected (controls). However, these methods typically require more than one affected dog and the collection of samples from cohorts ranging from two to thousands of dogs (**Figure 1**).

We first describe the most important positional approaches to mapping disease genes in the dog genome. We then briefly summarize sequence-based approaches that aim to directly identify disease-causing variants without prior mapping steps.

3.2.1. Genome-wide association studies. Genome-wide association studies (GWAS) are the method of choice when relatively many cases and control samples are available (16, 27, 28) (**Table 1**). Most inherited diseases in dogs are due to relatively young deleterious alleles segregating primarily within single breeds. Ideally, GWAS require unrelated cases and controls, with the provision that the cases nonetheless trace back to an unknown common ancestor. The idealistic scenario with unrelated cases and controls is rarely or never met in canine disease gene mapping studies. With the advent of better statistical methods for the correction of confounders, such as cryptic relationships between the samples, use of related dogs has become standard in GWAS. Avoiding the use of first-degree relatives in the study design is still recommended, but more distantly related dogs are commonly used in canine GWAS. The required >50,000 marker genotypes per dog can be obtained easily using either commercially available single-nucleotide variant (SNV) microarrays or sequenced-based approaches.

Table 1 Empirically determined sample number requirements for genome-wide association studies in dogs for traits with high penetrance (modified from 16)

Mode of inheritance	Minimum number of cases and controls
Monogenic recessive (Mendelian)	10 + 10
Monogenic dominant (Mendelian)	20 + 20
Complex (>fivefold increased risk)	>100 + >100

Successful GWAS have the potential to map disease genes to relatively small critical intervals. Mapping resolution depends mostly on the age of the disease-causing allele but also on the number and representative distribution of the available samples. For young deleterious alleles that arose in the past 50 years, the mapping resolution may be limited to several megabases. For diseases for which the underlying mutation event happened several hundred or even thousands of years ago, numerous recombination events and the resulting breakdown of LD enable a much more precise mapping to a few hundred or even tens of kilobases. For disease alleles shared between several breeds, the mapping resolution can be improved significantly if cases and controls from several breeds can be included in an across-breed GWAS (16). However, this requires careful matching of cases and controls and/or special attention to the confounding effect of the stratification on the observed p-values.

Importantly, GWAS relies upon segregation of the disease status within the population or breed. When accessing whether a disease phenotype segregates within a single population (e.g., breed), it is important to consider penetrance: If disease predisposition is fixed within a breed, but penetrance is incomplete, only modifying loci will be detected by intrabreed GWAS. For some traits that are fixed within breeds, across-breed mapping approaches have successfully identified the causative variants. These are most successful when there are many breeds with and without the trait of interest. This approach relies on the alleles being identical by descent (IBD) across breeds. Examples of successful across-breed mapping approaches include body size, morphology, coat color, and tail type (25, 29–35).

Identical by descent (IBD): identical haplotypes inherited through the maternal and paternal lines and resulting in long runs of homozygosity after inbreeding events

Logarithm of the odds (LOD): measure for the significance of results in linkage analyses (commonly accepted thresholds: LOD $\leq -2 \rightarrow$ linkage excluded; $-2 <$ LOD $< 3 \rightarrow$ ambiguous result; LOD $\geq 3 \rightarrow$ linkage established)

3.2.2. Linkage mapping. Linkage mapping is the method of choice if samples from complete families are available. This is particularly relevant for rare monogenic diseases, for which often only a single litter with affected dogs may be available. The favorable family structure of dogs, with relatively large litters that may comprise anywhere between 1 and ~12 puppies, is well-suited for linkage analysis. Linkage analysis does not require the high marker density of a GWAS. However, as the formerly used microsatellite markers have been largely replaced by SNVs due to their robust and cost-efficient genotyping, modern linkage analyses in dogs are typically done with the same microarray- or sequence-based SNVs as GWAS (36).

Linkage analyses in small families require that samples from both parents and at least two offspring are available. With single families and small numbers of samples, it is often not possible to reach logarithm of the odds (LOD)-scores of ≥ 3, the generally accepted significance threshold to confirm linkage to a given locus. However, single families are often sufficient to reach LOD-scores of ≤ -2, which allow reliable exclusion of parts of the genome. The mapping resolution of linkage analyses is typically in the range of several or even tens of megabases. Nonetheless, if 50 Mb of the dog genome shows positive or ambiguous LOD-scores, while the remaining >2,300 Mb or 98% of the genome can be safely excluded, a major step toward successful disease gene identification has been accomplished.

3.2.3. Autozygosity/homozygosity mapping. Dogs with a recessively inherited disease are very often inbred to a founder animal. This means that they inherited the disease-causing allele together with adjacent haplotype segments IBD. The resulting homozygosity can be efficiently exploited to map disease genes (37). Homozygosity mapping is a simple and powerful fine-mapping method to delimit a critical interval. This method can be applied to as few as two affected dogs and does not require samples from unaffected controls. The combination of homozygosity mapping with linkage analysis and/or GWAS has become a standard approach to map the disease-causing genes for monogenic autosomal recessive traits in dogs.

3.2.4. Haplotype analyses for fine mapping. The identification of shared IBD haplotype segments for fine-mapping purposes is not limited to homozygosity mapping in monogenic

autosomal recessive diseases. Phasing of genotype data and the separation of the two individual haplotypes also allow searches for shared haplotypes in cases affected by diseases with dominant or even complex modes of inheritance (38, 39). During the mapping of the gene for ectodermal dysplasia in hairless dogs, the identification of a shared haplotype between 140 dogs of 3 different breeds delimited the critical interval to only 102 kb (39).

3.2.5. Direct methods for integrated disease gene and causative variant identification. Advances in WGS have enabled approaches that directly aim to identify causal variants, obviating the strict need for a dedicated disease gene identification step. Today, causative variants for rare monogenic diseases can be identified directly by searching for private variants in the genome of an affected dog, contrasted against a few hundred or a few thousand control genomes from genetically diverse control dogs (40). Data sets with hundreds of canine genomes are publicly available (35, 41).

Causal variants for dominant pathogenic alleles that have arisen from de novo mutation events can be identified quickly via WGS of the affected offspring and both unaffected parents, the so-called trio-sequencing approach. Extrapolating from humans and cattle, one might expect approximately 100 de novo mutation events per trio of dogs, of which on average only one will represent a protein-changing variant (42–44). Successful causal variant identification by trio sequencing in dogs has been reported (15).

4. EXPERIMENTAL APPROACHES FOR VARIANT IDENTIFICATION

4.1. Reference Genome Versus Reference-Free Variant Detection/Annotation

Sequencing reads and genotypes are relatively meaningless without contextualization. The reference genome provides this information through two essential components: the assembly and ancillary data sets, called annotation. The dog reference genome is accessed graphically or can be downloaded from various sources, among the most popular being the National Center for Biotechnology Information (NCBI) Genome Data Viewer, Ensembl, and the University of California, Santa Cruz (UCSC) Genome Browser (45).

The assembly contains the DNA sequence of a single animal. Although tools and techniques involved in producing an assembly have evolved rapidly in the past decade, the fundamental steps remain the same. The reference genome is constructed from long DNA sequences, which are assembled into contiguous overlapping reads called contigs. Even modern assembly software struggles to incorporate reads from centromeres, telomeres, and other areas of the genome that are enriched for long stretches of highly repetitive sequences. To bridge these gaps, contigs are placed and oriented into scaffolds. In high-quality assemblies, the scaffolds often span the length of chromosomes. The reference's remaining DNA content is represented as unplaced scaffolds and contigs. In the context of mapping diseases and traits, the latter are usually ignored. Traditionally, the DNA sequence of a diploid assembly is represented as a consensus haplotype. As such, the true haplotypes of chromosome pairs are lost, and wastefully, half of the genome information is ignored. To address this, the production of haplotype-resolved assemblies will become increasingly more common (46).

For years, the assembly of a female boxer, called Tasha, served as the dog research community's genome (47). However, third-generation sequencing technologies along with scaffolding techniques, such as optical mapping and definition of DNA interactions through Hi-C, have revolutionized production of genome assemblies. To date, chromosome-level assemblies have been produced from a variety of breed dogs, including the boxer Tasha (48), basenji (49), German shepherd (50, 51), Great Dane (52), and Labrador retriever (L. Eory, W. Zhang, J. Aguilar, M. Jackson, D.D. Ozdemir, et al., manuscript in preparation). Wild canids, in the form of dingoes and wolf, are also represented by chromosome-length reference assemblies (53, 54). The NCBI compared five

of the dog assemblies to highlight their strengths and weaknesses (55). Ultimately, the appropriate choice of assembly depends on the experimental goals.

