LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs

Anne-Sophie Lequarré a,⁎, Leif Andersson b, Catherine André c, Merete Fredholm d, Christophe Hitte c, Tosso Leeb e, Hannes Lohi f, Kerstin Lindblad-Toh b,g, Michel Georges a

a Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, Avenue de l'Hôpital 1, 4000 Liège, Belgium
b Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
c Institut de Génétique et Développement de Rennes, CNRS-UMR6061, Université de Rennes 1, F-35000 Rennes, France
d Faculty of Life Sciences, Division of Genetics and Bioinformatics, Department of Basic Animal and Veterinary Sciences, University of Copenhagen, Frederiksberg C, Denmark
e Institute of Genetics, Vetsuisse Faculty, University of Bern, Switzerland
f Department of Veterinary Biosciences, Department of Medical Genetics, Program in Molecular Medicine, Folkhälsoan Institute of Genetics, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland
g Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA

ARTICLE INFO
Keywords:
Canine
Inherited disorders
Genetics
LUPA
Single nucleotide polymorphism arrays

ABSTRACT
The domestic dog offers a unique opportunity to explore the genetic basis of disease, morphology and behaviour. Humans share many diseases with our canine companions, making dogs an ideal model organism for comparative disease genetics. Using newly developed resources, genome-wide association studies in dog breeds are proving to be exceptionally powerful. Towards this aim, veterinarians and geneticists from 12 European countries are collaborating to collect and analyse the DNA from large cohorts of dogs suffering from a range of carefully defined diseases of relevance to human health. This project, named LUPA, has already delivered considerable results. The consortium has collaborated to develop a new high density single nucleotide polymorphism (SNP) array. Mutations for four monogenic diseases have been identified and the information has been utilised to find mutations in human patients. Several complex diseases have been mapped and fine mapping is underway. These findings should ultimately lead to a better understanding of the molecular mechanisms underlying complex diseases in both humans and their best friend.

© 2011 Elsevier Ltd. All rights reserved.

Introduction
Purebred dogs have been strongly selected for specific morphological, physiological and behavioural traits. The level of diversity observed between dog breeds is unique amongst mammalian species and many specific features are inherited in very close limits within a breed (Parker and Ostrander, 2005). This was largely achieved by the extensive use of founder dogs expressing the desired traits, which inadvertently also increased the frequency of the undesirable genes carried by those founders (Asher et al., 2009; Chase et al., 2009). Recessive as well as dominant defects with incomplete penetrance or late-onset and complex diseases are numerous in dogs. Breeding practices have raised concerns about the health status and well being of purebred dogs (Higgins and Nicholas, 2008), urging the search for solutions.

Inherited canine disorders include structural or metabolic defects, cardiovascular or neurological diseases, inflammatory disorders with immune system failures and cancers. All these conditions are equally common in humans and the clinical signs in dogs often mimic human diseases closely (Tsai et al., 2007). As a result of the relatively low effective population size within many dog breeds (compared with humans), the within-breed genetic architecture is predicted to be simplified and to involve larger gene effects (since these are less effectively counter-selected in smaller populations), which are likely to be easier to detect (Sutter et al., 2004; Karlsson et al., 2007).

Susceptibility genes identified in the dog will not only lead to diagnostic tools for breeders, but will give novel insights in disease pathogenesis and hence therapeutic opportunities, as well as to generate hypotheses to investigate the genetic susceptibility of similar diseases in humans. The latter is greatly aided by the fact that dogs and humans share roughly the same gene repertoire (Lindblad-Toh et al., 2005; Durrien et al., 2009).

Anticipating these opportunities, the dog was the fourth mammal for which a high quality genome sequence was published. Concomitantly, a large number of single nucleotide polymorphism (SNPs; >2 million) were discovered and the architecture of the genome for which a high quality genome sequence was published. Concomitantly, a large number of single nucleotide polymorphism (SNPs; >2 million) were discovered and the architecture of the genome
Linkage disequilibrium (LD) was shown to extend over much longer distances within breeds (as a result of the founder effects accompanying breed formation), hence drastically reducing the number of markers needed for genome-wide association studies (GWAS) compared to humans. However, across breeds, LD was shown to decay more than one breed (Sutter et al., 2004; Lindblad-Toh et al., 2005; Karlsson and Lindblad-Toh, 2008).

**LUPA: A European initiative**

To enhance the use of the dog as an effective model to study common complex diseases in human, a consortium of canine geneticists and clinicians representing 12 countries was formed and has been funded since 2008 by the European Commission (EC). The project is named LUPA, after the female wolf which fed the twins Romulus and Remus, the founders of Roma, to highlight how humans may benefit from genetic studies performed on their best friend. The 22 collaborating veterinary faculties and research centres have collected large dog cohorts for more than 20 carefully defined diseases of relevance to human health.

