Leader of the pack: gene mapping in dogs and other model organisms

Elinor K. Karlsson[*‡] *and Kerstin Lindblad-Toh*[*§]

Abstract | The domestic dog offers a unique opportunity to explore the genetic basis of disease, morphology and behaviour. We share many diseases with our canine companions, including cancer, diabetes and epilepsy, making the dog an ideal model organism for comparative disease genetics. Using newly developed resources, whole-genome association in dog breeds is proving to be exceptionally powerful. Here, we review the different trait-mapping strategies, some key biological findings emerging from recent studies and the implications for human health. We also discuss the development of similar resources for other vertebrate organisms.

Population bottleneck
A marked reduction in population size followed by the survival and expansion of a small random sample of the original population.

Linkage disequilibrium
(LD). Non-random association of alleles at two or more loci.

Haplotype block
A haplotype is the combination of alleles observed for one or more consecutive markers on a chromosome. A haplotype block is the region of a chromosome that contains no recombination.

*Broad Institute of Harvard and Massachusetts Institute of Technology, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.
‡FAS Center for Systems Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA.
§Department of Medical Biochemistry and Microbiology, Uppsala University, BOX 597, 751 24 Uppsala, Sweden.
Correspondence to K.L.-T.
e-mail: kersli@broad.mit.edu
doi:10.1038/nrg2382

Understanding the genetic basis of human disease is one of the greatest challenges of biomedical research[1]. Using an appropriate model organism avoids many difficulties inherent to human-based studies. Since the beginning of the twentieth century, the foremost model for genetic studies in mammals has been the mouse, which boasts an impressive array of experimental resources[2–5]. As a model for complex human disease, however, the mouse has significant limitations. Many diseases that occur spontaneously in humans must be induced in laboratory mice. Whereas complex human disease is polygenic, the genetic manipulations in mice test the effect of a single major gene. Conditional and partial knockouts provide some flexibility, but studying the interaction between multiple genes remains difficult. The mouse is ideal for studying a putative mutation, once it has been found, but it is less useful for finding those mutations to start with.

When considering other model organisms, the domestic dog immediately springs to mind as an ideal model for mapping human complex diseases, as it has both a similar spectrum of diseases and an advantageous population structure[6] (BOX 1). The modern dog is the most physically diverse domesticated species[7], encompassing over 400 breeds[8]. Each of these breeds is defined by specific behavioural and physical characteristics that have been driven to exceptionally high frequency by population bottlenecks and strong artificial selection[9]. This process had unintended consequences on the health of pure-bred dogs, with high rates of specific diseases in certain breeds reflecting enrichment of risk alleles owing to random fixation during bottlenecks, hitch-hiking of mutations near desirable traits and pleiotropic effects of selected variants[10,11]. Most breeds are less than 200 years old, and thus have long linkage disequilibrium (LD) and long haplotype blocks, making them particularly amenable to genome-wide association (GWA) mapping with fewer markers and fewer individuals compared with humans. However, the domestic dog population is ancient, dating back approximately 15,000 years, and thus has short LD and short haplotype blocks. Examining an associated locus in several breeds carrying the same mutation can identify a small region that is comparable in size to those found in human studies.

Geneticists have long recognized the exceptional characteristics of the domesticated dog and by the 1990s had begun mapping traits. Originally, large multigenerational canine pedigrees were used to focus on highly penetrant Mendelian phenotypes. Using simple sequence length polymorphism linkage mapping[12], canine geneticists found loci for progressive retinal atrophy[13,14], copper toxicosis[15,16], renal cancer[17,18] and narcolepsy[19,20]. Traits with complex inheritance, however, remained out of reach.

When developing an experimental organism, it is important to carefully design strategies and tools on the basis of the traits of interest in the model in question, as well as its population history and genome structure. Thus, the Dog Genome Sequencing Project, completed in 2005 (REF. 21), focused not just on sequencing the genome but also on generating the resources needed for association mapping, which include an understanding of canine haplotype structure and a dense SNP map. Currently, the tools for high-throughput SNP genotyping of the dog have been developed and several genes have already been mapped, including microphthalmia-associated transcription factor (*MITF*)

Genome-wide association
(GWA). An approach that tests the whole genome for a statistical association between a marker and a trait in unrelated cases and controls.

Simple sequence length polymorphism
Short tandem repeats of DNA that vary in length.

Linkage mapping
A mapping method which uses pedigrees to find broad genomic regions (10–20 centimorgans) that adhere to an inheritance model proposed for the trait of interest.

for white coat colour and a duplicated cluster of fibroblast growth factor (FGF) genes underlying a defect in spinal development[22,23].

Given that humans and dogs share many diseases, finding the causative loci in dogs can identify genes and pathways relevant to the disease in humans, potentially offering novel treatment targets. In addition, dissecting the network of interacting loci that cause diseases with complex, polygenic inheritance in dogs, and defining their severity, will give insight into the considerably more complex web underlying human diseases.

In this article, we review the steps necessary to take full advantage of the domestic dog for both quantitative and medical trait mapping, and the implications of the results for human biology and health. We also discuss the process of generating the appropriate genetic and genomic tools for the dog and how this translates to other vertebrate organisms with different biology and population history.

Developing the tools for trait mapping

Dog breeds have fascinated geneticists since at least the beginning of the 1900s, when the first inherited disorders in dogs were described[24]. Between 1910 and 1950, the *Journal of Heredity* alone published numerous articles on dogs, covering topics as diverse as coat colour, hairlessness, barking and haemophilia[25–28]. Following advances in genetics, canine geneticists incorporated cytogenetic and chromosomal techniques into this search for the cause of dog phenotypes. In 1989, DNA sequencing identified the first genetic variant for an inherited canine disorder — a point mutation underlying haemophilia B[24,29]. In 1997, linkage mapping using microsatellite markers identified a marker for a canine disease, copper toxicosis, in Bedlington terriers[15]. By the end of that year, the first genetic linkage map of the dog genome, containing 150 microsatellite markers, was complete[30]. The identification of a gene for canine narcolepsy in 1999 intensified interest in the dog as a model for human disease[20]. The following years saw several larger-scale projects reach completion, following in the footsteps of the Human Genome Project in resource building[31–34]. Finally, with the Dog Genome Sequencing Project (described in detail below), scientists were in a position to carefully consider what state-of-the-art tools should be developed to map traits in the dog.

Today, a robust set of canine genetics resources exist that facilitate the search for genetic variants underlying diseases and other phenotypes at all stages. These include: defining phenotypes and selecting sample sets; identifying trait loci through linkage, association or selection mapping; identifying gene function and genomic variants; and testing likely variants for function.

Box 1 | The dog as a model for human disease

The similarities between canine and human diseases are striking and have been thoroughly reviewed[70]. Here we summarize a few key features in which dogs excel, including similarity in disease manifestation, genome content, genetic predisposition, treatment and clinical trials.

In dogs, just as in humans, diseases occur spontaneously over the course of their lives and include many common human diseases, such as cancers, diabetes, heart disease, eye diseases, epilepsy, deafness and even psychiatric diseases such as obsessive compulsive disorder[71–75]. Clinical manifestations in dogs and humans are often similar. For example, the progression of osteosarcoma is nearly identical in dogs and humans, with the primary tumour occurring in the metaphyseal region of the long bones and metastasizing to the lungs, with death usually resulting from diffuse pulmonary metastasis[76].

The dog genome is less diverged from the human than the mouse genome, with an average nucleotide divergence of ~0.35 substitutions per site (compared to ~0.51 between humans and mice). The genome is also relatively compact (2.4 Gb), with less overall repeat insertion and segmental duplication than many mammals. Finally, dogs have approximately the same number of genes as humans, most of which are 1:1 orthologues[21].

In humans, family history is one of the strongest risk factors for nearly all diseases[77]. Likewise, in dogs, the exceptionally high prevalence of particular diseases in some breeds suggests a strong heritable component. For example, haemangiosarcomas affect ~15% of golden retrievers in the United States[78], whereas 5% of miniature wire-haired dachshunds in the United Kingdom suffer from epilepsy[52]. These substantially increased risks in breeds, which are populations with little genetic diversity that have arisen over a short period of time, suggest that just a few loci are involved, each with strong effect — making genetic dissection potentially more tractable in dogs than in humans.

Dogs routinely receive medical treatment for many common diseases, including cancer and epilepsy. With a lifespan that is approximately seven times shorter than humans, diseases of old age, such as cancer, manifest earlier and typically run their course in a few years. As treatment also proceeds more rapidly, a clinical trial that might take 5–15 years in humans takes only 1–3 years in dogs[79,80]. Thus, dogs can provide a useful testing ground for novel therapies. In one example, gene therapy that has successfully treated blindness in dogs might in time benefit human patients[81–83].

A genome sequence and SNP map. When the canine genetics community and the Broad Institute of Harvard and Massachusetts Institute of Technology proposed sequencing the dog genome, they considered carefully how to best complement existing resources and make full use of the dog and its advantageous population history (FIG. 1). Thus, the Dog Genome Sequencing Project[21] included meticulously designed experiments that explored the haplotype structure of the dog genome both in breeds and in the whole population. In addition, although the genome sequence represents a single pure-bred dog, additional sequencing of nine additional breeds and five canids produced data for a dense SNP map across the genome. Both elements proved crucial to successful GWA studies in dogs.

The current dog genome has 7.5 times genome coverage and includes ~99% of the euchromatic genome of a female boxer dog (using a female provides full coverage of the X chromosome)[21]. She was chosen on the basis of tests showing low rates of heterozygosity, which simplifies genome assembly and improves overall genome sequence accuracy. The euchromatic portion of the dog genome contains ~2.4 billion bases, spread across 38 autosomal chromosomes and chromosome X. The genome assembly has exceptional continuity. As the N50 contig size is 180 kb, most genes contain no sequence gaps, and the sequence for most chromosomes is ordered and oriented on just one or two supercontigs (the N50 supercontig size is 45 Mb).