Annotation provides context to the assembly's DNA by specifying regions or motifs of interest. These are assigned physical coordinates, specified in base-pair positions, on a chromosome or scaffold. Unlike the assembly, annotation is produced from a multitude of samples, molecular sources including DNA, RNA, and proteins. To reflect differential biology, annotation data can be derived from different tissues/cell types and disease states. The most commonly accessed annotations are gene and transcript models, produced by sequence analysis pipelines maintained by NCBI's RefSeq and Ensembl. Other annotation, to name but a few, includes coding and noncoding genes (56), variant calls (41), short-read RNA sequencing (51), various epigenetic marks (57–60), multispecies conservation (61), transcriptional start sites, and isoform sequencing (62). Given the current paucity in functional annotation of dog assemblies' noncoding features, the reporting of disease-causing variants has probably been biased toward protein-changing coding variants. The identification of noncoding functional variants remains challenging despite some recent advances (32, 63–66). It is expected that canine genetics will closely follow human standards with respect to the identification of noncoding disease variants (67).

The dog reference genome is the foundation for resequencing-based disease variant discovery. First, short reads produced by WGS/whole exome sequencing are mapped against the assembly to assign reads' physical coordinates. During variant calling, physical coordinates are assigned to genetic variants. The assembly is also used to polarize genetic variants so that alleles are assigned reference and alternate allele subtypes; this helps highlight base changes of potential interest. Finally, variant calls typically are coded to highlight overlap or proximity to genes, transcripts, allele frequencies, and possibly other genome annotations.

In aggregate, dogs are genetically diverse. Therefore, the process of disease variant discovery using short reads presents challenges regardless of the genome reference chosen for mapping. As alluded to above, less than 50% of a dog's diploid DNA content is represented in a 2.4-Gb haploid reference assembly. If all the DNA content present among all dogs is considered (e.g., the dog pangenome), then the DNA content represented within a single assembly is even less complete. In practice, the incompleteness of an assembly results in unmapped reads and thus undermines disease variant discovery. Moreover, the assembly itself introduces bias because it favors calling genetic variants from reads that map to it. As an alternative to linear reference assemblies, there is growing interest in producing variant-aware graph genomes (68), which incorporate the genetic variation and sequence uncertainty from the dog pangenome (**Figure 2**). Aside from increasing the overall mappable portion of eukaryote genomes by tens to hundreds of megabases, the use of graph genomes promises to enhance variant-calling sensitivity and accuracy, particularly for structural variants (SVs) (69–71). Another promising feature of graph genomes is the ability to tailor variant calling toward the genetic variation relevant to certain breeds or subpopulations of interest (72). The production of multiple high-quality canid assemblies will serve as a good start toward producing graph genomes. However, to fully leverage graph genomes, the dog research community must invest in efforts that catalog SVs from discovery approaches based on long-read sequencing.

Whole exome sequencing: typically interrogates 50–70 Mb of the 2,400-Mb dog genome

Pangenome: the union of all genomes within a clade (e.g., dogs), including information on the existing variants and haplotypes

Structural variant (SV): a sequence variant involving more than 20 consecutive nucleotides (others set the threshold higher, e.g., 100, 500, or 1,000 nt)

Small variant: a sequence variant involving 20 or fewer consecutive nucleotides; small variants include SNVs and small indel variants

4.2. Variant Categories

SNVs and small indel variants are the most common types of variants in the genome. Large-scale genome-sequencing efforts have identified close to 100 million of these variants in the canine genome (35, 41). However, the less common SVs, which we define here as changes involving more than ~20 nucleotides, still comprise the majority of variable nucleotides in the genome. The distinction between small variants and SVs is important, as current methodologies differ in their sensitivity and specificity to detect different classes of sequence variants (73, 74). Although

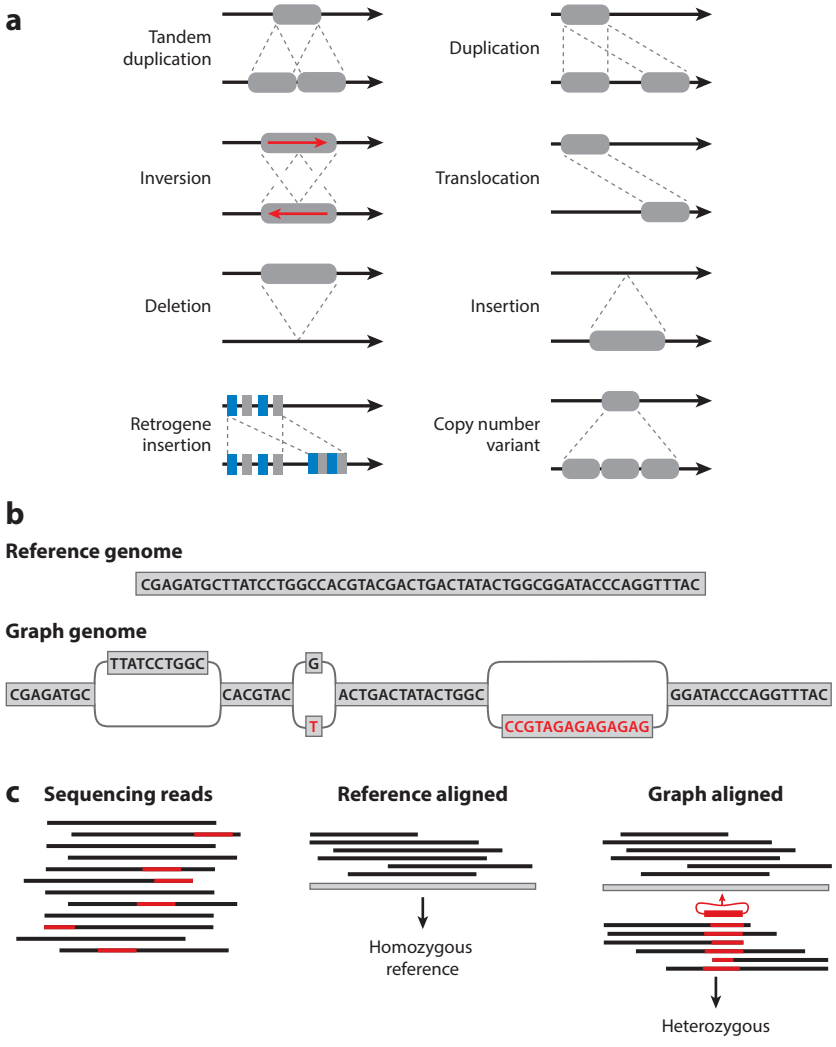

Figure 2

Structural variants (SVs). (*a*) Different SV types are indicated schematically. The insertion of intronless retrogenes is more common in dogs than in other mammalian species (145). (*b*) Schematic representations of a conventional linear reference genome assembly and a variant-aware graph assembly. (*c*) Current mapping and variant-calling workflows using linear reference assemblies have a low sensitivity to detect SVs, especially insertions that do not align to the reference. This may easily lead to false genotyping calls. The graph-aligned genotype calling gives the correct genotype for known variants contained in a variant-aware graph reference genome.

earlier research, due to technological limitations, may have been biased toward the identification of small pathogenic variants, it is now clear that all types of variants must be considered in canine disease genetics.

Dogs have a sizeable population of active long interspersed nuclear element (LINE)-1 insertions (52). Both short interspersed nucleotide element (SINE) and LINE-1 insertions are a common cause of inherited diseases in dogs (75–84). Retrogene insertions represent one particular class of structural sequence variation that has occurred several times in dogs (85) and that

may be relevant for diseases, for example, as in chondrodysplasia and chondrodystrophy (25, 26). Purely positional approaches to map a critical interval for a disease locus work perfectly fine in the case of retrogene insertions. However, researchers must be aware that if the disease is caused by a retrogene insertion, the annotated genes in a critical interval may be unrelated to the studied disease. In this case, the disease-causing variant also does not necessarily affect the coding sequence of one of the annotated genes.

4.3. Sanger Sequencing

The traditional approach to variant identification consists of an initial PCR amplification of the region of interest, followed by the direct Sanger sequencing of the PCR products. This standard approach is widely available to many laboratories and can be completed within one to two weeks without the need for complex bioinformatics and a high-performance computing infrastructure. Today, Sanger sequencing is used only for very small target regions that do not exceed a few kilobases in size. A typical application might involve a highly specific disease phenotype with a single functional candidate gene. In this case, PCR primers for the amplification of ~300–600-bp products harboring individual exons and their immediately flanking intronic sequences will be designed. Sanger sequencing of the resulting amplicons has good sensitivity to detect SNVs and small indels present in either a homozygous or heterozygous state (86, 87). However, the sensitivity to detect SVs is low. Especially if an SV is present in a heterozygous state, there is a high risk that the PCR will amplify only the reference allele but not the alternate allele, resulting in a false-negative variant call and an incorrectly determined genotype.

