The Working Dog Project: Investigating the genomics of working dog performance.

Significance: The Working Dog Project will apply cutting-edge genomics to investigate how genetic variation influences two key working dog behaviors: scenting and retrieving. Through this new initiative, we propose to promote, and enable, large-scale, collaborative research into the genetics of dog behavior, with the goal of providing the working dog community with the new tools they need to accelerate the breeding and training of successful working dogs.

Problem Statement:

Dogs and people have lived side by side for tens of thousands of years in a mutually beneficial relationship, with humans selectively breeding dogs to perform useful jobs. Even today, working dogs fill critical roles in a wide range of fields, from ensuring public safety to easing anxiety.

While demand for highly skilled animals is only growing, their availability is limited. Despite decades of pedigree analysis and focused breeding programs, the training success rate for service dogs hovers at only around 50%; for high-performing scent dogs, success rates are even lower. This leads to significant issues including wasted training dollars, puppies pushed into roles in which they cannot succeed, unmet client needs, and, in the military, dogs that cannot fulfill the elite requirements to protect and serve.

Genetic tests focused on behavior or temperament would help solve the central challenge: accurate prediction at an early age of potential for successful training. They could help identify which dogs should be trained for particular jobs, and would support more successful breeding programs. Indeed, the potential power of behavioral genetics in dogs was a major factor motivating the Dog Genome Project[1], which led to the publication of a reference dog genome in 2005[2].

Yet, despite this early enthusiasm, only a handful of genetic loci have been associated with canine behaviors in the last ten years. **Currently, there are no useful genetic tests for behavioral traits, suggesting a fundamentally new approach is needed.**

We believe that a major obstacle impeding canine behavioral genetics is that the studies have simply been far too small. Unlike Mendelian-inherited traits and diseases, behavior is genetically complex. Recent work both in people[3] and in dogs[4,5] shows that behavioral traits are shaped by changes in many different genes, rather than just one, and these changes are influenced by environmental factors. While we could map variants underlying traits like white coat color[6], hairlessness[7], and ridge[8] with just a few dozen dogs, to understand the genetics of behavior, we will need to study thousands or tens of thousands of dogs. Yet, individual working dog organizations and breeding programs can typically provide DNA samples for, at most, hundreds of dogs.

To successfully develop predictive genetic tests for working dogs, we need a new, highly collaborative initiative that brings together working dog organizations, and follows the powerful open science model pioneered by the human genetics community. We also need new tools for DNA sequencing and analysis tailored to the unique characteristics and history of working dogs. Now, for the very first time, new technologies make this possible, but it is far from simple.

Rationale:

Using a collaboration-focused approach, we will tackle the genetics of working dog behavior. The project will provide the foundation for developing trait prediction tools for working dog breeders, while also creating an unparalleled resource for canine genomics that is accessible to the global community of researchers and the general public. This open science model will activate a new era of large-scale dog genetic research that will ultimately benefit all dogs.

Innovation. This project will have the benefit of leading-edge technology from collaborating scientists at the forefront of dog genomic research, and will harness resources and technology not available just months prior. Through this work, we will establish a knowledge base that will fuel future expansion of the project into new behavioral traits and dog populations.

Scale. We envision this project as the first step in a comprehensive program, the scale and scope of which is unprecedented in dog genetics. We believe that this is the only approach that will succeed in providing the dog community with the results they seek. Human genomics has yielded priceless benefits for human health and well being, but only after decades of research and at a cost of billions of dollars. We know from centuries of scientific exploration that, while small studies fail, projects that achieve large sample sizes work, and revolutionize entire fields of research.

Collaboration. Because each individual working dog population is relatively small, we need new methods that can combine the information from many different populations to achieve very large sample sizes. While this is complicated, because of the complex genetics of dog breeds, lack of existing resources, and diverse approaches to phenotyping behaviors, we believe new technology now makes it possible.

By intersecting innovation, scale and collaboration, we can harness two ongoing technological revolutions. First, rapid and inexpensive next-generation sequencing has led to huge catalogs of genetic variation in dogs, allowing us to generate nearly complete genome sequences for every dog, rather than relying on sparse SNP genotyping methods that only capture a tiny percent (less than 0.1%) of the actual genetic information. Until now, this approach has not been an option for dog genomic studies because it requires a reference dataset with full genome sequences for over 1000 dogs. Dr. Elaine Ostrander (NIH) has now generated this resource and, through the data sharing platforms at the Broad Institute, we will be making it available to the dog research community.

