ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

REVIEW

doi: 10.1111/age.12460

The canine era: the rise of a biomedical model

F. G. van Steenbeek*, M. K. Hytönen†, P. A. J. Leegwater* and H. Lohi†

*Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 104, 3508 TD, Utrecht, the Netherlands. †Research Programs Unit, Molecular Neurology, Department of Veterinary Biosciences 00014, Folkhälsan Research Center, University of Helsinki, Helsinki, Finland.

Summary

Since the annotation of its genome a decade ago, the dog has proven to be an excellent model for the study of inherited diseases. A large variety of spontaneous simple and complex phenotypes occur in dogs, providing physiologically relevant models to corresponding human conditions. In addition, gene discovery is facilitated in clinically less heterogeneous purebred dogs with closed population structures because smaller study cohorts and fewer markers are often sufficient to expose causal variants. Here, we review the development of genomic resources from microsatellites to whole-genome sequencing and give examples of successful findings that have followed the technological progress. The increasing amount of whole-genome sequence data warrants better functional annotation of the canine genome to more effectively utilise this unique model to understand genetic contributions in morphological, behavioural and other complex traits.

Keywords animal model, dog, genetic tools, next generation sequencing

Unique population history and genetic architecture facilitate gene mapping

Domestication of the dog occurred thousands of years ago. The exact period in history and the region of origin have been extensively discussed and are still under debate (Savolainen *et al.* 2002; Pang *et al.* 2009; Boyko 2011; vonHoldt *et al.* 2011; Axelsson *et al.* 2013; Thalmann *et al.* 2013; Wang *et al.* 2013, 2016; Freedman *et al.* 2014; Shannon *et al.* 2015). The dog was used as a sentry, a food source and for hunting, but its role as companion animal swiftly resulted in the removal of the dog from its pack to become a pet. The French Philosopher Voltaire was the first to launch the term 'man's best friend':

'It seems that nature has given the dog to man for his defense and for his pleasure. Of all the animals it is the most faithful: it is the best friend man can have' (Voltaire, *Dictionnaire Philosophique,* 1764).

Selection and breed formation have resulted in a broad variety of dog phenotypes. Due to man's influence, the appearance of the dog varies greatly; for instance, there is up to a 40-fold difference in size between the Great Dane and the Chihuahua (Wayne & Ostrander 2007). The vast differences among breeds are the result of just a small number of gene variants that define, for instance, coat colour (Karlsson *et al.* 2007; Dreger *et al.* 2013), coat texture (Salmon Hillbertz *et al.* 2007; Cadieu *et al.* 2009) and size (Sutter *et al.* 2007), indicating that the genetic architecture is simplified and involves large gene effects. For example, the background of coat colour of Labrador Retrievers, Golden Retrievers and Flat-coated Retrievers is defined by two genomic loci (Fig. 1) (Lavrijsen *et al.* 2014a, b). The first locus on chromosome 5 is in close vicinity to the *melanocortin 1 receptor (alpha melanocyte-stimulating hormone receptor)* gene (*MC1R*), and the second locus on chromosome 11 includes the *tyrosinase-related protein 1* gene (*TYRP1*). Both genes are known to be involved in mammalian coat coloration (Everts *et al.* 2000; Newton *et al.* 2000; Schmutz *et al.* 2002).

Several examples demonstrate how selection for phenotypic characteristics results in severe health issues for canines. A brachycephalic phenotype selected for in breeds such as the Boxer, Boston Terrier, Pekingese and Bulldog causes breathing abnormalities because of the soft tissue blockage of the airways during respiration (Packer *et al.* 2015). Cavalier King Charles Spaniels are commonly affected by syringomyelia by which fluid-filled cavities develop within the spinal cord near the brain (Parker *et al.* 2011). This disease results from the foreshortening of the caudal skull, termed Chiari-like malformation (Cerda-Gonzalez *et al.* 2009).

Address for correspondence

F. G. Van Steenbeek, Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 104, 3508 TD, Utrecht, the Netherlands.
E-mail: f.g.vansteenbeek@uu.nl

Accepted for publication 18 April 2016

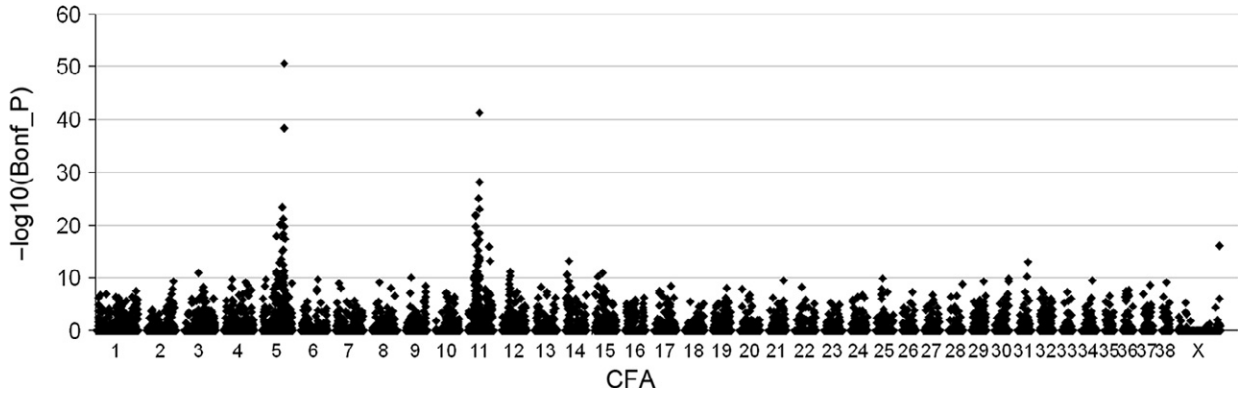

Figure 1 Manhattan plot for coat colour presenting the results of the comparison of 86 brown retrievers (73 Labrador Retrievers and 13 Flat-coated Retrievers) and 167 yellow retrievers (118 Labrador Retrievers and 49 Golden Retrievers) genotyped on Illumina CanineSNP20 BeadChip. Two regions containing genome-wide significant SNPs were observed. Both regions include genes known to be involved in coat coloration (*MC1R* on chromosome 5 and *TYRP1* on chromosome 11).

