Sequencing the Genome of the Domestic Dog *Canis familiaris*

By

Elaine A. Ostrander[1], Kerstin Lindblad-Toh[2], Eric S. Lander[2]

[1]Fred Hutchinson Cancer Research Center
1100 Fairview Avenue N., D4-100
Seattle, WA 98109-1024
206-667-6979
eostrand@fhcrc.org

[2]Whitehead/MIT Center for Genome Research
320 Charles Street
Cambridge, MA 02139
617-252-1906
kersli@genome.wi.mit.edu
eric@genome.wi.mit.edu

And

The Canine Genome Mapping Community:

Gregory M. Acland, Cornell University
Gustavo D. Aguirre, Cornell University
Matthew M. Binns, Animal Health Trust, England
Matthew Breen, Animal Health Trust, England
Linda C. Cork, Stanford University
Francis Galibert, University of Rennes, France
Urs Giger, University of Pennsylvania
Paula S. Henthorn, University of Pennsylvania
Gary F. Johnson, University of Missouri
K. Gordon Lark, University of Utah
Frode Lingaas, Norwegian School of Veterinary Science, Norway
W. Richard McCombie, Cold Spring Harbor Laboratory
James R. Mickelson, University of Minnesota
Keith E. Murphy, Texas A and M University
Donald F. Patterson, University of Pennsylvania
Rainer Storb, Fred Hutchinson Cancer Research Center
Patrick J. Venta, Michigan State University
Robert K. Wayne, University of California, Los Angeles

INTRODUCTION

Suppression, selection, and mixing within the wolf gene pool have yielded hundreds of breeds of domestic dog, nearly 300 of which are alive in the world today. Dogs are unique among mammalian species in the extent of variation they show in morphological traits such as height, weight, mass, shape, and behavior, yet within each breed, key traits are inherited within extremely narrow limits (1-6). The Chihuahua is less than six inches high at the shoulder; the Irish wolfhound close to three feet. The Pomeranian weighs between four and five pounds; the St. Bernard may weigh 150 pounds. The Collie has a small, narrow head like that of a fox; the Pug has a massive head with a short, blunt muzzle. Dog breeds exist which have been purpose bred for guarding, hunting, herding, driving, pulling, etc. No other mammalian species presents natural variation on a scale to rival these, yet individuals from nearly any breeds can be mated to yield fertile offspring.

Given the aggressive breeding programs needed to reproducibly generate animals of distinctive size, shape and behavior, it is not surprising that purebred dog fanciers have also produced closed breeding populations, characterized by over 400 inherited disorders. Genetic diseases are predicted to occur with high frequency in populations with closed gene pools and in which breeding of close relatives is used to propagate desired traits. Breeds established from a small number of founders and expanded rapidly to meet breeders' and consumers' demands suffer the most. Autosomal recessive and complex traits present the biggest problem as the status of asymptotic carriers may not be suspected until several litters have been produced. This includes diseases such as cancer, heart disease, deafness, blindness, motor neuron disease, skin disorders, and a host of autoimmune disorders, each of which has been difficult to study in humans (1-6).

In the following White Paper we argue for the development of a 6x assembled sequence of the dog, much as is currently available for the mouse, to facilitate the identification of genes that explain the enormous diversity of mammalian variation observed in nature today, and to better understand the genetic basis of complex diseases affecting both human and dog. In Section A we discuss some advantages of genetic mapping in dogs and provide examples where canine genetics is best suited to solve problems associated with mapping complex traits. We next turn to biomedical issues highlighting the contribution of the dog in the development of protocols for bone marrow transplantation and gene therapy. Finally, we address the limitations that the canine genome mapping community currently face in exploiting this canine diversity, and consider the benefits and advances that an essentially complete canine genome sequence are anticipated to yield.

In Section B we consider the strategic issues associated with sequencing the dog genome. These include a discussion of major projects currently funded that would be informed by the 6x sequence, the size and commitment of the canine research community, a proposed approach for sequencing, and a summary of current genomic resources that will prove useful for the sequencing effort. We comment, finally, on the likely very enthusiastic public response to this effort with a series of supportive letters included in the Appendicies.

A. Specific Biological Rationales for New Sequence Data

A.1. Utility of the Canine System for Genetic Mapping of Disease Traits

A.1.a. Mapping and Cloning Genetic Disorders in Dogs that Parallel Human Disease

In humans, common diseases show complex modes of inheritance, and as a result, they are frequently largely refractory to genetic analysis. Rodent systems are more tractable genetically, but the mutations typically represent induced rather than naturally arising alleles, and results are often of limited direct relevance to human disease because of profound differences in physiology. In contrast, the physiology, disease presentation and clinical response of dogs often mimic human diseases closely. Thus, positional cloning of genetic diseases in the dog can inform us about the genetic susceptibility to similar diseases in humans.

The structure of the canine population offers specific advantages for genetic mapping studies. Over one million purebred dogs are newly registered in the U.S. each year (7). These are divided into some 300 partially inbred genetic isolates called “breeds.” Gene flow between breeds is restricted by the pedigree barrier—registering a dog as a member of a particular breed requires that both of the dog’s parents be registered

members of that same breed. Most modern dog breeds are relatively young, with the majority having been developed within the last 300 years (8, 9 and refs therein). Many of these were derived from a small number of founders—as few as six, for instance, in the case of the modern Irish Wolfhound—that best represented the physical or behavioral traits breeders wished to feature. The natural history of some breeds has further restricted their genetic diversity over what is expected from breeding strategies alone. Catastrophic events in the last 100 years, such as the two world wars and the American depression, have produced severe bottlenecks in many breeds, at times reducing the effective breeding stock to only a few dogs. At the end of WWI, for instance, only five dogs of the Leonburger breed remained alive in Europe (9), and all Leonburgers alive today are believed to be descendents of those five. Genetic diversity in some breeds is further reduced by the presence of "popular sires." These dogs have physical features which make them particularly successful in the show ring and hunting or performance events, and as a result they may produce over 100 litters in their lifetime. For many breeds, therefore, modern purebred dogs represent a limited genetic pool, with disease predispositions derived from one or a small number of recent genetic founders. Thus, modern dog breeds offer all the advantages of geographically isolated human populations, but with a higher degree of isolation, narrower bottlenecks, and much better genealogical records.

The top 10 diseases in purebred dogs include several that are of major health concerns to humans such as cancer, epilepsy, allergy, retinal disease, cataracts, and heart disease. To date, disease loci have been genetically mapped or otherwise localized in the dog for many disorders including blindness (10-12), kidney cancer (13), narcolepsy (14), rheumatoid arthritis (15), SCID (16), keratin-associated diseases (17), cystinuria (18), bleeding disorders (19, 20), ceroid lipofuscinosis (21), and copper toxicosis (22, 23). In most cases investigators have been able to localize a disease locus to a modestly-sized interval of 10-20 million bp, although for some disorders such as several retinopathies (24, 25), narcolepsy (14), cystinuria (18), SCID (16), bleeding disorders (19, 20), and copper toxicosis (26) the underlying mutations have been identified. In most of the latter situations, candidate gene approaches proved fruitful. Testing specific candidates, however, is unlikely to be useful for a majority of other loci that have been mapped. Towards that end, a 6x assembled sequence of the dog genome is vital.