A dense SNP map, containing 2.5 million SNPs (on average 1 SNP per 1,000 bases), was constructed by comparing this high-quality genome for a boxer dog with the partial sequence of a standard poodle[34] (genome

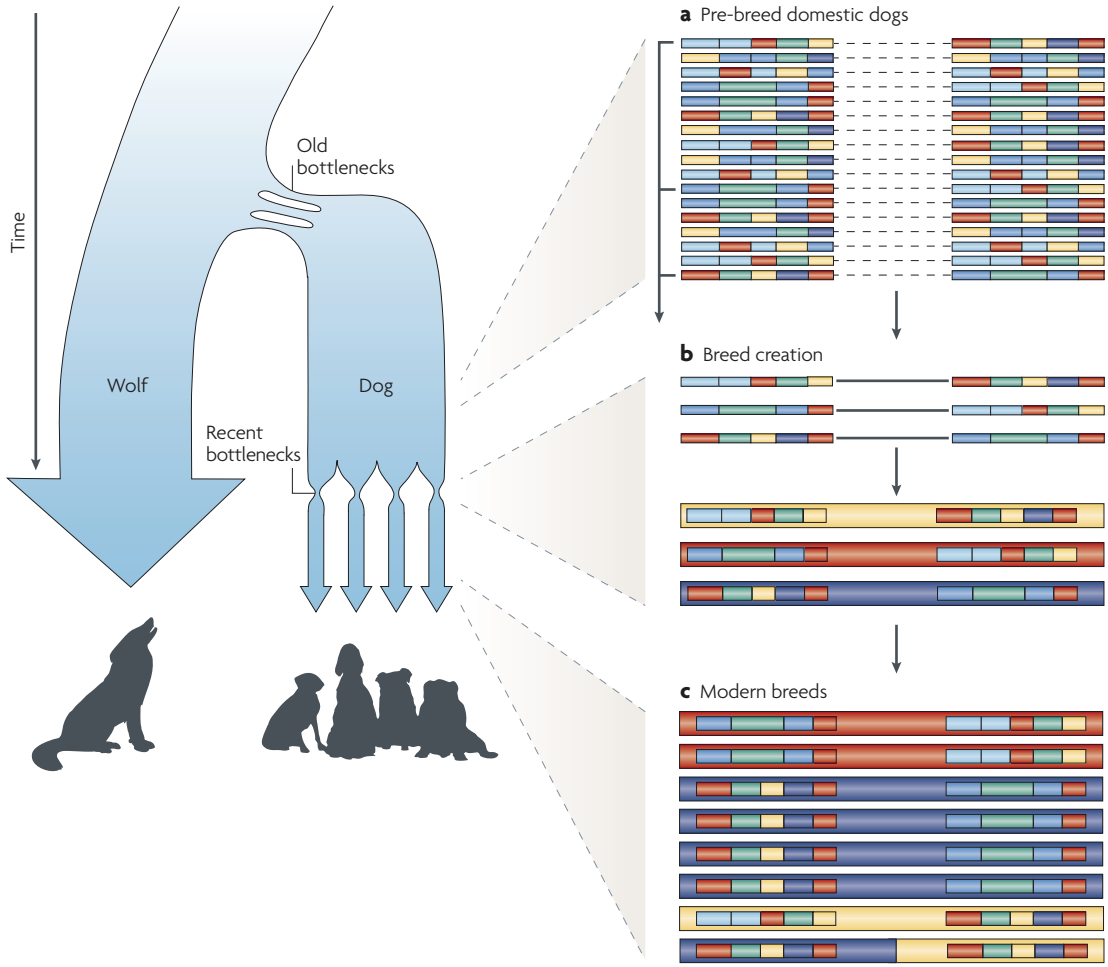


Figure 1 | **Haplotype structure of the dog.** Two population bottlenecks in dog population history, one old and one recent, shaped haplotype structure in modern dog breeds. First, the domestic dog diverged from wolves ~15,000 years ago, probably through multiple domestication events. Within the past few hundred years, modern dog breeds were created. Both bottlenecks influenced the haplotype pattern and linkage disequilibrium (LD) of current breeds. **a** | Before the creation of modern breeds, the dog population had the short-range LD that would be expected given its large size and the long time period since the domestication bottleneck. **b** | In the creation of modern breeds, a small subset of chromosomes was selected from the pool of domestic dogs. The long-range patterns that were carried on these chromosomes became common within the breed, thereby creating long-range LD. **c** | In the short time since breed creation, these long-range patterns have not yet been substantially broken down by recombination. Long breed haplotypes, however, still retain the underlying short ancestral haplotype blocks from the domestic dog population, and these are revealed when one examines chromosomes across many breeds. This figure is modified, with permission, from *Nature Genetics* REF. 22 © (2007) Macmillan Publishers Ltd. All rights reserved.

1:1 orthologues
Pairs of single genes in two different species that are descended from the same ancestral gene.

Selection mapping
A mapping design that finds trait loci by searching for selective sweeps. A selective sweep describes the reduction or elimination of genetic variation in a region owing to strong selection.

Genome coverage
The number of times, on average, that each base is sequenced.

Genome assembly
The consensus sequence of many short reads put together (a read is a fragment of sequenced DNA).

N50 contig size
A contig is a segment of the genome assembly that contains no gaps. An N50 contig size means that half of all bases reside in contigs of this size or longer.

Supercontig
Consecutive contigs that are separated by gaps of known size and connected by paired end-reads.

Validation rate
The rate at which genotypes are confirmed using a different technology.

coverage of 1.5 times) and 100,000 sequence reads for each of nine dogs from nine different breeds[21]. The SNPs have an average validation rate of 98%, although the SNPs discovered by comparing the poodle and boxer alone are slightly less reliable. Although SNPs have been discovered in just 11 different breeds, about 70% of them are polymorphic and thus useful in other breeds, with the precise set of SNPs varying by breed. Thus, the genome project successfully produced a SNP set that is ideal for trait mapping, with a dense and even distribution across the genome and utility in any dog breed.

Analysis of haplotype structure. Canine population history includes two population bottlenecks, the more recent of which occurred in just the last few hundred years when humans created genetically isolated dog breeds (FIG. 1). It is well established that bottlenecks create long LD in populations[35], facilitating trait mapping[36]. Such a pattern is experimentally confirmed in dog breeds, with LD estimated at several megabases, 40–100 times longer than in humans[21,37]. After a bottleneck event, the long LD breaks down over time[38]. Thus, in the whole domesticated dog population LD is shorter than in humans[21]. This bimodal pattern, with short LD in the population and long LD within breeds, suggests a two-stage mapping strategy in which the disease locus is first identified in a breed and then narrowed using multiple breeds (FIG. 2).

Coalescence modelling
Retrospective modelling of population history, used to generate expectations on genomic variation.

To assess the feasibility of a two-stage approach, the genome project completed a detailed analysis of the haplotype structure in breeds and in the dog population. In breeds, each haplotype block is 0.5–1 Mb long (a reflection of the long LD) and has three to six haplotypes specific to each breed[21,39]. In the whole dog population, haplotype blocks and LD are much shorter, about 10 kb. Only three to five common ancestral haplotypes are observed for each haplotype block, and haplotypes are shared between distantly related breeds.

This sharing of ancestral haplotypes across breeds has important implications for the design of trait-mapping studies. Most disease mutations are likely to pre-date breed creation, as dog breeds are young and have had little time to accumulate novel mutations. Thus, when several breeds suffer from high prevalence of the same disease, they might share the same causative mutation, carried into each breed on the same ancestral haplotype. However, detecting these shared ancestral blocks in a cross-breed study requires a marker density sufficient to detect 5–10 kb ancestral haplotypes, probably a SNP every 1–2 kb. Thus, a genome-wide scan is most efficiently done in a single breed, with more breeds added in the fine-mapping stage (FIG. 2).