4.4. Targeted Next-Generation Sequencing

This approach involves the enrichment of larger target regions (up to several megabases in size) followed by next-generation sequencing. Target regions may consist of either functionally related genes (gene panels) or the complete exome. These approaches have become very popular in human (diagnostic) genetics but have not been readily adopted in canine genetics. This is largely because the reagents required for the enrichment (highly multiplexed PCR primers or single-stranded probes for hybridization-based enrichment) are quite expensive and become economically attractive only if fairly large batches of samples can be processed simultaneously.

Variation in GC content may lead to biases in the representation of enriched libraries and consequently to uneven sequence coverage of the targeted regions. Whereas human exome enrichment reagents have gone through several rounds of optimization to ensure comparable representation of all targeted sequences, the available dog exome enrichment reagents have not yet reached the same quality level, reflecting the much smaller market volume. However, unlike for most other domestic animal species, canine exome enrichment reagents are commercially available and represent useful tools, e.g., when high sequence coverage to detect somatic variation is required (88). The use of canine exome sequencing to detect germline variants for Mendelian diseases has also been reported occasionally, e.g., for the discovery of a *LAMP3* variant in dogs with fatal neonatal interstitial lung disease (89).

Analyzing gene panel or exome sequencing data requires basic bioinformatic skills and moderately powerful computer hardware. However, the requirements (and costs) for data analysis and storage are substantially lower than for WGS.

4.5. Whole-Genome Sequencing

WGS using the Illumina short-read platform has become the method of choice for identifying sequence variants for most canine disease studies during the last years. It works best when

high-quality genomic DNA is available and a PCR-free library with 350–550-bp insert size can be prepared. The generation of PCR-free libraries minimizes GC biases and helps to achieve even sequence coverage across the genome. For reliable detection of germline sequence variants, WGS is typically performed at 20–30× coverage, which corresponds to the collection of approximately 180–270 million read pairs of 2 × 150 bp from the dog to be sequenced.

The raw sequence data are then mapped and aligned to a reference genome assembly, and variants with respect to the reference are called (90). The variant calling process must be done separately for small variants and SVs. Sensitivity and specificity for small variants are excellent with current software tools. Calling of SVs from Illumina short-read data is much less reliable, and sensitivity and specificity are currently insufficient to fully rely on automated SV-calling pipelines. The problems in SV calling are due mostly to three reasons: (*a*) The short read lengths preclude the unambiguous mapping of reads from repetitive regions, which is required to make full use of the read-pair information and to exploit the information from split-read mappings. (*b*) Gaps in imperfect reference genome assemblies often prompt false-positive SV calls. (*c*) And finally, insertion variants lead to sequence reads that cannot be correctly aligned to the reference genome. Because insertions often represent repetitive sequences (e.g., retroposons such as SINEs and LINEs), they are rarely accurately called by the currently available software tools. Genome-wide automated bioinformatic calling of SVs is possible, but, as mentioned above, many false-negative and false-positive SV calls have to be expected. Visual inspection of short-read alignments in the Integrative Genomics Viewer (IGV) browser has a very good sensitivity and specificity to detect SVs (91, 92). However, this approach is extremely time consuming and therefore limited to intervals of a few megabases at most. According to our experience, a trained researcher cannot visually analyze more than 1 Mb during an 8-h working day.

Illumina WGS is based on a highly standardized workflow that is relatively simple in the laboratory (and identical to human WGS). The costs of producing the raw data range from $500 to $800 for a single dog genome as of May 2022. The main challenge of WGS lies in the data analysis, which requires a very powerful high-performance computing infrastructure and sufficient storage capacities. Although thousands of dog WGS data sets are publicly available, only very few laboratories have sufficient IT resources to effectively work with such massive data volumes. Hopefully, additional public resources will be developed to facilitate the democratization of large-scale WGS analyses for the entire dog community. The availability of variant data from hundreds of thousands of human genomes and exomes in the gnomAD browser significantly advanced human disease genetics (93; **https://gnomad.broadinstitute.org**). A comparable canine gnomAD resource could be envisioned for the near future.

Further developments in sequencing technology are expected to lead to major improvements. Third-generation, single-molecule, long-read technologies such Pacific Biosciences and Oxford Nanopore sequencing enable de novo assemblies and reference-independent variant calling. These technologies also allow the accurate calling of SVs (94). Although they are still too expensive for routine use, these or other new technologies may eventually replace Illumina short-read sequencing.

5. MONOGENIC VERSUS COMPLEX DISEASES

5.1. Monogenic Diseases

As of May 2022, the OMIA (Online Mendelian Inheritance in Animals) database lists 324 canine diseases with at least one known causal variant (**https://www.omia.org**; 95, 96). The vast majority are fully penetrant monogenic diseases with unambiguous genotype–phenotype correlation. Monogenic diseases allow the distinction of two phenotypic categories, cases and controls, and

the unequivocal prediction of the underlying genotypes at the disease-causing variant for at least the cases or the controls (depending on the mode of inheritance). This greatly facilitates the identification of causal variants, especially if the causal variant belongs to one of the more challenging categories, such as noncoding regulatory variants.

A single dog with a discordant genotype–phenotype correlation is sufficient to exclude a candidate variant as disease causing (or to reject the hypothesis of a fully penetrant monogenic mode of inheritance). This represents a major difference in the genetic analyses of monogenic versus complex diseases.

Allelic heterogeneity: the same phenotype is caused by different variants in the same gene in different individuals

Locus heterogeneity: the same phenotype is caused by different variants in different genes in different individuals

5.2. Complex Diseases that Represent a Group of Monogenic Diseases (Heterogeneity)

Recent research has unveiled allelic and locus heterogeneity for some inherited diseases, even within single breeds. Heterogeneity is of course well known and widespread in human genetics. However, in purebred dogs, with their closed populations and significant inbreeding, it is relatively uncommon. Canine researchers must consider the possibility that a clinically homogeneous phenotype shared between two or more dogs may in fact represent a group of genetically different problems (**Figure 3**).

This is further illustrated with two examples. In the Leonberger breed, polyneuropathy resulting in denervation and secondary muscle atrophy, particularly of the pelvic limbs, was recognized almost 20 years ago (97). The mode of inheritance appeared to be complex (98). A

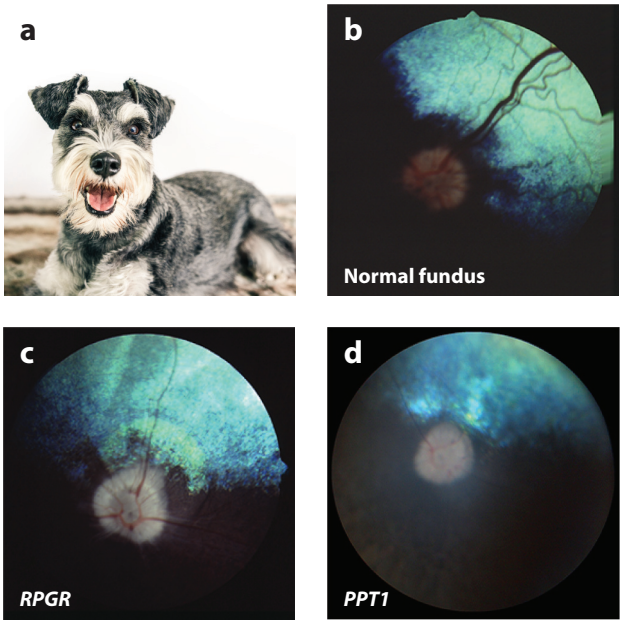

Figure 3

Locus heterogeneity in progressive retinal atrophy (PRA). (*a*) In the Miniature Schnauzer breed, several genetically distinct PRA forms exist. (*b*) Fundus image showing a healthy retina with well-defined blood vessels. (*c*) Advanced disease in a dog affected with XLPRA2 due to a variant in *RPGR* showing degeneration of the blood vessels, a pale optic disc, and increased reflection from the tapetum lucidum, indicating thinning of the retina (105, 106). (*d*) Advanced disease in a dog with PRA$_{PPT1}$ due to a variant in *PPT1/HIVEP3* (106, 107). The retinal findings in both diseases at all stages of the disease are identical. Photos in panels *b* and *c* were provided courtesy of Dr. Gus Aguirre, and the photo in panel *d* was provided by Dr. Julien Charron.

series of molecular genetic investigations then demonstrated that Leonberger polyneuropathy is a heterogeneous group of diseases with at least three pathogenic variants in *ARHGEF10*, *GJA9*, and *CNTNAP1* (99–101). Some cases of polyneuropathy in Leonbergers cannot be explained by any of these variants, suggesting that further pathogenic variants segregate in the population. The elucidation of causative variants in Leonberger polyneuropathy helped to better define the clinical variability, and differences in the genotype–phenotype correlations with, e.g., slightly different (but overlapping) ranges for the age of onset and the specific anatomical distribution of the nerve defects have now become clear (100). As soon as the causative variant for a subset of heterogeneous cases is known, it becomes easier to search for the remaining unknown disease-causing variants, as genotyping for known causative variants enables a rational stratification of the case cohort.