Simultaneously, the rapid spread of handheld devices (smart phones and tablets) is an unprecedented opportunity to solicit detailed behavioral phenotype information from dog owners. While owner provided phenotypes may be less rigorous than those provided by animal behavior professionals, we can assemble very large cohorts of dogs using this approach.

Project Goal:

We propose to develop a new model for genome-wide association of behavioral traits and disorders in dogs that will allow us to integrate the information from diverse dog populations, and include both working dogs and pet dogs. Using this model, we will seek the precise genetic

changes influencing key working dog behaviors.

We will look for genetic determinants of scenting and retrieving ability. This work will provide a starting point for developing predictive genetic tests for scenting ability, and establish collaborative use of open-access data sets as a paradigm for the study of dog behavior. We will use an approach for combining multiple datasets that is based on new statistical methods that have proven exceptionally powerful for investigating psychiatric disorders and other complex traits in people.

Work Plan:

Executive Summary

There are three key objectives:

1) Identify genetic variants that influence scenting and retrieving behaviors

2) Develop and test methods needed for subsequent phases of the project

3) Assess the feasibility of the collaborative approach needed to achieve the large sample sizes proposed for subsequent phases of the project

Specifically, we will search for genetic variants that influence two behaviors: scenting and retrieving using DNA samples and behavioral phenotypes from three distinct dog populations. We have selected these two behaviors because we will have measures for both in each of the included populations. In addition, our preliminary data suggests that behaviors that originate in the wolf predatory sequence, like scenting and retrieving, likely appeared early in dog domestication and were subjected to strong artificial selection. As artificial selection pushes variants with large effects on phenotype to high prevalence in populations, we believe we may have sufficient power to find genetic variants that influence these behaviors with just 800-1000 dogs.

To achieve the above objectives, we will focus on three components of work:

1. Collect detailed behavioral phenotypes and DNA samples

2. Generate low cost, full genome sequence information

3. Develop and apply new statistical methods for meta-analysis

Project Details

Component 1. Collect detailed behavioral phenotypes and DNA samples. We will include dogs from three different populations: (1) We will collect saliva samples and behavioral assessments from 100 NEADS assistance dogs, including both dogs in training and dogs that have been placed; (2) the Canine Performance Sciences group at Auburn University, led by Dr. Paul Waggoner, has already assembled a databank with DNA samples and full performance evaluations for 140 dogs from their breeding program, and will be a partner on this project; (3) Our Darwin's Dogs citizen science project (darwinsdogs.org) has enrolled over 13,000 dogs and

collected detailed owner-reported behavioral phenotypes. We are currently generating dense genotype data for 600 of these dogs through an NIH R21 grant, and we will include these dogs in our project at no additional cost. We have already established collaborations with NEADS and Auburn University and can obtain those samples quickly.

Component 2. Generate low-cost, full genome sequence information on each dog using Broad Institute multiplex sequencing and whole genome imputation.

Association studies in dogs to date have used low density genotyping arrays (with up 200,000 markers) that capture just a small proportion of the genetic variation in each dog, severely limiting the development of useful genetic tools. This is in stark contrast to current studies of complex traits in humans, where millions of markers are tested and a technique called genomic imputation is applied. With this approach, the state of millions of additional genetic markers can be then determined. Genomic imputation requires low-cost, whole genome sequence information and large, freely available reference databases. Imputation has helped human geneticists combine information from genetically diverse populations, identify causal genetic variation, and develop predictive genetic tests. We expect similar results for dogs through genomic imputation.

In component 2, we propose to generate full genome sequence information on each dog by applying genome imputation to low-coverage, whole genome sequencing data. At a price point comparable to using genotyping arrays, we will produce a dataset with far more statistical power, and one in which we can detect interesting genetic variation missed in the smaller panel of markers chosen for the array.

Until now, we have not had the tools needed to use this method in dogs, but that is changing. Dr. Elaine Ostrander and her colleagues have done full genome sequencing on over 1000 dogs, and are making this resource freely available through the data sharing platforms at the Broad. In addition, Broad Genomics has pioneered new methods for high-throughput, highly multiplexed sequencing. With these two resources, we will generate data for millions of genomic markers for under $200 a dog [9].

We will apply these methods to dogs from all three populations, yielding a dataset with over 5 million data points per dog, for all 840 dogs in our study. This approach will be critical to later phases of the project, allowing us to quickly generate nearly complete genetic data for thousands of dogs.

Component 3. Develop and apply new statistical methods for meta-analysis of both working dog populations and a cohort from Darwin's Dogs.