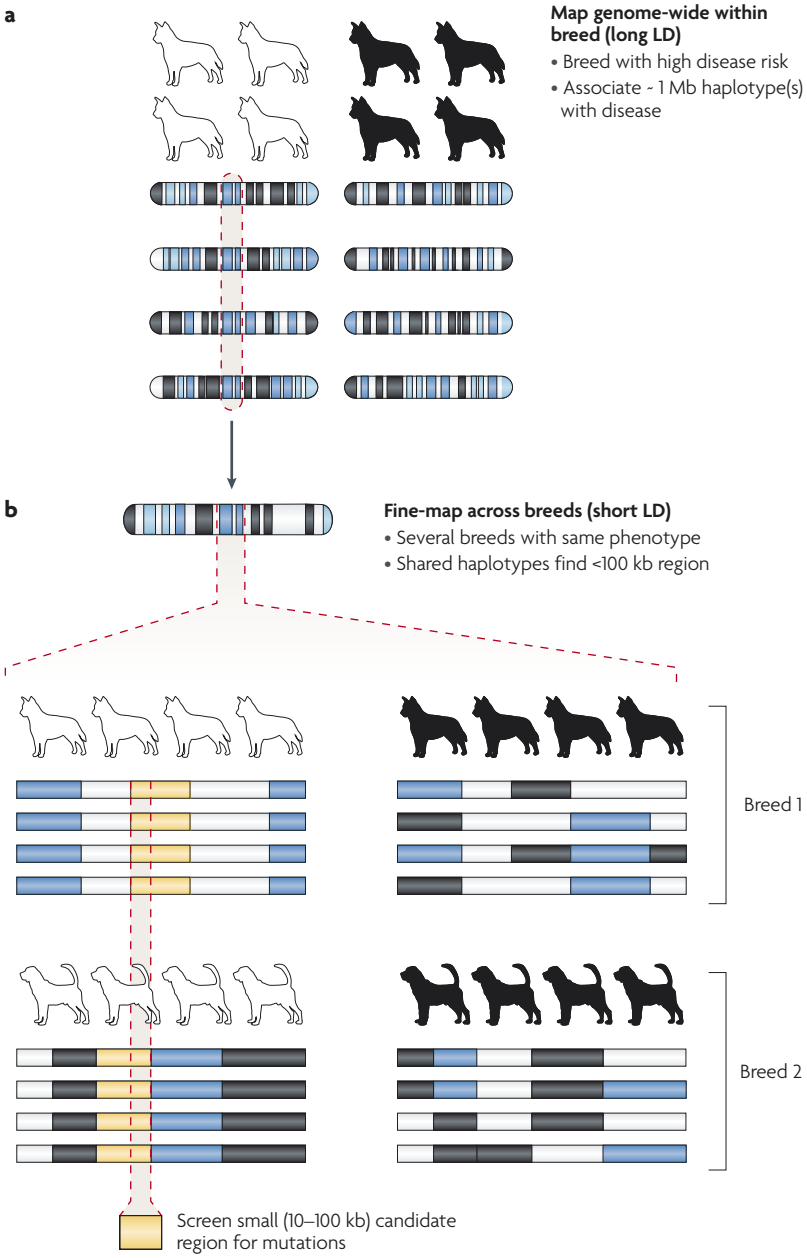
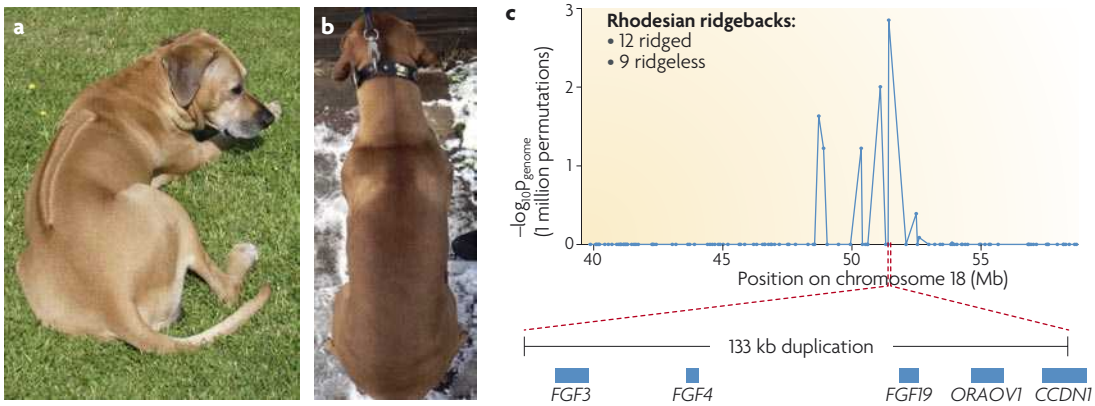


Figure 2 | **Two-stage mapping strategy.** A two-stage approach takes full advantage of the long linkage disequilibrium (LD) within breeds and the short ancestral haplotypes shared across breeds, allowing traits to be mapped with relatively few samples[39]. In this example, dogs with a mutation for loss of pigmentation (white) are compared with normally pigmented dogs (black). **a** | In stage 1, genome-wide association analysis within a breed uses dozens to hundreds of samples and ≥15,000 SNPs to identify one or more associated regions that are ~1 Mb long. **b** | In stage two, fine-mapping with a much denser set of SNPs in multiple breeds that share the same phenotype refines the association to a discrete region of 10–100 kb. This candidate region, which corresponds to the shared ancestral haplotype that carries the causative mutation, is screened for functional variants.

Genome-wide association in breeds. As part of the dog genome project, the number of SNPs and the number of dogs needed for GWA in a breed were estimated using data simulated by coalescence modelling[21,39]. Genome-wide, 15,000 SNPs proved sufficient for association mapping; using 30,000 added no more power, but sensitivity dropped when 7,500 were used. The number of dogs needed in a study using 15,000 SNPs varies depending on the inheritance pattern. For a simple Mendelian recessive trait with high penetrance and no phenocopies, mapping the disease allele requires just 20 cases and 20 controls; for a dominant trait, 50 cases and 50 controls suffice. Polygenic traits are more difficult to model accurately, as the power to detect disease-predisposing alleles varies depending on the relative risk conferred by the allele, the allele frequency and any interaction with other alleles and the environment. Using a simple multiplicative risk model, 100 cases and 100 controls detect a fivefold risk allele in 98% of data sets, whereas a twofold risk allele is found approximately half the time. The data simulated by the coalescence model was insufficient to test data sets of more than 100 cases and 100 controls, but we estimate that ~500 affected and ~500 unaffected dogs will provide sufficient power to map an allele conferring a twofold increased risk. Although alleles conferring a fivefold increased risk might be uncommon in human populations, the reduced genetic diversity of breed populations, and the remarkably high disease prevalence, suggests it is a fair estimate in pure-bred dogs.

To enable GWA mapping using ~15,000 SNPs, the Broad Institute collaborated with Affymetrix to generate a genome-wide SNP genotyping array[22]. Today, three different genome-wide arrays are available with sufficient SNP coverage for GWA in a breed: a 26,578 (27K) SNP Affymetrix array, a 49,663 (50K) SNP Affymetrix array and a 22,362 SNP Illumina array. There is an overlap of 14,985 SNPs between all three arrays, allowing a meta-analysis across platforms. In selecting genotyping platforms, the Broad Institute

Box 2 | **Mapping the ridge**



The Rhodesian ridgeback breed is characterized by a 'ridge' of inverted hair growth along the spine (panel **a**), which is inherited as an autosomal dominant trait; ridgeless dogs lack the mutation (panel **b**)[84]. The ridge allele predisposes dogs to a closed neural-tube defect known as dermoid sinus (DS), which is similar to dermal sinus in humans and suggests the involvement of a mutation that affects secondary neurulation[85,86]. Using array data (from the Affymetrix 27K array) for just nine ridgeless Rhodesian ridgebacks and 12 ridged controls (11 of which had DS), genome-wide association mapping found a 750 kb haplotype block on chromosome 18 that is strongly correlated with the trait (chi-squared p-value of 9.6×10^{-8} and genome-wide p-value of 1.4×10^{-3}; panel **c**). The genome-wide p-value is measured using a 'max(T)' permutation procedure that corrects for multiple tests[87], a crucial step in evaluating data for an association study that involves tens of thousands of data points. With 100,000 permutations, the association measured at the top-scoring locus in the ridge data set was 100-fold stronger than for any other region in the genome.

A closer analysis of the array data revealed an odd pattern of calls in the associated haplotype. At one particular SNP, every ridged dog was genotyped as being heterozygous, a highly significant deviation from Hardy–Weinberg proportions. A sequence duplication would explain this phenomenon — if the SNP in the duplication has a different allele from the original copy, then every dog with the duplication would genotype as heterozygous. Testing for a copy number polymorphism using multiple ligation-dependent genome amplification[88] confirmed a 133 kb duplication with perfect genotype–phenotype correlation. Every ridged dog tested had the duplication, including those from a second breed, the Thai ridgeback, and it was not found in a single one of the 47 tested ridgeless dogs[23]. Dogs homozygous for the duplication were more likely to have DS. Resequencing of the ends of the duplication confirmed that the Rhodesian and Thai ridgebacks share exactly the same breakpoints, strongly suggesting the two distantly related breeds share the same ancestral ridge allele[23].

The duplicated region contains four complete genes: fibroblast growth factor 3 (*FGF3*), *FGF4*, *FGF19* and oral cancer overexpressed 1 (*ORAOV1*), as well as the 3′ end of cyclin D1 (*CCND1*) (panel **c**). The histology of the ridge in ridgeback dogs suggests a mild defect of the planar cell polarity system, required for both normal hair follicle orientation and neural tube closure[85,89]. As FGF gene family members have crucial roles during embryonic development, including during hair follicle morphogenesis and skin development[90], dysregulated FGF expression along the dorsal midline during embryonic development might lead to disorganized hair follicles and an increased risk of DS in ridgeback dogs. The duplication misses, by just 10 kb, an enhancer element of crucial importance for embryonic expression of *FGF3* (REF. 56) — possibly explaining why the ridge phenotype is comparatively mild.

Panel **b** is reproduced, with permission, from REF. 24 © (2006) Cold Spring Harbor Laboratory Press. Panel **a** is courtesy of N. H. C. Salmon Hillbertz, Swedish University of Agricultural Sciences, Uppsala, Sweden.

Penetrance
The proportion of individuals carrying a genetic variant who express the trait connected with that variant.

Phenocopy
Describes an individual without the trait mutation who nonetheless exhibit the trait owing to environmental or other causes.

Multiplicative risk
An inheritance model whereby disease risk increases by λ in heterozygotes and λ^2 in homozygotes.

Corrects for multiple tests
The adjusting of p-values for statistical tests that include many markers, when the probability that significant values will occur by random chance is increased.

Assay conversion rate
The fraction of assays that work on a certain genotyping platform.

Call rate
The fraction of individuals that give genotyping calls for a particular SNP.