LUPA targets five partially overlapping disease categories: cancers, cardiovascular disorders, inflammatory disorders, neurological disorders and monogenic disorders. Mammary tumours, melanomas, soft tissue sarcomas and mast cell tumours are the cancers ‘under the microscope’. Among cardiovascular disorders, myxomatous mitral valve disease, as well as dilated cardiomyopathy, are currently being explored in several large breeds. A large study on the genetic basis of common physiological cardiovascular parameters in healthy dogs from 10 different breeds is a novel and promising approach to look for genetic variants underlying the variability of blood pressure, cardiac morphology, levels of hormones and metabolites. Among inflammatory disorders, the project is tackling different forms of diabetes (mainly dioestrus/gestational diabetes and autoimmune insulin deficiency diabetes), atopic dermatitis, thyroiditis and resistance to leishmaniasis. Neurological diseases include epilepsies in a range of breeds, as well as some behavioural disorders.

Finally, the LUPA programme includes the study of several defects assumed to be monogenic, which should rapidly lead to the discovery of causative mutations. It is expected that this will demonstrate the effectiveness of the dog model, provide diagnostic tests for the elimination of certain defects in the afflicted breeds and possible assist in resolving ‘orphan’ genetic diseases in humans.

The collection of cohorts is based on rigorous clinical characterisation using standardised protocols shared by the clinical teams involved in the study. For complex disorders, simulations demonstrated that 100–200 cases with the same number of healthy dogs should provide adequate power to detect loci conferring a 3–5-fold increased risk (Lindblad-Toh et al., 2005). DNA from cohorts reaching the required size are sent to one centralised, high-throughput SNP genotyping facility (Centre National de Genotypage, CNG, Paris) to ensure homogeneous genotypes. The initial 50 K Affymetrix array was replaced in 2009 by the now preferred 172 K array.

**Generation of a high density canine SNP array**

The generation of a novel high density canine SNP array ranks amongst the major early achievements of the LUPA project. Genomic regions that were underrepresented in the previous SNP collection (Lindblad-Toh et al., 2005), including the X chromosome, were captured out of multibreed DNA pools and resequenced (A. Vaysse personal communication). A selection of existing and newly discovered SNPs was included in a set of 172 K SNPs for the design of the new Illumina high-density array. The resulting SNP collection provides more uniform and dense coverage (average distance 13 kb between adjacent SNPs), The array performs remarkably well with regards to call rate, reproducibility and Mendelian consistency (A. Vaysse personal communication).

**Monogenic disorders**

Other early successes of LUPA include the elucidation of the molecular basis of four monogenic diseases.

**Primary ciliary dyskinesia**

Primary ciliary dyskinesia (PCD) is a respiratory disease segregating in Old English sheep dogs, amongst other breeds. PCD is a heterogeneous genetic disease resulting from defects in the structure and/or motility of cilia. The analysis of five affected and 15 healthy dogs identified a 15 Mb segment on chromosome 34 for which all affected dogs shared autozygosity. The analysis of six candidate genes, selected with information from ciliome databases from the 151 genes in the interval, led to the identification of a disrupting mutation in CDC39, a gene not previously implicated in PCD.

We studied the integrity of the human orthologue in unresolved human PCD cases with similar ultrastructural defects in cilia. Altogether, 15 different loss-of-function mutations were found, explaining half of the examined cases, or approximately 8% of all ‘orphan’ human PCD cases (Merveille et al., 2011). The prevalence of this disease appears to be high in European Old English sheep dogs (A.C. Merveille personal communication), but the diagnostic test that has been developed should be effective to control the disease if applied thoughtfully.

**Ichthyosis**

Following a similar approach, the gene causing ichthyosis in Golden retrievers was found using 20 cases and 20 controls. The disease is characterised by mild to severe abnormal desquamation across the whole body. A 2 Mb haplotype on canine chromosome 12 was identified, for which all affected dogs shared homozygosity. It contained 80 genes, of which three were good candidates. A disruptive mutation in one of these was shown to be perfectly associated with the disease, yet never found in any other breed, including other retrievers.

A canine genetic test has been developed, with the submission of an international patent. Sequencing the human orthologue in 10 unrelated unresolved cases revealed two different mutations in two families with lamellar ichthyosis (C. André, personal communication). The gene, which had not previously been associated with any dermatological condition, was shown to play an important role in the cutaneous barrier function.