The second example involves early-onset ataxia in Belgian shepherd dogs. Here, an initial compilation of six related litters with affected puppies suggested a monogenic autosomal-recessive disease. However, separate linkage analyses for each of the litters pointed to different chromosomes. So far, four disease-causing variants leading to identical or at least similar clinical phenotypes have been identified in *KCNJ10*, *ATP1B2*, *YARS2*, and *SELENOP* (77, 102–104). Similar to the first example, additional unknown disease-causing variants may exist, as the four known variants still do not explain all known cases of Belgian shepherd dogs with early-onset ataxia (104). These examples underscore the value of precise and accurate phenotyping. The genetic analysis of inherited diseases is facilitated greatly if heterogeneous groups of cohorts can be assigned to more specific homogeneous subphenotypes based on consistent phenotypic criteria.

5.3. Truly Complex Diseases

Complex diseases, also known as polygenic diseases, result from the confluence of multiple alleles located throughout the carrier's genome. The collective effects of these alleles determine disease risk, onset, and penetrance. Unlike those of fully penetrant monogenic diseases, the occurrence and severity of polygenic diseases are often influenced by extrinsic factors such as environment (nutrition, climate, chemical/biological exposure) and lifestyle (exercise, food intake). Excluding infectious diseases and injuries, the majority of human societies' health burden is caused by complex diseases. Among the most common are cardiovascular disease, stroke, diabetes, and depressive/anxiety disorders (108). Dogs are equally burdened by complex diseases. In the United Kingdom, obesity, periodontal disease, osteoarthritis, incontinence, and dermatitis ranked among the most common diagnoses made by first-opinion practices (109).

Like those for monogenic diseases, the odds ratios for particular complex diseases differ substantially from breed to breed, and often disease risk is associated with morphology. For example, the prevalence of extrahepatic portosystemic shunts (ePSS), a congenital condition that occurs when aberrant vasculature bypasses the liver, is highest among toy and small breeds (110). Among breeds with the highest disease risk is the Yorkshire terrier ($n = 483$, odds ratio = 58.7, confidence interval = 42.9–80.2). A case-control GWAS of Yorkshire terriers, as well as other small breeds, failed to generate a significant association for ePSS, though a suggestive association was reported (111).

ePSS are found almost exclusively in toy or small breeds of dogs. Conversely, canine hip dysplasia (CHD) is a disease of medium- to giant-sized dogs (112, 113). In affected dogs, the hip joint becomes painfully arthritic, leading to loss of mobility. Various screening methods, particularly palpitation and radiographs, are routinely deployed to aide selection of breeding stock. Numerous studies have leveraged screening data to identify genetic associations of CHD. Analysis of 1,179 United Kingdom–based Labrador retrievers yielded associations related to the cranial acetabular edge but not other joint laxity phenotypes, including the popular Norberg angle (114).

In a combined approach using Norberg angle phenotypes from 69 breeds ($n = 921$), a genetic association within proximity to the gene *CTBP2* was identified. Interestingly, the effect of the locus appeared relevant to golden and Labrador retrievers but absent from German shepherd dogs (another susceptible breed) (111). Among British-, Finnish-, and Swedish-based German shepherd dogs, CHD was studied within single geography-restricted populations as well as by meta-analysis. Although five associations were identified, including those within proximity to genes *MED13* and *PLEKHA7* (115), none of the associated loci overlapped between single populations (116). Further discordance among genetic association studies is documented (114–122). Discrepancies between association studies have led to the view that CHD is strongly influenced by the genomic architectures of affected populations, though surely other confounding variables, such as CHD screening methodology, age, neuter status, and obesity, could be at play. Of note, Mikkola et al.'s (119) multibreed case-control study of more than 1,600 dogs with normal and extreme hip scores replicated association of markers described previously, bringing the validated number of CHD loci to 10.

We posit that some breeds are in fact defined by multiple deleterious alleles. The most obvious of these are bulldog- and pug-type breeds, whose reduced face length results in a pronounced overbite, a rounded cranium, and hypertelorism. This iconic skull shape is known as brachycephaly ("short-headed") and appears morphologically analogous to certain types of human craniofacial anomalies. Mounting evidence suggests that the genetic underpinnings of brachycephaly may act pleiotropically. Brachycephalic carriers of a *DVL2* deletion have reduced protein expression and bear stunted, corkscrew-shaped tails caused by hemivertebrae. Thoracic hemivertebrae also occur at a lower frequency, suggesting incomplete penetrance of the mutant *DVL2* allele (14, 123). Given their unusual skull morphology, brachycephalic breeds are at risk of developing secondary conditions, including corneal ulceration (124) and the devastating respiratory condition brachycephalic obstructive airway syndrome (125). The severity of canine brachycephaly is influenced by skeletal size, as evidenced by toy breeds such as pugs, Pekingese, and Boston terriers that exhibit the most extreme forms of brachycephaly (32, 126). Among candidate loci suspected of contributing to extreme brachycephaly is a missense variant in *BMP3*, encoding a member of the TGFβ superfamily of signaling ligands (126). To help parse the effects of skull size and face length, quantitative trait mapping was conducted using geometric morphometrics phenotypes and genotypes from 372 dogs of pedigree, mixed, and random-bred backgrounds. This analysis revealed a LINE-1 retrotransposon insertion that segregates with reduction in facial skeleton length. The LINE-1 insertion disrupts *SMOC2* splicing and gene expression (32). Notably, the implication of *SMOC2* in human craniofacial anomalies was subsequently supported by findings using mice (32, 127) and begs further reexamination of human data from patients with reported *SMOC2* variants (128).

Our final example of complex genetics is behavior. Though not a disease, there is an unmet need to address the heritable component of behaviors that disqualify working dogs from service (129–132), prevent pet rehoming (133), or emerge with aging/cognitive decline (8). Like in CHD and skull morphology, success in this area will require rigorous, individualized phenotyping and dense genotyping to capture rare alleles. Given the heterogeneity of dog behaviors, the anticipated small effect sizes of alleles, and the need to control for environmental interactions, study designs require very large cohort sizes and may benefit from the inclusion of random-bred dogs (134).

6. HOW TO PROVE THE CAUSALITY FOR CANDIDATE VARIANTS

Similar to the situation in other species, proof of causality for a candidate variant is often difficult to obtain in dogs. Great care must be taken to correctly interpret the strength of data supporting the causality of a given variant. The extensive LD within dog breeds can challenge the identification of the causal variant because there will be many variants within a critical interval in very strong

Table 2 Complementary lines of evidence for the causality of a variant

Type of data	Examples
Association data	Population data (association, absence, or extremely low frequency of the alternate allele in control cohorts, 1 breed versus multiple breeds) Co-segregation in families
Computational predictions	Type of variant (loss of function, missense, predicted regulatory, other) Evolutionary conservation 3D structural modeling of altered proteins or RNAs Expression data (tissue, cell type, developmental stage)
Functional validation	In vitro (e.g., biochemically demonstrated loss of enzyme activity, cell culture assays) In vivo (e.g., genetically engineered mice, other model organisms) Experimental expression data (RNA, protein, chromatin changes)
Other types of evidence	Demonstration of de novo mutation event Specificity of the phenotype (e.g., neurodegenerative disease versus a specific inborn error of metabolism with documented enzyme deficiency) ≥2 independent variants in the same gene leading to the same phenotype

or even perfect association. Due to the extensive LD, association data in dogs must be analyzed with special caution. The existing scientific literature also must be appropriately cited. It cannot be emphasized enough that association does not necessarily equal causation.

Table 2 provides an overview of experimental and computational approaches that can be accumulated to support or possibly even prove the causality of candidate variants. In most studies, several lines of evidence are combined, which may help to strengthen the argumentation.

6.1. Constraints When Working with Dogs

Because the vast majority of dogs used in genetics studies are pets, one limitation is the difficulty of sample collection for many tissues. If the condition is lethal, owners may be willing to allow sample collection, but delays collecting tissue or lack of a priori knowledge about what to target can still be limiting for full functional evaluation. DNA samples, on the other hand, are readily available in the form of buccal swabs or blood or semen samples. In developed countries, dogs are increasingly medically insured, which can provide epidemiological data on disease that cross veterinary practices. In addition, multiple hospital ownership with electronic medical databases can now be used for genetic studies (135; **https://www.rvc.ac.uk/VetCOMPASS**). These are normally limited to one country but nonetheless provide a means of ascertaining health across a large range of samples. Differences between countries regarding the rules governing canine research may further limit sample collection from privately owned dogs. The beagle is the predominant purpose-bred laboratory dog used for canine research around the world (136).

