

The Canine Genome Sequence and Its Implications for Finding Disease Genes

Kerstin Linblad-Toh, PhD; Broad Institute, MIT

The canine genome was sequenced for two reasons; to enable disease gene mapping in dogs and to help identify all the important features, including genes and regulatory elements, in the human and other mammalian genomes.

The genome sequence of a female boxer was generated. Each position in the genome was sampled ~7.5 times, which means that the genome is relatively complete (~99%). The genome has been compared to other mammalian genomes such as human, mouse and rat.

The genes have been identified and the majority of these have a corresponding gene in other mammalian genomes. Thus, a disease gene identified in one species can be studied also in other mammals, where it is likely to cause similar disease. A number of possible regulatory elements have been identified by comparison of the mammals. Mutations in these elements as well as in genes are more likely to cause disease than those found elsewhere in the genome.

A map of several million single base variants (markers or SNPs) has been generated. Small amounts of sequence generated from ten dogs from other breeds, 4 wolves and one coyote were compared to the boxer sequence to generate this map.

The structure of variation within breeds suggests that mapping disease genes in breeds that have a high incidence of a particular disease will be feasible. In fact, a disease with a high breed predisposition (>5 fold higher than average breeds) could be mapped using 100-200 affected and 100-200 unaffected controls within that breed, whereas fully penetrant autosomal recessive traits would require only a few dozen samples. The structure of variation also suggests that many genes predisposing to disease will be shared among breeds. Thus, once a region predisposing to disease within one breed has been identified, several breeds with the same phenotype should be used to pinpoint the actual disease gene. Efforts to find genes in one breed are therefore likely to benefit multiple breeds.

A mapping array containing ~15,00 SNPs has been generated and is currently being evaluated. This array will be commercially available for all scientists and should be an excellent tool for mapping disease genes.

The Broad Institute is currently collecting samples from dogs affected with osteosarcoma, hemangiosarcoma, melanoma, mammary carcinoma, lymphoma and mast cell tumors. We also need older unaffected dogs to use as controls (see www.broad.mit.edu/mammals/dog/ for more details). With genomic tools in place, collection of well-diagnosed samples is the next key step towards identifying canine disease genes.