**Neuronal ceroid lipofuscinosis**

A disease similar to a neuronal ceroid lipofuscinosis (NCL), which segregated in a pedigree of American Staffordshire terriers, was fine-mapped using the 172 K array (Abitbol et al., 2010). In hu-
mans, NCLs represent a heterogeneous group of genetic neurodegenerative diseases due to storage of intermediate products of biochemical pathways in lysosomes of primarily neurones. The aetiology of adult forms remains to be established. American Staffordshire terriers suffer from a late onset form of ataxia starting between 3 and 5 years of age and characterised by a cerebellar atrophy. Histopathological analysis revealed marked loss of Purkinje cells, with intracellular accumulation of autofluorescent storage material, a hallmark of NCL. The disease segregates as an autosomal recessive trait.

Through combined linkage and association analysis using approximately 300 microsatellites on 60 cases and 60 controls, the NCL locus was mapped to chromosome 9. Ten control and 12 NCL dogs were genotyped afterwards with the 172 K Illumina array to identify additional informative markers. A 1039 kb critical region was defined that spanned nine protein-coding genes. A breed-specific non-synonymous substitution in exon 2 of the arylsulphatase G (ARSG) gene was detected in two copies in all affected dogs and in one copy in healthy carriers. Functional assays demonstrated that the resulting mutation of a highly-conserved amino acid leads to a 75% decrease in the enzymatic activity of the ARSG protein. The ARSG gene, not yet involved in human NCLs, is presently being investigated as a candidate in human NCL patients.

Channelopathies

LUPA partners identified a mutation in a non-ion channel gene causing juvenile epilepsy in the Lagotto Romagnolo breed (Jokinen et al., 2007; Seppälä et al., 2011). Focal seizures in affected dogs occur at 5–9 weeks of age and resolve spontaneously by 13 weeks of age. This pattern is reminiscent of common human childhood epilepsies with remission and the identified gene is currently being screened in cohorts of human patients. In addition to being the first gene for idiopathic epilepsy found in the dog, this study identified a novel pathway causing epilepsy when perturbed and provides a useful large animal model to study the molecular mechanisms leading to common remitting epilepsies in the developing brain (Seppälä et al., 2011).

Complex diseases

Using SNP genotypes acquired with the novel 172 K array on larger cohorts, significant genome-wide associations have already been obtained for several complex disorders.

Dilated cardiomyopathy

Dilated cardiomyopathy is a condition observed both in humans and several large dog breeds. LUPA is studying this disease in four breeds: Great Danes, Newfoundlands, Dobermans and Irish wolfhounds. In the Doberman pinscher, the disease is common and is apparently inherited in an autosomal dominant manner (Meurs et al., 2007). Heart remodelling is preceded by detectable modifications of muscular electroactivity (Wess et al., 2010). Using old unaffected dogs as controls, a highly significant association with dilated cardiomyopathy has been detected on chromosome 5 in Dobermans (Mausberg et al., 2011).

Myxomatous mitral valve disease

Myxomatous mitral valve disease (MMVD) is another cardiac condition that is very frequent in humans. This syndrome is characterised by the enlargement of the mitral valve leading eventually to its prolapse into the left atrium on ventricular contraction (mitral regurgitation). The causes of MMVD in humans remain unknown, but the disease is clustered in families. A similar disease has a very high prevalence in Cavalier King Charles spaniels and GWAS conducted with 100 cases and 100 controls have recently identified two novel candidate regions. Studies are currently underway to identify the causative gene and mutation within the regions (Madsen et al., 2011).

Intervertebral disc disease

Another promising study pertains to intervertebral disc herniation. This condition commonly affects Dachshunds, with a continuous spectrum of disc degeneration, suggesting a complex aetiology. Disease severity can be quantified from the number of calcified discs. Using 50 cases and 50 controls, a major locus significantly associated with the condition has been identified on chromosome 12. The region is currently being re-sequenced to identify variants and to study their involvement in the pathology (Mogensen et al., 2011).

Idiopathic epilepsy

Several idiopathic epilepsies (IE) in the dog are being studied and novel IE loci have been mapped to several different chromosomes for various breeds, including Belgian shepherds, Norwich and Border terriers, the Schipperke and the Finnish Spitz. Sequence capture and next-generation sequencing are currently being applied to screen the associated regions for mutations. The ongoing studies indicate that most IEs in dogs are genetically complex and require large cohorts and careful phenotyping to yield validated results. Interestingly, the newly mapped loci do not contain any known epilepsy genes. This study may therefore lead to the discovery of novel genes for common forms of human epilepsy.

Other studied diseases, including inflammatory disorders and cancers, show either suggestive or significant associations to multiple loci and are currently undergoing additional research. In some cohorts, the stratification that is sometimes present even within canine breeds may impede the appearance of clear signals. Case-control studies are usually designed so that only dogs that are unrelated at the parental or grandparental level are included, but this may not completely control the effect of the complex pedigree structure in dogs. Cautious genetic design and analytical approaches that explicitly take account of the pedigree structure, such as GenAbel (Aulchenko et al., 2007), EMMA-X (Kang et al., 2008) and GCTA (Yang et al., 2011) are currently being exploited.