The characteristic wrinkled skin of Shar-Pei is associated with familial Shar-Pei fever (FSF), characterised by periodic fever and inflammation similar to human familial Mediterranean fever (French FMF Consortium 1997). High concentrations of hyaluronan in the skin cause the wrinkling phenotype (Zanna *et al.* 2008). FSF was found to be associated with the expression level of the gene *hyaluronan synthase 2* (*HAS2*), and the copy number of a 16.1-kb repeat fragment close to the gene has been correlated with both *HAS2* expression and disease (Olsson *et al.* 2011). The repeat is expected to contain a regulatory element that regulates the gene expression of *HAS2*. An increased copy number of the duplication gives rise to an increased copy number of the potential enhancer elements, increasing the expression of *HAS2* and consequently the elevating levels of hyaluronan. The large amount of hyaluronan is postulated to be responsible for the fever and inflammation and, as a result, selection for appearance directly causes a selection for disease risk.

The extent of white colour in Dalmatians correlates with congenital sensorineural deafness (Stritzel *et al.* 2009). Involvement of microphthalmia-associated transcription factor (MITF) defines the extent of extreme white found in the coat (Karlsson *et al.* 2007; Baranowska Korberg *et al.* 2014) due to its function in melanogenesis. *MITF* is also known to be secondarily associated with hearing loss in relation to the regulation of pigmentation (Waardenburg/Tietz syndrome) (TIETZ 1963) and deafness in mice (Motohashi *et al.* 1994; Yoshida *et al.* 1996). A causative mutation for deafness in Dalmatians has not yet been found.

The dog is a relatively large animal as compared to mice, and many canine disorders resemble the corresponding human conditions better than do induced rodent models. Numerous examples exist across disease groups, including epilepsy (Lohi *et al.* 2005), narcolepsy (Lin *et al.* 1999), developmental conditions (Bannasch *et al.* 2010; Hytonen *et al.* 2012; Wolf *et al.* 2015), skin diseases (Grall *et al.* 2012; Tengvall *et al.* 2013), eye disorders (Sargan *et al.* 2007; Ahonen *et al.* 2013) and cancer (Arendt *et al.* 2015). Unique population history and breed structure have resulted in a genetic architecture that facilitates the exploration of the genetic basis of simple and complex disorders (Ostrander & Kruglyak 2000). Extensive linkage disequilibrium combined with low haplotype diversity within breeds requires a lower number of markers to test the genome and also fewer samples for statistical power in genome-wide association studies (GWAS) as compared to human GWAS (Sutter & Ostrander 2004; Sutter *et al.* 2004). An informative example is the canine systemic lupus erythematosus study that identified a limited set of risk loci but with each of them having a high contribution to the disease (Wilbe *et al.* 2010). Additionally, interbreed mapping with shared conditions provides a powerful approach to narrow down the associated loci for the identification of the causal variant, as demonstrated by the identification of a *FOXI3* frameshift mutation in canine ectodermal dysplasia (Drogemuller *et al.* 2008). A 1.7-Mb disease locus in Chinese Crested Dogs was narrowed down to a 102-kb region by interbreed analysis in Peruvian and Mexican Hairless breeds. To date, there are over 200 known pathogenic variations across disorders in dogs, and this number is rapidly increasing (OMIA database).

Evolving genetic tools

The annotation of the canine genome has allowed for a rapid development of powerful genomic tools for gene discovery (Fig. 2). Sparse microsatellite experiments (Clark *et al.* 2004) have been replaced by high-density SNP arrays and resequencing experiments. One of the first successes of the microsatellite approach was the localisation of the gene causing copper toxicosis in Bedlington Terriers with a set of 213 microsatellite markers (Yuzbasiyan-Gurkan *et al.* 1997). Eventually, this led to the