6.2. Functional Experiments in Canine Cells

Perhaps the most difficult challenge of advancing canine genetics research is demonstrating causality through functional experimentation. The challenge begins by prioritizing which variants to test experimentally. This process is often aided by positional overlap with annotation based on functional genomics data and/or evolutionary constraint (see Section 4.1). Generally, functional experiments aim to contrast tissues or cells, such as before and after disease onset or by comparing the biological activity in the presence versus absence of a putatively causal variant. Such experimentation inevitably requires biological material that is nearly impossible to source from healthy dogs or additional cases for ethical and logistical reasons. Even when sourcing biological material is possible, the decision of which tissues or cell types to source is typically unclear for many

diseases. Although mice and other species can serve as surrogate sources for biomaterial, these experiments are biased toward testing simple coding variants. Moreover, they cannot replicate species-specific genomic landscapes (e.g., chromatin conformation, epigenetics) and the complex interactions a putative causal variant might exert on its local genomic landscape.

Cell lines derived from dogs offer the possibility of testing genetic function conspecifically. However, the availability of primary and immortalized canine cell lines for research purposes is extremely limited, consisting of mostly select adult primary cells and immortalized cancer lines. Embryonic and fetus-derived material are lacking. A search of a popular distributor's canine cell products returned approximately 21 lines (8 tissues sources, 10 neoplastic). In contrast, human cell products exceeded 1,800 cell lines, including nearly 60 primary cell products derived from 20 tissue sources.

To overcome the limitations of sourcing control, diseased, and embryonic-like biomaterial from dogs, we view routine production of induced pluripotent stem cells (iPSCs) from healthy and diseased dogs as an essential, unmet need of the research community. Given their ability to self-renew indefinitely and their intrinsic potential to differentiate into any cell type, iPSCs are often discussed in the context of personalized and regenerative medicine. Production of iPSCs involves reprogramming embryonic or adult primary cell lines. Reprogramming typically involves delivery and overexpression of Yamanaka factors (*OCT4*, *SOX2*, *KLF4*, *MYC*) as well as other genes and specialized culture medium containing cocktails of growth factors and inhibitors, to reverse the epigenetic imprinting of terminal differentiated cells that are transfected (137).

In humans and mice, iPSCs can be produced efficiently from a variety of embryonic and adult cell types. Although numerous studies describe production of canine iPSCs (138–140), production of bona fide iPSCs from dogs remains challenging due to low reprogramming efficiency and the inability of cell lines to retain pluripotency over subsequent passages (141). Dogs are not unique in this regard; research investments to improve production animal reprogramming will inevitably unlock clues that will enable routine iPSC production from dog biomaterial. The therapeutic promise of iPSCs remains elusive even in humans, but iPSCs' ability to cryopreserve rare patient genomes and biology, potential to produce various terminal cell types, and amenability to gene editing will elevate canine genetics and disease research, as they have done for human basic and translational research.

6.3. American College of Medical Genetics/Association for Molecular Pathology Criteria

The American College of Medical Genetics (ACMG) and the Association for Molecular Pathology (AMP) defined a transparent set of criteria that greatly help to classify the strength of supporting evidence for the causality of genetic variants in human patients. The ACMG/AMP consensus recommendations also include specific terminology to differentiate five functional classes of variants: pathogenic, likely pathogenic, uncertain significance, likely benign, and benign (142). These classifications are now increasingly integrated in databases of clinically relevant human variants (**https://www.ncbi.nlm.nih.gov/clinvar/**).

The ACMG/AMP criteria and the human functional classification can be applied with minor adaptations to dogs and other domestic animals (87). This should facilitate better standardization and comparability among future dog disease studies.

6.4. Genetic Proof for Causality

The pedigree structures in purebred animals sometimes provide extremely valuable constellations that allow direct comparison of ancestral (healthy) haplotypes with their derived pathogenic

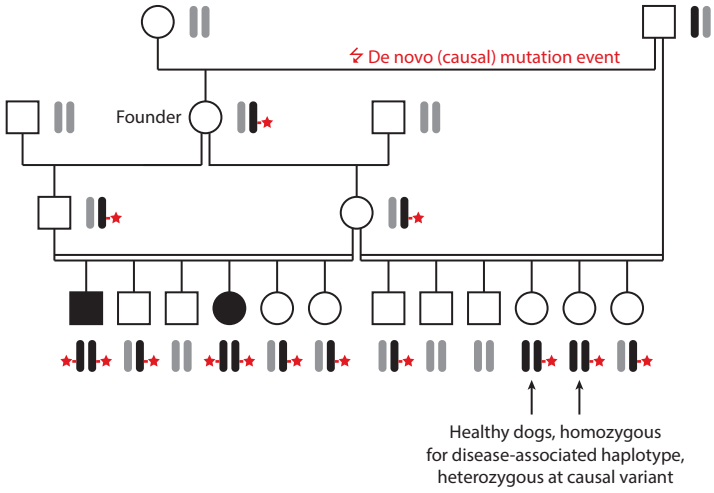

Figure 4

Obtaining genetic proof for the causality of a pathogenic variant. This hypothetical pedigree exemplifies a fully penetrant monogenic autosomal recessive disease in purebred dogs. The causal mutation event happened on the black chromosome/haplotype of the animal at the top right corner of the pedigree. The resulting pathogenic allele was transmitted to the founder animal and is indicated by a red asterisk on this black chromosome. Haplotypes indicated in gray represent wild-type chromosomes. The litter with affected puppies on the bottom left is inbred to the founder animal. The affected animals are homozygous for the pathogenic allele and the disease-associated haplotype. The litter on the bottom right is inbred to the sire of the founder animal. Two offspring received the black haplotype in its ancestral state from the sire (*arrows*). They also received the black haplotype with the pathogenic allele through their maternal line. Thus, the haplotypes in these two animals are identical by descent with the exception of the causal variant. Identifying unaffected dogs with such a genetic makeup provides extremely strong support for the causality of the heterozygous variant.

counterparts. These situations arise when an animal is inbred to an ancestor of the founder. In this case, the disease-associated haplotype could be transmitted IBD from the ancestor to the descendant, resulting in a long run of homozygosity. However, this long run of homozygosity may be interrupted by heterozygous positions that were generated by de novo mutation events anywhere in the inbreeding loop. The identification of just one single de novo mutation event in a previously mapped critical interval can provide definitive proof for the causality of the variant with the heterozygous genotype. This method was first described during the identification of the causal variant for the *callipyge* phenotype in sheep (143).

In the case of recessive diseases, it is therefore worthwhile to pay special attention to unaffected dogs that carry a disease-associated marker haplotype in a homozygous state (**Figure 4**). Such dogs are most likely inbred to an ancestor of the founder and offer the possibility to obtain genetic proof for causality, as has been reported in the case of a *VLDLR* variant in Eurasier dogs with cerebellar hypoplasia (144).

DISCLOSURE STATEMENT

T.L. is listed as an inventor on two patents regarding genetic testing of specific traits in dogs (revenues go to the University of Bern) and holds a grant from the Swiss National Science Foundation entitled "Genetic Analysis of Inherited Skin Diseases in Domestic Animals" (310030_200354, 2021–2025). J.J.S. was previously a paid consultant for Roslin Technologies, Ltd.

ACKNOWLEDGMENTS

T.L. is supported by a grant from the Swiss National Science Foundation (310030_200354). J.J.S. receives institute strategic funding from the institutional strategic program from the Biotechnology and Biosciences Research Council (BBS/E/D/10002070, BBS/E/D/10002071, BBS/E/D/30002275, and BBS/E/D/30002276).