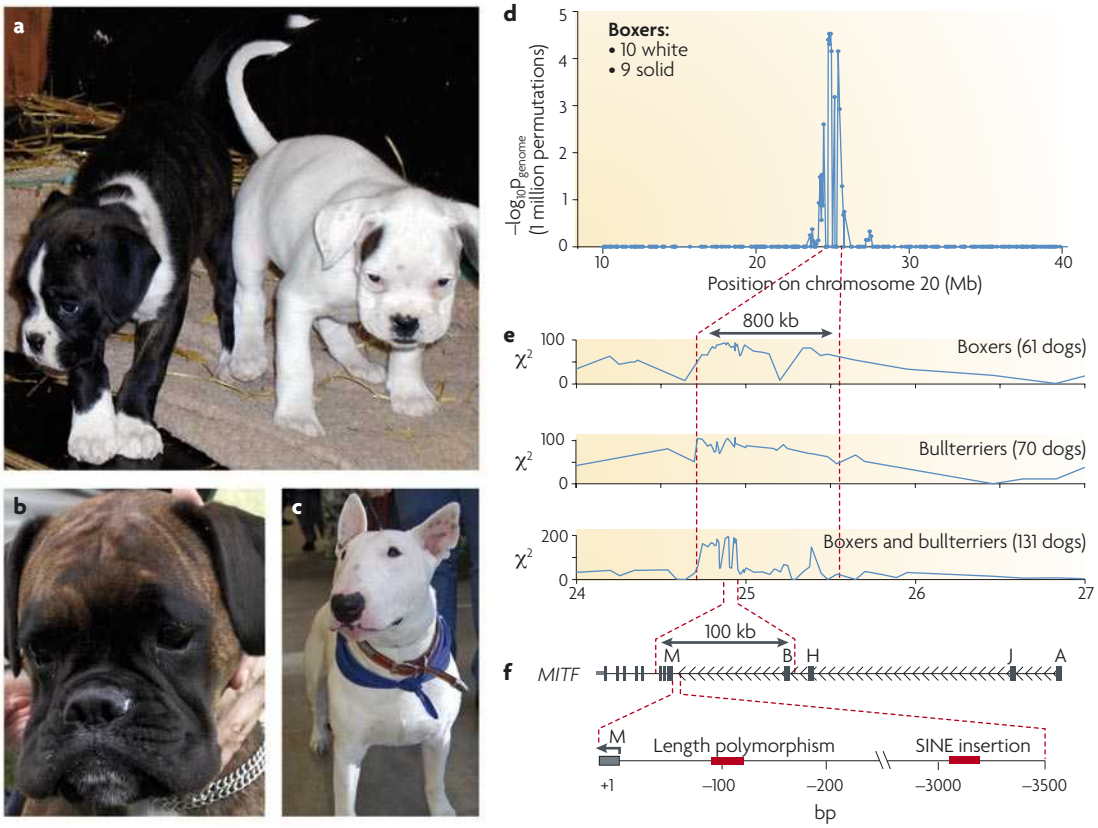
Copy number variant
(CNV). A genomic region that is longer than 1 kb and occurs a variable number of times.

considered both technical specifications (for example, assay conversion rates) and price compatibility with canine genetics funding levels. Although all three arrays are fully functional for GWA, they offer slightly different advantages. The Illumina array has extremely high call rates and uniform genome coverage, whereas the 50K Affymetrix array has much higher SNP density. The match–mismatch probe design of the 27K Affymetrix array might in principle allow more sophisticated analysis of intensity data to reveal copy number variants (CNVs)[40]. Still, all three arrays have low SNP densities that make mapping old selective sweeps[22] or surveying the genome for CNVs difficult, as many CNVs are on the order of 10 kb in size and would require multiple consecutive SNPs to detect.

Successful trait mapping

The two proof-of-principle studies using the canine SNP arrays for GWA — mapping the white spotting in boxers and bull terriers and the dorsal hair ridge in ridgeback breeds[22,23] — are described, respectively, in BOX 2 and BOX 3. These are the first two studies published, and numerous others are underway across the canine genetics community. The American Kennel Club Canine Health Foundation and the Morris Animal Foundation are currently funding mapping studies of cancers, autoimmune diseases, kidney diseases, liver diseases, epilepsy and other neurological diseases, orthopaedic diseases, endocrine disorders, behavioural disorders, eye diseases, gastrointestinal diseases and reproductive disorders. The LUPA consortium, funded by the European Union, aims to

Box 3 | **Mapping white spotting**



The absence of skin and coat pigmentation in white boxers is a semi-dominantly inherited trait in which heterozygous dogs are part pigmented, part white (termed flash; panel **a**, the solid phenotype is shown in panel **b**). White boxers suffer increased rates of deafness, reminiscent of the human auditory-pigmentary disorders Waardenburg and Tietz syndromes[91,92]. White bull terriers have an identical phenotype (panel **c**). Breeding studies in the 1950s designated the white-coat variant as the extreme-white, or s^w, allele of the major white spotting locus (*S*)[93]. Other alleles assigned to this locus are Irish spotting (s^i), seen in Basenji and Bernese mountain dogs, and piebald spotting (s^p), seen in beagles, fox terriers and English Springer spaniels. Previous research excluded several strong candidate genes[94,95].

The locus for this trait was mapped using a two-stage protocol. In the first stage, the Affymetrix 27K SNP array data for just 10 white and 9 solid boxers mapped the s^w allele to a region of less than 1 Mb containing just one gene, microphthalmia-associated transcription factor (*MITF*), an intricately regulated developmental gene involved in pigmentary and auditory disorders in humans and mice (panel **d**)[54,96,97]. The most strongly associated SNP ($p = 7 \times 10^{-10}$; $p_{genome} = 3 \times 10^{-5}$) was 1,000-fold more strongly associated than any other region in the genome. The 800 kb associated haplotype at this locus was homozygous in all white boxers and absent from solid dogs, correlating perfectly with the semi-dominant inheritance pattern[22].

In the second stage of the protocol, the associated 800 kb haplotype was fine-mapped using ~70 SNPs in boxers and a second breed, bull terriers. After confirming the association independently in the two breeds, the data sets were combined and a narrower ~100 kb long region of association was defined (panel **e**). This region is similar in size to those found in cross-breed studies of progressive rod-cone degeneration[98] and collie eye anomaly[59] and contains the first exon of the melanocyte-specific isoform of the *MITF* gene. Exon resequencing revealed that there were no coding changes in the associated region, instigating a lengthy and difficult search for possible regulatory mutations.

First, every possible candidate mutation was identified by a comparison of finished sequence from each chromosome of a heterozygous dog. Then, each of the 124 polymorphisms found were resequenced in fully pigmented dogs from many breeds, and in white boxers and bull terriers. Those not concordant with phenotype, 78 in total, were removed from further consideration. Of the 46 remaining polymorphisms, two are found in sequence conserved across mammals and thus are most likely to affect gene function. Both mutations — a SINE insertion (3 kb upstream of the melanocyte-specific start site of *MITF*, M) and a length polymorphism (<100 bp upstream of M) lie upstream of M and downstream of other known start sites (A, J, H and B) (panel **f**). Although it therefore seems plausible that one or both of these polymorphisms cause the white coat-colour phenotype, this hypothesis awaits experimental confirmation.

Semi-dominant
A Mendelian inheritance pattern in which heterozygous individuals exhibit a phenotype that is intermediate to the two homozygous phenotypes.

find genes responsible for at least 18 diseases, including four cancers, four inflammatory disorders and three heart diseases, in the next 4 years. To do so, they are collecting DNA, health information and GWA data for 8,000 dogs, thus creating a data resource with enormous future potential. As researchers find the suspected genetic causes for cancers, the newly established Canine Comparative Oncology and Genomics Consortium (CCOGC) aims to provide samples for functional studies. It recently announced funding for six institutions to begin collecting tumour and normal tissue for a shared repository focusing on cancers that have clear human relevance, including osteosarcoma, lymphoma and melanoma. Through these combined efforts, this new era in canine disease research offers enormous potential for advancing medical research.

Considerations for genome-wide association mapping. In general, designing mapping studies within dog breeds involves the same considerations as in human studies[22], but the emphasis differs between the GWA and fine-mapping stages. For both stages, phenotypes for both cases and controls must be carefully defined. Using detailed pathology and veterinary records, the disease status of the dog can be determined with considerable accuracy, and medically necessary surgery might provide tissue biopsies with no additional adverse impact on the dog's health. Although including individuals with ambiguous phenotypes might reduce the power of a sample set, an unnecessarily strict definition can make sample collection onerous.

For the GWA step, it is crucial that both the cases and controls are drawn from as broad a population as possible and are geographically and genealogically matched to minimize population stratification. Otherwise, the association might yield false hits that, rather than identifying disease loci, simply reflect spurious differences between the cases and controls. The extensive genealogical and health records maintained by breed clubs, veterinarians and pet-insurance companies facilitate the design and implementation of trait-mapping studies.

Determining the relative risk and number of samples required is vital. Under-powering studies, by including insufficient numbers of cases and controls, will make it difficult to identify true regions of association. Thus far, only studies mapping traits with Mendelian inheritance have been published, and for these the power calculations that were done for the canine genome project have proven accurate[22]. In addition, unpublished preliminary data for complex traits, such as several cancers and immunological diseases, confirm that finding loci with strong effect requires roughly 100 cases and 100 controls.

At the fine-mapping stage, in which additional samples from the original breed should be included, there is a lower risk of problems with random population stratification, as only a few loci are considered. However, it is essential that additional breeds share the same phenotype. Although more closely related breeds might be more likely to share the same risk alleles, haplotypes are commonly shared even between distantly related breeds[21]. For example, the same ridge mutation was found in an Asian breed, the Thai ridgeback, and an African breed, the Rhodesian ridgeback[23]. Thus, although using more related breeds might be preferable, any breed sharing the phenotype is a candidate for fine-mapping. For complex traits, defining the same phenotype across breeds might be more difficult as identical phenotypes might nonetheless have different genetic components. Thus, fine-mapping of a complex trait should involve more than two breeds.

Alternative mapping strategies in the dog. Even though GWA in dogs might be the most powerful mapping approach for disease research, other trait-mapping strategies have been proposed. Here we briefly review and contrast the various approaches, which include GWA mapping, quantitative trait locus (QTL) mapping, across-breed mapping and selection mapping.

For association mapping of disease genes, dog breeds offer the power of a genetically isolated population, including extensive LD and a more uniform genetic background that simplifies phenotype definition[21,36,37,39]. For example, when diabetes is common in a breed a single form predominates, unlike the complex spectrum of diabetic diseases seen in large human populations[41]. Because of the tight recent population bottlenecks, the inbreeding coefficient in dog breeds is roughly 12%. At each ~500 kb locus there are just 3–4 common haplotypes[21], making it easier to detect the association of one of those haplotypes with the disease. By contrast, extensive inbreeding also has negative consequences for mapping. The presence of widespread homozygosity — roughly 6% of the genome resides in homozygous regions that are >0.5 Mb — means that large genomic regions are essentially invisible to trait mapping. Even a strong risk factor, if fixed in a breed, will be undetectable by standard association mapping that uses genetic segregation. If a mutation arises on a haplotype that is common in a breed, it can also be difficult to detect unless the actual disease mutation is queried.