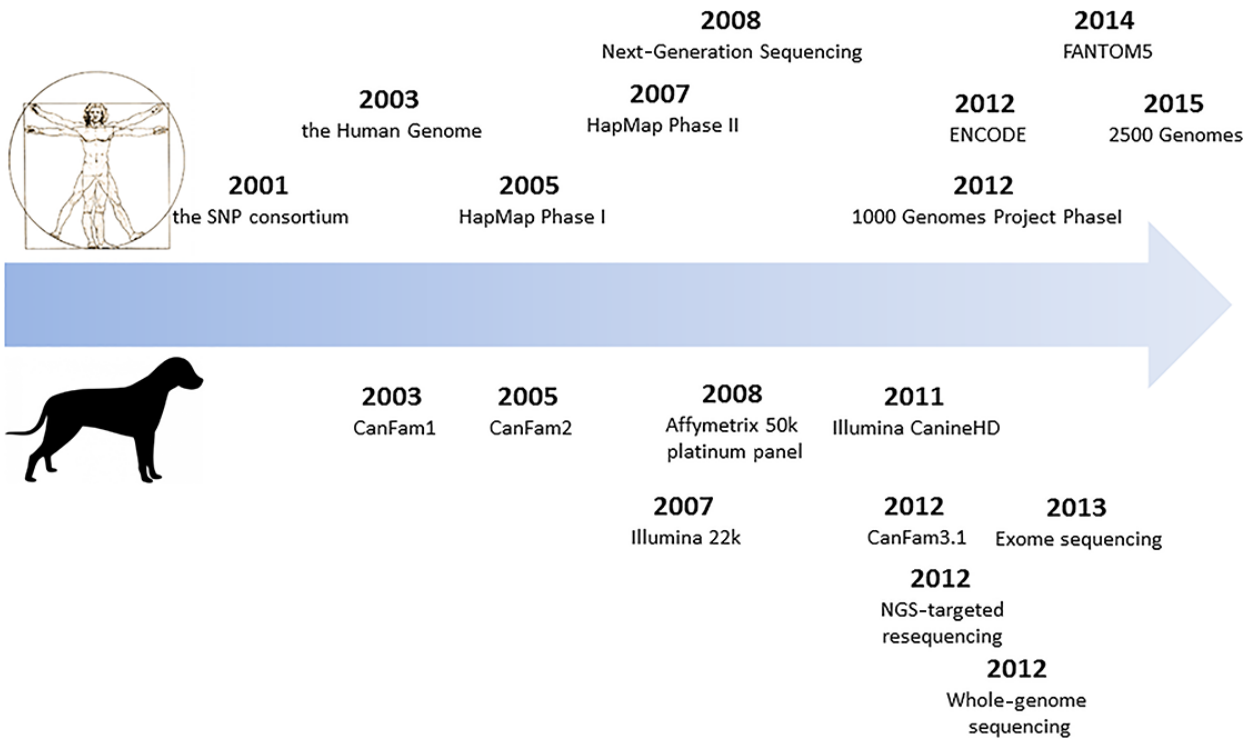

Figure 2 Developments in human and canine genetics; the different milestones of development of genetic tools in dog and human genetics.

discovery of the causative mutation in *COMMD1* (van De Sluis *et al.* 2002). Other microsatellite experiments were used to localise the first canine epilepsy gene (Lohi *et al.* 2005) and the dwarfism mutation discovery in German Shepherds (Voorbij *et al.* 2011). The first higher coverage canine DNA sequence became available in 2004. Sequencing of the DNA of the Boxer Tasha by the Broad Institute resulted in a 7.5× coverage genome, which was a significant step forward in canine genetic research (Lindblad-Toh *et al.* 2005). More than 30 million DNA sequence reads were generated in just 6 months at a cost of 30 million dollars and resulted in the dog being one of the first mammals to have its genome sequenced. Owing to the availability of the reference genome, the *IGF1* gene could be rapidly identified as a determinant of dog size (Sutter *et al.* 2007).

Prior to the boxer reference sequence, a low coverage poodle genome was already available (Kirkness *et al.* 2003). The annotations of the canine genomes and the subsequent identification of large numbers of SNPs have enabled the development of series of SNP arrays with an increased resolution in gene mapping. The first arrays contained only approximately 1500 SNPs, but provided faster experiments and comparable power to that of the existing panel of 256 microsatellite markers, as evidenced by the identification of white coat spotting in Boxers on chromosome 20 (Leegwater *et al.* 2007). The development of higher-density SNP arrays allowed genome-wide association analyses beyond family relationships and across breed populations. Again, breed-defining traits such as ridgeback, white spotting, chondrodysplasia and coat type were among the first to be mapped with these arrays (Karlsson *et al.* 2007; Cadieu *et al.* 2009; Parker *et al.* 2009). The further development of higher-resolution SNP assays (Ke *et al.* 2011) resulted in the successful application of the Affymetrix 50k platinum panel (Awano *et al.* 2009; Roque *et al.* 2011, 2012; Seppala *et al.* 2012) and the Illumina 22k array (Kyostila *et al.* 2012), both of which in turn were replaced by the Illumina CanineHD array (Vaysse *et al.* 2011) with 173k markers (eurolupa.eu).

These developments have significantly facilitated the gene discovery across traits and breeds, with high relevance for human medicine (Lequarre *et al.* 2011). A recent example is the identification of the *CCDC39* defect first in canine primary ciliary dyskinesia (PCD), leading to the discovery of many loss-of-function mutations in the human orthologue and PCD (Merveille *et al.* 2011). Another example, this time in Golden Retrievers, is the identification of the *PNPLA1* mutation in autosomal recessive congenital ichthyosis. This canine discovery led to the discovery of two causative mutations in the human orthologue in human ichthyosis (Grall *et al.* 2012).

Is there still a need for higher-resolution SNP arrays to facilitate disease gene mapping in dogs? The long linkage disequilibrium structure and the limited number of haplotypes within breeds allow for the efficient testing of dog genomes with much fewer markers in comparison with the human genome (Lindblad-Toh *et al.* 2005). The current

173k HD SNP array has 80 000–120 000 informative markers in each breed after quality control, which provides a reasonable coverage across genomes. However, the current design is not necessarily optimal for all breeds and will not tag all possible haplotypes for the most efficient study of complex disorders. Accordingly, a recent study by Hayward *et al.* (2016), addressing complex disorders in a cohort of over 4000 dogs with a tailored-made 180k canine SNP array, demonstrates the need for higher resolution as well as the need for increased sample size for each complex trait. The availability of an increasing number of whole-genome sequences across breeds provides the resources to redesign a higher-resolution SNP array for more optimal and powerful analyses.