LITERATURE CITED

1. Shearin AL, Ostrander EA. 2010. Leading the way: canine models of genomics and disease. *DMM Dis. Models Mech.* 3(1–2):27–34
2. Khanna C, Lindblad-Toh K, Vail D, London C, Bergman P, et al. 2006. The dog as a cancer model. *Nat. Biotechnol.* 24(9):1065–66
3. Pinho SS, Carvalho S, Cabral J, Reis CA, Gärtner F. 2012. Canine tumors: a spontaneous animal model of human carcinogenesis. *Transl. Res.* 159(3):165–72
4. Gardner HL, Fenger JM, London CA. 2016. Dogs as a model for cancer. *Annu. Rev. Anim. Biosci.* 4:199–222
5. Hakim CH, Kumar SRP, Pérez-López DO, Wasala NB, Zhang D, et al. 2021. Cas9-specific immune responses compromise local and systemic AAV CRISPR therapy in multiple dystrophic canine models. *Nat. Commun.* 12(1):6769
6. Marcó S, Haurigot V, Jaén ML, Ribera A, Sánchez V, et al. 2021. Seven-year follow-up of durability and safety of AAV CNS gene therapy for a lysosomal storage disorder in a large animal. *Mol. Ther. Methods Clin. Dev.* 23:370–89
7. Aguirre GD, Cideciyan AV, Dufour VL, Ripolles-García A, Sudharsan R, et al. 2021. Gene therapy reforms photoreceptor structure and restores vision in NPHP5-associated Leber congenital amaurosis. *Mol. Ther.* 29(8):2456–68
8. Creevy KE, Akey JM, Kaeberlein M, Promislow DEL, Barnett BG, et al. 2022. An open science study of ageing in companion dogs. *Nature* 602(7895):51–57
9. Zangerl B, Goldstein O, Philp AR, Lindauer SJP, Pearce-Kelling SE, et al. 2006. Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics* 88(5):551–63
10. Drögemüller M, Jagannathan V, Becker D, Drögemüller C, Schelling C, et al. 2014. A mutation in the *FAM83G* gene in dogs with hereditary footpad hyperkeratosis (HFH). *PLOS Genet.* 10(5):e1004370
11. Drögemüller C, Becker D, Brunner A, Haase B, Kircher P, et al. 2009. A missense mutation in the *SERPINH1* gene in Dachshunds with osteogenesis imperfecta. *PLOS Genet.* 5(7):e1000579
12. Grall A, Guaguère E, Planchais S, Grond S, Bourrat E, et al. 2012. *PNPLA1* mutations cause autosomal recessive congenital ichthyosis in golden retriever dogs and humans. *Nat. Genet.* 44(2):140–47
13. Vernau KM, Runstadler JA, Brown EA, Cameron JM, Huson HJ, et al. 2013. Genome-wide association analysis identifies a mutation in the thiamine transporter 2 (*SLC19A3*) gene associated with Alaskan Husky encephalopathy. *PLOS ONE* 8(3):e57195
14. Mansour TA, Lucot K, Konopelski SE, Dickinson PJ, Sturges BK, et al. 2018. 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(12):e1007850
15. Bauer A, Waluk DP, Galichet A, Timm K, Jagannathan V, et al. 2017. A *de novo* variant in the *ASPRV1* gene in a dog with ichthyosis. *PLOS Genet.* 13(3):e1006651
16. Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. *Nat. Rev. Genet.* 9(9):713–25
17. Dreger DL, Rimbault M, Davis BW, Bhatnagar A, Parker HG, Ostrander EA. 2016. Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Dis. Model Mech.* 9(12):1445–60
18. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, et al. 2015. A global reference for human genetic variation. *Nature* 526(7571):68–74

19. Ali MB, Evans JM, Parker HG, Kim J, Pearce-Kelling S, et al. 2020. Genetic analysis of the modern Australian labradoodle dog breed reveals an excess of the poodle genome. *PLOS Genet*. 16(9):e1008956
20. Bannasch D, Famula T, Donner J, Anderson H, Honkanen L, et al. 2021. The effect of inbreeding, body size and morphology on health in dog breeds. *Canine Med. Genet.* 8:12
21. Lepori V, Mühlhause F, Sewell AC, Jagannathan V, Janzen N, et al. 2018. A nonsense variant in the *ACADVL* gene in German Hunting Terriers with exercise induced metabolic myopathy. *G3* 8(5):1545–54
22. Has C, Bauer JW, Bodemer C, Bolling MC, Bruckner-Tuderman L, et al. 2020. Consensus reclassification of inherited epidermolysis bullosa and other disorders with skin fragility. *Br. J. Dermatol.* 183(4):614–27
23. Garcia TM, Kiener S, Jagannathan V, Russell DS, Leeb T. 2020. A *COL7A1* variant in a litter of neonatal Basset Hounds with dystrophic epidermolysis bullosa. *Genes* 11(12):1458
24. Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, et al. 2007. A β-defensin mutation causes black coat color in domestic dogs. *Science* 318(5855):1418–23
25. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, et al. 2009. An expressed *Fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325(5943):995–98
26. Brown EA, Dickinson PJ, Mansour T, Sturges BK, Aguilar M, et al. 2017. *FGF4* retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. *PNAS* 114(43):11476–81
27. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, et al. 2007. Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nat. Genet.* 39(11):1321–28
28. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, et al. 2021. Genome-wide association studies. *Nat. Rev. Methods Primers* 1:59
29. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Pielberg GR, et al. 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLOS Genet.* 7(10):e1002316
30. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, et al. 2010. A simple genetic architecture underlies morphological variation in dogs. *PLOS Biol.* 8(8):49–50
31. Bannasch D, Young A, Myers J, Truvé K, Dickinson P, et al. 2010. Localization of canine brachycephaly using an across breed mapping approach. *PLOS ONE* 5(3):e9632
32. Marchant TW, Johnson EJ, McTeir L, Johnson CI, Gow A, et al. 2017. Canine brachycephaly is associated with a retrotransposon-mediated missplicing of *SMOC2*. *Curr. Biol.* 27(11):1573–84.e6
33. Anderson H, Honkanen L, Ruotanen P, Mathlin J, Donner J. 2020. Comprehensive genetic testing combined with citizen science reveals a recently characterized ancient *MC1R* mutation associated with partial recessive red phenotypes in dog. *Canine Med. Genet.* 7:16
34. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, et al. 2009. Coat variation in the domestic dog is governed by variants in three genes. *Science* 326(5949):150–53
35. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, et al. 2019. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* 10:1489
36. Kyöstilä K, Syrjä P, Jagannathan V, Chandrasekar G, Jokinen TS, et al. 2015. A missense change in the *ATG4D* gene links aberrant autophagy to a neurodegenerative vacuolar storage disease. *PLOS Genet.* 11(4):e1005169
37. Lander ES, Botstein D. 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236(4808):1567–70
38. Marchant TW, Dietschi E, Rytz U, Schawalder P, Jagannathan V, et al. 2019. An *ADAMTS3* missense variant is associated with Norwich Terrier upper airway syndrome. *PLOS Genet.* 15(5):e1008102
39. Drögemüller C, Karlsson EK, Hytönen MK, Perloski M, Dolf G, et al. 2008. A mutation in hairless dogs implicates *FOXI3* in ectodermal development. *Science* 321(5895):1462
40. Backel KA, Kiener S, Jagannathan V, Casal ML, Leeb T, Mauldin EA. 2020. A *DSG1* frameshift variant in a Rottweiler dog with footpad hyperkeratosis. *Genes* 11(4):469
41. Jagannathan V, Drögemüller C, Leeb T, Aguirre G, André C, et al. 2019. A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Anim. Genet.* 50(6):695–704

42. Besenbacher S, Liu S, Izarzugaza JMG, Grove J, Belling K, et al. 2015. Novel variation and *de novo* mutation rates in population-wide *de novo* assembled Danish trios. *Nat. Commun.* 6:5969
43. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–75
44. Harland C, Charlier C, Karim L, Cambisano N, Deckers M, et al. 2017. Frequency of mosaicism points towards mutation-prone early cleavage cell divisions in cattle. bioRxiv 079863. **https://doi.org/10.1101/079863**
45. Wang J, Kong L, Gao G, Luo J. 2013. A brief introduction to web-based genome browsers. *Brief. Bioinform.* 14(2):131–43
46. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18(2):170–75
47. Lindblad-Toh K, Wade Claire M, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–19
48. Jagannathan V, Hitte C, Kidd JM, Masterson P, Murphy TD, et al. 2021. Dog10K_Boxer_Tasha_1.0: a long-read assembly of the dog reference genome. *Genes* 12(6):847
49. Edwards RJ, Field MA, Ferguson JM, Dudchenko O, Keilwagen J, et al. 2021. Chromosome-length genome assembly and structural variations of the primal Basenji dog (*Canis lupus familiaris*) genome. *BMC Genom.* 22:188
50. Field MA, Rosen BD, Dudchenko O, Chan EKF, Minoche AE, et al. 2020. 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(4):giaa027
51. Wang C, Wallerman O, Arendt M-L, Sundström E, Karlsson Å, et al. 2021. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Commun. Biol.* 4:185
52. Halo JV, Pendleton AL, Shen F, Doucet AJ, Derrien T, et al. 2021. Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes. *PNAS* 118(11):e2016274118
53. Sinding M-HS, Gopalakrishnan S, Raundrup K, Dalén L, Threlfall J, Gilbert T. 2021. The genome sequence of the grey wolf, *Canis lupus* Linnaeus 1758. *Wellcome Open Res.* 6:310
54. Field MA, Yadav S, Dudchenko O, Esvaran M, Rosen BD, et al. 2022. The Australian dingo is an early offshoot of modern breed dogs. *Sci. Adv.* 8(16):eabm5944
55. Natl. Cent. Biotechnol. Inf. (NCBI). 2020. *Canis lupus familiaris annotation release 106*. Annot. Rep., NCBI, Bethesda, MD. **https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Canis_lupus_familiaris/106/accessed06-May-2022**
56. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, et al. 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45(8):e57
57. Yamazaki J, Matsumoto Y, Jelinek J, Ishizaki T, Maeda S, et al. 2021. DNA methylation landscape of 16 canine somatic tissues by methylation-sensitive restriction enzyme-based next generation sequencing. *Sci. Rep.* 11:10005
58. Megquier K, Genereux DP, Hekman J, Swofford R, Turner-Maier J, et al. 2019. BarkBase: epigenomic annotation of canine genomes. *Genes* 10:433
59. Ballester B, Medina-Rivera A, Schmidt D, González-Porta M, Carlucci M, et al. 2014. Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *eLife* 3:e02626
60. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328:1036–40
61. Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, et al. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* 587:240–45
62. Zhang W, Eory L, Salavati M, Clark E, Smith J, et al. 2022. Applications of single-molecule real-time isoform sequencing (Iso-Seq) for unravellign complexity of dog transcriptomes. In *Proceedings of the 12th World Congress of Genetics Applied to Livestock Production*. Edinburgh: Roslin Inst. In press
63. Bannasch DL, Kaelin CB, Letko A, Loechel R, Hug P, et al. 2021. Dog colour patterns explained by modular promoters of ancient canid origin. *Nat. Ecol. Evol.* 5(10):1415–23