Copy number variation (CNV) in dogs

An additional initiative of the consortium has been to characterise the number and position of copy number variants (CNVs) in dog breeds studied by LUPA. The study aimed to map CNVs in >50 dogs from 17 breeds and three wolves by comparative genome hybridisation (aCGH). A high-density Nimblegen array with 2.1 million probes was used, giving a resolution of approximately 1 kb spacing across the canine genome. Structural changes may underlie phenotypic variation, as extensively documented in domestic animals. For instance, dominant white colour in pigs (Moller et al., 1996), the hair ridge in Rhodesian Ridgeback dogs (Salmon-Hillertz et al., 2007), greying with age in horses (Rosengren Pielberg et al., 2008) and familial Shar-Pei fever (Olsson et al., 2011) are all caused by duplications.

Preliminary analysis of the aCGH data has revealed thousands of CNVs. Analyses are being conducted to assess the number of CNVs shared between breeds or specific to a given breed, as well as to identify the possible biological impact of CNVs based on gene content and biological pathways. The aCGH information aids in the
interpretation of fluorescence intensities from SNP array data to determine the CNV genotype of individual dogs and its possible correlation with phenotype and/or disease status.

Detecting signatures of selection

The LUPA database, already containing genotypes from 10,000 dogs, can be used to perform GWAS across breeds, with the aim to identify variants underlying traits shared by several breeds, as has already been documented in other studies (Jones et al., 2008; Cadieu et al., 2009; Akey et al., 2010; Bannasch et al., 2010; Boyko et al., 2010). An alternative approach searches for selective sweeps: regions with reduced heterozygosity associated with a specific trait within a single or within a group of breeds that share the same trait (A. Vaysse personal communication).

A recent LUPA collaborative effort (Olsson et al., 2011) has taken advantage of this selective-sweep approach to identify the region underlying the thickened and wrinkled skin of Chinese Shar-Pei dogs. This approach was combined with a traditional GWAS to identify the gene for familial Shar-Pei fever, a form of periodic fever. The region is located on chromosome 13 near the HAS2 gene, a rate-limiting enzyme for the synthesis of hyaluronic acid (HA). HA is a major constituent of mucin in the thickened skin typically observed in Shar-Peis. Shar-Peis harbour a duplication 350 kb upstream of the HAS2 gene; a high copy number of this duplication is associated with both increased HAS2 expression and the periodic fever syndrome. These results call for studies exploring the role of hyaluronic in human periodic fevers.

Conclusions

The 4-year European LUPA project has already fulfilled many of its objectives, elucidating the molecular basis of four monogenic diseases, including the discovery of novel genes shown to be involved in the equivalent human pathologies. Canine tests have been conducted concomitantly to help reduce the frequency of these diseases in the relevant breeds. LUPA has also improved canine genetics tools by designing a new highly efficient 172 K Illumina array. Both the array and CNV characterisation help to better mine disease susceptibility concealed in the canine genome.

Using these tools, LUPA partners have already identified loci associated with susceptibility to several complex disorders. More results should follow in the coming months and years. Fine-mapping of the associated genomic regions and high throughput sequencing experiments to find the causative genes and mutations are ongoing. Even if it may take some time to identify the actual causative mutations and/or precise molecular mechanisms leading to disease (especially when dealing with regulatory mutations), identifying the causative genes or pathways, which is relatively easier, may on its own pave the way for the investigation of similar pathways in the corresponding human disorders (Wilke et al., 2010), as well as for the identification of novel drug targets. Personalised medicine for dogs, based on the genes and pathways underlying the development of specific disorders, may be a future possibility for veterinary medicine. We also envisage clinical trials in dogs before translating them to humans, in order to test new therapies for the benefit of both species.

A major achievement of LUPA is to have greatly improved the dialogue between veterinary clinicians and geneticists throughout Europe and the rest of the world. The benefit of collaborating and building cohorts in a multicentric fashion has become obvious for everyone involved in the project. This was quite difficult before LUPA. The initiative has also been well received by breeders and kennel clubs, whose collaboration is essential for improving detection of disease syndromes and for sample collection. Dog owners and the general public have been far better informed, encouraging them to contribute to ongoing and new studies. Human geneticists and clinicians have shown increasing interest in the remarkable genetic relevance of the dog to its masters.

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

Acknowledgements

The work presented in this paper was carried out as a part of the LUPA Project, which was funded by the EC under the Health theme of the Seventh Framework Program (FP7, GA: 201370). The authors acknowledge all the members of the LUPA consortium, as well as the veterinarians, the breeders and the owners who actively participated in sample collection.

References