Next generation disease model

Targeted resequencing

The advent of next generation sequencing (NGS) has clearly also benefited canine genetic research. Studies in small pedigrees often result in the identification of megabase-sized regions with many candidate genes to examine for causality. Mutation discovery can be greatly aided in these cases with NGS approaches. An illustrative example is the gene discovery in canine mucopolysaccharidosis type 7 (Hytonen *et al.* 2012). The original GWAS identified a 13-Mb shared region in the affected dogs with over 220 genes, many of which appeared to be functional candidates for the skeletal phenotype. By using targeted resequencing of the entire region in only two cases filtered against two unaffected dogs, the authors were able to identify a mutation in the functional domain of *GUSB* as a cause of the disease. Even in complex disorders, such as canine mast cell tumours, this approach has been proven successful. By genotyping cases from Golden Retrievers originating from different populations, distinct chromosomal regions associated with mast cell tumours could be identified (Arendt *et al.* 2015). Sequence capture of the different associated regions and the subsequent sequencing identified a SNP in *GNAI2*. In addition, several genes involved in hyaluronan synthesis have been implicated in the development of these tumours. Comparable sequencing approaches have been successfully applied in canine ophthalmologic disorders (Downs & Mellersh 2014), orthopaedic research (Hytonen *et al.* 2012; Lavrijsen *et al.* 2014a,b) and other conditions (Vaysse *et al.* 2011; Forman *et al.* 2012, 2013; Kim *et al.* 2012; Yokoyama *et al.* 2012; Tengvall *et al.* 2013; Agler *et al.* 2014; Massey *et al.* 2014).

Whole exome sequencing

The first published whole-exome experiment in dogs revealed the mutation for progressive renal atrophy in the Papillon and Phalène breeds (Ahonen *et al.* 2013). The study utilised a combined approach of genetic linkage analysis and exome sequencing to map the locus and to identify the causative variant in *CNGB1*. Another exome sequencing example suggested that variants in *NEB* could be causal for Basset Hounds suffering from primary angle closure glaucoma (Ahram *et al.* 2015). Exome sequencing in a trio of one affected individual and its healthy parents resulted in the identification of four non-synonymous variants within the identified chromosome 19 region of interest, one of which is postulated to be causative.

Exome designs have been based on the CanFam2.0 reference. The older reference versions have limited genome coverage and lack many coding, small RNA and regulatory regions that are included in the recent CanFam3.1-based designs (Broeckx *et al.* 2015). The new exome capture reagents provide powerful and cost-efficient resequencing tools for gene discovery in many Mendelian disorders that are likely to harbour coding mutations.

Whole-genome sequencing

The cost of whole-genome sequencing (WGS) continues to decrease and offers a comprehensive analysis of the entire DNA sequence. At the same time, WGS avoids technical enrichment biases that are present in capture reagents and methods in targeted analyses, including exome studies. The increasing amount of available WGS data provides powerful filtering strategies and facilitates mutation discovery in coding and non-coding regions. Several examples already exist. A breed-wide study investigating cranial development combined GWAS and WGS approaches followed by a functional validation in zebrafish (Schoenebeck *et al.* 2012). Genetic mapping in a cohort of 576 dogs from 72 breeds in four different categories of skull shape identified four different associated loci. Genome-wide sequencing was performed in 11 dog breeds and focused on the most promising, previously unexplored regions. The 190-kb region included two candidates, of which *BMP3* revealed a missense mutation. The BMP3 morpholino-injected zebrafish showed the loss of jaw structures and frontal bossing as well as cartilage abnormalities.

Other examples of studies with locus-mapping strategies prior to WGS include the identification of *ADAMTS20* as having a risk variant for cleft lip and palate (Wolf *et al.* 2015), the discovery of four risk loci involved in both B-cell lymphoma and hemangiosarcoma (Tonomura *et al.* 2015) and the categorisation of genomic alterations in canine transmissible venereal tumours by comparing these to 186 canine whole-genome sequences (Decker *et al.* 2015). In the case of the discovery of the frameshift mutations involved in Imerslund–Grasbeck syndrome in Beagles and Border Collies (Drogemuller *et al.* 2014), the genes *CUBN* and *AMN* were known from human cases. Together, these genes

represent 80 exons. Performing genome-wide sequencing of single cases was cheaper and faster than was classical DNA sequencing of all exons one by one. Therefore, genome-wide sequencing is also considered to be a powerful and cost- and time-efficient tool if strong candidate genes are known (Drogemuller *et al.* 2014; Guo *et al.* 2014; Willet *et al.* 2015).

One can easily foresee that eventually DNA sequence analysis will make SNP arrays redundant. Genotyping by sequencing may potentially provide a single method in the future for gene mapping and sequencing purposes (Cooke *et al.* 2016). The first example of the combination of mapping and mutation discovery by WGS was used to provide insight into signatures of selection that are associated with high altitude. The strongest Z-transformed F_{ST} values when comparing breeds living at high, middle and low altitudes were found within a region comprising *EPAS1*, carrying a most likely causal mutation in its PAS domain (Liu *et al.* 2014). In order to obtain an insight into extensive genomic alterations in canine spontaneous mammary cancer, a series of mammary carcinomas (seven simple and four complex) have been sequenced at whole-genome, whole-exome and whole-transcriptome resolutions. Based on this elegant combination of three high-throughput screenings, it was postulated that simple carcinomas originate from genomic aberrations and complex carcinomas are caused by epigenetic changes. Combining multiple NGS strategies strongly enhances the comprehension of the underlying biology of disorders.

Future perspectives

1000 Canomes (exomes and genomes)

Ongoing disease-related whole-exome sequencing and WGS efforts are rapidly progressing in the canine genetics community. These efforts will yield sequences from hundreds of dogs across breeds and wolf populations to reveal millions of common and rare variants. It is important to further encourage researchers to share their data among groups to aid and empower individual studies. A corresponding successful collaborative example includes the 1000 bull genome project that facilitates monogenic and complex disease mapping in bovine (Daetwyler *et al.* 2014). The ultimate goal in the canine community should be the establishment of a public database that includes all genome sequences and variants together with the sufficient phenotypic information from the sequenced dogs to allow the efficient use of data. Fortunately, such an international effort involving 10 000 dog genomes has been recently initiated (http://www.dog10kgenomes.org/), and individual research groups have already started to share genome data on a smaller scale to support each other's mapping efforts. The publicly available human 1000 genomes data (1000 Genomes Project Consortium *et al.* 2010) and more recent 2500 genomes data (Sudmant *et al.* 2015) as well as the Genome of the Netherlands Project (1000 Genomes Project Consortium *et al.* 2010; Genome of the Netherlands Consortium 2014) have been instrumental in characterising geographic, structural and functional spectrums of genetic variation that contributes to human diseases. 'The 1000 Canomes' would serve the same purpose in dogs, significantly expediting gene discovery and comparative studies across species.

Functional annotation of the canine genome

The sequence data itself may have limited use without a comprehensive functional annotation of the dog genome. Several improvements have been accomplished in the CanFam2.0 reference genome containing several sequencing errors and gaps (Hoeppner *et al.* 2014). Combined BAC and RNA sequencing efforts from 10 different canine tissues resulted in the elimination of most of the gaps and annotation of novel protein-coding genes and additional isoforms per gene in the current CanFam3.1 reference. However, although the current reference is useful for many studies, it still suffers from gaps and incomplete and inaccurate annotation of many genes, particularly in the 5′ end and regulatory regions of the genes. The use of advanced long read sequencing technologies, such as the PacBio long sequencing read approach, would help to fill the remaining gaps, as recently demonstrated for a canine epilepsy locus (Koskinen *et al.* 2015). The current short read WGS technology is not optimal for discovering structural variants in genomes and could also benefit from long read WGS approaches. Although comparative genomics between canine and human or mouse genomes may help to specify candidate functional variation in the regulatory regions, more comprehensive annotation of the entire genome would aid genetic and epigenetic studies in dogs, particularly in complex diseases such as cancer and behavioural traits. Ongoing efforts are expected to provide the next generation of improvements for the reference, further facilitating the use of dogs as comparative models for many traits.

Over the past decade, canine genetic research has focused largely on coding region variations (Merveille *et al.* 2011; Grall *et al.* 2012). Whether this is a useful strategy depends on the type of the disorder being investigated. The success rate of focusing on only coding variants is higher in monogenic Mendelian disorders than in complex disorders. In complex disorders, causative variants have been reported frequently in introns, influencing exon skipping (Lee *et al.* 2012). Furthermore, intergenic variants have been found to regulate tissue-specific expression (Grundberg *et al.* 2012). The first examples of non-coding variants and their role in complex disorders in dogs have been published, confirming the power of the canine model (Karlsson *et al.* 2013; Tang *et al.* 2014). The fact that roughly 5% of the

human genome is evolutionarily conserved when compared with rodents implies that these regions are likely functionally important. Based on the finding that only one-third of these genomic regions code for proteins, it has been hypothesised that intergenic regions do not consist merely of so-called junk DNA, but do indeed harbour functional elements.

Based on this theory, The Encyclopedia of DNA Elements (ENCODE) Consortium had its kickoff in 2004 (ENCODE Project Consortium 2004) to discover regulatory elements. Numerous datasets covering many different cell types were used for the search of functional DNA elements, resulting in a series of fundamental findings (Pennisi 2012; Skipper *et al.* 2012; Stamatoyannopoulos 2012). The consortium managed to systematically map regions of transcription, transcription factor associations, chromatin structures and histone modifications on the human genome. These findings opened up a discussion as to which parts of the genome are truly functional. According to the consortium, approximately 80% of the genome was found be functional. Rands *et al.* (2014) questioned these data and postulated that only 8.2% is subject to negative selection and therefore functional. The ENCODE data describe protein–DNA interactions and define transcribed bases as functional. Proteins in nucleosomes that pack DNA, transcription binding factors and other gene expression regulating elements are included in these data, and this is why the estimate of the ENCODE Consortium for the fraction of functional DNA reached 80%.

Another major effort to identify functional areas of the genome was the Functional Annotation of the Mammalian Genome 5 (FANTOM5) Project, which provided comprehensive expression profiles, transcription start sites, enhancer maps and functional annotation of mammalian cell-type-specific transcriptomes with wide applications in biomedical research (FANTOM Consortium and the RIKEN PMI and CLST (DGT) *et al.* 2014). Similar ENCODE and FANTOM approaches to functional annotation of the canine genome would significantly facilitate the identification of causal variants across the genome to fully utilise the dog model. Extreme morphological and behavioural variation as well as the high number of complex diseases in dogs suggests a strong involvement of regulatory variation contributing to traits. Solving complex disorders could benefit greatly from such a CANCODE approach.

References

1000 Genomes Project Consortium, Abecasis G.R., Altshuler D., Auton A., Brooks L.D., Durbin R.M., Gibbs R.A., Hurles M.E. & McVean G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73.

Agler C., Nielsen D.M., Urkasemsin G. *et al.* (2014) Canine hereditary ataxia in Old English Sheepdogs and Gordon Setters is associated with a defect in the autophagy gene encoding RAB24. *PLoS Genetics* **10**, e1003991.

Ahonen S.J., Arumilli M, Lohi H *et al.* (2013) A *CNGB1* frameshift mutation in Papillon and Phalene dogs with progressive retinal atrophy. *PLoS One* **8**, e72122.

Ahram D.F., Grozdanic S.D., Kecova H., Henkes A., Collin R.W. & Kuehn M.H. (2015) Variants in *nebulin* (*NEB*) are linked to the development of familial primary angle closure glaucoma in Basset Hounds. *PLoS One* **10**, e0126660.

Arendt M.L., Melin M., Tonomura N. *et al.* (2015) Genome-wide association study of Golden Retrievers identifies germ-line risk factors predisposing to mast cell tumours. *PLoS Genetics* **11**, e1005647.

Awano T., Johnson G.S., Wade C.M. *et al.* (2009) Genome-wide association analysis reveals a *SOD1* mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2794–9.

Axelsson E., Ratnakumar A., Arendt M.L., Maqbool K., Webster M.T., Perloski M., Liberg O., Arnemo J.M., Hedhammar A. & Lindblad-Toh K. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–4.

Bannasch D., Young A., Myers J. *et al.* (2010) Localization of canine brachycephaly using an across breed mapping approach. *PLoS One* **5**, e9632.

Baranowska Korberg I., Sundstrom E., Meadows J.R. *et al.* (2014) A simple repeat polymorphism in the *MITF*-M promoter is a key regulator of white spotting in dogs. *PLoS One* **9**, e104363.

Boyko A.R. (2011) The domestic dog: man's best friend in the genomic era. *Genome Biology* **12**, 216.

Broeckx B.J., Hitte C., Coopman F. *et al.* (2015) Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Scientific Reports* **5**, 12810.

Cadieu 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.

Cerda-Gonzalez S., Olby N.J., Broadstone R., McCullough S. & Osborne J.A. (2009) Characteristics of cerebrospinal fluid flow in Cavalier King Charles Spaniels analyzed using phase velocity cine magnetic resonance imaging. *Veterinary Radiology & Ultrasound* **50**, 467–76.

Clark L.A., Tsai K.L., Steiner J.M., Williams D.A., Guerra T., Ostrander E.A., Galibert F. & Murphy K.E. (2004) Chromosome-specific microsatellite multiplex sets for linkage studies in the domestic dog. *Genomics* **84**, 550–4.

Cooke T.F., Yee M.C., Muzzio M. *et al.* (2016) GBStools: A statistical method for estimating allelic dropout in reduced representation sequencing data. *PLoS Genetics* **12**, e1005631.

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.

van de Sluis B., Rothuizen J., Pearson P.L., van Oost B.A. & Wijmenga C. (2002) Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Human Molecular Genetics* **11**, 165–73.

Decker B., Davis B.W., Rimbault M. *et al.* (2015) Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Research* **25**, 1646–55.

Downs L.M. & Mellersh C.S. (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**, e93990.

Dreger D.L., Parker H.G., Ostrander E.A. & Schmutz S.M. (2013) Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in Basset Hounds and Pembroke Welsh Corgis. *Journal of Heredity* **104**, 399–406.

Drogemuller C., Karlsson E.K., Hytonen M.K., Perloski M., Dolf G., Sainio K., Lohi H., Lindblad-Toh K. & Leeb T. (2008) A mutation in hairless dogs implicates *FOXI3* in ectodermal development. *Science* **321**, 1462.

Drogemuller M., Jagannathan V., Howard J., Bruggmann R., Drogemuller C., Ruetten M., Leeb T. & Kook P.H. (2014) A frameshift mutation in the *cubilin* gene (*CUBN*) in Beagles with Imerslund-Grasbeck syndrome (selective cobalamin malabsorption). *Animal Genetics* **45**, 148–50.

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–40.

Everts R.E., Rothuizen J. & van Oost B.A. (2000) Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (*MC1R*) in Labrador and Golden Retrievers with yellow coat colour. *Animal Genetics* **31**, 194–9.

FANTOM Consortium and RIKEN PMI and CLST (DGT) *et al.* (2014) A promoter-level mammalian expression atlas. *Nature* **507**, 462–70.

Forman O.P., Penderis J., Hartley C., Hayward L.J., Ricketts S.L. & Mellersh C.S. (2012) Parallel mapping and simultaneous sequencing reveals deletions in *BCAN* and *FAM83H* associated with discrete inherited disorders in a domestic dog breed. *PLoS Genetics* **8**, e1002462.

Forman O.P., De Risio L. & Mellersh C.S. (2013) Missense mutation in *CAPN1* is associated with spinocerebellar ataxia in the Parson Russell Terrier dog breed. *PLoS One* **8**, e64627.

Freedman A.H., Gronau I., Schweizer R.M. *et al.* (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics* **10**, e1004016.

French FMF Consortium (1997) A candidate gene for familial Mediterranean fever. *Nature Genetics* **17**, 25–31.

Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics* **46**, 818–25.

Grall A., Guaguere E., Planchais S. *et al.* (2012) *PNPLA1* mutations cause autosomal recessive congenital ichthyosis in Golden Retriever dogs and humans. *Nature Genetics* **44**, 140–7.

Grundberg E., Small K.S., Hedman A.K. *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* **44**, 1084–9.

Guo J., Johnson G.S., Brown H.A., Provencher M.L., da Costa R.C., Mhlanga-Mutangadura T., Taylor J.F., Schnabel R.D., O'Brien D.P. & Katz M.L. (2014) A *CLN8* nonsense mutation in the whole genome sequence of a mixed breed dog with neuronal ceroid lipofuscinosis and Australian Shepherd ancestry. *Molecular Genetics and Metabolism* **112**, 302–9.

Hayward J.J., Castelhano M.G., Oliveira K.C. *et al.* (2016) Complex disease and phenotype mapping in the domestic dog. *Nature Communications* **7**, 10460.

Hoeppner M.P., Lundquist A., Pirun M. *et al.* (2014) An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS One* **9**, e91172.

vonHoldt B.M., Pollinger J.P., Earl D.A *et al.* (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research* **21**, 1294–305.

Hytonen M.K., Arumilli M., Lappalainen A.K., Kallio H., Snellman M., Sainio K. & Lohi H. (2012) A novel *GUSB* mutation in Brazilian Terriers with severe skeletal abnormalities defines the disease as mucopolysaccharidosis VII. *PLoS One* **7**, e40281.

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.

Karlsson E.K., Sigurdsson S., Ivansson E. *et al.* (2013) Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near *CDKN2A/B*. *Genome Biology* **14**, R132.

Ke X., Kennedy L.J., Short A.D. *et al.* (2011) Assessment of the functionality of genome-wide canine SNP arrays and implications for canine disease association studies. *Animal Genetics* **42**, 181–90.

Kim R.N., Kim D.S., Choi S.H. *et al.* (2012) Genome analysis of the domestic dog (Korean Jindo) by massively parallel sequencing. *DNA Research* **19**, 275–87.

Kirkness E.F., Bafna V., Halpern A.L. *et al.* (2003) The dog genome: survey sequencing and comparative analysis. *Science* **301**, 1898–903.

Koskinen L.L., Seppala E.H., Belanger J.M. *et al.* (2015) Identification of a common risk haplotype for canine idiopathic epilepsy in the *ADAM23* gene. *BMC Genomics* **16**, 465.

Kyostila K., Cizinauskas S., Seppala E.H., Suhonen E., Jeserevics J., Sukura A., Syrja P. & Lohi H. (2012) A *SEL1L* mutation links a canine progressive early-onset cerebellar ataxia to the endoplasmic reticulum-associated protein degradation (ERAD) machinery. *PLoS Genetics* **8**, e1002759.

Lavrijsen I.C., Leegwater P.A., Martin A.J., Harris S.J., Tryfonidou M.A., Heuven H.C. & Hazewinkel H.A. (2014a) Genome wide analysis indicates genes for basement membrane and cartilage matrix proteins as candidates for hip dysplasia in Labrador Retrievers. *PLoS One* **9**, e87735.

Lavrijsen I.C., Leegwater P.A., Wangdee C. *et al.* (2014b) Genome-wide survey indicates involvement of loci on canine chromosomes 7 and 31 in patellar luxation in Flat-coated Retrievers. *BMC Genetics* **15**, 64.

Lee Y., Gamazon E.R., Rebman E., Lee Y., Lee S., Dolan M.E., Cox N.J. & Lussier Y.A. (2012) Variants affecting exon skipping contribute to complex traits. *PLoS Genetics* **8**, e1002998.

Leegwater P.A., van Hagen M.A. & van Oost B.A. (2007) Localization of white spotting locus in Boxer dogs on CFA20 by genome-wide linkage analysis with 1500 SNPs. *Journal of Heredity* **98**, 549–52.

Lequarre A.S., Andersson L., Andre C., Fredholm M., Hitte C., Leeb T., Lohi H., Lindblad-Toh K. & Georges M. (2011) LUPA: a European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *Veterinary Journal* **189**, 155–9.

Lin L., Faraco J., Li R., Kadotani H., Rogers W., Lin X., Qiu X., de Jong P.J., Nishino S. & Mignot E. (1999) The sleep disorder canine narcolepsy is caused by a mutation in the *hypocretin (orexin) receptor 2* gene. *Cell*, **98**, 365–76.

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.

Liu D., Xiong H., Ellis A.E., Northrup N.C., Rodriguez C.O. Jr, O'Regan R.M., Dalton S. & Zhao S. (2014) Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. *Cancer Research* **74**, 5045–56.

Lohi H., Young E.J., Fitzmaurice S.N. *et al.* (2005) Expanded repeat in canine epilepsy. *Science* **307**, 81.

Massey J., Short A.D., Catchpole B., House A., Day M.J., Lohi H., Ollier W.E. & Kennedy L.J. (2014) Genetics of canine anal furunculosis in the German Shepherd dog. *Immunogenetics* **66**, 311–24.

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.

Motohashi H., Hozawa K., Oshima T., Takeuchi T. & Takasaka T. (1994) Dysgenesis of melanocytes and cochlear dysfunction in mutant microphthalmia (mi) mice. *Hearing Research* **80**, 10–20.

Newton J.M., Wilkie A.L., He L., Jordan S.A., Metallinos D.L., Holmes N.G., Jackson I.J. & Barsh G.S. (2000) *Melanocortin 1 receptor* variation in the domestic dog. *Mammalian Genome* **11**, 24–30.

Olsson M., Meadows J.R., Truve K. *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 Genetics* **7**, e1001332.

Ostrander E.A. & Kruglyak L. (2000) Unleashing the canine genome. *Genome Research* **10**, 1271–4.

Packer R.M., Hendricks A., Tivers M.S. & Burn C.C. (2015) Impact of facial conformation on canine health: brachycephalic obstructive airway syndrome. *PLoS One* **10**, e0137496.

Pang J.F., Kluetsch C., Zou X.J. *et al.* (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Molecular Biology and Evolution* **26**, 2849–64.

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.

Parker J.E., Knowler S.P., Rusbridge C., Noorman E. & Jeffery N.D. (2011) Prevalence of asymptomatic syringomyelia in Cavalier King Charles Spaniels. *Veterinary Record* **168**, 667.

Pennisi E. (2012) Genomics. ENCODE project writes eulogy for junk DNA. *Science*, **337**, 1159–61.

Rands C.M., Meader S., Ponting C.P. & Lunter G. (2014) 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genetics* **10**, e1004525.

Roque J.B., O'Leary C.A., Duffy D.L., Kyaw-Tanner M., Latter M., Mason K., Vogelnest L. & Shipstone M. (2011) IgE responsiveness to Dermatophagoides farinae in West Highland White Terrier dogs is associated with region on CFA35. *Journal of Heredity* **102** (Suppl 1), S74–80.

