



A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves

V. Jagannathan , C. Drögemüller , T. Leeb  and Dog Biomedical Variant Database Consortium (DBVDC)[†]

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

Summary

The domestic dog serves as an excellent model to investigate the genetic basis of disease. More than 400 heritable traits analogous to human diseases have been described in dogs. To further canine medical genetics research, we established the Dog Biomedical Variant Database Consortium (DBVDC) and present a comprehensive list of functionally annotated genome variants that were identified with whole genome sequencing of 582 dogs from 126 breeds and eight wolves. The genomes used in the study have a minimum coverage of 10× and an average coverage of ~24×. In total, we identified 23 133 692 single-nucleotide variants (SNVs) and 10 048 038 short indels, including 93% undescribed variants. On average, each individual dog genome carried ~4.1 million single-nucleotide and ~1.4 million short-indel variants with respect to the reference genome assembly. About 2% of the variants were located in coding regions of annotated genes and loci. Variant effect classification showed 247 141 SNVs and 99 562 short indels having moderate or high impact on 11 267 protein-coding genes. On average, each genome contained heterozygous loss-of-function variants in 30 potentially embryonic lethal genes and 97 genes associated with developmental disorders. More than 50 inherited disorders and traits have been unravelled using the DBVDC variant catalogue, enabling genetic testing for breeding and diagnostics. This resource of annotated variants and their corresponding genotype frequencies constitutes a highly useful tool for the identification of potential variants causative for rare inherited disorders in dogs.

Keywords animal model, bioinformatics, *Canis lupus familiaris*, functional annotation, genetic diversity, Mendelian, precision medicine, rare disease, variant database, whole genome sequencing

Introduction

The domestic dog (*Canis lupus familiaris* or *Canis familiaris*) was the first domesticated animal species. In the last few

hundred years, more than 400 genetically isolated breeds of dogs have been created through human selection and breeding in closed populations. Dogs are the species with the greatest intra-species phenotypic diversity among vertebrates. Body size as well as limb and skull proportions differ noticeably among breeds, with Chihuahuas at one end of the spectrum measuring 20 cm in height and 2 kg in weight compared with Great Danes on the other end exceeding 76 cm in height and 90 kg weight. In addition to morphology, behavioural variation exists across breeds, with many breeds being highly specialized for single tasks such as herding, hunting, retrieving, guarding, detecting scent or providing companionship. The genetic bases for phenotypic variation among breeds have been reported, including body size (Sutter *et al.* 2007; Boyko *et al.* 2010; Vaysse *et al.* 2011; Rimbault & Ostrander 2012; Hayward *et al.* 2016), skull shape (Schoenebeck *et al.* 2012), short legs (Parker *et al.* 2009; Brown *et al.* 2017), hair

Address for correspondence

V. Jagannathan, Institute of Genetics, Vetsuisse Faculty, University of Bern, Bremgartenstrasse 109a, Bern 3001, Switzerland.
E-mail: vidhya.jagannathan@vetsuisse.unibe.ch

Dog Biomedical Variant Database Consortium (DBVDC): Gustavo Aguirre, Catherine André, Danika Bannasch, Doreen Becker, Brian Davis, Cord Drögemüller, Kari Ekenstedt, Kiterie Faller, Oliver Forman, Steve Friedenber, Eva Furrow, Urs Giger, Christophe Hitte, Marjo Hytönen, Vidhya Jagannathan, Tosso Leeb, Hannes Lohi, Cathryn Mellersh, James R. Mickelson, Leonardo Murgiano, Anita Oberbauer, Sheila Schmutz, Jeffrey Schoenebeck, Kim Summers, Frank van Steenbeek, Claire Wade.

[†]The full affiliations are listed in S4Appendix S1.

Accepted for publication 21 June 2019

morphology (Drögemüller *et al.* 2008; Cadieu *et al.* 2009; Hytönen & Lohi 2019) and aggression or fear (Zapata *et al.* 2016; Sarviaho *et al.* 2019). These studies provided important insights for understanding the genetic regulation of analogous traits in humans.

Many genetic diseases are analogous to those in humans, for example susceptibility to certain types of cancer and a large number of Mendelian diseases. The list of known diseases in dogs is greater than that in any other domestic animal species (Ostrander *et al.* 2017). The Online Mendelian Inheritance in Animal (OMIA) database lists 426 potential dog models for human diseases (<https://omia.org>; accessed 14 May 2019). Canine heritable diseases not only show the clinical and pathological features of their human counterparts but also harbour underlying causative genetic variants in the same genes as in humans (Hytönen & Lohi 2016). In addition, the identification of disease-causing variants in previously uncharacterized genes provides novel candidates for rare human diseases. Examples include rare diseases like retinitis pigmentosa, osteogenesis imperfecta, congenital ichthyosis, nasal parakeratosis, footpad hyperkeratosis, neurodegenerative vacuolar storage disease, myoclonic epilepsy and lethal acrodermatitis (Zangerl *et al.* 2006; Drögemüller *et al.* 2009, 2014a; Merveille *et al.* 2011; Grall *et al.* 2012; Jagannathan *et al.* 2013; Kyöstilä *et al.* 2015; Hytönen *et al.* 2016; Wielaender *et al.* 2017; Bauer *et al.* 2018a).

Given the breeding history of dogs, with extreme genetic bottlenecks during breed formation and the subsequent maintenance of closed populations starting from a small number of founding animals, it is not surprising that individual breeds also exhibit breed-specific patterns of disease susceptibility. In the last 15 years, geneticists have mapped hundreds of Mendelian trait loci using smaller sample sizes than are required in human disease association studies. Instances include *MITE*, which is involved in white spotting, found with nine cases and 10 controls (Karlsson *et al.* 2007) and the epilepsy gene, *LG12*, that was mapped with 11 cases and 11 controls (Seppälä *et al.* 2011). For simple Mendelian traits, only 15 000 informative single-nucleotide variants (SNVs) spanning the genome were reported as being adequate for successful genome-wide association studies, whereas for the same study in humans, 300 000 SNVs might be required (Lindblad-Toh *et al.* 2005). This is due mainly to long-ranging intra-breed linkage LD, which extends to roughly 1 Mb in dogs compared with a few kb in humans (Sutter *et al.* 2004; Lindblad-Toh *et al.* 2005). Because of the long-ranging LD, the identification of association signals is easier in dogs compared with humans. However, fine mapping and the identification of the underlying causal variant(s) for a given association signal are often quite challenging in dogs. Within dog breeds, typically many variants close to the causal variant are in strong or even perfect LD and hence it may be difficult to discriminate among them.