64. Karyadi DM, Karlins E, Decker B, vonHoldt BM, Carpintero-Ramirez G, et al. 2013. A copy number variant at the *KITLG* locus likely confers risk for canine squamous cell carcinoma of the digit. *PLOS Genet.* 9(3):e1003409
65. Park K, Kang J, Subedi KP, Ha JH, Park C. 2008. Canine polydactyl mutations with heterogeneous origin in the conserved intronic sequence of *LMBR1*. *Genetics* 179(4):2163–72
66. Olsson M, Meadows JRS, Truvé K, Pielberg GR, Puppo F, et al. 2011. A novel unstable duplication upstream of *HAS2* predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. *PLOS Genet.* 7(3):e1001332
67. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, et al. 2022. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med.* 14(1):73
68. Church DM, Schneider VA, Steinberg KM, Schatz MC, Quinlan AR, et al. 2015. Extending reference assembly models. *Genome Biol.* 16:13
69. Sherman RM, Forman J, Antonescu V, Puiu D, Daya M, et al. 2019. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* 51(1):30–35
70. Talenti A, Powell J, Hemmink JD, Cook EAJ, Wragg D, et al. 2022. A cattle graph genome incorporating global breed diversity. *Nat. Commun.* 13:910
71. Crysnanto D, Leonard AS, Fang ZH, Pausch H. 2021. Novel functional sequences uncovered through a bovine multiassembly graph. *PNAS* 118(20):e2101056118
72. Crysnanto D, Pausch H. 2020. Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery. *Genome Biol.* 21:184
73. Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12(5):363–76
74. Ho SS, Urban AE, Mills RE. 2020. Structural variation in the sequencing era. *Nat. Rev. Genet.* 21(3):171–89
75. Wang W, Kirkness EF. 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res.* 15(12):1798–808
76. Linek M, Doelle M, Leeb T, Bauer A, Leuthard F, et al. 2020. *ATP2A2* SINE insertion in an Irish Terrier with Darier disease and associated infundibular cyst formation. *Genes* 11(5):481
77. Mauri N, Kleiter M, Dietschi E, Leschnik M, Högler S, et al. 2017. A SINE insertion in *ATP1B2* in Belgian Shepherd dogs affected by spongy degeneration with cerebellar ataxia (SDCA2). *G3* 7(8):2729–37
78. Kehl A, Haaland AH, Langbein-Detsch I, Mueller E. 2021. A sine insertion in *F8* gene leads to severe form of hemophilia A in a family of Rhodesian Ridgebacks. *Genes* 12(2):134
79. Pelé M, Tiret L, Kessler JL, Blot S, Panthier JJ. 2005. SINE exonic insertion in the *PTPLA* gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. *Hum. Mol. Genet.* 14(11):1417–27
80. Wiedmer M, Oevermann A, Borer-Germann SE, Gorgas D, Shelton GD, et al. 2016. A *RAB3GAP1* SINE insertion in Alaskan Huskies with polyneuropathy, ocular abnormalities, and neuronal vacuolation (POANV) resembling human Warburg micro syndrome 1 (WARBM1). *G3* 6(2):255–62
81. Downs LM, Mellersh CS. 2014. An intronic SINE insertion in *FAM161A* that causes exon-skipping is associated with progressive retinal atrophy in Tibetan Spaniels and Tibetan Terriers. *PLOS ONE* 9(4):e93990
82. Clark LA, Wahl JM, Rees CA, Murphy KE. 2006. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *PNAS* 103(5):1376–81
83. Wolf ZT, Leslie EJ, Arzi B, Jayashankar K, Karmi N, et al. 2014. A LINE-1 insertion in DLX6 is responsible for cleft palate and mandibular abnormalities in a canine model of Pierre Robin sequence. *PLOS Genet.* 10(4):e1004257
84. Credille KM, Minor JS, Barnhart KF, Lee E, Cox ML, et al. 2009. Transglutaminase 1-deficient recessive lamellar ichthyosis associated with a LINE-1 insertion in Jack Russell terrier dogs. *Br. J. Dermatol.* 161(2):265–72
85. Batcher K, Dickinson P, Maciejczyk K, Brzeski K, Rasouliha SH, et al. 2020. Multiple *FGF4* retrocopies recently derived within canids. *Genes* 11(8):839

86. Leuthard F, Lehner G, Jagannathan V, Leeb T, Welle M. 2019. A missense variant in the *NSDHL* gene in a Chihuahua with a congenital cornification disorder resembling inflammatory linear verrucous epidermal nevi. *Anim. Genet.* 50(6):768–71
87. Christen M, Austel M, Banovic F, Jagannathan V, Leeb T. 2020. NSDHL frameshift deletion in a mixed breed dog with progressive epidermal nevi. *Genes* 11(11):1297
88. Megquier K, Turner-Maier J, Swofford R, Kim JH, Sarver AL, et al. 2019. Comparative genomics reveals shared mutational landscape in canine hemangiosarcoma and human angiosarcoma. *Mol. Cancer Res.* 17(12):2410–21
89. Dillard KJ, Ochs M, Niskanen JE, Arumilli M, Donner J, et al. 2020. Recessive missense *LAMP3* variant associated with defect in lamellar body biogenesis and fatal neonatal interstitial lung disease in dogs. *PLOS Genet.* 16(3):e1008651
90. Koboldt DC. 2020. Best practices for variant calling in clinical sequencing. *Genome Med.* 12:91
91. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29(1):24–26
92. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant review with the integrative genomics viewer. *Cancer Res.* 77(21):e31–34
93. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581(7809):434–43
94. Giani AM, Gallo GR, Gianfranceschi L, Formenti G. 2019. Long walk to genomics: history and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* 18:9–19
95. Nicholas FW. 1998. Genetic databases: online catalogues of inherited disorders. *OIE Rev. Sci. Tech.* 17(1):346–50
96. Nicholas FW. 2021. Online Mendelian Inheritance in Animals (OMIA): a record of advances in animal genetics, freely available on the Internet for 25 years. *Anim. Genet.* 52(1):3–9
97. Shelton GD, Podell M, Poncelet L, Schatzberg S, Patterson E, et al. 2003. Inherited polyneuropathy in Leonberger dogs: A mixed or intermediate form of Charcot-Marie-Tooth disease? *Muscle Nerve* 27(4):471–77
98. Hultin Jäderlund K, Baranowska Körberg I, Nødtvedt A. 2011. Inherited polyneuropathy in Leonberger dogs. *J. Vet. Intern. Med.* 25(5):997–1002
99. Becker D, Minor KM, Letko A, Ekenstedt KJ, Jagannathan V, et al. 2017. A *GJA9* frameshift variant is associated with polyneuropathy in Leonberger dogs. *BMC Genom.* 18:662
100. Letko A, Minor KM, Friedenberg SG, Shelton GD, Salvador JP, et al. 2020. A *CNTNAP1* missense variant is associated with canine laryngeal paralysis and polyneuropathy. *Genes* 11(12):1426
101. Ekenstedt KJ, Becker D, Minor KM, Shelton GD, Patterson EE, et al. 2014. An *ARHGEF10* deletion is highly associated with a juvenile-onset inherited polyneuropathy in Leonberger and Saint Bernard dogs. *PLOS Genet.* 10(10):e1004635
102. Mauri N, Kleiter M, Leschnik M, Högler S, Dietschi E, et al. 2017. A missense variant in *KCNJ10* in Belgian Shepherd dogs affected by spongy degeneration with cerebellar ataxia (SDCA1). *G3* 7(2):663–69
103. Gurtner C, Hug P, Kleiter M, Köhler K, Dietschi E, et al. 2020. *YARS2* missense variant in Belgian Shepherd dogs with cardiomyopathy and juvenile mortality. *Genes* 11(3):313
104. Christen M, Högler S, Kleiter M, Leschnik M, Weber C, et al. 2021. Deletion of the *SELENOP* gene leads to CNS atrophy with cerebellar ataxia in dogs. *PLOS Genet.* 17(8):e1009716
105. Parshall CJ, Wyman M, Nitroy S, Acland GM, Aguirre GD. 1991. Photoreceptor dysplasia: an inherited progressive retinal atrophy of Miniature Schnauzer dogs. *Prog. Vet. Comp. Ophthalmol.* 1(3):187–203
106. Murgiano L, Becker D, Torjman D, Niggel JK, Milano A, et al. 2019. Complex structural *PPT1* variant associated with non-syndromic canine retinal degeneration. *G3* 9(2):425–37
107. Aguirre GD, Lohi H, Kaukonen M, Murgiano L. 2020. Formal commentary. *PLOS Genet.* 16(11):e1009059
108. World Health Organ. 2022. *Global health estimates: leading causes of DALYs*. Doc., World Health Organ., Geneva, accessed May 6, 2022. **https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/global-health-estimates-leading-causes-of-dalys**
109. O'Neill DG, James H, Brodbelt DC, Church DB, Pegram C. 2021. Prevalence of commonly diagnosed disorders in UK dogs under primary veterinary care: results and applications. *BMC Vet. Res.* 17:69