Roque J.B., O'Leary C.A., Duffy D.L., Kyaw-Tanner M., Gharahkhani P., Vogelnest L., Mason K., Shipstone M. & Latter M. (2012) Atopic dermatitis in West Highland White Terriers is associated with a 1.3-Mb region on CFA 17. *Immunogenetics*, **64**, 209–17.

Salmon Hillbertz N.H., Isaksson M., Karlsson E.K. *et al.* (2007) Duplication of *FGF3, FGF4, FGF19* and *ORAOV1* causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nature Genetics* **39**, 1318–20.

Sargan D.R., Withers D., Pettitt L., Squire M., Gould D.J. & Mellersh C.S. (2007) Mapping the mutation causing lens luxation in several terrier breeds. *Journal of Heredity* **98**, 534–8.

Savolainen P., Zhang Y.P., Luo J., Lundeberg J. & Leitner T. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610–3.

Schmutz S.M., Berryere T.G. & Goldfinch A.D. (2002) *TYRP1* and *MC1R* genotypes and their effects on coat color in dogs. *Mammalian Genome* **13**, 380–7.

Schoenebeck J.J., Hutchinson S.A., Byers A. *et al.* (2012) Variation of *BMP3* contributes to dog breed skull diversity. *PLoS Genetics* **8**, e1002849.

Seppala E.H., Koskinen L.L., Gullov C.H. *et al.* (2012) Identification of a novel idiopathic epilepsy locus in Belgian Shepherd dogs. *PLoS One* **7**, e33549.

Shannon L.M., Boyko R.H., Castelhano M. *et al.* (2015) Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13639–44.

Skipper M., Dhand R. & Campbell P. (2012) Presenting ENCODE. *Nature* **489**, 45.

Stamatoyannopoulos J.A. (2012) What does our genome encode? *Genome Research* **22**, 1602–11.

Stritzel S., Wohlke A. & Distl O. (2009) A role of the microphthalmia-associated transcription factor in congenital sensorineural deafness and eye pigmentation in Dalmatian dogs. *Journal of Animal Breeding and Genetics* **126**, 59–62.

Sudmant P.H., Rausch T., Gardner E.J. *et al.* (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81.

Sutter N.B. & Ostrander E.A. (2004) Dog star rising: the canine genetic system. *Nature Reviews Genetics* **5**, 900–10.

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 R., Noh H.J., Wang D. *et al.* (2014) Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biology* **15**, R25.

Tengvall K., Kierczak M., Bergvall K. *et al.* (2013) Genome-wide analysis in German Shepherd dogs reveals association of a locus on CFA 27 with atopic dermatitis. *PLoS Genetics* **9**, e1003475.

Thalmann O., Shapiro B., Cui P *et al.* (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–4.

Tietz W. (1963) A syndrome of deaf-mutism associated with albinism showing dominant autosomal inheritance. *American Journal of Human Genetics* **15**, 259–64.

Tonomura N., Elvers I., Thomas R *et al.* (2015) Genome-wide association study identifies shared risk loci common to two malignancies in Golden Retrievers. *PLoS Genetics* **11**, e1004922.

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.

Voorbij A.M., van Steenbeek F.G., Vos-Loohuis M., Martens E.E., Hanson-Nilsson J.M., van Oost B.A., Kooistra H.S. & Leegwater P.A. (2011) A contracted DNA repeat in *LHX3* intron 5 is associated with aberrant splicing and pituitary dwarfism in German Shepherd dogs. *PLoS One* **6**, e27940.

Wang G.D., Zhai W., Yang H.C *et al.* (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nature Communications* **4**, 1860.

Wang G.D., Zhai W., Yang H.C. *et al.* (2016) Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Research* **26**, 21–33.

Wayne R.K. & Ostrander E.A. (2007) Lessons learned from the dog genome. *Trends in Genetics* **23**, 557–67.

Wilbe M., Jokinen P., Truve K. *et al.* (2010) Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nature Genetics* **42**, 250–4.

Willet C.E., Makara M., Reppas G., Tsoukalas G., Malik R., Haase B. & Wade C.M. (2015) Canine disorder mirrors human disease: exonic deletion in *HES7* causes autosomal recessive spondylocostal dysostosis in miniature Schnauzer dogs. *PLoS One* **10**, e0117055.

Wolf Z.T., Brand H.A., Shaffer J.R. *et al.* (2015) Genome-wide association studies in dogs and humans identify *ADAMTS20* as a risk variant for cleft lip and palate. *PLoS Genetics* **11**, e1005059.

Yokoyama J.S., Lam E.T., Ruhe A.L. *et al.* (2012) Variation in genes related to cochlear biology is strongly associated with adult-onset deafness in border collies. *PLoS Genetics* **8**, e1002898.

Yoshida H., Kunisada T., Kusakabe M., Nishikawa S. & Nishikawa S.I. (1996) Distinct stages of melanocyte differentiation revealed by analysis of nonuniform pigmentation patterns. *Development* **122**, 1207–14.

Yuzbasiyan-Gurkan V., Blanton S.H., Cao Y., Ferguson P., Li J., Venta P.J. & Brewer G.J. (1997) Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington Terriers. *American Journal of Veterinary Research* **58**, 23–7.

Zanna G., Fondevila D., Bardagi M., Docampo M.J., Bassols A. & Ferrer L. (2008) Cutaneous mucinosis in Shar-Pei dogs is due to hyaluronic acid deposition and is associated with high levels of hyaluronic acid in serum. *Veterinary Dermatology* **19**, 314–8.