Discovery of most genetic risk variants in human complex diseases have come from large-scale meta-analyses of genomewide association studies, a trend that is accelerating[10]. Meta-analysis synthesizes information from multiple independent studies to increase statistical power and reduce false-positive findings. One of the biggest to date, a very successful meta-analysis of genome wide association studies (GWAS) of depression published in *Nature Genetics*, combined carefully phenotyped patient populations with a cohort for which they had only self-

reported depression diagnoses[11]. This design is similar to our proposal to combine information from working dogs, with professional behavior evaluations, and pet dogs, with owner reported behavioral phenotypes.

By meta-analyzing full genome sequencing data, we will be able to pinpoint the precise genetic variation responsible for behavioral differences. This is in contrast to earlier genetic mapping studies in dogs, which, because they were using sparser data, were only able to find regions of association. The regions usually contain many genes and thousands of different variants, and are not specific enough for precision use. Our ability to identify specific, causal factors will support the development of high-quality predictive genetic tests later in the project.

Using the data from component 2, we will develop the methods for applying meta-analysis to large, very dense genomic datasets for dogs. We will adapt tests used in human populations to account for factors unique to dogs, including the limited genetic diversity within breeds, the large genetic differentiation between dog breeds, and the mixing of breeds seen in many pet dogs.

We will apply the dog specific methods we develop to the data from component 2 and search for genetic variants associated with scenting and/or retrieving ability.

To facilitate development, we will first assess our overall statistical power to detect any causal variants using the genome-wide distribution of association scores. We will examine in depth the top loci and their effect on phenotype. As we have detailed information for each dog, we will integrate environmental factors that might influence the behaviors of interest into the analysis through a joint meta-analysis for main and interaction effects. We will also use cross-phenotype meta-analysis to check for multiple associations at a single marker for both retrieving and scenting ability, to assess whether they share a common genetic background or are entirely distinct abilities. Finally, we will carry out exploratory cross-phenotype checks, to see if any of our significant markers are also associated with other phenotypes.

Long term vision:

The Working Dog Project, with the support of the Theriogenology Foundation, will serve as a proof-of-principle demonstration that we can combine information from diverse dog populations, and that a large-scale, collaborative approach to working dog genetics is feasible. Through this project, we will also seek to identify the first genetic variants significantly associated with working dog performance. However, this project alone will not yield a working toolset for predicting behavioral outcomes in dogs.

Human genomics have yielded enormous benefit for health and safety, but only as a result of decades of research. We will shortcut this process in dogs by leveraging our experience in humans, but it can not be done in one single project.

The Working Dog Project is the first project in a comprehensive program that engages working dog organizations, scientists and dog breeders. This program will leverage the power of genomics to support better working dog breeding and training programs. Work for subsequent projects will be planned based on results from previous projects, so we can integrate what has been learned into the plan. Based on our current available data and methodologies, we propose

three follow-on projects to complete the program.

Working Dog Project #2. We will bring together a large consortium of researchers, working dog organizations, and citizen science initiatives to build the largest dog behavioral genomics databank ever assembled. Using this resource of tens of thousands of dogs, and the tools developed by the Working Dog Project, we will pinpoint the precise genetic changes influencing dog behavior.

Working Dog Project #3. We will develop, and validate, polygenic predictive tests for working dog performance. In this phase, we will test predictive models developed in earlier projects on a new set of dogs. This critical step will allow us to directly assess the accuracy of the predictive models, and ensure that these new tools have broad applicability to diverse dog working dog populations.

Working Dog Project #4. We will build software and/or genomic tools, and make them easily accessible for working dog breeders. Using the information from the previous projects, we will design algorithms and diagnostics that will improve performance outcomes while maintaining the health of the breeding population. We will develop tools, and then validate and improve these tools, in partnership with all dog breeders.

By engaging all stakeholders, and following the powerful open science model pioneered by the human genetics community, we can use modern genomics to make the ancient partnership between people and dogs even stronger.

Principal Investigator

Elinor K. Karlsson Ph.D.

As the Principal Investigator, Dr. Karlsson will be responsible for overall management of the project. She is the Director of Vertebrate Genomics at the Broad Institute, and Assistant Professor in the Bioinformatics and Integrative Biology Program at the University of Massachusetts Medical School (UMMS). Dr. Karlsson is a computational geneticist with over 12 years of experience in canine genomics and mammalian sequence analysis. She developed widely used statistical methodologies for association mapping and other genetic analyses in dogs, and has expertise developing algorithms to detect genetic signatures of adaptation and trait association and methods for functionally investigating candidate variants. Dr. Karlsson led the first genome-wide association studies of complex traits in dogs, linking particular genetic mutations to osteosarcoma[12] and compulsive disorders[4,5,13]. She will oversee the completion of all work related to the three components discussed in this proposal.

Facilities

The Broad Institute is a unique scientific community of diverse talents that brings together world-class faculty, professional staff, postdoctoral fellows and students from throughout the MIT and Harvard communities and beyond in a collaborative environment. The Institute is organized

around scientific programs and platforms, a transparent infrastructure that allows biology- and technology-focused scientists to jointly build, apply, and share cutting-edge tools and knowledge with scientists worldwide. The Broad currently maintains six specialized scientific platforms, which provide centers for technological innovation and large scale genomics projects –Chemical Biology, Genomics (comprised of Samples, Genetic Arrays, and Sequencing), Imaging, Metabolite Profiling, Proteomics, and RNAi.

The Broad Institute will facilitate the success of our grant proposal in several key ways.

1. A vibrant and diverse community conducts collaborative science by multi-disciplinary teams. Our community consists of biologists, clinicians, experimental technologists and engineers, computational scientists, and software engineers, and involves staff scientists, graduate students, postdoctoral fellows, platform directors, and Core and Associate Member faculty. As a result, we have developed an institutional culture of respect and a mutual scientific 'language' that facilitates effective research across disciplines. In particular, our team-based approach to genomics ensures that computational efforts are driven by biological questions, that computational scientists impact the conception and design of experiments, and that biologist users can quickly and critically evaluate computational solutions, and affect their development.

2. Direct and sustained interactions between academic research labs and technology platforms. The technology platforms (1) provide exceptional capabilities and scale to conduct genomics research; (2) are the 'next step' when a technology that initially emerged in a research lab needs to be scaled up; and (3) are partners with the research labs in developing new technologies, bringing know-how, technological innovation and exceptional human resources and equipment to this process.

We will work closely with the Broad Genomics platform, one of the largest produces of human genomic information in the world. They have pioneered methods development, rapid and nimble implementation of new instruments, and continuous improvement of massively parallel sequencing, enabling high-throughput, rapid sequencing of many samples at low per-base cost. From receiving the first alpha Solexa instrument (installed 9 years ago) to the first HiSeqX10 platform, the group has driven technology innovation by improving processes, inventing applications, and creating rigorous quality measures – and shared these developments with the scientific community.

3. A nimble culture of innovation and scientific entrepreneurship. The non-hierarchical culture of the Broad, combined with unprecedented expertise and resources, encourages nimble science and fast development and adoption of new approaches. Open space and mixed lab organization facilitates the sharing of ideas, know how, technologies and equipment, thus increasing the pace of discoveries and leading to exciting new directions for research. In particular, it encourages grass-roots initiatives, by trainees (graduate students and postdocs) and young investigators, providing them the resources and backbone to conduct large scale genomics research.

1. Ostrander EA, Lindblad-Toh K, Lander ES, Others. Sequencing the genome of the domestic dog Canis familiaris. URL: http://www genome gov/Pages/Research/Sequencing/SeqProposals/CanineSEQedited pdf, National Human Genome Research Institute (last accessed June. 2005; Available: https://www.genome.gov/pages/research/sequencing/seqproposals/canineseqedited.pdf

2. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438: 803–819.

3. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511: 421–427.

4. Dodman NH, Karlsson EK, Moon-Fanelli A, Galdzicka M, Perloski M, Shuster L, et al. A canine chromosome 7 locus confers compulsive disorder susceptibility. Mol Psychiatry. 2010;15: 8–10.

5. Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, et al. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. Genome Biol. 2014;15: R25.

6. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, Anderson N, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet. 2007;39: 1321–1328.

7. Drögemüller C, Karlsson EK, Hytönen MK, Perloski M, Dolf G, Sainio K, et al. A mutation in hairless dogs implicates FOXI3 in ectodermal development. Science. 2008;321: 1462.

8. Hillbertz NHCS, Isaksson M, Karlsson EK, Hellmén E, Pielberg GR, Savolainen P, et al. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat Genet. Nature Publishing Group; 2007;39: 1318–1320.

9. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat Genet. 2012;44: 631–635.

10. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013;14: 379–389.

11. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016;48: 1031–1036.

12. Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. Genome Biol. 2013;14: R132.

13. Noh HJ, Tang R, Flannick J, Dushlaine CO, Swofford R, Howrigan D, et al. Evolutionary and functional data power search for obsessive-compulsive disorder genes [Internet]. bioRxiv. 2017. p. 107193. doi:10.1101/107193

13. Noh HJ, Tang R, Flannick J, Dushlaine CO, Swofford R, Howrigan D, et al. Evolutionary and functional data power search for obsessive-compulsive disorder genes [Internet]. bioRxiv. 2017. p. 107193. doi:10.1101/107193