Pure-bred dogs are also well suited to mapping QTLs that control large phenotypic differences, such as longevity, size and morphology. In one ongoing study, a large cohort of dogs from the Portuguese water dog breed is being collected and carefully examined for different morphological parameters. This study has already identified a large number of morphological QTLs[42]. One QTL that was linked to size contained the insulin-like growth factor 1 (*IGF1*) gene, at which multiple small breeds show evidence of a selective sweep of a single shared allele[43]. By contrast, the inability to make large canine crosses limits the effectiveness of linkage mapping for complex traits in dogs; livestock populations, in which QTLs are easily mapped and epistasis between loci is easily studied, might offer more power. Information on morphological traits, whether genetically defined or primarily environmental, can also add power to canine disease studies. For example, an investigation of diabetes could incorporate body mass. Although somewhat difficult, pursuing such studies in dog breeds would still be considerably easier than similar efforts in humans.

Population stratification
The presence of multiple population subgroups that show limited interbreeding. When such subgroups differ both in allele frequency and in disease prevalence, this can lead to erroneous results in association studies.

Quantitative trait locus
(QTL). A stretch of DNA that is closely linked to a continuously variable phenotype.

Inbreeding coefficient
The probability that two alleles are identical by descent.

Traits that have been driven to fixation by drift or artificial selection within a dog breed cannot be mapped within that breed. This includes many of the most striking phenotypic differences that distinguish breeds, such as the spots of the Dalmatian or the extremely wrinkled skin in the Chinese Shar-Pei breed. Such phenotypes require either cross-breed association mapping or selection mapping.

Across-breed association studies compare many different breeds with and without the trait of interest. Although any pair of breeds will have countless differences, with enough breeds only trait-related alleles will consistently differ between cases and controls. In recent work, QTLs were mapped for size and other phenotypes using ten dogs from each of 148 dog breeds[44]. Researchers found that the more breeds that were included, the greater the power, especially if close to 50% of the breeds were affected by the trait of interest. Thus, although their novel approach seems to be effective for common selected phenotypes, it is not applicable to traits that are found in just one or a few breeds, such as the Shar-Pei's wrinkled skin. It is also crucial that p-values are adjusted for genome-wide significance. Owing to widespread differences in allele frequencies between different breeds, artificially large p-values can arise and should be rigorously tested using, for example, genome-wide permutation.

Selection mapping is based on the assumption that genes controlling a strongly selected phenotype will lie within selective sweeps. Several different genome-wide tests for selection have been developed in humans and they search for regions with an abnormally long haplotype of increased frequency. These tests probably will not work in a single dog breed, as studies would be confounded by recent origins, long haplotypes and extensive homozygosity[45]. An across-breed version of these tests, however, could prove more effective. If a set of breeds shares the same trait with the same genetic origin, the genes responsible will lie within selective sweeps that occur over the same genomic region in all of them, dramatically reducing the background noise created by the bottlenecks that occurred at breed creation. A selective-sweep test, unlike across-breed association, could be applicable to traits that are seen only in a few breeds.

Both these cross-breed approaches hold great potential for mapping the many traits fixed in breeds, from coat colour to size to behaviours such as herding. Owing to the high frequency of defining phenotypes in a breed, just a small number of individuals are sufficient to represent the entire breed. By eliminating the expense of sampling deeply in each population, more breeds can be included, with a resulting increase in statistical power. Several caveats apply to both approaches. First, they assume that breeds with the same trait have inherited the same causative variants from the ancestral dog population — phenotypes with more heterogeneous origins will be far more difficult to map. Second, the current genome-wide canine SNP arrays might prove to be insufficient for such studies. Across breeds, the haplotype structure and diversity of dogs is similar to humans, and truly effective across-breed studies could require an array with a marker density similar to the >500,000 SNP human arrays.

Lessons for the future

The biology of the mutations found. Few of the actual sequence mutations underlying complex traits have been found in either humans or dogs. However, mapping of human complex traits finds many mutations that lie outside known genes and are likely to alter gene regulation rather than the coding sequence[46]. This seems likely to hold true in dogs as well — unpublished data for several diseases find the strongest association outside the coding portion of genes — and intuitively this makes sense. Deleterious coding mutations alter the protein, often causing severe phenotypic changes, whereas regulatory mutations might only affect gene functions by smaller dosage effects or in specific tissues, thereby allowing the individual to reproduce normally.

The mutations known for Mendelian canine traits implicate several types of sequence variants (TABLE 1). Most involve point mutations in coding sequence, possibly because such mutations are the easiest to identify; but many of the most recent studies find more complex mutations. The insertion of a canine-specific short interspersed nuclear element (SINE)[47–49] causes inherited narcolepsy in Dobermans[20], centronuclear myopathy in the Labrador retriever[50], and grey or 'merle' coat colouring in several breeds[51]. The expansion of an unstable dodecamer repeat in the *NHLRC1* (NHL repeat containing 1; also known as *EPM2B*) gene causes Lafora disease, a form of epilepsy, in miniature wire-haired dachshunds[52]. This was the first example of a disease-causing repeat expansion outside of humans, in which trinucleotide repeat expansion is associated with several neurological disorders[53].

For both canine traits mapped through GWA (ridge and white coat colour; BOXES 2,3), the mutations are non-coding and otherwise similar to those described above: a CNV for one and a short repeat length polymorphism and a SINE insertion for the other[22,23]. The genes involved are among the most fundamental transcription factors, with carefully regulated expression patterns and essential roles during development. Coding mutations almost certainly would seriously impair development, making a regulatory change the only viable option. Mice with coding mutations in *MITF*, the coat colour gene, have under-developed eyes, reduced fertility and are nearly always deaf[54]. Mice lacking any of the FGF genes duplicated in ridgeback dogs suffer high levels of embryonic lethality; even when viable their ears, heart, nervous system and skeleton do not develop normally[55–57]. Although dog breeders are willing to tolerate the low rates of deafness and dermoid sinus seen in white and ridged dogs, respectively, more serious developmental deficiencies would be purged from the gene pool. Thus, mutations affecting the regulation of important proteins might underlie many traits in domestic animals[58–61]. In one recent example, researchers found that a *cis*-acting regulatory mutation in syntaxin 17 (*STX17*), a widely expressed gene involved in intercellular membrane trafficking, causes premature hair greying and increased susceptibility to melanoma in the domesticated horse[62].

Short interspersed nuclear element
(SINE). Retrotransposons ~200 bases long that are derived from a tRNA–Lysine and occur frequently throughout the canine genome.

Table 1 | **A subset of the canine diseases in which the mutation is known (part 1)**

Disease	Gene	Breed	Mutation (date published)	Refs
Central nervous system or lysosomal storage				
Epilepsy (Lafora type)	Malin (*NHLRC1*)	Minature wire-haired dachshund	Tandem 12-bp repeat expansion* (2005)	52,99
Shaking puppy (generalized tremor)	Proteolipid protein 1 (*PLP1*)	English Springer spaniel	Missense mutation (1990)	100
Narcolepsy	Hypocretin (orexin) receptor 2 (*HCRTR2*)	Dachshund	Intronic SINE insertion* (1999)	101
Ceroid lipofuscinosis	Ceroid-lipofuscinosis, neuronal 8 (*CLN8*)	English setter	14 bp coding deletion (1999)	102
Eye diseases — PRA				
Rod–cone dysplasia 1	Phosphodiesterase 6B, cGMP-specific, rod, beta (*PDE6B*)	Irish setter	Nonsense mutation (1993)	103
Rod–cone dysplasia 3	Phosphodiesterase 6A, cGMP-specific, rod, alpha (*PDE6A*)	Cardigan Welsh corgi	1 bp coding deletion (1999)	104
Autosomal dominant PRA	Rhodopsin (*Rho*)	English mastiff	Missense mutation (2002)	105
Eye diseases — other				
Cone degeneration	Cyclic nucleotide gated channel beta 3 (*CNGB3*)	Alaskan malamute	Gene deletion* (2002)	14
		German short-haired pointer	Missense mutation (2002)	14
Congenital night blindness	Retinal pigment epithelium-specific protein 65 kDa (*RPE65*)	Briard	4 bp coding deletion (1999)	106
Gastrointestinal, liver and endocrine disorders				
Imerslund–Grasbeck disorder (cobalamin malabsorption)	Amnionless (*AMN*)	Giant schnauzer	33 bp coding deletion (2004)	107
Copper toxicosis	Copper metabolism (Murr1) domain containing 1 (*MURR1*)	Bedlington terrier	One exon deleted* (2002)	16
Genodermatoses				
Epidermolysis bullosa (dystrophic form)	Collagen, type VII, alpha 1 (*COL7A1*)	Golden retriever	Missense mutation (2001)	108
Epidermolysis bullosa (junctional form)	Laminin, alpha 3 (*LAMA3*)	German short-haired pointer	Intronic repeat insertion* (2005)	109
Hereditary blood disorders				
Haemophilia B (factor IX deficiency)	Coagulation factor IX (*F9*)	Mixed breed	Missense mutation (1997)	29
Von Willebrand disease type II	Von Willebrand factor (*VWF*)	German pointers	Missense mutation (2004)	110
Von Willebrand disease type III	Von Willebrand factor (*VWF*)	Scottish terrier	1 bp coding deletion (2000)	111
Muscular and skeletal disorders				
Centronuclear myopathy	Protein tyrosine phosphatase-like, member a (*PTPLA*)	Labrador retriever	SINE insertion in exon* (2005)	112
X-linked dystophin muscular dystrophy	Dystrophin (*DMD*)	Golden retriever	Splice-junction point mutation* (1992)	113
Pharmacogenetic problems				
Drug sensitivity (Ivermectin)	Multidrug resistance 1 (*MDR1*)	Collie	4 bp coding deletion (2001)	114
Primary immunodeficiencies				
Leukocyte adhesion deficiency	Integrin, beta 2 (*ITGB2*)	Irish setter	Missense mutation (1999)	115
Severe combined immunodeficiency	Interleukin 2 receptor, gamma (*IL2RG*)	Basset hound	4 bp coding deletion (1994)	116
Cyclic hematopoiesis (cyclic neutropaenia)	Adaptor-related protein complex 3, beta 1 subunit (*AP3B1*)	Collie	1 bp coding insertion (2003)	117
Renal Disorders				
Alport syndrome	Collagen, type IV, alpha 5 (*COL4A5*)	Samoyed	Nonsense mutation (1994)	118
Renal cystadenocarcinoma and nodular dermatofibrosis	Folliculin (*FLCN*)	German shepherd	Missense mutation (2003)	18

*Involves sequence that is not coding. bp, base pairs; PRA, progressive retinal atrophy; SINE, short interspersed nuclear element.