110. Tobias KM, Rohrbach BW. 2003. Association of breed with the diagnosis of congenital portosystemic shunts in dogs: 2,400 cases (1980–2002). *J. Am. Vet. Med. Assoc.* 223(11):1636–39
111. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, et al. 2016. Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* 7:10460
112. Coopman F, Verhoeven G, Saunders J, Duchateau L, van Bree H. 2008. Prevalence of hip dysplasia, elbow dysplasia and humeral head osteochondrosis in dog breeds in Belgium. *Vet. Rec.* 163(22):654–58
113. Schnelle GB. 1937. Bilateral congenital subluxation of the coxo-femoral joints in a dog. *Univ. Pa. Bull. Sch. Vet. Med. Vet. Ext. Q.* 37:15–16
114. Sánchez-Molano E, Woolliams JA, Pong-Wong R, Clements DN, Blott SC, Wiener P. 2014. Quantitative trait loci mapping for canine hip dysplasia and its related traits in UK Labrador Retrievers. *BMC Genom.* 15:833
115. Zhou Z, Sheng X, Zhang Z, Zhao K, Zhu L, et al. 2010. Differential genetic regulation of canine hip dysplasia and osteoarthritis. *PLOS ONE* 5(10):e13219
116. Wang S, Strandberg E, Arvelius P, Clements DN, Wiener P, Friedrich J. 2021. Genome-wide association studies for canine hip dysplasia in single and multiple populations—implications and potential novel risk loci. *BMC Genom.* 22:636
117. Pfahler S, Distl O. 2012. Identification of quantitative trait loci (QTL) for canine hip dysplasia and canine elbow dysplasia in Bernese mountain dogs. *PLOS ONE* 7(11):e49782
118. Fels L, Marschall Y, Philipp U, Distl O. 2014. Multiple loci associated with canine hip dysplasia (CHD) in German shepherd dogs. *Mamm. Genome* 25(5–6):262–69
119. Mikkola L, Kyöstilä K, Donner J, Lappalainen AK, Hytönen MK, et al. 2021. An across-breed validation study of 46 genetic markers in canine hip dysplasia. *BMC Genom.* 22:68
120. Friedenberg SG, Zhu L, Zhang Z, van den Foels WB, Schweitzer PA, et al. 2011. Evaluation of a fibrillin 2 gene haplotype associated with hip dysplasia and incipient osteoarthritis in dogs. *Am. J. Vet. Res.* 72(4):530–40
121. Fels L, Distl O. 2014. Identification and validation of quantitative trait loci (QTL) for canine hip dysplasia (CHD) in German Shepherd dogs. *PLOS ONE* 9(5):e96618
122. Lavrijsen ICM, Leegwater PAJ, Martin AJ, Harris SJ, Tryfonidou MA, et al. 2014. Genome wide analysis indicates genes for basement membrane and cartilage matrix proteins as candidates for hip dysplasia in Labrador Retrievers. *PLOS ONE* 9(1):e87735
123. Niskanen JE, Reunanen V, Salonen M, Bannasch D, Lappalainen AK, et al. 2021. Canine *DVL2* variant contributes to brachycephalic phenotype and caudal vertebral anomalies. *Hum. Genet.* 140(11):1535–45
124. O'Neill DG, Lee MM, Brodbelt DC, Church DB, Sanchez RF. 2017. Corneal ulcerative disease in dogs under primary veterinary care in England: epidemiology and clinical management. *Canine Genet. Epidemiol.* 4:5
125. Harvey CE. 1989. Inherited and congenital airway conditions. *J. Small Anim. Pract.* 30(3):184–87
126. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, et al. 2012. Variation of *BMP3* contributes to dog breed skull diversity. *PLOS Genet.* 8(8):e1002849
127. Takahata Y, Hagino H, Kimura A, Urushizaki M, Kobayashi S, et al. 2021. Smoc1 and Smoc2 regulate bone formation as downstream molecules of Runx2. *Commun Biol.* 4:1199
128. Alfawaz S, Fong F, Plagnol V, Wong FSL, Fearne J, Kelsell DP. 2013. Recessive oligodontia linked to a homozygous loss-of-function mutation in the SMOC2 gene. *Arch. Oral Biol.* 58(5):462–66
129. Lofgren SE, Wiener P, Blott SC, Sanchez-Molano E, Woolliams JA, et al. 2014. Management and personality in Labrador Retriever dogs. *Appl. Anim. Behav. Sci.* 156:44–53
130. Chen FL, Zimmermann M, Hekman JP, Lord KA, Logan B, et al. 2021. Advancing genetic selection and behavioral genomics of working dogs through collaborative science. *Front. Vet. Sci.* 8:662429
131. Friedrich J, Strandberg E, Arvelius P, Sánchez-Molano E, Pong-Wong R, et al. 2019. Genetic dissection of complex behaviour traits in German Shepherd dogs. *Heredity* 123(6):746–58
132. Guide Dogs. 2022. *Born to guide*. **https://www.guidedogs.org.uk/born-to-guide**
133. Murray JK, Kinsman RH, Lord MS, da Costa REP, Woodward JL, et al. 2021. "Generation Pup"—protocol for a longitudinal study of dog behaviour and health. *BMC Vet. Res.* 17:1
134. Morrill K, Hekman J, Li X, McClure J, Logan B, et al. 2022. Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science* 376(6592):eabk0639

135. Salt C, Morris PJ, German AJ, Wilson D, Lund EM, et al. 2017. Growth standard charts for monitoring bodyweight in dogs of different sizes. *PLOS ONE* 12(9):e0182064
136. Hubrecht R, Kirkwood J, eds. 2010. *The UFAW Handbook on the Care and Management of Laboratory and Other Research Animals: Eighth Edition*. Herts, UK: Univ. Fed. Anim. Welf.
137. Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–76
138. Kimura K, Tsukamoto M, Tanaka M, Kuwamura M, Ohtaka M, et al. 2021. Efficient reprogramming of canine peripheral blood mononuclear cells into induced pluripotent stem cells. *Stem Cells Dev.* 30(2):79–90
139. Yoshimatsu S, Edamura K, Yoshii Y, Iguchi A, Kondo H, et al. 2021. Non-viral derivation of a transgene-free induced pluripotent stem cell line from a male beagle dog. *Stem Cell Res.* 53:102375
140. Yoshimatsu S, Nakajima M, Iguchi A, Sanosaka T, Sato T, et al. 2021. Non-viral induction of transgene-free iPSCs from somatic fibroblasts of multiple mammalian species. *Stem Cell Rep.* 16(4):754–70
141. Tobias IC, Kao MMC, Parmentier T, Hunter H, LaMarre J, Betts DH. 2020. Targeted expression profiling reveals distinct stages of early canine fibroblast reprogramming are regulated by 2-oxoglutarate hydroxylases. *Stem Cell Res Ther.* 11(1):102375
142. Richards S, Aziz N, Bale S, Bick D, Das S, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17(5):405–24
143. Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, et al. 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* 12(10):1496–506
144. Gerber M, Fischer A, Jagannathan V, Drögemüller M, Drögemüller C, et al. 2015. A deletion in the *VLDLR* gene in Eurasier dogs with cerebellar hypoplasia resembling a Dandy-Walker-like malformation (DWLM). *PLOS ONE* 10(2):e0108917
145. Batcher K, Varney S, York D, Blacksmith M, Kidd JM, et al. 2022. Recent, full-length gene retrocopies are common in canids. *Genome Res.* 32(8):1602–11

Annual Review of
Animal Biosciences

Volume 11, 2023

Contents

Errata

An online log of corrections to *Annual Review of Animal Biosciences* articles may be found at http://www.annualreviews.org/errata/animal