Genome wide screens are ongoing for many diseases including: inherited retinopathies (Cornell, Animal Health Trust), hip dysplasia (Michigan State, Cornell), motor neuron disease (Stanford University), metabolic disorders (University of Pennsylvania), conotruncal heart defect (University of Pennsylvania), deafness and skin diseases (Texas A and M), other heart disorders such as subaortic stenosis (Oregon Health Sciences University, Ohio State), mammary cancer (Michigan State), cataracts (University of Missouri) and epilepsy (University of Missouri). In some cases individual labs are undertaking the scans. Others, like the scan for hip dysplasia in a Labrador/Greyhound cross, are being done at Marshfield Institute (http://research.marshfieldclinic.org/genetics/), which recently indicated a willingness to undertake genome scans of the dog.

In the following three sections, we consider the specific examples of retinitis pigmentosa, cancer and narcolepsy. These highlight, respectively, the ways in which pathways can be dissected by mapping several similar diseases in several distinct breeds; the utility of the domestic dog for mapping and cloning disease genes that are common in the human population but genetically intractable because of locus heterogeneity in humans; and, finally, the ways in which studying rare diseases in dogs can inform us about the biochemistry of common diseases in humans.

A.1.b. Mapping Genes for Retinitis Pigmentosa: Understanding of Multigene Disorders

Retinitis Pigmentosa (RP) is a blinding human retinal degeneration affecting approximately 1 in 4,000 people. Autosomal dominant, autosomal recessive, X-linked, and digenically inherited forms of RP are recognized, and linkage studies have identified at least 37 RP loci in the human genome, for which 22 genes have been cloned. However, despite these recent advances, current estimates indicate that there are at least as many genes remaining to be identified as are so far known to cause human RP. Naturally-occurring hereditary retinal degenerations in dogs, termed Progressive Retinal Atrophies (PRAs), are widespread and have provided several models of autosomal recessive, X-linked, and autosomal dominant RP (10, 24, 27-29). These have provided valuable opportunities to investigate not only the cell biology and pathogenesis of these diseases, but also to begin genetic therapeutic interventions (30). Thus far, loci have been identified for six forms of canine PRA (10-12, 31-33). Despite these advances, for the great majority of the at least 80 different breeds of dog

affected by PRA, no causal gene or mutation has been identified. On the other hand, some of the forms of PRA so far mapped appear to represent genes not previously recognized as causal in human RP (11). One would predict that as these PRA genes are identified they will yield new candidates for RP genes in humans.

Availability of an essentially complete genome sequence of the dog will dramatically accelerate the search for causative genes in PRA, and thus in RP. Moreover, it will be extremely valuable in advancing gene therapy of retinal degenerative disorders. Comparison of the complete genomic sequence for homologous genes in dog, human, mouse and other species will expedite the recognition of potentially important regulatory, enhancer, and promoter elements that will enable construction of vectors optimized for expression in the correct cell (e.g. rod or cone photoreceptor, or retinal pigment epithelial cell).

A.1.c. Mapping Cancer Susceptibility Loci: Simplifying Locus Heterogeneity and Phenocopy

Cancer is the most frequent disease of dogs, with 1 in 3 dogs expected to get cancer at some point in their life. If many common alleles were involved in susceptibility to any given cancer, we would expect all breeds to show similar incidence patterns, and they clearly do not. Therefore in any given breed it is likely that there are a small number of (or even one) disease alleles of strong effect. Indeed, for most cancers we can identify breeds or sets of breeds in which the disease is present in excess. Larger dogs (*e.g.*, Boxers, Airedales, Great Danes, Saint Bernards) have twice as many soft tissue tumors as the general canine population (34). Breeds with excessive mammary cancer include Cocker Spaniels, and Skye and Boston terriers. Large and giant breeds such as Great Danes, Deerhounds, St. Bernards, and Irish Setters develop osteosarcoma frequently (35-38). A high breed-specific incidence of lymphoma is reported for Boxers and Pointers (38), and more recently, for Golden Retrievers and Rottweilers. Melanoma is observed in excess in Boxers and Scottish Terriers (39).

Studies of individual canine pedigrees provide further support for the idea that single cancer susceptibility genes can be mapped in dogs. Specific pedigrees or lines of dog have been described with increased risk for osteosarcoma (St. Bernards) (40), lymphoma (Bull Mastiffs) (41), malignant histiocytosis (Bernese Mountain Dogs) (42), and kidney cancer (German Shepherd dogs) (43). Recently, kidney cancer was mapped in a large informative pedigree of German Shepherd dogs to canine chromosome 5 by the Ostrander and Lingaas groups (13). Ongoing efforts are aimed at positional cloning of this disease gene, an effort that could be greatly enhanced by the availability of 6x sequence of the dog. Other studies are aimed at finding genes for malignant histiocytosis in the Flat-Coated Retriever (Animal Health Trust, England), osteosarcoma in the hounds (University of Michigan), and lymphoma in Golden Retrievers (Fred Hutchinson Cancer Research Center).

A.1.d. Narcolepsy: Mapping Rare Traits in Dogs Informs Common Diseases in Humans

While the study of common diseases like cancer can aid in the identification of disease genes that are likely mutated in similar human diseases, the study of rare disorders in dogs is a useful way to better understand the molecular biology of common disease processes. A good example is the recent mapping and cloning of a gene for inherited narcolepsy in a colony of Doberman pinschers (14). While narcolepsy itself is a relatively rare disease, sleep disorders as a whole are common in humans. Yet little is known about the molecular biology of sleep. In Dobermans, a highly penetrant and rare form of narcolepsy was found to be caused by mutation in the hypocretin 2 receptor gene (14). Clinical studies quickly established that hypocretin deficiency is associated with most cases of narcolepsy in humans (44-46). Ongoing studies of hypocretin indicate that it may have a key role in circadian clock-dependent alertness as well as integrating hypothalamic signals for neuroendocrine release, metabolic rate, appetite, and mood as well as sleep. These findings present the hypocretin system as a therapeutic target not only in the treatment of narcolepsy, but also more common sleep disturbances.

The positional cloning of the narcolepsy gene was greatly facilitated by the fortuitous fact that the gene was localized to a region of canine chromosome 12 that corresponded to a relatively gene poor region of human 6p21. However, timely and cost effective cloning of other loci in gene-rich regions will require assembled and aligned sequence of the dog genome.

A.1.e. Mapping QTLs for Morphology in the Canine System

Finally, we consider the example of using the dog to study the genetics of diversity. The very fact that such an array of phenotypically stable breeds has been created in so few generations may imply that a relatively small number of key loci are responsible for the characteristic features that define particular breeds. A recent analysis of canid morphology in the Portuguese Water Dog (PWD) has demonstrated the value of particular dog populations for studies in quantitative genetics (47). The PWD population is derived from a small founding population. A working collaboration was established between scientists and dog owners, similar to the logistical relationships used in human genetics. A total of 330 dogs were phenotyped using 90 metrics derived from x-rays, and genotyped using 550 microsatellite markers.

Principal Component (PC) analysis was used to dissect genetic networks regulating PWD skeletal shape and size. PC analysis classifies phenotypic variation into independent systems of correlated components. Because PCs are phenotypes, Quantitative Trait Loci (QTLs) that inform them can be identified. Significant QTLs were identified for four distinct PCs that addressed distinct aspects of skeletal morphology. For example, the first principal component (PC1) represented the overall size of the skeleton. Two QTLs associated with this component were identified, one of which appears to regulate the growth factor IGF-1. The PWD results strongly support the hypothesis that the QTLs associated with these four PCs represent major regulatory genes. Selection operating on such genes could result in the rapid transformation of skeletal types into different breeds such as the Greyhound and Pit Bull. Similar changes have been observed in hominid evolution suggesting that these or similar loci may be common to most mammals. The 6x sequence of the dog is awaited before large scale positional cloning studies will be undertaken to follow up these exciting results.

A.2. Biomedical Applications: The Canine System for Marrow Transplantation

The canine system is widely regarded for its 30 year role in developing bone marrow transplantation protocols, as documented by several thousand peer-reviewed publications. As a random-bred species, dogs closely approximate issues of tissue typing and clinical management encountered in humans undergoing marrow transplant. Compared to primates, dogs are more readily available, less expensive, more disease-free, and easier to work with. Compared to mice, dogs are large enough for serial blood and tissue samples and continuous intravenous infusions. Many of the principles and therapeutic approaches developed in the canine marrow and stem cell transplantation model have been directly translated to the clinic. These are reviewed in (48-53). For example, studies in dogs first showed the predictive value of *in vitro* major histocompatibility (MHC) testing for graft outcome, and the feasibility of phenotypically MHC-matched unrelated marrow grafts. Dog studies first drew attention to graft-versus-host-disease (GVHD) in the setting of genotypic MHC compatibility; as a consequence, most of the state-of-the-art immunosuppression to prevent and treat GVHD used clinically was developed in dogs. Other conditioning programs developed in the dog model include fractionated total body irradiation (TBI) in the mid-1970s and, over the past 10 years, the use of monoclonal antibodies coupled to a ß-emitting radioactive isotope, iodine-131, now in the clinic. Finally, the dog has been the setting for studies of the recovery of the immune system after transplant, mechanisms of graft-host tolerance, and late radiation sequelae, including secondary malignancies. The above work has been slowed, in part, by the need to laboriously clone and sequence an endless list of canine growth factors, hormone receptors, MHC related loci, etc. that are purported to play a role in immune function. Efforts to further exploit the canine system for development of marrow transplant protocols for humans would be greatly enhanced by the availability of the 6x dog sequence.

A.3. Biomedical Applications: Successful Gene Therapy in the Canine System

A variety of canine systems have been used to study and improve the efficacy of gene therapy protocols. These include the use of marrow transplantation for SCID (54) and other immune diseases (55), transrectal gene therapy of the prostate (56), treatment of mucopolysaccharidosis VII (57), restoration of vision by gene therapy to the retina (30), and liver sustained gene therapy to treat hemaglobinopathies (58). Among the most recent dramatic results are those of Haskins et al. who have administered a retrovirus vector containing the canine beta-glucuronidase (GUSB) cDNA to seven neonatal dogs with mucopolysaccharidosis VII and have maintained serum GUSB activity from 50% of normal to 70 times normal for up to a year and a half (the current time). Untreated dogs are growth retarded, cannot stand by 6 months of age, have corneal clouding, and cardiac

murmurs. Treated dogs, however, maintained for 15-18 months, are of near normal body weight, continue to be able to run, have clear corneas, and no cardiac murmurs. In a similarly dramatic study, Acland et al. have used a recombinant adeno-associated virus (AAV) carrying wild-type RPE65 to restore visual function in a naturally occurring animal model, the RPE65 -/- dog, which suffers from early and severe visual impairment similar to that seen in humans with Leber congenital amaurosis (30). Their results indicate that visual function was restored and, at 18 months post treatment, treated dogs are still sighted (Acland et al., personal communication). A host of other canine gene therapy studies are underway or planned, each of which would be aided by the ability to quickly isolate genes and variants of interest from the 6x canine sequence.

A.4. Current Limitations Faced by the Canine Genome Mapping Community; Anticipated Benefits From a Canine Genome Sequencing Project

The projects described in the preceding sections have largely been enabled by recent advances in canine genome mapping (see below). The progress achieved encourages additional research by attracting investigators to the field, and by convincing funding agencies that such research can be fruitful. However, further efforts are considerably restricted in their scope, and hampered in their implementation, by lack of access to the canine genomic sequence.

Most projects, to date, have exploited canine traits for which either direct candidate genes could be proposed and evaluated, or for which large informative pedigrees were available to enable linkage mapping to identify candidate regions. A major component of such research efforts comprises the cloning, sequencing and mapping of individual canine homologs of genes either proposed as candidate genes, or expected to be located in candidate regions. This is necessary to identify new informative polymorphisms (e.g. SNPs, microsatellites) for high resolution mapping of candidate regions, and to examine each exon, exon/intron boundary, and promoter region for positional candidates. Availability of canine genome sequence data from the public domain database would dramatically shortcut this effort and increase investigators' efficiency. In addition, comparison of human and canine genomic sequences permits recognition of conserved noncoding segments that would have otherwise been overlooked or noted only after laborious sequencing efforts.

Almost entirely unexamined so far have been those canine traits that occur widely in the population, for which large numbers of individuals expressing the traits of interest are potentially available, but for which ideal meiotic linkage mapping pedigrees are not available. Because of the structure of the canine population, especially the clustering into breeds, it is anticipated that this broader canvas of traits would be ideally suited for analysis using SNP/haplotype mapping using Identity-By-Descent (IBD) and Linkage Disequilibrium (LD) methods. For this to begin however, we need an efficient way to build SNP maps of the dog. A reasonably complete draft canine genome sequence is the vital step forward.

A.5. The Canine Genome Sequence Will Inform the Human and Mouse Sequences

The mammalian comparative genomics field is still in its infancy, as we are only now nearing completion of the human, mouse and rat genome sequences. The sequencing of the dog genome could strongly add to the value of these genomes as well as benefit from them. Preliminary comparative analysis of the human and mouse genomes suggests that approximately 3-5% of those genomes are strongly conserved; half corresponding to coding and half to non-coding sequences. Having a third non-rodent, non-primate mammalian genome would be highly informative for identifying the non-coding conserved elements that are biologically relevant. This should open the doors to a better understanding of gene-regulation in humans, dogs and other mammals. Simultaneously, the careful annotation of the human and mouse genomes, already in progress, will inform positional cloning efforts in the dog by providing candidate genes or possibly even variant alleles that could be linked to disease.

Finally, it will also be possible to study the presence and absence of specific gene family members in the different species. An excellent example is provided by the work of Francis Galibert in describing the canine olfactory receptor (OR) gene family. An analysis of the canine 1x sequence (described below) revealed some 654 OR sequences of which 100 were already known. The 651 unique canine ORs have been mapped and compared to the human OR gene family. Ninety-seven belong to class 1 and are clustered on canine chromosome 21; the remaining 554 belong to class 2 and are located at 31 regions on 22 distinct dog chromosomes. The three largest canine clusters, on CFA18, 20 and 21, are larger than their human counterparts

located on HSA 11, 19, 21. Thus, while the overall organization is very similar between human and dog, the size of the dog clusters appears larger. In addition, while the number of pseudogenes is thought to be about 70% in human and 20% in mouse, it is presently estimated to be 15% in the canine genome. Clearly, much can be learned about function, organization, and evolution from comparative analysis of dog, mouse, rat and human large gene families.

B. Strategic Issues in Acquiring New Sequence Data

B.1. Size, Current Funding Status and Needs of the Research Community

B.1.a. Current Federally Funded Canine Grants

A large number of federally funded grants exist that would benefit from 6x dog sequence (Table 1). A search of the CRISP database reveals that there are currently 453 such grants. As is evident in Appendix C, many are funded by the National Heart, Lung and Blood Institute and the National Institute of Diabetes & Digestive & Kidney Diseases.

Table 1. CRISP Database Grant Listing by Organism

Organism	CRISP Database Hits
Pig	466
Dog	453
Guinea Pig	427
Cat	389
Rhesus	267
Chicken	248
Bovine	244
Chimpanzee	18
Gorilla	2

B.1.b. Current Nonfederally Funded Efforts that would Benefit from the Canine Sequence

The American Kennel Club Canine Health Foundation (AKC-CHF) and the Morris Animal Foundation fund a large number of grants each year specific to canine research (see letters of support from both in Appendix A). Currently there are 79 active AKC-CHF grants, nearly three fourths of which are dedicated to genetics or genome map development in the dog. The AKC-CHF distributes $1.5 million a year in research funding aimed at canine health. The Morris Animal Foundation distributes $1.4 million a year in funding for dogs. Both foundations have indicated a willingness to contribute to the funding of trainees participating in the sequencing effort. The current distribution of AKC-CHF grants, the diseases of interest, and a subset of relevant breeds is listed in Appendix B.

B.2. Current Status of Genomic Resources in the Canine Community

In this section we summarize the current state of the canine genome project, and briefly state the available resources that would be useful towards the development of 6x assembled dog sequence.

B.2.a. Canine Maps

Fully integrated meiotic linkage, radiation hybrid (RH), cytogenetic and comparative maps encompassing 1800 markers have been constructed for the canine genome (59), incorporating individual mapping efforts of the last several years (60-63). The most recently published canine RH map includes 1078 anonymous microsatellite markers, 320 genes or ESTs, and a set of cosmid "anchor probes" that have been RH and linkage mapped as well as ordered by FISH to facilitate integration of all maps. We note that 50% of polymorphic markers have Het or PIC values >0.5, allowing us to develop a minimal mapping set for undertaking initial genome wide scans in pedigrees of interest (64). The average inter-marker distance on the RH map is 17 cR_{5000}. From maximum likelihood predictions, we predict the size of the canine genome to be 26.5 Morgan (65). Taking into account the predicted cR/kb correspondence and the correspondence of 1 Mb

per 1 cM, the total size of the RH map is 23.5 Morgan. This map is estimated to cover more than 90% of the canine genome. Details can be found on our respective web sites: (http://www.fhcrc.org/science/dog_genome/dog.html and http://www-recomgen.univ-rennes1.fr/doggy.html). In addition to the above, using a set of differentially labeled whole chromosome paint probes for reciprocal painting of dog and human chromosomes (66), 68-75 evolutionarily conserved segments between the dog and the human genomes are now identified (67-69). These in turn are aligned on the DAPI banded karyotype of the dog (59).

Drs. Ostrander and Galibert have recently completed lab work towards the construction of a 3400 marker RH map of the dog. The additional 1900 markers above and beyond those in the 1500 marker map include 700 BAC-end sequences, 500 new microsatellites, and several hundred new canine specific genes and ESTs. Final construction of the map figures is underway and manuscript preparation is in progress.

B.2.b. Newly Funded Efforts: Evolutionary Conservation of Chromosomal Segments

While RH mapping and reciprocal paint studies suggest that the canine genome is organized into about 70 large segments corresponding to portions of human chromosomes, the question remains as to how well gene order is maintained within such large evolutionarily conserved regions. For instance, are the genes on canine chromosome 5 maintained in the same order as they appear on human chromosomes 11q, 17p, 1p, and 16q? Currently available data are encouraging. RH and linkage mapping of more than 1000 genes and canine specific ESTs (59, 62, 63, 70) suggest that, within large evolutionarily conserved blocks, the expected genes are found with few exceptions. Gene order appears well maintained at the resolution for which we can currently build RH maps, about 650 Kb (59, 70, 71). Finally, for regions where gene order has been systematically and thoroughly investigated, such as canine chromosome 5, individual deviations from predicted gene order are few, although inversions in blocks of genes are noted and, again, resolution is still low. Canine chromosome 5 notwithstanding, however, there are few places in the canine genome where we have a significant amount of detailed comparative map data. To remedy this, Drs. Elaine Ostrander at the FHCRC, Francis Galibert at the University of Rennes, and Ewen Kirkness at TIGR were recently funded to map 10,000 independent gene sequences from the canine 1x sequence on a newly constructed 10,000 rad canine RH panel. Genes are being selected from the 1x sequence to span all human chromosome arms. The 3–year grant will be completed in July 2005. Enthusiasm is high on the part of all investigators involved to contribute this information to the assembly of the canine genome sequence.

B.2.c. Other Canine Genome Resources

A number of other resources are available that may aid in the assembly of the dog genome sequence. First, there exists a dense, high quality canine BAC library, made under the direction of Pieter de Jong (72). The library contains approximately 165,888 BAC clones with a mean insert size of 155 kb, predicting an 8.1–fold coverage of the canine genome. Several canine EST projects have also been initiated in the community. These will provide a valuable resource for annotating the canine draft sequence. The largest of these is ongoing at the Cold Spring Harbor (CSH) labs under the direction of Dr. W. Richard McCombie *(A Dog EST Resource for Comparative Functional Genomics*). Dr. McCombie has begun efforts to construct nearly a dozen normalized or subtracted cDNA libraries and generate sequence for 270,000 ESTs. Thus far some 8,000 canine specific ESTs have been deposited into GenBank. He has also focused on developing an extensive web–based database for maintaining data related to these and all of the canine community EST projects which can be viewed at (http://www.cshl.org/genbin/cgi-bin/golden_retriever.cgi). Once sequences of interest are found they can be retrieved or their position relative to the human genome displayed on the UCSC Genome Browser. This infrastructure will add considerably to the utility of canine sequences for comparative genomics. Finally, Dr. McCombie has outlined plans to undertake full length cDNA sequencing of canine genes in upcoming efforts (see McCombie Letter in Appendix A). This work will greatly aid in the validation and annotation of the canine genome sequence.

We note also that Eric Green, NISC, have generated ~20 Mb of BAC full-shotgun dog sequence to date, and will likely double that amount over the next year. Much of that sequence will be finished. The sequenced BACs from this effort will prove useful for validating/QCing the whole-genome shotgun assemblies that are eventually produced, much as is currently being done for the mouse and rat sequencing efforts.

B.2.d. 1x Non-Public Canine Genome Sequence

The only large scale sequencing effort to date for the dog was carried out at Celera, and resulted in 1x sequence of the dog genome. In this effort, genomic DNA from a male Standard Poodle was used to prepare plasmid libraries of small- and medium-sized inserts (~2 kb and ~10 kb, respectively). Approximately equal numbers of small- and medium-insert templates were prepared and sequenced as described previously (73). The finished sequence data consists of 4.3 million reads (average read length, 675 bases), representing approximately 1x coverage of the haploid canine genome (74). Currently, the sequence data exists in a multiple fasta file format at TIGR. For random 1x coverage, a Lander-Waterman analysis predicts that 64% of a 2.8 Gb genome has been sequenced, and that the average gap size is 650 bp (75).

While the sequence is not publicly available, data from it has been made available to all researchers that have requested it via collaborations with Drs. Ewen Kirkness and Claire Fraser at The Institute for Genomic Research (TIGR). Investigators have generally requested the 1x dog sequence be screened for subsets of genes predicted to be in regions of interest that were selected based on predicted synteny between the human and dog genomes. These experiments have been both successful and informative. High quality reads have been available representing partial sequence of 84% of requested genes to date. More than 90% of the dog peptides display homology with human peptides that extended for greater than 75% of the length of the peptide. These regions of homology displayed average levels of amino acid identity and similarity of 83% and 89%, respectively. It is concluded that, at least for known dog peptides, the sequence similarity between orthologous dog and human peptides is not confined to discrete motifs, but is normally detectable over most of the length of the peptides. This ensures that true orthologues of human genes will be easily identified from the dog sequence.

B.3. Proposed Sequencing Approach and Communication with Sequencing Centers

The dog genome has a haploid content of 2.8×10^9 bp of DNA distributed over 38 autosomes plus the X and Y. As co-authors of this white paper, the Whitehead Genome Center has expressed strong support for undertaking the sequence of the canine genome. We have worked together to develop a plan that specifically meets the needs of the community. This is summarized, in brief, below.

B.3.a. Deep Shotgun Sequencing

We propose a whole genome shotgun sequencing (WGS) strategy designed to yield a long-range, high-quality assembly covering >95% of the dog genome. The WGS sequencing strategy employed for the mouse genome resulted in an assembly consisting of 89 ultracontigs placed on the 20 mouse chromosomes and covering ~96% of the genome. Similarly, for dog we propose collecting a total of 40M paired-end reads generated from different vectors and insert sizes (2, 4, 6, 10, 40 and 200 kb). By using different library sizes we hope to minimize cloning bias and to allow a hierarchical linking approach in the assembly process. We propose the below breakdown of insert sizes, which would result in an approximate 6-fold sequence coverage (Phred-20) and a 50-fold physical coverage of the dog genome allowing for a sequencing pass rate of approximately 80%:

Table 2. Sequence Strategy Summary

Insert size (kb)	Vector	Attempted reads (M)	PHRED20 Sequence coverage (x)	Physical coverage (x)
2-6	Plasmid	35.0	5.4	19.0
10	Plasmid	2.5	0.4	4.0
40	Fosmid	2.0	0.2	10.0
200	BAC	0.5	0.1	17.0
Total		40.0	6.1	50.0

The following steps are necessary:

- Generation of libraries by random shear (possible exception BAC library, but if not sheared, two libraries with different enzymes)
- Paired-end sequencing
- Assembly using ARACHNE or other WGS assembly program
- Linking of supercontigs to chromosomes using a genetic map

This procedure should result in an assembly with >95% coverage of the genome and long-range continuity.

B.3.b. A Finished Genome

We believe that producing a high-quality, long-range shotgun assembly is the main objective for this initiative. Should it be deemed desirable, however, to produce a finished genome sequence, this can be accomplished by picking a minimal number of BACs (or fosmids), covering the genome from the placement of BAC ends in the WGS assembly, and sequencing the chosen BACs at a low coverage to generate templates for finishing. This would, however, constitute a similar effort in terms on cost and capacity as obtaining the shot-gun assembly, and is not thought to be necessary.

B.3.c. SNP Discovery

To be able to study the genetic relationship of different breeds and to generate a large set of markers for high-resolution mapping of genetic traits, we propose generating low-coverage whole genome sequence data for 10 different breeds. Specifically, we propose generating 100,000 reads from each breed, resulting in approximately 25,000 novel SNPs for each breed given an average SNP rate of 1/1000 bp and that approximately 50% of reads can be placed uniquely in the genome. This methodology has proven successful for generating SNPs in three inbred mouse strains and we expect it to work well for the dog. We do expect, however, that the number of SNPs discovered in each breed will vary based on the level of polymorphism for that particular breed. Still, we would expect to generate a total of 250,000 SNPs with an average spacing of ~10 kb across the genome. Breeds selected will include those that have appeared as founders in multiple modern breeds such as the Mastiff, Bloodhound, Greyhound, and Pomeranian.

B.4. Decision Making Process to Select Breed/Animal to Sequence

B.4.a. Selection of a Breed for Sequencing

The decision regarding what breed of dog to sequence is complex, and will of necessity involve several considerations including degree of homozygosity of various breeds, biomedical studies associated with specific breeds, and maximization of current resources. These are discussed in turn.

It is well documented by existing sequencing efforts (Eric Green and Ewen Kirkness, Personal Communications) as well as studies of canid specific repeats that the dog, while lacking *Alu* sequences, contains a large number of well distributed polymorphic SINE sequences. It is possible, although not certain, that this degree of polymorphism may affect the ease with which genomic sequence can be assembled. As a result, the Whitehead Genome Center has expressed a preference to sequence a breed of dog of comparatively low heterozygosity. The canine genomics community has therefore assembled a list of candidate breeds that should be specifically considered. These include the Ibizan hound, Pharoah Hound, Samoyed, Saluki, Maltese, Leonburger, etc. These breeds, as they exist today, are believed to have resulted from generally closed breeding populations (7). At a recent meeting of the Dog Genome Community in St. Louis (May 16-19, 2002), it was agreed that representative DNA samples from each of the breeds under consideration would be sent to the Ostrander lab where they would be SNP typed by BAC-end resequencing to investigate relative levels of heterozygosity. Samples are also being sent from several veterinary school breeding colonies where unusually inbred dogs have resulted from breeding programs aimed at propagating dogs with specific diseases. These data will be available for consideration in the final decision process.

Additional considerations are largely biomedical or resource driven. The breed in which the greatest amount of significant biomedical research has been done to date is the Beagle. Hence, enthusiasm for sequencing the Beagle genome is high. The final consideration is the degree to which resources exist to date. The canine 8.1 fold BAC library, constructed in 1999 by Pieter de Jong, utilized DNA from a Doberman

Pinscher produced from the colony of narcoleptic dogs at Stanford (72). Approximately 1000 BACs from the canine BAC library have been RH mapped, and 4000 others have been end-sequenced. The community requests that the information collected to date not be lost and that the mapped BACs be fingerprinted and included in the new sequencing efforts. If a dog other than the Doberman is chosen for sequencing, a new BAC library will likely need to be constructed, and Pieter de Jong has expressed his willingness to do so.

It is also of note that the 1x Celera sequence was completed on a Standard Poodle, which is not a particularly inbred breed. Unless additional sequence were to be generated using the same dog initially used to generate the 1x sequence, there would be no particular advantage in selecting a Standard Poodle for sequencing. TIGR has, however, stated a potential willingness to contribute the 1x dog sequence to the public effort and generate additional sequence on the initial Poodle used.

B.4.b. Canine Genome Mapping Community Committee for Selection

The canine community has a long history of collaboration and collegiality. There is a strong desire from many in the community to participate in the decision making regarding what breed/breeds are sequenced. Based on participation and work to date, we propose a committee of the following representatives to work with the sequencing centers regarding all dog specific issues:

Canine Genome Leadership Group
Paula Henthorn, Ph.D., Associate Professor, University of Pennsylvania
Matthew Binns, Ph.D., Professor, Animal Health Trust, England
Francis Galibert, Ph.D., Professor, University of Rennes and CNRS, France
Patrick Venta, D.V.M., Ph.D., Associate Professor, University of Michigan
Greg Acland, D.V.M., Senior Research Associate, Cornell University
Elaine Ostrander, Ph.D., Member, Fred Hutchinson Cancer Research Center

B.5. Public Relations and Support from the Companion Animal Community

In preparing the community for potential sequencing of the canine genome we have spent considerable time meeting with representatives from Kennel Clubs, Companion Animal Support groups and Foundations that support research on canine health to ensure that there is strong public support for this effort. Letters are included in Appendix A.

We have also approached several of these organizations to determine their interest in providing financial support for the sequencing effort. Many, such as the AKC-CHF, Morris Animal Foundation, Seeing Eye, and Nestlé Purina currently provide modest grants (average $50,000 a year) or fellowships to trainees working on canine health issues. Many have indicated a willingness to contribute to the sequencing effort in the form of salary for trainees in specific labs, pending detailed discussions. Some of these are detailed in the letters in Appendix A.

Summary

The dog enjoys a genetic diversity unrivaled by any other mammalian species. A thousand centuries of directed breeding by humans has channeled that diversity into an unequaled variety of morphologies and behaviors, and also into a storehouse of inherited diseases. The availability of a high-quality genome sequence and associated SNP map would provide the key allowing us to exploit that genetic heritage to advance our understanding of normal biology and disease of both dogs and men.

REFERENCES:

1. Ostrander EA, Giniger E. Semper fidelis: what man's best friend can teach us about human biology and disease. Am J Hum Genet *61:* 475-480, 1997.
2. Ostrander EA, Giniger E. Let sleeping dogs lie? Nat Genet *23:* 3-4, 1999.
3. Ostrander EA, Galibert F, Patterson DF. Canine genetics comes of age. Trends in Genetics *16:* 117-123, 2000.
4. Patterson DF, Haskins ME, Jezyk PF, Giger U, Meyers-Wallen VN, Aguirre G, Fyfe JC, Wolfe JH. Research on genetic diseases: reciprocal benefits to animals and man. J Am Vet Med Assoc *193:* 1131-1144, 1988.
5. Patterson D. Companion animal medicine in the age of medical genetics. J Vet Internal Med *14:* 1-9, 2000.
6. Galibert F, Andre C, Cheron A, Chuat JC, Hitte C, Jiang Z, Jouquand S, Priat C, Renier C, Vignaux F. The importance of the canine model in medical genetics. Bull Acad Natl Med *182:* 811-821, 1998.
7. American Kennel Club. Breed Registry Data Base. Vol. 1926-1998, New York, 1999.
8. Wayne RK, Ostrander EA. Origin, genetic diversity, and genome structure of the domestic dog. Bioessays *21:* 247-257, 1999.
9. Ostrander E.A., Kruglyak L. Unleashing the canine genome. Genome Res *10*: 1271-1274, 2000.
10. Acland GM, Ray K, Mellersh CS, Gu W, Langston AA, Rine J, Ostrander EA, Aguirre GD. 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 *96:* 3048-3053, 1998.
11. Acland GM, Ray K, Mellersh CS, Gu W, Langston AA, Rine J, Ostrander EA, Aguirre GD. A novel retinal degeneration locus identified by linkage and comparative mapping of canine early retinal degeneration. Genomics *59:* 134-142, 1999.
12. Sidjanin DJ, Lowe JK, McElwee JL, Milne BS, Phippen RM, Sargan DR, Aguirre GD, Acland GM, Ostrander EA. Canine *CNGB3* mutations establish cone degeneration as orthologous to the human achromatopsia locus *ACHM3*. Human Molec Genet, *In Press*, 2002.
13. Jonasdottir TJ, Mellersh CS, Moe L, Heggebo R, Gamlem H, Ostrander EA, Lingaas F. Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. Proc Natl Acad Sci USA *97:* 4132-4137, 2000.
14. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de Jong PJ, Nishino S, Mignot, E. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. Cell *98:* 365-376, 1999.
15. Ollier WE, Kennedy LJ, Thomson W, Barnes AN, Bell SC, Bennett D, Angles JM, Innes JF, Carter SD. Dog MHC alleles containing the human RA shared epitope confer susceptibility to canine rheumatoid arthritis. Immunogenetics *53:* 669-673, 2001.
16. Henthorn PS, Somberg RL, Fimiani VM, Puck JM, Patterson DF, Felsburg PJ. 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.
17. Keller RC, Switonski M, Jorg H, Ladon D, Arnold S, Schelling,C. Chromosomal assignment of two putative canine keratin gene clusters. Anim Genet *29:* 141-143, 1998.
18. Henthorn PS, Liu J, Gidalevich T, Fang J, Casal ML, Patterson DF, Giger U. Canine cystinuria: polymorphism in the canine SLC3A1 gene and identification of a nonsense mutation in cystinuric Newfoundland dogs. Hum Genet *107:* 295-303, 2000.
19. Chao H, Samulski R, Bellinger D, Monahan P, Nichols T, Walsh C. Persistent expression of canine factor IX in hemophilia B canines. Gene Ther *6:* 1695-1704, 1999.
20. Venta PJ, Li J, Yuzbasiyan-Gurkan V, Brewer GJ, Schall WD. Mutation causing von Willebrand's disease in Scottish Terriers. J Vet Intern Med *14:* 10-19, 2000.
21. Lingaas F, Aarskaug T, Sletten M, Bjerkas I, Grimholt U, Moe L, Juneja RK, Wilton AN, Galibert F, Holmes NG, Dolf G. Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs. Anim Genet *29:* 371-376, 1998.

22. Yuzbasiyan-Gurkan V, Blanton SH, Cao V, Ferguson P, Li J, Venta PJ, Brewer GJ. Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers. Am J Vet Res *58:* 23-27, 1997.
23. van de Sluis BJ, Breen M, Nanji M, van Wolferen M, de Jong P, Binns MM, Pearson PL, Kuipers J, Rothuizen J, Cox DW, Wijmenga C, van Oost BA. Genetic mapping of the copper toxicosis locus in Bedlington terriers to dog chromosome 10, in a region syntenic to human chromosome region 2p13- p16. Hum Mol Genet *8:* 501-507, 1999.
24. Aguirre GD, Baldwin V, Pearce-Kelling S, Narfstrom K, Ray K, Acland GM. Congenital stationary night blindness in the dog: common mutation in the RPE65 gene indicates founder effect. Mol Vis *4:* 23, 1998.
25. Veske A, Nilsson SE, 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.
26. van De Sluis B, Rothuizen J, Pearson PL, van Oost BA, Wijmenga C. Identification of a new copper metabolism gene by positional cloning in a purebred dog population. Hum Mol Genet *11:* 165-173, 2002.
27. Aguirre GD, Baldwin V, Weeks KM, Acland GM, Ray K. Frequency of the codon 807 mutation in the cGMP phosphodiesterase beta-subunit gene in Irish setters and other dog breeds with hereditary retinal degeneration. J Hered *90:* 143-147, 1999.
28. Zhang Q, Acland GM, Zangerl B, Johnson JL, Mao Z, Zeiss CJ, Ostrander EA, Aguirre GD. Invest Ophthalmol Vis Sci *42:* 2466-2471, 2001.
29. Kijas J, Cideciyan A, Aleman T, Pianta M, Pearce-Kelling S, Miller B, Jacobson S, Aguirre G, Acland G. Naturally-occurring *rhodopsin* mutation in the dog causes retinal dysfunction and degeneration mimicking human dominant retinitis pigmentosa. Proc Natl Acad Sci USA, *In Press*, 2002.
30. Acland GM, Aguirre GD, Ray J, Zhang Q, Aleman TS, Cideciyan AV, Pearce-Kelling SE, Anand V, Zeng Y, Maguire AM, Jacobson SG, Hauswirth WW, Bennett J. Gene therapy restores vision in a canine model of childhood blindness. Nat Genet *28:* 92-95, 2001.
31. Petersen-Jones SM, Entz DD, Sargan DR. cGMP phosphodiesterase-alpha mutation causes progressive retinal atrophy in the Cardigan Welsh corgi dog. Invest Ophthalmol Vis Sci *40:* 1637-1644, 1999.
32. Aguirre G, Lolley R, Farber D, Fletcher T, Chader G. Rod-cone dysplasia in Irish Setter dogs: A defect in cyclic GMP metabolism in visual cells. Science *201:* 1133, 1978.
33. Zeiss CJ, Ray K, Acland GM, Aguirre GD. Mapping of X-linked progressive retinal atrophy (XLPRA), the canine homolog of retinitis pigmentosa 3 (RP3). Hum Mol Genet *9:* 531-537, 2000.
34. Priester WA, Mantel N. Occurrence of tumors in domestic animals. Data from 12 United States and Canadian colleges of veterinary medicine. J Natl Cancer Inst *47:* 1333-1344, 1971.
35. Tjalma RA. Canine bone sarcoma: Estimation of relative risk as a function of body size. J Natl Cancer Inst *36:* 1137-1150, 1966.
36. Misdorp W. Skeletal osteosarcoma. Animal model: canine osteosarcoma. Am J Pathol *98:* 285-288, 1980.
37. Withrow SJ, Powers BE, Straw RC, Wilkins RM. Comparative aspects of osteosarcoma. Dog versus man. Clin Orthop *270:* 159-168, 1991.
38. Dorn RC, Priester WA. Epidemiology. *In:* G. H. Theilen and B. R. Madewell, eds. (eds.), Veterinary Cancer Medicine, pp. 27-52. Philadelphia: Lea and Febiger, 1987.
39. Theilen G, Madewell B. Veterinary Cancer Medicine, 2nd edition, p. 233-325. Philadelphia: Lea and Febiger, 1987.
40. Bech-Nielsen S, Haskins M, Reif J, Brodey R, Patterson D, Spielman R. Frequency of osteosarcoma among first-degree relatives of St. Bernard dogs. J Natl Cancer Inst *60:* 349-353, 1978.
41. Onions DE. A prospective survey of familial canine lymphosarcoma. J Natl Cancer Inst *72:* 909-912, 1984.
42. Padgett G, Madewell B, Keller, E. Inheritance of histiocytosis in Bernese mountain dogs. Journal of Small Animal Practice *36:* 93-98, 1995.
43. Moe L, Lium B. Hereditary multifocal renal cystadenocarcinomas and nodular dermatofibrosis in 51 German shepherd dogs. Journal of Small Animal Practice *38:* 498-505, 1997.
44. Nishino S, Ripley B, Overeem S, Lammers GJ, Mignot E. Hypocretin (orexin) deficiency in human narcolepsy [letter]. Lancet *355:* 39-40, 2000.
45. Peyron C, Faraco J, Rogers W, Ripley B, Overeem S, Charnay Y, Nevsimalova S, Aldrich M, Reynolds D, Albin R, Li R, Hungs M, Pedrazzoli M, Padigaru M, Kucherlapati M, Fan J, Maki R, Lammers GJ, Bouras

C, Kucherlapati R, Nishino S, Mignot E. A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. Nat Med *6:* 991-997, 2000.
46. Thannickal TC, Moore RY, Nienhuis R, Ramanathan L, Gulyani S, Aldrich M, Cornford M, Siegel J. Reduced number of hypocretin neurons in human narcolepsy. Neuron *27:* 469-474, 2000.
47. Chase K, Carrier DR, Adler FR, Jarvik T, Ostrander EA, Lorentzen TD, Lark KG. Genetic basis for systems of skeletal quantitative traits: Principal component analysis of the Canid skeleton. Proc Natl Acad Sci USA, *In Press*, 2002.
48. Thomas ED, Storb R, Clift RA, Fefer A, Johnson FL, Neiman PE, Lerner KG, Glucksberg H, Buckner CD. Bone-marrow transplantation. N Engl J Med *292:* 832-843, 895-902, 1975.
49. Thomas ED, Storb R. The development of the scientific foundation of hematopoietic cell transplantation based on animal and human studies. *In:* E. D. Thomas, K. G. Blume, and S. J. Forman (eds.), Hematopoietic Cell Transplantation, 2nd edition, pp. 1-11. Boston: Blackwell Science, 1999.
50. Storb R, Thomas ED. Allogeneic bone-marrow transplantation. Immunol Rev *71:* 77-102, 1983.
51. Storb R, Thomas ED. Graft-versus-host disease in dog and man: the Seattle Experience. *In:* M. G (ed.) Immunological Reviews No. 88, pp. 215-238. Copenhagen: Munksgaard, 1985.
52. Storb R, Deeg HJ. Failure of allogeneic canine marrow grafts after total-body irradiation. Allogeneic "resistance" versus transfusion-induced sensitization. Transplantation *42:* 571-580, 1986.
53. Storb R, Yu C, McSweeney P. Mixed chimerism after transplantation of allogeneic hematopoietic cells. *In:* E.D. Thomas, K.G. Blume, and S.J. Forman (eds.), Hematopoietic Cell Transplantation, 2nd edition, pp. 287-295. Boston: Blackwell Science, 1999.
54. Hartnett BJ, Yao D, Suter SE, Ellinwood NM, Henthorn PS, Moore PE, McSweeney PA, Nash RA, Brown JD, Weinberg KI, Felsburg PJ. Transplantation of X-linked severe combined immunodeficient dogs with CD34+ bone marrow cells. Biol Blood Marrow Transplant *8:* 188-197, 2002.
55. Licht T, Haskins M, Henthorn P, Kleiman SE, Bodine DM, Whitwam T, Puck JM, Gottesman MM, Melniczek JR. Drug selection with paclitaxel restores expression of linked IL-2 receptor gamma -chain and multidrug resistance (MDR1) transgenes in canine bone marrow. Proc Natl Acad Sci USA *99:* 3123-3128, 2002.
56. Weld KJ, Mayher BE, Allay JA, Cockroft JL, Reed CP, Randolph MM, Lu Y, Steiner MS, Gingrich JR. Transrectal gene therapy of the prostate in the canine model. Cancer Gene Ther *9:* 189-196, 2002.
57. Ponder KP, Melniczek JR, Xu L, Weil MA, O'Malley TM, O'Donnell PA, Knox VW, Aguirre GD, Mazrier H, Ellinwood NM, Sleeper M, Maguire AM, Volk S, Mango RL, Zweigle J, Wolfe JH, Haskins ME. Marked clinical improvements in mucopolysaccharidosis VII dogs after neonatal administration of a retroviral vector expressing ß-glucuronidase. Submitted.
58. Mount JD, Herzog RW, Tillson DM, Goodman SA, Robinson N, McCleland ML, Bellinger D, Nichols TC, Arruda VR, Lothrop CD, Jr, High KA. Sustained phenotypic correction of hemophilia B dogs with a factor IX null mutation by liver-directed gene therapy. Blood *99:* 2670-2676, 2002.
59. Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C, Holmes NG, Cheron A, Suter N, Vignaux F, Bristow AE, Priat C, McCann E, Andre C, Boundy S, Gitsham P, Thomas R, Bridge WL, Spriggs HF, Ryder EJ, Curson A, Sampson J, Ostrander EA, Binns MM, Galibert F. 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.
60. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, Ostrander EA. A linkage map of the canine genome. Genomics *46:* 326-336, 1997.
61. Werner P, Mellersh CS, Raducha GM, DeRose S, Acland GM, Prociuk U, Wiegand N, Aguirre GD, Henthorn PS, Patterson DF, Ostrander EA. Anchoring of canine linkage groups with chromosome specific markers. Mammalian Genome *10:* 812-823, 1999.
62. Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, Jouquand S, Werner P, André C, DeRose S, Patterson DF, Ostrander EA, Galibert F. An integrated linkage-radiation hybrid map of the canine genome. Mammalian Genome *11:* 120-130, 2000.
63. Priat C, Hitte C, Vignaux F, Renier C, Jiang Z, Jouquand S, Cheron A, Andre C, Galibert F. A whole-genome radiation hybrid map of the dog genome. Genomics *54:* 361-378, 1998.

64. Richman M, Mellersh CS, Andre C, Galibert F, Ostrander EA. Characterization of a minimal screening set of 172 microsatellite markers for genome-wide screens of the canine genome. J Biochem Biophys Methods *47:* 137-149, 2001.
65. Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, Aguirre GD, Ziegle JS, Ostrander EA, Rine J. A second-generation genetic linkage map of the domestic dog, canis familiaris. Genetics *151:* 803-820, 1999.
66. Langford CF, Fischer PE, Binns MM, Holmes NG, Carter NP. Chromosome-specific paints from a high-resolution flow karyotype of the dog. Chromosome Res *4:* 115-123, 1996.
67. Breen M, Thomas R, Binns MM, Carter NP, Langford CF. Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (Canis familiaris) and human. Genomics *61:* 145-155, 1999.
68. Yang F, O'Brien PC, Milne BS, Graphodatsky AS, Solanky N, Trifonov V, Rens W, Sargan D, Ferguson-Smith MA. A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. Genomics *62:* 189-202, 1999.
69. Sargan DR, Yang F, Squire M, Milne BS, O'Brien PC, Ferguson-Smith MA. Use of flow-sorted canine chromosomes in the assignment of canine linkage, radiation hybrid, and syntenic groups to chromosomes: refinement and verification of the comparative chromosome map for dog and human. Genomics *69:* 182-195, 2000.
70. Parker HG, Yuhua X, Mellersh CS, Khan S, Shibuya H, Johnson GS, Ostrander EA. Meiotic linkage mapping of 52 genes onto the canine map does not identify significant levels of microrearrangement. Mamm Genome *12:* 713-718, 2001.
71. Vignaux F, Priat C, Jouquand S, Hitte C, Jiang Z, Cheron A, Renier C, Andre C, Galibert F. Toward a dog radiation hybrid map. Journal of Hereditary *90:* 62-67, 1999.
72. Li R, Mignot E, Faraco J, Kadotani H, Cantanese J, Zhao B, Lin X, Hinton L, Ostrander EA, Patterson DF, de Jong PJ. Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library. Genomics *58:* 9-17, 1999.
73. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. The sequence of the human genome. Science *291:* 1304-1351, 2001.
74. Vinogradov AE. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry *31:* 100-109, 1998.
75. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics *2:* 231-239, 1988.