Table 1 | **A subset of the canine diseases in which the mutation is known (part 2)**

Disease	Gene	Breed	Mutation (date published)	Refs
Phenotypic traits				
Coat colour	Melanocortin 1 receptor (*MC1R*)	Many breeds	Nonsense mutation (2000)	119,120
Merle coat, deafness and ophthalmologic abnormalities	Silver homologue (*SILV*)	Many breeds	SINE insertion at splice junction* (2006)	121
Gross muscle hypertrophy	Myostatin (*MSTN*)	Whippet	Nonsense mutation (2007)	122
White coat colour and deafness	Microphthalmia-associated transcription factor (*MITF*)	Boxer and bull terrier	SINE insertion and/or length polymorphism in regulatory sequence* (2007)	22
Ridge and dermal sinus	Fibroblast growth factors *FGF3*, *FGF4*, *FGF19*, and oral cancer overexpressed 1 (*ORAOV1*)	Rhodesian and Thai ridgeback	133 kb duplication containing four genes* (2007)	22
Black coat colour	Beta-defensin 103 (*CBD103*)	Many breeds	Amino-acid deletion* (2007)	123

*Involves sequence that is not coding. SINE, short interspersed nuclear element.

Unfortunately, searching for non-coding mutations can be a laborious process — especially when the causative variant is a short insertion or deletion, or a copy number polymorphism. For canine disease mutations that are shared between several breeds, the candidate region should be less than 100 kb. The challenge will be much greater for mutations that are specific to a single breed, in which case GWA mapping will yield a region up to 2 Mb long. For complex traits, resequencing the entire region in many dogs might be necessary to find the most likely polymorphisms. Until recently this task would have proven prohibitively expensive, but is now feasible owing to recent advances in sequencing technology[63]. In addition, ongoing projects are using these new technologies to carefully define the functional regions of genomes, making it easier to anticipate which polymorphisms are most likely to disrupt function[64–66].

Determining functional consequences. The challenging task of assaying the functional consequences of a sequence variant can be tackled from several different directions. In dogs, the options are limited. The expression of a gene of interest can be compared in cases and controls using expression arrays or quantitative RT-PCR. However, the appropriate tissue at the appropriate time-point must be available. For cancers, the CCOGC should be helpful in supplying carefully controlled tumour samples, accompanied by any progression and clinical trial data from the Comparative Oncology Trials Consortium (COTC). Even so, for many diseases appropriate tissues and cell lines are still sparse.

Ethical and practical considerations make *in vivo* tests of mutations in dogs a poor option. For such studies, standard model organisms such as the mouse and zebrafish will continue to be the models of choice.

Application to other vertebrates. In many other vertebrate organisms, the full potential of association mapping remains under-used. Whereas genome-wide SNP mapping panels exist for several domestic animal species, the haplotype characterization has been less thorough. However, for several species, such as cattle and chickens, linkage analysis has been used successfully for QTL mapping, taking advantage of the large families and careful characterization of quantitative traits.

Thus, the key considerations when developing the tools for mapping in a new species are: definition of the project goals, whether disease mapping, quantitative trait mapping or identification of selective sweeps; ascertaining the population history and its effect on the genome; determination of the haplotype structure, which requires examining several random genomic regions in detail; estimation of the number of markers needed, whether SNPs or other types of variants, and the best method of discovery; and identifying the tools needed, and how to design them for the widest possible range of applications. Such resource development is underway in numerous domestic animals, including chickens, cattle, cats and horses, and is proposed for several fish, including stickleback and tilapia. Here we discuss three ongoing genome projects, in the horse, cow and stickleback, to illustrate how the distinctive properties of each species guided this process (TABLE 2).

The domestic horse has many of the same trait-mapping advantages as the dog, and likewise shares many diseases with people, including allergies, neuromuscular diseases and metabolic diseases. The domestic horse, like the dog, is comprised of breeds shaped by population bottlenecks and artificial selection. Compared with dog breeds, however, in horse breeds the selective pressure was lower and cross-breeding more common, suggesting somewhat shorter LD in breeds and more haplotype sharing between breeds. To assess the extent of LD and haplotype diversity in horses, the ongoing horse genome project is surveying shorter regions at higher density than the dog genome project. The trait-mapping power calculations will be based on a population model that produces the observed extent of LD. Most probably, the horse GWA genotyping arrays will require denser SNP coverage than the dog arrays.

Cattle is an agriculturally important species that has been crucial for the development of many human societies. Since its domestication about 8,000 years ago, cattle have been providing us with draft power, meat and milk, and thus were subjected to strong artificial selection

Table 2 | **Considerations for model organisms**

Consideration	Mouse	Dog	Horse	Cattle	Stickleback
Traits of interest	Various diseases (often monogenic), some morphology and behaviour	Morphology, behaviour and diseases (including cancer, epilepsy and diabetes)	Diseases (including allergies and neurological diseases) and exercise physiology	Commercially important quantitative traits	Fresh-water adaptations involving morphology, habitats and behaviour
Similarity in phenotypes to human	Some	Many	Many	Some	Few
Veterinary care and detailed records	Only for inbred mice	Yes	Yes	Yes	No
Shared environment with humans	No	Yes	Limited	Limited	No
Pedigrees	Yes	Yes	Yes	Yes	No (but isolated by location)
Suitability for functional studies (transgenes and knockouts)	Yes	No	No	No	Not currently done
Population types	Inbred strains	Breeds	Breeds	Breeds	Geographical populations
Bottlenecks	100 years ago	~15,000 and 200 years ago	Only weak	Only weak	~15,000 years ago
Evolutionary influences	Human inbreeding and genetic manipulation	Human artificial selection and partial inbreeding	Human artificial selection and partial inbreeding	Human artificial selection and partial inbreeding	Adaptation
Extent of linkage disequilibrium	Long	Long and short	Intermediate (expected)	Intermediate to short (large population size)	Short (expected)
SNP map	Yes	Yes	Needed	Yes, additional work in progress	Might be needed
Mapping strategy	Crosses, QTL mutagenesis	Two stage association mapping: within and across breed, QTL across breed and selection mapping	Linkage and association mapping: within and across breed	Crosses, QTL, linkage and some association mapping	Surveying multiple populations for allele frequency differences
Tools	SNP and haplotype map, GWA array	GWA array	GWA array	GWA array and linkage tools	New technology resequencing

GWA, genome-wide association.

for the most productive animals[67]. Extensive pedigrees for lineages that were established in the mid 1800s[68], and careful phenotypic records, make cattle ideal for examining quantitative traits, especially those of agricultural importance. However, although a draft genome assembly and a simple sequence length polymorphism linkage map are available, along with several smaller SNP mapping panels, no thorough and well designed haplotype analysis has been completed. The distinctive population history and differing mapping potential of this commercially important species makes trait mapping in cattle distinct from both dogs and horses, and an in-depth characterization of genome structure and breed relationships is crucial for developing appropriate resources and tools.

The stickleback fish underwent a remarkable adaptive radiation ~15,000 years ago as countless small populations of fish were trapped, by changing sea levels, in freshwater environments across the world. The history of the stickleback populations suggests a mapping strategy distinct from those used in horse and dog breeds, which were created through artificial selection in a short time. In sticklebacks, natural selection drove phenotypic changes acquired over a much longer timescale. Each population of sticklebacks experienced approximately similar environmental stresses (a freshwater ecosystem), leading to repeated selection for the same ancestral allele at the same locus[69]. Given the age of the freshwater populations the haplotype blocks and LD will probably be short, and mapping these loci by GWA would require a very dense set of markers. With a much smaller genome size (460 Mb) than mammals, the best approach in sticklebacks is to use new sequencing technology to survey the whole genomes of individuals from multiple marine and freshwater populations and search for variants selected in freshwater populations. This not only would ensure detection of selected loci no matter how short the LD, but would also identify a plethora of SNPs and other sequence polymorphisms, including the actual functional variants for some phenotypes.

In the near future the new sequencing technologies should allow the stickleback mapping approach to be used to survey the genomes of domesticated animals, such as dogs, chickens, cattle, horses and pigs, revealing selective sweeps that are undetectable with the less dense genotyping technologies that are currently in common use.

Adaptive radiation
The evolution, through adaptation to different ecological niches, of phenotypic differences between individuals derived from a single species.

Conclusion

With the necessary tools and resources completed, the domestic dog is now a fully fledged genetic model that is ideally suited to trait mapping. The search is on for genes shaping morphology and behaviour, as well as disease susceptibility, progression and outcome. The first whole-genome association studies show that monogenic traits can be mapped with remarkable power using just a few dozen samples, and probably the same approach will soon find genes for complex traits, such as cancer, epilepsy and diabetes. Scientists have also begun developing cross-breed mapping strategies to find the loci controlling the remarkable phenotypic variation between breeds. Whatever the approach, mapping the associated genomic loci is simply the first step — finding the genes and variants that are responsible will be challenging, especially for polygenic traits caused by regulatory mutations. Nonetheless, the similarity of the dog and human genomes, as well as their close social relationship, means that this research will undoubtedly advance not just canine research, but also our understanding of human development and disease.

Many other model organisms have significant trait-mapping potential and await development of the necessary resources. The work in the dog clearly shows that by considering the unique population history and biology of each species, tools of exceptional power and efficiency can be designed. In coming years, the horse, cow, chicken, stickleback and many other species will join the dog in helping scientists find the genomic mechanisms that define complex traits. As the trail-blazing model organism for trait mapping, the domestic dog has, once again, proven itself man's best friend.

1. Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265**, 2037–2048 (1994). **A good description of the requirements for complex trait mapping.**
2. Bucan, M. & Abel, T. The mouse: genetics meets behaviour. *Nature Rev. Genet.* **3**, 114–123 (2002).
3. Cook, M. C., Vinuesa, C. G. & Goodnow, C. C. ENU-mutagenesis: insight into immune function and pathology. *Curr. Opin. Immunol.* **18**, 627–633 (2006).
4. Copeland, N. G., Jenkins, N. A. & Court, D. L. Recombineering: a powerful new tool for mouse functional genomics. *Nature Rev. Genet.* **2**, 769–779 (2001).
5. Paigen, K. A miracle enough: the power of mice. *Nature Med.* **1**, 215–220 (1995). **A discussion of the mouse as the model organism of choice.**
6. Ostrander, E. A. & Kruglyak, L. Unleashing the canine genome. *Genome Res.* **10**, 1271–1274 (2000).
7. Wayne, R. K. Limb morphology of domestic and wild canids: the influence of development on morphologic change. *J. Morphol.* **187**, 301–319 (1986).
8. Fogle, B., Morgan, T. & Fogle, B. *The New Encyclopedia of the Dog* (Dorling Kindersley, New York, 2000).
9. American Kennel Club. *The Complete Dog Book* (eds Crowley, J. & Adelman, B.) (Howell Book House, New York, 1998).
10. Patterson, D. F. *et al.* Research on genetic diseases: reciprocal benefits to animals and man. *J. Am. Vet. Med. Assoc.* **193**, 1131–1144 (1988).
11. Sargan, D. R. IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. *Mamm. Genome* **15**, 503–506 (2004).
12. Sargan, D. R., Aguirre-Hernandez, J., Galibert, F. & Ostrander, E. A. An extended microsatellite set for linkage mapping in the domestic dog. *J. Hered.* **98**, 221–231 (2007).
13. Acland, G. M. *et al.* Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (*prcd*) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. *Proc. Natl Acad. Sci. USA* **95**, 3048–3053 (1998).
14. Sidjanin, D. J. *et al.* Canine *CNGB3* mutations establish cone degeneration as orthologous to the human achromatopsia locus *ACHM3*. *Hum. Mol. Genet.* **11**, 1823–1833 (2002).
15. Yuzbasiyan-Gurkan, V. *et al.* Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers. *Am. J. Vet. Res.* **58**, 23–27 (1997).
16. van De Sluis, B., Rothuizen, J., Pearson, P. L., van Oost, B. A. & Wijmenga, C. Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Hum. Mol. Genet.* **11**, 165–173 (2002).
17. Jonasdottir, T. J. *et al.* Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. *Proc. Natl Acad. Sci. USA* **97**, 4132–4137 (2000).
18. Lingaas, F. *et al.* A mutation in the canine BHD gene is associated with hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog. *Hum. Mol. Genet.* **12**, 3043–3053 (2003).
19. Mignot, E. *et al.* Genetic linkage of autosomal recessive canine narcolepsy with a mu immunoglobulin heavy-chain switch-like segment. *Proc. Natl Acad. Sci. USA* **88**, 3475–3478 (1991). **This paper describes the identification of the first canine SINE mutation.**
20. Lin, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365–376 (1999).
21. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005). **A paper that described the canine genome sequence and haplotype structure.**
22. Karlsson, E. K. *et al.* Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nature Genet.* **39**, 1321–1328 (2007). **This paper describes the proof-of-principle studies for GWA mapping in the dog.**
23. Salmon Hillbertz, N. H. *et al.* Duplication of *FGF3*, *FGF4*, *FGF19* and *ORAOV1* causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nature Genet.* **39**, 1318–1320 (2007). **This paper reports the second mutation to be identified from the proof-of-principle experiments for GWA in the dog.**
24. Giger, U., Sargan, D. R. & McNiel, E. A. in *The Dog and its Genome* (eds Ostrander, E. A., Giger, U. & Lindblad-Toh, K.) 249–289 (Cold Spring Harbor Laboratory Press, New York, 2006).
25. Little, C. C. Coat color in pointer dogs. *J. Hered.* **5**, 244–248 (1914).
26. Wright, S. The hairless dog. *J. Hered.* **8**, 519–520 (1917).
27. Whitney, L. F. Heredity of the trail barking propensity in dogs. *J. Hered.* **20**, 561–562 (1929).
28. Hutt, F. B., Rickard, C. G. & Field, R. A. Sex-linked hemophilia in dogs. *J. Hered.* **39**, 3–9 (1948).
29. Evans, J. P., Brinkhous, K. M., Brayer, G. D., Reisner, H. M. & High, K. A. Canine hemophilia B resulting from a point mutation with unusual consequences. *Proc. Natl Acad. Sci. USA* **86**, 10095–10099 (1989).
30. Mellersh, C. S. *et al.* A linkage map of the canine genome. *Genomics* **46**, 326–336 (1997).
31. Breen, M. *et al.* Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Res.* **11**, 1784–1795 (2001).
32. Guyon, R. *et al.* A 1-Mb resolution radiation hybrid map of the canine genome. *Proc. Natl Acad. Sci. USA* **100**, 5296–5301 (2003).
33. Breen, M. *et al.* An integrated 4249 marker FISH/RH map of the canine genome. *BMC Genomics* **5**, 65 (2004).
34. Kirkness, E. F. *et al.* The dog genome: survey sequencing and comparative analysis. *Science* **301**, 1898–1903 (2003).
35. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates, Sunderland, Massachusets, 1997).
36. Peltonen, L., Palotie, A. & Lange, K. Use of population isolates for mapping complex traits. *Nature Rev. Genet.* **1**, 182–190 (2000). **An excellent review of the power of isolated populations for mapping genes that underlie complex traits.**
37. Sutter, N. B. *et al.* Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* **14**, 2388–2396 (2004).
38. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).
39. Wade, C. M., Karlsson, E. K., Mikkelsen, T. S., Zody, M. C. & Lindblad-Toh, K. in *The Dog and its Genome* (eds Ostrander, E. A., Giger, U. & Lindblad-Toh, K.) 179–208 (Cold Spring Harbor Laboratory Press, New York, 2006).
40. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
41. Florez, J. C., Hirschhorn, J. & Altshuler, D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu. Rev. Genomics Hum. Genet.* **4**, 257–291 (2003).
42. Lark, K. G., Chase, K. & Sutter, N. B. Genetic architecture of the dog: sexual size dimorphism and functional morphology. *Trends Genet.* **22**, 537–544 (2006).
43. Sutter, N. B. *et al.* A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112–115 (2007).
44. Jones, P. *et al.* Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics* **179**, 1033–1044 (2008).
45. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
46. Altshuler, D. & Daly, M. Guilt beyond a reasonable doubt. *Nature Genet.* **39**, 813–815 (2007). **A good recent review of human whole-genome association results**
47. Minnick, M. F., Stillwell, L. C., Heineman, J. M. & Stiegler, G. L. A highly repetitive DNA sequence possibly unique to canids. *Gene* **110**, 235–238 (1992).
48. Bentolila, S. *et al.* Analysis of major repetitive DNA sequences in the dog (*Canis familiaris*) genome. *Mamm. Genome* **10**, 699–705 (1999).
49. Vassetzky, N. S. & Kramerov, D. A. CAN — a pan-carnivore SINE family. *Mamm. Genome* **13**, 50–57 (2002).
50. Pele, M., Tiret, L., Kessler, J. L., Blot, S. & Panthier, J. J. SINE exonic insertion in the *PTPLA* gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. *Hum. Mol. Genet.* **14**, 1417–1427 (2005).
51. Clark, L. A., Wahl, J. M., Rees, C. A. & Murphy, K. E. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proc. Natl Acad. Sci. USA.* **103**, 1376–1381 (2006).
52. Lohi, H. *et al.* Expanded repeat in canine epilepsy. *Science* **307**, 81 (2005). **This article describes the identification of the first canine repeat length polymorphism.**
53. Kazemi-Esfarjani, P., Trifiro, M. A. & Pinsky, L. Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. *Hum. Mol. Genet.* **4**, 523–527 (1995).