With whole-genome sequencing (WGS), it has become feasible to at least comprehensively access the existing

genome variation. Identification of causal variants for heritable traits in a WGS approach involves the mapping of case and control sequence reads to a reference genome sequence (Bourneuf *et al.* 2017). The currently used CanFam3.1 reference genome sequence is derived from a female Boxer with contig and scaffold N50 sizes of 180 kb and 45 Mb respectively (Lindblad-Toh *et al.* 2005). Any given dog genome typically has several million variants compared with the reference genome. Therefore, a meticulous approach to reliable variant calling and hierarchical filtering strategies is required to reduce the millions of variants per sample to a manageable list of candidate genes and causative variants. Many heritable traits in dogs have been successfully solved using such a WGS approach for both diseases and morphological features; examples include a *de novo* variant in *ASPRV1* for ichthyosis and variants in *GJA9* for polyneuropathy, *RBP4* in congenital eye disease and *DVL2* in dogs with screw tails (Bauer *et al.* 2017a; Becker *et al.* 2017; Kaukonen *et al.* 2018; Mansour *et al.* 2018). More recently, Plassais *et al.* (2019) identified candidate causative variants for several phenotypes.

The cost of a mammalian WGS including data analysis and data storage is still above US\$1000. Hence, WGS experiments in individual research laboratories directed at identifying the causal variants for Mendelian traits are carried out with small numbers of cases and controls. The efficiency of such WGS approaches can be drastically improved by including variants from dogs sequenced for other projects from around the world. Comprehensive variant databases containing accurately called and genotyped variants from hundreds or thousands of dog genomes enable a straightforward filtering option based on allele and/or genotype frequencies within or across selected cohorts. This is a powerful approach to identifying and excluding common variants that are unlikely to cause rare inherited diseases. A comprehensive set of variants together with their allele and genotype frequencies can thus greatly help to distinguish functionally relevant from neutral variants. The use of such filtering strategies has become standard practice in human medical genetics, and for example, the gnomAD database (<https://gnomad.broadinstitute.org/>) has become indispensable for the identification of candidate pathogenic variants in human genetics (Lek *et al.* 2016). A similar approach is used within the cattle community and the 1000 Bull Genome project (Hayes & Daetwyler 2019).

Currently, the canine variations reported in the National Center for Biotechnology Information (NCBI) genetic variation database hosted at the European Variation Archive contain ~2.9 million SNVs. We submitted a comprehensive list of variants from 238 purpose-bred dogs, containing 18 639 483 SNVs and 9 293 851 short indels, to the European Variation Archive in 2017 (PRJEB24066), but this dataset was not processed until May 2019. Comprehensive variation data from 132 canids and 90 village dogs were reported previously (Marsden *et al.* 2016; Taylor *et al.*

2016). The iDog database released a variation list from 127 individual dogs and grey wolves (Bai *et al.* 2015; Tang *et al.* 2019). Lastly, Plassais *et al.* (2019) reported the analysis of 722 canine whole-genome sequences to discover variants associated with 16 different phenotypes, including body weight variation observed across modern dog breeds but absent in wild canids.

We report here a comprehensive list of high-quality variants from 590 genome sequences comprising 582 dogs from 126 different breeds and eight wolves (Table S1). We included 178 sequences from public databases that had been produced by other groups, whereas the remaining 412 whole-genome sequences were produced and submitted to public databases by members of the Dog Biomedical Variant Database Consortium (DBVDC) from more than 20 institutions around the world. The DBVDC was established in 2013 to aggregate WGS data from a variety of projects examining dog biomedical and morphological phenotypes. The DBVDC proceeds in incremental 'runs', with a run taking place approximately every 3–6 months.

Materials and methods

Datasets

Whole genome sequences of 582 dogs and eight wolves available at the European Nucleotide Archive and the Short Read Archive were used for the analysis. See Table S1 for details of the project and sample accessions. The CanFam3.1 reference genome was downloaded from Ensembl (http://www.ensembl.org/Canis_familiaris/Info/Index) and was used for alignment of reads.

Samples, read filtering and alignment

A total of 590 individual genomes were sequenced with Illumina sequencing technology NextSeq, HiSeq or NovaSeq. The 590 individuals consisted of 467 males and 123 females. The samples were selected based on their availability in public databases and with a minimum coverage of 10×. Raw sequencing reads were filtered for adaptors and trimmed based on the quality score using FASTP (Chen *et al.* 2018). Quality filtered reads in FASTQ format were then aligned with the reference genome assembly CanFam3.1 using BWA (version 0.7.13; Li & Durbin 2010). GATK CALLABLELOCI was used to collect statistics on callable, uncallable, poorly mapped and other parts of the genome. The following arguments were used --minBaseQuality 20, --minDepth 4 and --minMappingQuality 10.

Variant calling and filtering

Aligned reads stored in SAM format were co-ordinate sorted and converted to bam format using SAMTOOLS (version 0.1.18; Li *et al.* 2009). Duplicates were marked with PICARD tools ([\[broadinstitute.github.io/picard\]\(http://broadinstitute.github.io/picard\)\). Best practices established for the GENOME ANALYSIS TOOLKIT \(GATK version 3.8; DePristo *et al.* 2011\) were used for calling SNVs and short indels. Base quality recalibration was performed using BaseRecalibrator from GATK, where the recalibration report was formed using the default setting for covariates and the NCBI dbSNP database \(Build 151\) as the database for known sites. For each sample, the HaplotypeCaller program was used to call variants from the recalibrated bam file. The output, a GVCF file \(with g.vcf extension\), contained raw, unfiltered SNV and indel calls for all sites in the genome, variant or invariant. The GenotypeVCF tool was then used for performing joint genotyping on the per-sample GVCF files generated by HaplotypeCaller and produced a single VCF for the cohort. The single cohort VCF file was passed to the VariantFiltration tool for filtering based on several metrics provided by GenotypeVCF tool. Hard filtering was done using the following criteria, as suggested by GATK documentation: QUAL > 30.0, QD > 2.0, MQ > 40.0, FS < 60.0, HaplotypeScore < 13.0, MQRankSum > -12.5 and ReadPosRankSum > -8.0.](http://</p>
</div>
<div data-bbox=)

Annotation of variants

The filtered VCF file was annotated using SnpEFF (version 4.3; Cingolani *et al.* 2012) and NCBI annotation release 105 of the CanFam3.1 build. Analysis of 'protein-changing' variants included variants annotated by SnpEFF with the following sequence ontology terms: missense_variant, start_lost, stop_gained, stop_lost, stop_retained_variant, splice_acceptor_variant, splice_donor_variant, conservative_inframe_deletion, conservative_inframe_insertion, disruptive_inframe_deletion, disruptive_inframe_insertion, exon_loss_variant, frameshift_variant, and gene_fusion. SnpEFF classifies missense variants, conservative in-frame insertions and conservative in-frame deletions as having a 'moderate impact'. 'High-impact' variants include all other protein-changing variants. SnpEFF predicts loss-of-function (LoF) variants that include stop-gains (nonsense), splice site-disrupting SNVs, frameshift indels in a coding sequence or larger deletions that remove coding exons. SnpEFF also filters putative LoF variants identified in the last 5% of the coding region. VCFtools was used for studying the statistics of the variants (Danecek *et al.* 2011).

Phylogenetic tree

The phylogenetic tree was constructed using SnpPhylo (Lee *et al.* 2014). We used only exonic biallelic variants and used an LD threshold of 0.7 to remove redundancy owing to LD, leaving a total of 288 452 variants that were used for the tree construction.

Enrichment analysis

The genes carrying tolerant LoF variants were used for enrichment analysis using the Database for Annotation,

Visualization and Integrated Discovery (DAVID) functional annotation tool (Huang *et al.* 2009). The DAVID functional annotation clustering uses an algorithm to explore relationships among the annotation terms from various annotation sources and then presents a score for an enriched group of terms. In the current version of DAVID, the annotation tool includes more than 40 annotation categories including GO terms, KEGG Pathways and UniProt.

Identification of potential embryonic lethal alleles

To identify and prioritize LoF variants in putative embryonic lethal (EL) genes, we used data from recent publications on genome-wide screens for genes that cause embryonic lethality in mouse (Dickinson *et al.* 2016), cattle (Agerholm *et al.* 2001, 2006; Charlier *et al.* 2012; Fritz *et al.* 2013, 2018; Sonstegard *et al.* 2013; Cooper *et al.* 2014; Daetwyler *et al.* 2014; Venhoranta *et al.* 2014; Adams *et al.* 2016; Schütz *et al.* 2016; Schwarzenbacher *et al.* 2016; Michot *et al.* 2017), pig (Derks *et al.* 2019) and human (Shamseldin *et al.* 2015; Lord *et al.* 2019). The non-redundant curated list of EL genes from each of the studies was used to obtain orthologous gene ids from the NCBI dog genome CanFam3.1 annotation release 105 (Table S2). Biallelic variants in these orthologous genes were considered to cause potential EL alleles, if annotated by *SNPEFF* as high impact and not present in the homozygous variant state in any of the individuals. Additionally, we restricted our list to variants with an average sequencing depth of more than 5 using the GATK variant filter 'select DP >'. DP values in the VCF file represent the number of reads passing quality control used to calculate the genotype at a specific site in a specific sample, with higher values for DP generally leading to more accurate genotype calls.

Identification of potential developmental disorder alleles

We also used the developmental disorder (DD) genes included in the DDG2P panel (<https://decipher.sanger.ac.uk/ddd>; accessed 25 March 2019) (Deciphering Developmental Disorders Study 2015). This included 1846 unique genes. Mapping them onto CanFam3.1 NCBI Annotation release 105 found 1809 orthologous genes. Variants in these 1809 genes were filtered in the same manner as explained above (Table S2).

Results and discussion

We analysed short-read whole genome sequencing data from 582 dogs and 8 wolves. The read coverage for the 590 genomes ranged between 10× and 66× with a mean coverage of 24× (Table S1). The exclusive usage of genomes with relatively high coverage improved the accuracy of variant detection and the assignment of correct genotype calls in each analysed dog genome. All WGS data available in the DBVDC collection are processed through the

DBVDC pipeline, which detects sequence variants in the form of SNVs and short indels from alignments of the sequences. Each animal with a whole genome sequence is genotyped for all variants detected. The aim of the consortium is to make summary data and variant (VCF) files available to the wider scientific community.

Diploid calls were confidently made for an average 94.0% of the autosomal bases in the reference genome, with a range of 80–97% across the 590 genomes. These numbers are similar to those reported in 132 individuals from five canid species (Taylor *et al.* 2016). After filtering, a total of >33 million variants were identified consisting of 23 133 692 high-confidence SNVs and 10 048 038 indels (Table 1). Only 7% of these variants were contained in the NCBI dbSNP database (Build 151). The average SNV transition-to-transversion ratio was 2.08, and the SNV heterozygote-to-homozygote ratios were in the range of 0.5–3.3 (average 1.2). We identified 6 297 746 distinct short insertions (range 1–704 bp) and 3 750 292 distinct short deletions (range 1–329 bp), with an average of 1 427 335 indel variants per individual genome. The estimated heterozygote-to-homozygote ratios were 0.4–2.5 for short indels (average 1.02).

We used NCBI Annotation Release 105 to functionally annotate the variants (Appendix S1) and identified SNVs in 29 413 genes and indels in 29 440 genes out of a total of 29 831. As expected, only 2% of the variants were located within protein-coding regions/exonic sequences. The number of homozygous and heterozygous variants varied greatly between individual genomes. We speculate that this may reflect in part the different levels of inbreeding (Table 1).

Phylogenetic tree

A total of 288 452 biallelic exonic variants were used to construct a phylogenetic tree of the 590 genomes

Table 1 Variants in a typical dog genome

	Average	Median	Range
SNVs			
Homozygous	1 950 938	1 957 262	1 127 141–3 073 757
Heterozygous	2 189 278	2 153 775	311 003–3 912 095
Indels			
Homozygous	800 793	816 114	324 094–1 388 497
Heterozygous	803 006	791 225	170 327–1 122 579
High-impact variants¹			
SNVs	1371	1379	527–2200
Indels	4553	4709	1778–12 710
Moderate impact variants¹			
SNVs	36 315	36 810	10 460–53 789
Indels	1808	1838	515–3525

SNV, single-nucleotide variant.

Note: The table gives average numbers and ranges of identified variants and their predicted effects in an individual as calculated from the comprehensive analysis of 590 genomes.

¹Variant impact classification as provided by the *SNPEFF* output (see Appendix S1).

Table 2 Causative variants for inherited traits in dogs that were identified with the help of the Dog Biomedical Variant Database Consortium dataset

Gene	Phenotype	Breed	OMIA	Reference
ACADVL	Exercise-induced metabolic myopathy	German Hunting Terrier	002140-9615	Lepori <i>et al.</i> (2018)
ACP4	Amelogenesis imperfecta	Akita	002177-9615	Hytönen <i>et al.</i> (2019)
ADAMTS3	Upper airway syndrome	Norwich Terrier	n.a.	Marchant <i>et al.</i> (2019)
ALPL	Hypophosphatasia	Karelian Bear Dog	002162-9615	Kyöstilä <i>et al.</i> (2019)
ANLN	Respiratory distress syndrome	Dalmatian	000101-9615	Holopainen <i>et al.</i> (2017)
ASPRV1	Ichthyosis	German Shepherd Dog	002099-9615	Bauer <i>et al.</i> (2017a)
ATG4D	Neurodegenerative vacuolar storage disease	Lagotto Romagnolo	001954-9615	Kyöstilä <i>et al.</i> (2015)
ATP13A2	Neuronal ceroid lipofuscinosis (CLN12)	Australian Cattle Dog	001552-9615	Schmutz <i>et al.</i> (2019)
ATP1B2	Spongy degeneration with cerebellar ataxia 2 (SDCA2)	Belgian Shepherd	002110-9615	Mauri <i>et al.</i> (2017a)
CHRNE	Myasthenic syndrome, congenital, owing to CHRNE	Heideterrier	000685-9615	Herder <i>et al.</i> (2017)
CLCN1	Myotonia congenita	Labrador Retriever	000698-9615	Quitt <i>et al.</i> (2018)
CLN8	Neuronal ceroid lipofuscinosis	Alpenländische Dachsbracke	001506-9615	Hirz <i>et al.</i> (2017)
COL11A2	Skeletal dysplasia 2 (SD2)	Labrador Retriever	001772-9615	Frischknecht <i>et al.</i> (2013)
COL1A2	Osteogenesis imperfecta	Lagotto Romagnolo	002112-9615	Letko <i>et al.</i> (2019a)
COL5A1	Ehlers–Danlos syndrome	Labrador Retriever	n.a.	Bauer <i>et al.</i> (2019a)
COL5A1	Ehlers–Danlos syndrome	mixed breed	n.a.	Bauer <i>et al.</i> (2019a)
COL6A1	Muscular dystrophy, Ullrich type	Landseer	001967-9615	Steffen <i>et al.</i> (2015)
COL7A1	Epidermolysis bullosa, dystrophic	Central Asian Shepherd	000341-9615	Niskanen <i>et al.</i> (2017)
CUBN	Intestinal cobalamin malabsorption owing to CUBN mutation	Beagle	001786-9615	Drögemüller <i>et al.</i> (2014b)
CUBN	Intestinal cobalamin malabsorption owing to CUBN mutation	Border Collie	001786-9615	Owczarek-Lipska <i>et al.</i> (2013)
DIRAS1	Epilepsy, generalized myoclonic, with photosensitivity	Rhodesian Ridgeback	002095-9615	Wielander <i>et al.</i> (2017)
EDA	X-linked hypohidrotic ectodermal dysplasia	Dachshund	000543-9615	Hadji Rasouliha <i>et al.</i> (2018)
EDA	X-linked hypohidrotic ectodermal dysplasia	mixed breed	000543-9615	Waluk <i>et al.</i> (2016)
ENAM	Amelogenesis imperfecta	Parson Russell Terrier	001805-9615	Hytönen <i>et al.</i> (2019)
EXT2	Osteochondromatosis	American Staffordshire Terrier	001214-9615	Friedenberg <i>et al.</i> (2018)
FAM20C	Dental hypomineralization	Border Collie		Hytönen <i>et al.</i> (2016)
FAM83G	Hyperkeratosis, palmoplantar	Irish Terrier, Kromfohländer	001327-9615	Drögemüller <i>et al.</i> (2014a)
GJA9	Polyneuropathy (LPN2)	Leonberger	002119-9615	Becker <i>et al.</i> (2017)
GP9	Bernard–Soulier syndrome	Cocker Spaniel	001198-9615	Gentilini <i>et al.</i> (unpublished data)
KCNJ10	Spongy degeneration with cerebellar ataxia 1 (SDCA1)	Malinois	002089-9615	Mauri <i>et al.</i> (2017b)
KRT71	Curly hair	many	000245-9615	Bauer <i>et al.</i> (2019a, 2019b) and Salmela <i>et al.</i> (2019)
MC1R	Cream coat colour	Australian Cattle Dog	001199-9615	Dürrig <i>et al.</i> (2018)
MFSD12	Phaeomelanin dilution	Many	n.a.	Hédan <i>et al.</i> (2019)
MFSD8	Neuronal ceroid lipofuscinosis	Chihuahua	001962-9615	Karli <i>et al.</i> (2016)
MKLN1	Lethal acrodermatitis	Bull Terrier and Miniature Bull Terrier	002146-9615	Bauer <i>et al.</i> (2018a)
MLPH	Coat colour dilution	Chow Chow, Sloughi, Thai Ridgeback	000031-9615	Bauer <i>et al.</i> (2018b)
NAPEPLD	Leukoencephalomyelopathy	Great Dane and Rottweiler	001788-9615	Minor <i>et al.</i> (2018)
NECAP1	Progressive retinal atrophy	Giant Schnauzer	n.a.	Hitti <i>et al.</i> (2019)
NHLRC1	Myoclonus epilepsy of Lafora	Chihuahua	000690-9615	Barrientos <i>et al.</i> (2019)
NME5	Primary ciliary dyskinesia	Alaskan Malamute	n.a.	Anderegg <i>et al.</i> (2019)
NSDHL	CHILD-like syndrome	Labrador Retriever	002117-9615	Bauer <i>et al.</i> (2017b)
OCA2	Oculocutaneous albinism II	German Spitz	002130-9615	Caduff <i>et al.</i> (2017a)
OLFML3	Goniodysgenesis	Border Collie	001223-9615	Pugh <i>et al.</i> (2019)
PLN	Dilated cardiomyopathy	Welsh Springer Spaniel	002195-9615	Yost <i>et al.</i> (2019)
PPT1	Photoreceptor dysplasia	Miniature Schnauzer	001311-9615	Murgiano <i>et al.</i> (2019)
PTPRQ	Deafness	Doberman Pinscher	002148-9615	Guevar <i>et al.</i> (2018)
RAB3GAP1	Polyneuropathy, ocular abnormalities and neuronal vacuolation	Alaskan Husky	001970-9615	Wiedmer <i>et al.</i> (2015)
SCN8A	Spinocerebellar ataxia	Alpenländische Dachsbracke	002194-9615	Letko <i>et al.</i> (2019b)
SCARF2	Van den Ende–Gupta syndrome	Wirehaired Fox Terrier	002016-9615	Hytönen <i>et al.</i> (2016)
SGK3	Hypotrichosis, recessive	Scottish Deerhound	001279-9615	Hytönen & Lohi (2019)
SIX6	Eye malformation, congenital	Golden Retriever	n.a.	Hug <i>et al.</i> (2019)

Table 2 (Continued)

Gene	Phenotype	Breed	OMIA	Reference
<i>SLC19A3</i>	Leigh-like subacute necrotising encephalopathy	Yorkshire Terrier	001097-9615	Drögemüller <i>et al.</i> (2019)
<i>SLC37A2</i>	Craniomandibular osteopathy	West Highland White Terrier, Scottish Terrier, Cairn Terrier	000236-9615	Hytönen <i>et al.</i> (2016)
<i>SLC45A2</i>	Coat colour, albinism, oculocutaneous type IV	Bull Mastiff	001821-9615	Caduff <i>et al.</i> (2017b)
<i>SMOC2</i>	Brachycephaly	Many	001551-9615	Marchant <i>et al.</i> (2017)
<i>SUV39H2</i>	Nasal parakeratosis	Greyhound	001373-9615	Bauer <i>et al.</i> (2018c)
<i>SUV39H2</i>	Nasal parakeratosis	Labrador Retriever	001373-9615	Jagannathan <i>et al.</i> (2013)
<i>TECPR2</i>	Neuroaxonal dystrophy	Spanish Water Dog	001975-9615	Hahn <i>et al.</i> (2015)
<i>VLDLR</i>	Cerebellar hypoplasia	Eurasier	001947-9615	Gerber <i>et al.</i> (2015)

OMIA, Online Mendelian Inheritance in Animal.

Note: More details on these variants are given in Table S2.

analysed. This confirmed the expected breed clustering (Fig. S1) as reported in other studies (Plassais *et al.* 2019). Subtle differences include a single Dalmatian which was closer to the English Pointer in our tree, whereas Dalmatians were closer to Curly Coated Retrievers and Dachshunds in Plassais *et al.* (2019). Also, Greyhounds in our tree were closer to the Terriers, whereas in Plassais *et al.* (2019), they were closer to Borzoi and Bouvier des Flandres.

Loss-of-function-tolerant genes

The entire DBVCD dataset comprising 590 genomes contained 32 240 loss-of-function (LoF) variants (5614 SNVs and 26 626 indels). LoFs represent a more stringently filtered subset of high-impact variants, based on previously suggested parameters (MacArthur *et al.* 2012). Only 181 of these LoF variants were found in dbSNP (Build 151). The LoF variants were found in 40 847 RefSeq protein-coding transcripts (including XM xenoRefs) belonging to 8109 genes.

In order to look for LoF-tolerant genes, i.e. genes that are not essential for survival and reproduction, we subjected the putative LoF variants to a series of filtering criteria. The exclusion criteria were as follows: (i) LoF variants not occurring in homozygous variant state in at least one of the individuals, (ii) LoF variants for which all protein-coding transcripts of the gene were not affected, (iii) LoF variants that overlapped any repetitive sequences, (iv) variants affecting non-canonical splice sites, (v) indel variants affecting splice sites and (vi) known OMIA gene variants published in OMIA. After performing these filtering steps, we obtained 1897 genes harbouring LoF variants (Table S2) from a total of 13 603 genes annotated to contain LoF variants. Enrichment analysis (Appendix S1) revealed that the olfactory receptor genes are the only large class of functionally related genes that are tolerant of LoF variants.

Potential embryonic lethal and developmental disorder variants

We used our dataset to identify and prioritize variants that could potentially give rise to recessive alleles causing embryonic lethality (EL) in dogs. We obtained 528 dog orthologues to genes with published evidence for EL alleles in other species. After filtering for variants as detailed in Appendix S1, we identified 247 genes where a single or a small number of dogs were heterozygous whereas none of the 590 genomes were homozygous for the mutant allele (Table S2). Across the 590 genomes, each genome carried on average 30 potential EL alleles (range 6–181). This is slightly higher than what has been reported for 624 cattle (Charlier *et al.* 2016).

Loss-of-function variants were also analysed for 2211 dog orthologues to DD genes. We found potentially deleterious variants in 894 DD genes (Table S2). Across the 590 genomes, each genome carried on average 97 potential DD alleles (range 21–583).

The numbers of identified potential EL and DD alleles are probably overestimated, as the current NCBI Annotation Release 105 of the CanFam3.1 reference genome is still imperfect, resulting in erroneous S_{NPEFF} functional effect predictions. Future improvements in the reference genome assembly and its functional annotation will further increase the accuracy of the variant catalogue. We also acknowledge that our catalogue lacks most of the structural variation, which may be assessed by the use of third-generation long-read sequencing technologies.

Relevance for biomedical research

Work based on the DBVDC variant catalogue has already resulted in the identification of more than 50 causative variants for various inherited traits in dogs (Tables 2 & S2). These included many canine homologues of known human hereditary disorders, but also identified several novel genes

not previously known to be associated with human diseases, thus providing new candidate genes for those homologous human diseases.

Conclusion

The variant analysis of 590 canine genomes identified ~33 million functionally annotated variants. We made an effort to include only genome sequences with high coverage and applied stringent filtering criteria to ensure the high quality of the variant and genotype calls. This dataset should help to identify causative variants for monogenic disorders more efficiently. The addition of more genomes will eventually also aid in the identification of causal variants for complex traits.

Acknowledgements

The authors would like to thank Eva Andrist, Nathalie Besuchet Schmutz, Muriel Fragnière and Sabrina Schenk for expert technical assistance, the Next Generation Sequencing Platform of the University of Bern for performing many high-throughput sequencing experiments and the Interfaculty Bioinformatics Unit of the University of Bern for providing high-performance computing infrastructure. We also acknowledge all researchers of the canine community who deposited dog or wolf whole genome sequencing data into public databases. Funding information is listed in Appendix S1.

Data availability

The genome sequence accessions and metadata are available from Table S1. The variants are available at the European Variant Archive under project ID PRJEB32865 at <https://www.ebi.ac.uk/eva/?eva-study=PRJEB32865>.

References

Adams H.A., Sonstegard T.S., VanRaden P.M., Null D.J., VanTassell C.P., Larkin D.M. & Lewin H.A. (2016) Identification of a nonsense mutation in *APAF1* that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *Journal of Dairy Science* **99**, 6693–701.

Agerholm J.S., Bendixen C., Andersen O. & Arnbjerg J. (2001) Complex vertebral malformation in holstein calves. *Journal of Veterinary Diagnostic Investigation* **13**, 283–9.

Agerholm J.S., McEvoy F. & Arnbjerg J. (2006) Brachyspina syndrome in a Holstein calf. *Journal of Veterinary Diagnostic Investigation* **18**, 418–22.

Anderegg L., Im Gut Hof M., Hetzel U., Howerth E.W., Leuthard F., Jagannathan J. & Leeb T. (2019) *NME5* frameshift variant in Alaskan Malamutes with primary ciliary dyskinesia. *PLoS Genetics*. in press.

Bai B., Zhao W.M., Tang B.X. *et al.* (2015) DoGSD: the dog and wolf genome SNP database. *Nucleic Acids Research* **43**, D777–83.

Barrientos L., Maiolini A., Häni A., Jagannathan V. & Leeb T. (2019) *NHLRC1* dodecamer repeat expansion demonstrated by whole genome sequencing in a Chihuahua with Lafora disease. *Animal Genetics* **50**, 118–9.

Bauer A., Waluk D.P., Galichet A. *et al.* (2017a) A *de novo* variant in the *ASPRV1* gene in a dog with ichthyosis. *PLoS Genetics* **13**, e1006651.

Bauer A., De Lucia M., Jagannathan V., Mezzalana G., Casal M.L., Welle M.M. & Leeb T. (2017b) A large deletion in the *NSDHL* gene in Labrador Retrievers with a congenital cornification disorder. *G3 (Bethesda)* **7**, 3115–21.

Bauer A., Jagannathan V., Högl S. *et al.* (2018a) *MKLN1* splicing defect in dogs with lethal acrodermatitis. *PLoS Genetics* **14**, e1007264.

Bauer A., Kehl A., Jagannathan V. & Leeb T. (2018b) A novel *MLPH* variant in dogs with coat colour dilution. *Animal Genetics* **49**, 94–7.

Bauer A., Nimmo J., Newman R., Brunner M., Welle M.M., Jagannathan V. & Leeb T. (2018c) A splice site variant in the *SUV39H2* gene in Greyhounds with nasal parakeratosis. *Animal Genetics* **49**, 137–40.

Bauer A., Bateman J.F., Lamande S.R. *et al.* (2019a) Identification of two independent *COL5A1* variants in dogs with Ehlers Danlos syndrome. *BMC Veterinary Research* in preparation.

Bauer A., Hadji Rasouliha S., Brunner M.T. *et al.* (2019b) A second *KRT71* allele in curly coated dogs. *Animal Genetics* **50**, 97–100.

Becker D., Minor K.M., Letko A., Ekenstedt K.J., Jagannathan V., Leeb T., Shelton G.D., Mickelson J.R. & Drögemüller C. (2017) A *GJA9* frameshift variant is associated with polyneuropathy in Leonberger dogs. *BMC Genomics* **18**, 662.

Bourneuf E., Otz P., Pausch H. *et al.* (2017) Rapid discovery of *de novo* deleterious mutations in cattle enhances the value of livestock as model species. *Scientific Reports* **7**, 11466.

Boyko A.R., Quignon P., Li L. *et al.* (2010) A simple genetic architecture underlies morphological variation in dogs. *PLoS Biology* **8**, e1000451.

Brown E.A., Dickinson P.J., Mansour T. *et al.* (2017) *FGF4* retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 11476–81.

Cadiou E., Neff M.W., Quignon P. *et al.* (2009) Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150–3.

Caduff M., Bauer A., Jagannathan V. & Leeb T. (2017a) *OCA2* splice site variant in German Spitz dogs with oculocutaneous albinism. *PLoS ONE* **12**, e0185944.

Caduff M., Bauer A., Jagannathan V. & Leeb T. (2017b) A single base deletion in the *SLC45A2* gene in a Bull mastiff with oculocutaneous albinism. *Animal Genetics* **48**, 619–21.

Charlier C., Agerholm J.S., Coppieters W. *et al.* (2012) A deletion in the bovine *FANCI* gene compromises fertility by causing fetal death and brachyspina. *PLoS ONE* **7**, e43085.

Charlier C., Li W., Harland C. *et al.* (2016) NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Research* **26**, 1333–41.

Chen S., Zhou Y., Chen Y. & Gu J. (2018) *FASTP*: an ultra-fast all-in-one *FASTQ* preprocessor. *Bioinformatics* **34**, i884–90.

Cingolani P., Platts A., Wang L., Coon M., Nguyen T., Wang L., Land S.J., Lu X. & Ruden D.M. (2012) 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* 6, 80–92.
- Cooper T.A., Wiggans G.R., Null D.J., Hutchison J.L. & Cole J.B. (2014) Genomic evaluation, breed identification, and discovery of a haplotype affecting fertility for Ayrshire dairy cattle. *Journal of Dairy Science* 97, 3878–82.
- Daetwyler H.D., Capitan A., Pausch H. *et al.* (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics* 46, 858–65.
- Danecek P., Auton A., Abecasis G. *et al.* (2011) The variant call format and *vcftools*. *Bioinformatics* 27, 2156–8.
- Deciphering Developmental Disorders Study (2015) Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 519, 7542.
- DePristo M.A., Banks E., Poplin R., *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genetics* 43, 491–8.
- Derks M.F.L., Gjuvsland A.B., Bosse M. *et al.* (2019) Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genetics* 15, e1008055.
- Dickinson M.E., Flenniken A.M., Ji X. *et al.* (2016) High-throughput discovery of novel developmental phenotypes. *Nature* 537, 508–14.
- Drögemüller C., Karlsson E.K., Hytönen M.K., Perloski M., Dolf G., Sainio K., Lohi H., Lindblad-Toh K. & Leeb T. (2008) A mutation in hairless dogs implicates *FOXT3* in ectodermal development. *Science* 321, 1462.
- Drögemüller C., Becker D., Brunner A., Haase B., Kircher P., Seeliger F., Fehr M., Baumann U., Lindblad-Toh K. & Leeb T. (2009) A missense mutation in the *SERPINH1* gene in Dachshunds with osteogenesis imperfecta. *PLoS Genetics* 5, e1000579.
- Drögemüller M., Jagannathan V., Becker D. *et al.* (2014a) A mutation in the *FAM83G* gene in dogs with hereditary footpad hyperkeratosis (HFH). *PLoS Genetics* 10, e1004370.
- Drögemüller M., Jagannathan V., Howard J., Bruggmann R., Drögemüller C., Ruetten M., Leeb T. & Kook P.H. (2014b) A frameshift mutation in the cubilin gene (*CUBN*) in Beagles with Imerlund-Gräsbeck syndrome (selective cobalamin malabsorption). *Animal Genetics* 45, 148–50.
- Drögemüller M., Matisek K., Jagannathan V. *et al.* (2019) *SLC19A3* loss-of-function variant in Yorkshire Terriers with Leigh-like subacute necrotising encephalopathy. *Animal Genetics* in preparation.
- Dürig N., Letko A., Lepori V. *et al.* (2018) Two MC1R loss-of-function alleles in cream-coloured Australian Cattle Dogs and white Huskies. *Animal Genetics* 49, 284–90.
- Friedenberg S.G., Vansteenkiste D., Yost O. *et al.* (2018) A *de novo* mutation in the *EXT2* gene associated with osteochondromatosis in a litter of American Staffordshire Terriers. *Journal of Veterinary Internal Medicine* 32, 986–92.
- Frischknecht M., Niehof-Oellers H., Jagannathan V. *et al.* (2013) A *COL11A2* mutation in Labrador retrievers with mild disproportionate dwarfism. *PLoS ONE* 8, e60149.
- Fritz S., Capitan A., Djari A. *et al.* (2013) Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in *GART*, *SHBG* and *SLC37A*. *PLoS ONE* 8, e65550.
- Fritz S., Hoze C., Rebours E. *et al.* (2018) An initiator codon mutation in *SDE2* causes recessive embryonic lethality in Holstein cattle. *Journal of Dairy Science* 101, 6220–31.
- Gerber M., Fischer A., Jagannathan V. *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, e0108917.
- Grall A., Guaguère E., Planchais S. *et al.* (2012) *PNPLA1* mutations cause autosomal recessive congenital ichthyosis in Golden Retriever dogs and humans. *Nature Genetics* 44, 140–7.
- Guevar J., Olby N.J., Meurs K.M., Yost O. & Friedenberg S.G. (2018) Deafness and vestibular dysfunction in a Doberman Pinscher puppy associated with a mutation in the *PTPRQ* gene. *Journal of Veterinary Internal Medicine* 32, 665–9.
- Hadij Rasouliha S., Bauer A., Dettwiler M., Welle M.M. & Leeb T. (2018) A frameshift variant in the *EDA* gene in Dachshunds with X-linked hypohidrotic ectodermal dysplasia. *Animal Genetics* 49, 651–4.
- Hahn K., Rohdin C., Jagannathan V., Wohlsein P., Baumgärtner W., Seehusen F., Spitzbarth I., Grandon R., Drögemüller C. & Jäderlund K.H. (2015) *TECPR2* associated neuroaxonal dystrophy in Spanish water dogs. *PLoS ONE* 10, e0141824.
- Hayes B.J. & Daetwyler H.D. (2019) 1000 Bull Genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annual Reviews in Animal Biosciences* 7, 89–102.
- Hayward J.J., Castelhana M.G., Oliveira K.C. *et al.* (2016) Complex disease and phenotype mapping in the domestic dog. *Nature Communications* 7, 10460.
- Hédan B., Cadieu E., Botherel N. *et al.* (2019) Identification of a missense variant in *MFSD12* involved in dilution of phaeomelanin leading to white or cream coat color in dogs. *Genes* 10, 386.
- Herder V., Ciurkiewicz M., Baumgärtner W., Jagannathan V. & Leeb T. (2017) Frame-shift variant in the *CHRNE* gene in a juvenile dog with suspected myasthenia gravis-like disease. *Animal Genetics* 48, 625.
- Hirz M., Drögemüller M., Schänzer A. *et al.* (2017) Neuronal ceroid lipofuscinosis (NCL) is caused by the entire deletion of *CLN8* in the Alpenländische Dachsbracke dog. *Molecular Genetics and Metabolism* 120, 269–77.
- Hitti R.J., Oliver J.A.C., Schofield E.C. *et al.* (2019) Whole genome sequencing of Giant Schnauzer Dogs with progressive retinal atrophy establishes *NECAP1* as a novel candidate gene for retinal degeneration. *Genes* 10, 385.
- Holopainen S., Hytönen M.K., Syrjä P., Arumilli M., Järvinen A.K., Rajamäki M. & Lohi H. (2017) *ANLN* truncation causes a familial fatal acute respiratory distress syndrome in Dalmatian dogs. *PLoS Genetics* 13, e1006625.
- Huang D.W., Sherman B.T. & Lempicki R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* 37, 1–13.
- Hug P., Andereeg L., Dürig N., Lepori V., Jagannathan V., Spiess B., Richter M. & Leeb T. (2019) A *SIX6* nonsense variant in Golden Retrievers with congenital eye malformations. *Genes* 10, 454.
- Hytönen M.K. & Lohi H. (2016) Canine models of human rare disorders. *Rare Diseases* 4, e1241362.
- Hytönen M.K. & Lohi H. (2019) A frameshift insertion in *SGK3* leads to recessive hairlessness in Scottish Deerhounds: a candidate gene for human alopecia conditions. *Human Genetics* 138, 535–9.

- Hytönen M.K., Arumilli M., Lappalainen A.K. *et al.* (2016) Molecular characterization of three canine models of human rare bone diseases: Caffey, van den Ende-Gupta, and Raine syndromes. *PLoS Genetics* **12**, e1006037.
- Hytönen M.K., Arumilli M., Sarkiala E., Nieminen P. & Lohi H. (2019) Canine models of human amelogenesis imperfecta: identification of novel recessive *ENAM* and *ACP4* variants. *Human Genetics* **138**, 525–33.
- Jagannathan V., Bannoehr J., Plattet P. *et al.* (2013) A mutation in the *SUV39H2* gene in Labrador Retrievers with hereditary nasal parakeratosis (HNPK) provides insights into the epigenetics of keratinocyte differentiation. *PLoS Genetics* **9**, e1003848.
- Karli P., Oevermann A., Bauer A., Jagannathan V. & Leeb T. (2016) *MFSD8* single-base pair deletion in a Chihuahua with neuronal ceroid lipofuscinosis. *Animal Genetics* **47**, 631.
- Karlsson E.K., Baranowska I., Wade C.M. *et al.* (2007) Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nature Genetics* **39**, 1321–8.
- Kaukonen M., Woods S., Ahonen S., Lemberg S., Hellman M., Hytönen M.K., Permi P., Glaser T. & Lohi H. (2018) Maternal inheritance of a recessive *RBP4* defect in canine congenital eye disease. *Cell Reports* **23**, 2643–52.
- Kyöstilä K., Syrjä P., Jagannathan V. *et al.* (2015) A missense change in the *ATG4D* gene links aberrant autophagy to a neurodegenerative vacuolar storage disease. *PLoS Genetics* **11**, e1005169.
- Kyöstilä K., Syrjä P., Lappalainen A.K., Arumilli M., Hundi S., Karkamo V., Viitmaa R., Hytönen M.K. & Lohi H. (2019) A homozygous missense variant in the alkaline phosphatase gene *ALPL* is associated with a severe form of canine hypophosphatasia. *Scientific Reports* **9**, 973.
- Lee T.H., Guo H., Wang X., Kim C. & Paterson A.H. (2014) *SNPHYLO*: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 62.
- Lek M., Karczewski K.J., Minikel E.V. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–91.
- Lepori V., Mühlhause F., Sewell A.C. *et al.* (2018) Nonsense variant in the *ACADVL* gene in German Hunting Terriers with exercise induced metabolic myopathy. *G3 (Bethesda)* **8**, 1545–54.
- Letko A., Beineke A., Jagannathan V. & Drögemüller C. (2019a) A *de novo* in-frame duplication in the *COL1A2* gene of a Lagotto Romagnolo dog with osteogenesis imperfecta. *Animal Genetics* in preparation.
- Letko A., Dietschi E., Nieburg M., Jagannathan V., Gurtner C., Oevermann A. & Drögemüller C. (2019b) A missense variant in *SCN8A* in Alpine Dachshund dogs affected by spinocerebellar ataxia. *Genes* **10**, 362.
- Li H. & Durbin R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–95.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. & Durbin R. (2009) The sequence alignment/map format and *SAMTOOLS*. *Bioinformatics* **25**, 2078–9.
- Lindblad-Toh K., Wade C.M., Mikkelsen T.S. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–19.
- Lord J., McMullan D.J., Eberhardt R.Y. *et al.* (2019) Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. *Lancet* **393**, 747–57.
- MacArthur D.G., Balasubramanian S., Frankish A. *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–8.
- Mansour T.A., Lucot K., Konopelski S.E. *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 Genetics* **14**, e1007850.
- Marchant T.W., Johnson E.J., McTeir L. *et al.* (2017) Canine brachycephaly is associated with a retrotransposon-mediated missplicing of *SMOC2*. *Current Biology* **27**, 1573–84.
- Marchant T.W., Dietschi E., Rytz U. *et al.* (2019) An *ADAMTS3* missense variant is associated with Norwich Terrier upper airway syndrome. *PLoS Genetics* **15**, e1008102.
- Marsden C.D., Ortega-Del Vecchyo D., O'Brien D.P., Taylor J.F., Ramirez O., Vilà C., Marques-Bonet T., Schnabel R.D., Wayne R.K. & Lohmueller K.E. (2016) Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 152–7.
- Mauri N., Kleiter M., Dietschi E. *et al.* (2017a) A SINE insertion in *ATP1B2* in Belgian Shepherd dogs affected by spongy degeneration with cerebellar ataxia (SDCA2). *G3 (Bethesda)* **7**, 2729–37.
- Mauri N., Kleiter M., Leschnik M. *et al.* (2017b) A missense variant in *KCNJ10* in Belgian Shepherd dogs affected by spongy degeneration with cerebellar ataxia (SDCA1). *G3 (Bethesda)* **7**, 663–9.
- Merveille A.C., Davis E.E., Becker-Heck A. *et al.* (2011) *CCDC39* is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nature Genetics* **43**, 72–8.
- Michot P., Fritz S., Barbat A. *et al.* (2017) A missense mutation in *PFAS (phosphoribosylformylglycinamide synthase)* is likely causal for embryonic lethality associated with the MH1 haplotype in Montbéliarde dairy cattle. *Journal of Dairy Science* **100**, 8176–87.
- Minor K.M., Letko A., Becker D. *et al.* (2018) Canine *NAPEPLD*-associated models of human myelin disorders. *Scientific Reports* **8**, 5818.
- Murgiano L., Becker D., Torjman D. *et al.* (2019) Complex structural *PPT1* variant associated with non-syndromic canine retinal degeneration. *G3 (Bethesda)* **9**, 425–37.
- 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**, e0177527.
- Ostrander E.A., Wayne R.K., Freedman A.H. & Davis B.W. (2017) Demographic history, selection and functional diversity of the canine genome. *Nature Reviews Genetics* **18**, 705–20.
- Owczarek-Lipska M., Jagannathan V., Drögemüller C., Lutz S., Glanemann B., Leeb T. & Kook P.H. (2013) A frameshift mutation in the *cubilin* gene (*CUBN*) in Border Collies with Imerslund-Gräsbeck syndrome (selective cobalamin malabsorption). *PLoS ONE* **8**, e61144.
- Parker H.G., VonHoldt B.M., Quignon P. *et al.* (2009) An expressed *Fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* **325**, 995–8.
- Plassais J., Kim J., Davis B.W., Karyadi D.M., Hogan A.N., Harris A.C., Decker B., Parker H.G. & Ostrander E.A. (2019) Whole

- genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications* **10**, 1489.
- Pugh C.A., Farrell L.L., Carlisle A.J. *et al.* (2019) Arginine to glutamine variant in olfactomedin like 3 (OLFML3) is a candidate for severe goniodysgenesis and glaucoma in the Border Collie dog breed. *G3 (Bethesda)* **9**, 943–54.
- Quitt P.R., Hytönen M.K., Matiassek K., Rosati M., Fischer A. & Lohi H. (2018) Myotonia congenita in a Labrador Retriever with truncated *CLCN1*. *Neuromuscular Disorders* **28**, 597–605.
- Rimbault M. & Ostrander E.A. (2012) So many doggone traits: mapping genetics of multiple phenotypes in the domestic dog. *Human Molecular Genetics* **21**, R52–7.
- Salmela E., Niskanen J., Arumill M., Donner J., Lohi H. & Hytönen M.K. (2019) A novel *KRT71* variant in curly-coated dogs. *Animal Genetics* **50**, 101–4.
- Sarviaho R., Hakosalo O., Tiira K., Sulkama S., Salmela E., Hytönen M.K., Sillanpää M.J. & Lohi H. (2019) Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Translational Psychiatry* **9**, 18.
- Schmutz I., Jagannathan V., Bartenschlager F., Stein V.M., Gruber A.D., Leeb T. & Katz M.L. (2019) *ATP13A2* missense variant in Australian Cattle Dogs with late onset neuronal ceroid lipofuscinosis. *Molecular Genetics and Metabolism* **127**, 95–106.
- Schoenebeck J.J., Huthinson S.A., Byers A. *et al.* (2012) Variation of *BMP3* contributes to dog breed skull diversity. *PLoS Genetics* **8**, e1002849.
- Schütz E., Wehrhahn C., Wanjek M., Bortfeld R., Wemheuer W.E., Beck J. & Brenig B. (2016) The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of *TBFI1M* and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of *APOB*. *PLoS ONE* **11**, e0154602.
- Schwarzenbacher H., Burgstaller J., Seefried F.R. *et al.* (2016) A missense mutation in *TUBD1* is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. *BMC Genomics* **17**, 400.
- Seppälä E.H., Jokinen T.S., Fukata M. *et al.* (2011) *LGI2* truncation causes a remitting focal epilepsy in dogs. *PLoS Genetics* **7**, e1002194.
- Shamseldin H.E., Tulbah M., Kurdi W. *et al.* (2015) Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families. *Genome Biology* **16**, 116.
- Sonstegard T.S., Cole J.B., VanRaden P.M., VanTassell C.P., Null D.J., Schroeder S.G., Bickhart D. & McClure M.C. (2013) Identification of a nonsense mutation in *CWC15* associated with decreased reproductive efficiency in Jersey cattle. *PLoS ONE* **8**, e54872.
- Steffen F., Bilzer T., Brands J., Golini L., Jagannathan V., Wiedmer M., Drögemüller M., Drögemüller C. & Leeb T. (2015) A nonsense variant in *COL6A1* in Landseer dogs with muscular dystrophy. *G3 (Bethesda)* **5**, 2611–7.
- Sutter N.B., Eberle M.A., Parker H.G., Pullar B.J., Kirkness E.F., Kruglyak L. & Ostrander E.A. (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Research* **14**, 2388–96.
- Sutter N.B., Bustamante C.D., Chase K. *et al.* (2007) A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112–5.
- Tang B., Zhou Q., Dong L. *et al.* (2019) iDog: an integrated resource for domestic dogs and wild canids. *Nucleic Acids Research* **47**, D793–800.
- Taylor J.F., Whitacre L.K., Hoff J.L., Tizioto P.C., Kim J., Decker J.E. & Schnabel R.D. (2016) Lessons for livestock genomics from genome and transcriptome sequencing in cattle and other mammals. *Genetics Selection Evolution* **48**, 59.
- Vaysse A., Ratnakumar A., Derrien T. *et al.* (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genetics* **7**, e1002316.
- Venhoranta H., Pausch H., Flisikowski K. *et al.* (2014) In frame exon skipping in *UBE3B* is associated with developmental disorders and increased mortality in cattle. *BMC Genomics* **15**, 890.
- Waluk D., Zur G., Kaufmann R., Welle M.M., Jagannathan V., Drögemüller C., Müller E.J., Leeb T. & Galichet A. (2016) A splice defect in the *EDA* gene in dogs with an X-linked hypohidrotic ectodermal dysplasia (XLHED) phenotype. *G3 (Bethesda)* **6**, 2949–54.
- Wiedmer M., Overmann A., Borer-Germann S.E. *et al.* (2015) A *RAB3GAP1* SINE insertion in Alaskan Huskies with polyneuropathy, ocular abnormalities and neuronal vacuolation (POANV) resembling human Warburg Micro Syndrome 1 (WARBM1). *G3 (Bethesda)* **6**, 255–62.
- Wieland F., Sarviaho R., James F. *et al.* (2017) Generalized myoclonic epilepsy with photosensitivity in juvenile dogs caused by a defective *DIRAS* family GTPase 1. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 2669–74.
- Yost O., Friedenber S.G., Jesty S.A., Olby N.J. & Meurs K.M. (2019) The R9H phospholamban mutation is associated with highly penetrant dilated cardiomyopathy and sudden death in a spontaneous canine model. *Gene* **697**, 118–22.
- Zangerl B., Goldstein O., Philp A.R. *et al.* (2006) Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics* **88**, 551–63.
- Zapata I., Serpell J.A. & Alvarez C.E. (2016) Genetic mapping of canine fear and aggression. *BMC Genomics* **17**, 572.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.
Figure S1 Phylogenetic tree of 582 dogs and eight wolves based on exonic SNVs.

Table S1 Detailed descriptive statistics and accessions of 590 genomes.

Table S2 Catalogues of tolerated, potential embryonic lethal, potential development disorder and DBVDC causal variants.

Appendix S1 Affiliations of DBVDC members.