54. Steingrimsson, E., Copeland, N. G. & Jenkins, N. A. Melanocytes and the microphthalmia transcription factor network. *Annu. Rev. Genet.* **38**, 365–411 (2004).
55. Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. & Goldfarb, M. Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246–249 (1995).
56. Powles, N. *et al.* Regulatory analysis of the mouse Fgf3 gene: control of embryonic expression patterns and dependence upon sonic hedgehog (Shh) signalling. *Dev. Dyn.* **230**, 44–56 (2004).
57. Wright, T. J. *et al.* Mouse FGF15 is the ortholog of human and chick FGF19, but is not uniquely required for otic induction. *Dev. Biol.* **269**, 264–275 (2004).
58. Giuffra, E. *et al.* A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm. Genome* **13**, 569–577 (2002).
59. Parker, H. G. *et al.* Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**, 1562–1571 (2007).
A description of haplotype sharing in different breeds affected by Collie eye anomaly.
60. Van Laere, A. S. *et al.* A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836 (2003).
61. Eriksson, J. *et al.* Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* **4**, e1000010 (2008).
62. Rosengren Pielberg, G. *et al.* A *cis*-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genet.* **40**, 1004–1009 (2008).
63. Schuster, S. C. Next-generation sequencing transforms today's biology. *Nature Methods* **5**, 16–18 (2008).
64. Margulies, E. H. *et al.* An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *Proc. Natl Acad. Sci. USA* **102**, 4795–4800 (2005).
65. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
66. The Broad Institute. Mammalian Genome Project. [online], <http://www.broad.mit.edu/mammals> (2008).
67. Diamond, J. M. *Guns, Germs, and Steel: the Fates of Human Societies* (W. W. Norton & Co., New York, 1997).
68. Willham, R. L. From husbandry to science: a highly significant facet of our livestock heritage. *J. Anim. Sci.* **62**, 1742–1758 (1986).
69. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
70. Sutter, N. & Ostrander, E. Dog star rising: The canine genetic system. *Nature Rev. Genet.* **5**, 900–910 (2004).
71. Gershwin, L. J. Veterinary autoimmunity: autoimmune diseases in domestic animals. *Ann. N. Y Acad. Sci.* **1109**, 109–116 (2007).
72. Khanna, C. *et al.* The dog as a cancer model. *Nature Biotechnol.* **24**, 1065–1066 (2006).
73. Loscher, W., Schwartz-Porsche, D., Frey, H. H. & Schmidt, D. Evaluation of epileptic dogs as an animal model of human epilepsy. *Arzneimittelforschung* **35**, 82–87 (1985).
74. Overall, K. L. Natural animal models of human psychiatric conditions: assessment of mechanism and validity. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **24**, 727–776 (2000).
75. Vail, D. M. & MacEwen, E. G., Spontaneously occurring tumors of companion animals as models for human cancer. *Cancer Invest.* **18**, 781–792 (2000).
76. Mueller, F., Fuchs, B. & Kaser-Hotz, B. Comparative biology of human and canine osteosarcoma. *Anticancer Res.* **27**, 155–164 (2007).
77. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
78. Glickman, L., Glickman, N. & Thorpe, R. The Golden Retriever Club of America National Health Survey 1998–1999 [online], <http://www.grca.org/healthsurvey.pdf> (1999).
79. Hansen, K. & Khanna, C. Spontaneous and genetically engineered animal models; use in preclinical cancer drug development. *Eur. J. Cancer* **40**, 858–880 (2004).
80. Khanna, C. *et al.* A randomized controlled trial of octreotide pamoate long-acting release and carboplatin versus carboplatin alone in dogs with naturally occurring osteosarcoma: evaluation of insulin-like growth factor suppression and chemotherapy. *Clin. Cancer Res.* **8**, 2406–2412 (2002).
81. Acland, G. M. *et al.* Long-term restoration of rod and cone vision by single dose rAAV-mediated gene transfer to the retina in a canine model of childhood blindness. *Mol. Ther.* **12**, 1072–1082 (2005).
82. Acland, G. M. *et al.* Gene therapy restores vision in a canine model of childhood blindness. *Nature Genet.* **28**, 92–95 (2001).
83. Bennicelli, J. *et al.* Reversal of blindness in animal models of leber congenital amaurosis using optimized AAV2-mediated gene transfer. *Mol. Ther.* **16**, 458–465 (2008).
84. Hillbertz, N. H. & Andersson, G. Autosomal dominant mutation causing the dorsal ridge predisposes for dermoid sinus in Rhodesian ridgeback dogs. *J. Small Anim. Pract.* **47**, 184–188 (2006).
85. Copp, A. J., Greene, N. D. & Murdoch, J. N. The genetic basis of mammalian neurulation. *Nature Rev. Genet.* **4**, 784–793 (2003).
86. Ackerman, L. L. & Menezes, A. H. Spinal congenital dermal sinuses: a 30-year experience. *Pediatrics* **112**, 641–647 (2003).
87. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
88. Isaksson, M. *et al.* MLGA a rapid and cost-efficient assay for gene copy-number analysis. *Nucleic Acids Res.* **35**, e115 (2007).
89. Wang, Y., Badea, T. & Nathans, J. Order from disorder: self-organization in mammalian hair patterning. *Proc. Natl Acad. Sci.* **103**, 19800 (2006).
90. Ornitz, D. M. & Itoh, N. Fibroblast growth factors. *Genome Biol.* **2**, REVIEWS3005 (2001).
91. Dourmishev, A. L., Dourmishev, L. A., Schwartz, R. A. & Janniger, C. K. Waardenburg syndrome. *Int. J. Dermatol.* **38**, 656–663 (1999).
92. Tietz, W. A syndrome of deaf-mutism associated with albinism showing dominant autosomal inheritance. *Am. J. Hum. Genet.* **15**, 259–264 (1963).
93. Little, C. C. *The Inheritance of Coat Color in Dogs* (Comstock Pub. Associates, Ithaca, New York, 1957).
94. Metallinos, D. & Rine, J. Exclusion of EDNRB and KIT as the basis for white spotting in Border Collies. *Genome Biol.* **1**, RESEARCH0004 (2000).
95. van Hagen, M. A. *et al.* Analysis of the inheritance of white spotting and the evaluation of KIT and EDNRB as spotting loci in Dutch boxer dogs. *J. Hered.* **95**, 526–531 (2004).
96. Smith, S. D., Kelley, P. M., Kenyon, J. B. & Hoover, D. Tietz syndrome (hypopigmentation/deafness) caused by mutation of MITF. *J. Med. Genet.* **37**, 446–448 (2000).
97. Tassabehji, M., Newton, V. E. & Read, A. P. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nature Genet.* **8**, 251–255 (1994).
98. Goldstein, O. *et al.* Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. *Genomics* **18**, 18 (2006).
99. Bradbury, J. Canine epilepsy gene mutation identified. *Lancet Neurol.* **4**, 143 (2005).
100. Nadon, N. L., Duncan, I. D. & Hudson, L. D. A point mutation in the proteolipid protein gene of the 'shaking pup' interrupts oligodendrocyte development. *Development* **110**, 529–537 (1990).
101. Lin, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365–376 (1999).
102. Katz, M. L. *et al.* A mutation in the CLN8 gene in English Setter dogs with neuronal ceroid-lipofuscinosis. *Biochem. Biophys. Res. Commun.* **327**, 541–547 (2005).
103. Suber, M. L. *et al.* Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta-subunit gene. *Proc. Natl Acad. Sci. USA* **90**, 3968–3972 (1993).
104. Petersen-Jones, S. M., Entz, D. D. & Sargan, D. R. cGMP phosphodiesterase-alpha mutation causes progressive retinal atrophy in the Cardigan Welsh corgi dog. *Invest. Ophthalmol. Vis. Sci.* **40**, 1637–1644 (1999).
105. Kijas, J. W. *et al.* Naturally occurring rhodopsin mutation in the dog causes retinal dysfunction and degeneration mimicking human dominant retinitis pigmentosa. *Proc. Natl Acad. Sci. USA* **99**, 6328–6333 (2002).
106. Veske, A., Nilsson, S. E., Narfstrom, K. & Gal, A. Retinal dystrophy of Swedish briard/briard-beagle dogs is due to a 4-bp deletion in RPE65. *Genomics* **57**, 57–61 (1999).
107. Fyfe, J. C. *et al.* The functional cobalamin (vitamin B12)-intrinsic factor receptor is a novel complex of cubilin and amnionless. *Blood* **103**, 1573–1579 (2004).
108. Baldeschi, C. *et al.* Genetic correction of canine dystrophic epidermolysis bullosa mediated by retroviral vectors. *Hum. Mol. Genet.* **12**, 1897–1905 (2003).
109. Capt, A. *et al.* Inherited junctional epidermolysis bullosa in the German Pointer: establishment of a large animal model. *J. Invest. Dermatol.* **124**, 530–535 (2005).
110. Kramer, J. W. *et al.* A von Willebrand's factor genomic nucleotide variant and polymerase chain reaction diagnostic test associated with inheritable type-2 von Willebrand's disease in a line of German shorthaired pointer dogs. *Vet. Pathol.* **41**, 221–228 (2004).
111. Venta, P. J., Li, J., Yuzbasiyan-Gurkan, V., Brewer, G. J. & Schall, W. D. Mutation causing von Willebrand's disease in Scottish Terriers. *J. Vet. Intern. Med.* **14**, 10–19 (2000).
112. Pele, M., Tiret, L., Kessler, J. L., Blot, S. & Panthier, J. J. SINE exonic insertion in the *PTPLA* gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. *Hum. Mol. Genet.* **14**, 1417–1427 (2005).
113. Sharp, N. J. *et al.* An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. *Genomics* **13**, 115–121 (1992).
114. Mealey, K. L., Bentjen, S. A., Gay, J. M. & Cantor, G. H. Ivermectin sensitivity in collies is associated with a deletion mutation of the mdr1 gene. *Pharmacogenetics* **11**, 727–733 (2001).
115. Kijas, J. M. *et al.* A missense mutation in the beta-2 integrin gene (ITGB2) causes canine leukocyte adhesion deficiency. *Genomics* **61**, 101–107 (1999).
116. Henthorn, P. S. *et al.* IL-2R gamma gene microdeletion demonstrates that canine X-linked severe combined immunodeficiency is a homologue of the human disease. *Genomics* **23**, 69–74 (1994).
117. Benson, K. F. *et al.* Mutations associated with neutropenia in dogs and humans disrupt intracellular transport of neutrophil elastase. *Nature Genet.* **35**, 90–96 (2003).
118. Zheng, K., Thorner, P. S., Marrano, P., Baumal, R. & McInnes, R. R. Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked hereditary nephritis resulting from a single base mutation in the gene encoding the alpha 5 chain of collagen type IV. *Proc. Natl Acad. Sci. USA* **91**, 3989–3993 (1994).
119. Everts, R. E., Rothuizen, J. & van Oost, B. A. Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (MC1R) in Labrador and Golden retrievers with yellow coat colour. *Anim. Genet.* **31**, 194–199 (2000).
120. Newton, J. M. *et al.* Melanocortin 1 receptor variation in the domestic dog. *Mamm. Genome* **11**, 24–30 (2000).
121. Clark, L. A., Wahl, J. M., Rees, C. A. & Murphy, K. E. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proc. Natl Acad. Sci. USA* **103**, 1376–1381 (2006).
122. Mosher, D. S. *et al.* A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* **3**, e79 (2007).
123. Candille, S. I. *et al.* A β-defensin mutation causes black coat color in domestic dogs. *Science* **318**, 1418–1423 (2007).

Acknowledgements
We thank L. Andersson for helpful comments on the manuscript and our colleagues in the canine genetics community.

FURTHER INFORMATION
Kerstin Lindblad-Toh's homepage: http://www.broad.mit.edu/about/bios/bio-lindblad-toh.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF