ABSTRACT

KENNERLY, ERIN MICHELLE. Genomic Explorations of Infectious and Genetic Disease in Canids. (Under the direction of Gregory C. Gibson.)

Wild and domesticated dogs have long been of interest to humans as both companions and working animals. Selection for desired traits like behavior, size, and shapes has resulted in over 350 genetically distinct breeds, but inevitably less desirable traits like susceptibility to disease and congenital genetic disorders have been carried through in certain breeds of dogs. Because of their genetic history and close relation to humans, canids are an ideal model system to study the interaction between environment and genetics in the promotion of complex diseases and disorders. In this thesis, I explore the use of emerging genomic tools to study infectious disease, the effect of habitat change in red wolves, and pharmacogenetic response to epilepsy treatment across breeds of dogs.

In the first study, heterologous microarrays were used to examine the effect of captive and free ranging habitats on peripheral blood gene expression in red wolves. An algorithm was developed to simultaneously estimate both genetic and environmental factors that contribute to differential gene expression profiles. The first two principal components of overall expression variation defined effects predominantly of habitat and genetic relatedness. Genes associated with stress pathways were the most differentially expressed as determined by gene ontology analysis. I applied microarray technology to define the extent of immune related changes in gene expression associated with a 12 week hookworm infection in three beagle pups. Measured immunoglobulin levels indicated an active immune response, and gene expression profile changes involving 305 transcripts were evident between time points. However gene expression changes associated with Th1 and Th2 types of response were evident preventing a clear characterization of the type of immune response associated with a hookworm infection. Lastly, I use a pharmacogenetic approach

to identify genetic variation associated with Phenobarbital drug response in epileptic dogs. Using 384 genetic markers in 30 genes that have been implicated in drug metabolism and transport in a case control study between Phenobarbital responsive and nonresponsive epileptic dogs, I identified associations in *SCN2A2*, *KCNQ*3, *ABCC4*, *GABRA2* and *Epoxide hydrolase* gene, which may be predictive of drug response. Because of the strong effects of breed structure, these should be studied more thoroughly before they are considered as the basis of a genetic test for drug response. As with many genomic experiments these studies provide insight and the starting point for further studies into the relationship between genes and phenotypes.

Genomic Explorations of Infectious and Genetic Disease in Canids

by
Erin Michelle Kennerly

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Genetics

Raleigh, North Carolina

2008

APPROVED BY:

Dr. Gregory C. Gibson Chair of Advisory Committee Genetics	Dr. Trudy F. C. Mackay Co-Chair of Advisory Committee Genetics
Dr. Natasha Olby Committee member Clinical Sciences	Dr. Jorge Piedrahita Committee member Molecular Biomedical Sciences

DEDICATION

Thank you to all the people I love who never let me give up my dreams. Especially to my parents whose support and “hard pushes” taught me to never stop at closed doors no matter how big and sturdy they seemed. To Robbie for all the patience these last few years, I love you. Lastly, to my little brother who taught me how to keep humor in any situation.

BIOGRAPHY

Erin Michelle Kennerly was born on September 10, 1980 in Asheville North Carolina. At age 3 her parents moved to Greensboro, NC where they lived until she was 16 when they moved to Madison, NC. She attended Western Guilford High School in Greensboro through her sophomore year when she transferred to Dalton L. McMichael High School in Madison to complete high school in June of 1998. August of 1998, she started in the First Year College Program at North Carolina State University in Raleigh, NC and later transferred into the Animal Science degree program her sophomore year. While attending NCSU, Erin began working at the NCSU veterinary school as a technician in a canine gene therapy research program under the direction of Dr. Mark Jackson. It was during this time with Dr. Jackson that she first became interested in genetics, especially how it relates to understanding diseases and the treatment of those diseases. Deciding to switch from the traditional Animal Science degree program, Erin added genetics as a minor and took an advanced undergraduate genomics course taught by Dr. Greg Gibson. Dr. Gibson's course further inspired her to pursue a career in research and upon graduating with honors from NCSU in December 2002, she began working in Dr. Gibson's lab as a research technician. She quickly realized she actually enjoyed working inside all day in the lab and eventually she was overseeing the molecular biology part of Dr. Gibson's lab. In August of 2005, Erin continued to follow the obvious path and started in the genetics graduate program at NCSU under the direction of Dr. Greg Gibson. As part of her dissertation work Erin continued to apply genetics to understand immune response in hookworm infected dogs, the effect of the environment on captive wolves, as well as drug response in epileptic dogs.

ACKNOWLEDGEMENTS

I would like to thank the Department of Genetics for the giving me the opportunity to learn and develop as a scientist. My advisor and mentor Dr. Greg Gibson for always believing in me and pushing me to be better and think outside the box. A special thanks to my committee members, Dr. Trudy Mackay, Dr. Natasha Olby, and Dr. Jorge Piedrahita for their support and encouragement through my graduate career. I need to especially thank Dr. Olby who was influential in getting the epilepsy project off the ground and was very instrumental in its success. I should thank Dr. Olby and Dr. Breen for their help in making my very first research project a success. I would like to thank all the current and past members of the Gibson lab, including pseudo labmate Felix Araujo. A special thanks to Ian Dworkin and Lisa Goering for all the advice and the long talks over the years. I would like to thank Jenny Moser, Gisele, and Ana Patricia for helping to guide and teach me during my first years in the lab. To Youssef, I really don't know what to say, I guess you have finally left me speechless. I am grateful for the opportunity to meet Martha who taught me how to look on the bright side of life and sing like no one cares how bad it sounds. Martha also truly introduced me to Dr. Atchley who has since become a good friend. Thanks to Steph, Laura and Murray, who provide another place besides my house for lab functions and inherited me with all my tendencies from Lisa and Ian. Many thanks to all the lab members who have tried in vain to educate me about politics and change my stubborn mind, but in the end just convinced me even more that it is okay to agree to disagree. You guys' friendship has been invaluable and something I will cherish always. Thanks to Owen McMillan for tolerating me during the transition of lab space and for "not caring" as much as I do. Thanks to my

collaborator Prema Arasu for the guidance and talks about life. Also thanks to collaborators Karen Munana and Julie Osborne at the vet school for sample leads for the epilepsy project. Michael Stoskopf and Anne Ballman for help with the red wolf project. A special thanks to Stan Martin and Russ Wolfinger for help with JMP formatting and SAS models. I wouldn't be here today if I didn't have the support of all you guys.

Lastly, I would like to thank my parents, Rob and Laura, for their support through all the years. Thank you to my best friend Misty for always being there to talk and listen. Robbie, thank you for "finding" me and being my rock and hopefully we will have many happy years to come. Last but not least, I am grateful for my rat terrier (aka my shadow Winnie) Winston who always gives me love and devotion no matter how late I came home and how bad my mood. I only wish life could be as simple as the bond between me and my dog.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER FOUR

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER ONE

Introduction

Overview

With the emergence of the genomic era, researchers are taking a fresh angle on studies of complex traits. The human genome was completed in April of 2003 which expanded the ability of researchers to take more of a top down or a systems approach to identify genetic variants responsible for phenotypes or diseases. Further genomic sequencing of other model organisms has allowed for the direct comparison of genomes. Through the use of microarray technology, researchers are now able to more easily identify pathways and even indentify new gene networks involved with a complex trait. Whole genome association studies have allowed scientists to identify key genetic components in complex diseases like cancers (McGrath, *et al.,* 2008)). These studies have also allowed for the rapid expansion of fields like the newly emerging pharmacogenomics and metabolomics. In addition researchers are better able to characterize the effect of the environment on a whole genome level as well as better understand the extent of genetic and environmental interactions. To date whole genome sequencing has been done on 32 different eukaryotes, ranging from the malaria parasite Plasmodium to the chimpanzee. With the decreasing cost of sequencing, more genomes are being sequenced, allowing the use of non-traditional model organisms to be used for studying human diseases and to understand complex traits.

In this thesis, I investigate the application of genomic tools to answer questions about environmental effects, the immune response and drug response in dogs, a non-traditional model organism. I will demonstrate how microarray technology can be used to measure gene transcript changes associated with lifestyle in the red wolf, as well as genes involved in an immune response to

hookworm infection in beagle pups. Lastly, I will use association testing to identify genetic variants associated with drug response in epileptic dogs.

Canine Domestication

Dogs were the first domesticated animals and have served as companions to humans for centuries. It is estimated that over 40% of US households have at least one dog (Case, 2008). The first archeological evidence of a domesticated dog dates back 14,000-15,000 years ago; however DNA studies date the separation from the grey wolf (*Canis lupus*) at around 100,000 years ago (Vila, *et al.,* 1997; Morey, 1992). Based on mitochondrial sequence data the carnivore clade separated into the Canoidea and Feloidea clades ~40-50 million years ago followed by further separation of the Canoidea clade into 3 distinct groups ~10 million years ago, which includes the modern canids (Wayne, *et al.*, 1989; Dragoo and Honeycutt, 1997). The *Canis* phylogeny includes the grey wolf, coyote (*Canis latrans*) and the ethopian wolf (*Canis simensis*). Also within the monophyletic clade is a sister taxon of jackals and Asian wild dogs (figure 1). Even though all species within the *Canis* clade can hybridize and have the same number of chromosomes (*2n=78*), only the jackal and the grey wolf were suspected to be the ancestor of the domestic dog. Comparison of cytochrome *b* sequence and additional comparison of mtDNA control region markers in 140 dogs, 162 wolves from 27 populations worldwide, 5 coyotes, 2 golden jackals, 2 black backed jackals and 8 simien jackals identified the grey wolf as the main ancestor of the domestic wolf (Wayne, 1993; Vila, *et al*., 1997). However, since mitochondrial DNA is maternally inherited, a contribution of coyotes and jackals to

the domestic dog gene pool cannot be excluded without further studies.

Historically, grey wolves roamed much of Europe, Asia and North America making identification of the exact location of domestication, or determining whether there was a single domestication or multiple events, difficult. Unlike modern day dogs, few genetic subdivisions are seen within the wolf population. Because the grey wolf still exist as a fairly uniform species, studies can be done which compare the modern day dog to their ancient relative, thus allowing a better understanding of modifications during domestication. Ancient remains of both dog and wolves have been found throughout Europe, Asia, and the New World. This along with the wide variety of phenotypes among modern dogs suggests there may have been multiple domestication events with continued genetic exchange between wolves for many years (Vila, *et al*., 1997). Several studies comparing haplotypes between Old and New World breeds as well as different wolf populations throughout the world, provide evidence for an old world domestication event that occurred several times, or has been followed by several rounds of interbreeding between wolves and dogs (Vila, *et al.,* 1997; Salvolainen, *et al.,* 2002).

Sequence divergence between grey wolves and dogs is about 1.8% compared to 4% divergence between other closely related canids like the coyotes and Ethiopian wolves (Vila, *et al*., 1999; Wayne, 1993). Even though they share the majority of their genome with the grey wolf, domestic dogs are very morphologically diverse, as evident in the over 500 breeds worldwide. There have been several studies aimed at uncovering the source of variation used for the artificial selection of such a diverse range of phenotypes seen in dogs. One such study by Björnerfeldt, *et al.*

(2006) suggested that lifestyle changes associated with domestication relaxed selective pressures and allowed for a faster rate of accumulation of nonsynonymous changes. These changes allowed for the introduction of novel genetic variation which allowed for further selection. However it may be possible that the genetic variation needed to produce such a range of phenotypes was always present in the wolf.

A parallel approach to study of the genetic basis of domestication has been undertaken with a shared relative of dogs and wolves, the silver fox. The Institute of Cytology and Genetics of the Russian Academy of Science began an artificial domestication program aimed at characterizing the changes associated with canid domestications. These foxes have been selected for friendly behaviors towards humans for 40 generations (Trut, 1999). The tamed foxes' social behavior mimics that of the domestic dog. Even more interesting is that even though selection was only based on behavior, morphological and developmental differences were also evident in the tamed foxes (Trut, *et al.,* 2004: Wayne, 2001). These changes included floppy ears, rolled tails, shorter skull shape, coat color and increased opening of the eyelids (Wayne, 2001). Further studies are currently underway to identify regions of the genome where selection is acting.

In chapter two, I demonstrate the use of microarray technology to examine gene expression differences associated with lifestyle in captive and free range red wolves. Although not mentioned above in the evolutionary framework of the domestic dog, red wolves are a close relative to the modern dog. There is still debate on the exact origin of red wolves, but microsatellite and mitochondrial data indicate they likely derived from a coyote and grey wolf hybridization (Wayne,

1993; Wayne and Jenkins, 1991; Roy *et al.,* 1994). Despite this, conservation efforts to preserve the red wolf began in the 1980's which by then had become extinct in the wild. Today there are several captive breeding and release facilities throughout North America which include the facility at Alligator River, NC. (which was the source of our samples). The original captive breeding program was established from a colony of 40 wild caught wolves and extensive pedigrees have been kept since. Having the pedigree information is useful, because it allows us to track lineages and account for genetic relatedness among our samples. In this way we can account for the amount of variation in gene expression explained by both the environmental effects as well as the genetic effects.

The Canine as a Model System

The canine genome sequence was released in 2005 and has since provided a background structure to expand the use of the dog as a model system in systematic and population genetic based studies (Lindblah-Toh, *et al.,* 2005). The domestic dog (*Canis familiaris*) historically has not been used as a model system for human studies due to the lack of molecular tools, cost of rearing, and generation length, which lessened the feasibility of large scale experiments (Wayne and Ostrander, 2007). However dogs share a unique history with man that does not extend to other traditional model organisms. Dogs were the first organisms to be domesticated and were selected for various uses and roles in society. Dogs share not only a migratory history with humans but were also under similar selective pressures from their common environments and diets (Morey, 1994). Dogs also possess a wide range of diversity in body size, shape, coat color and behavior (reviewed in Wayne and Ostrander, 2007). In addition dogs have many commonalities in the clinical

manifestation of many diseases and syndromes in humans.

In chapter three, I use microarrays to examine gene expression changes in beagle pups during an acute phase infection with the canine hookworm, *Ancylstoma caninum*. Hookworm infection is the leading helminth infection of humans worldwide and leads to weight loss, anemia, and death (Fujiwara, *et al.,* 2006). Currently there are several other animal models used to study hookworm infections; however dogs are the only organisms where the infective hookworm larvae will fully develop and mature to blood feeding parasites (Loukas and Prociv, 2001). Dogs are also able to successfully acquire long term protective immunity against hookworm infections where humans do not (Fujiwara, *et al.,* 2006). Dogs are commonly used for preclinical and immunotoxicological screening of investigational drugs and their immune system shares many characteristics with humans, making dogs a more comparable model to study gene expression changes related to infection. Dogs, like humans, are both immunologically competent at the time of birth and their immune system reaches full maturity shortly after birth (Felsburg, 2002; Holsapple, *et al.,* 2003). In addition, studies comparing the common gamma chains, which are receptors for many interleukins and are important for lymphocyte development is essentially identical between dogs and humans (Flesburg, 2002). The similarities between the human and dog immune systems as well as their comparable infective hookworm states makes dogs an ideal model to study immune response to hookworm infections.

Population Structure

The population dynamics of dogs make them an ideal system to study both monogenic and polygenic diseases. The canine population has gone through two major bottlenecks; one with the domestication from the wolf and the second with the creation of modern breeds. The creation of breeds and more recently breed clubs with closed stud books served to reduce the effective population size, thus increasing the effects of genetic drift which further increased the divergence between breeds. Average nucleotide heterozygosity across many breeds is estimated at 8×10^{-4} similar to what is seen in human populations. However, over 27% of the total molecular genetic variation is between dog breeds, compared to 5-10% that is seen between human populations (Sutter and Ostrander, 2004). Parker *et al.* (2005) further demonstrated the considerable divergence between breeds by correctly assigning 99% of 414 dogs to their respective 85 breeds using only 96 microsatellite markers. In this same study, dogs separated into a hierarchy of 4 main groups corresponding with historical geographic origin, behavior and morphology (Parker, *et al.,* 2005). A more recent study including an additional 47 breeds showed a fifth cluster of large mountain breed dogs (figure 2) (Parker, *et al.,* 2007).

Further analysis of the canine genome led to the inference that ancestral haplotype blocks were relatively small (~10kb), each consisting of 4 to 5 different haplotypes for the entire canine population prior to the creation of breeds (Lindblad-Toh, *et al*., 2005). Long range previously uncorrelated blocks were selected upon during the creation of breeds, thus resulting in common long range blocks within breeds. However when looking across multiple breeds, traces of the ancestral haplotypes are still present as recombination has not had time to break down these blocks

(Lindbald-Toh, *et al.,* 2005). This population structure is particularly useful in association studies where linkage disequilibrium measurements are important for project design and SNP density. Linkage disequilibrium (LD), the non-random association between alleles, is estimated at about 1Mb in most breeds, but can extend out to 3.2 Mb as seen in breeds like the Pekingnese (Sutter, *et al.,* 2004).

These LD blocks are over 20 times longer than those seen in human populations. However when examining LD across multiple breeds, it breaks down closer to levels seen in human populations (~200kb) (Figure 3) (Lindbald-Toh, *et al.,* 2005; Sutter, *et al.,* 2004). By characterizing SNPs within these LD blocks we can capture the majority of the genetic variation within each haplotype block. The longer the extent of LD, the larger the block and likely fewer SNPs will be needed to capture the genetic diversity. The large LD blocks within breeds greatly reduce the SNP density needed to characterize genetic diversity, while smaller LD blocks across multiple breeds allows for more fine scale mapping of a region of interest. However, because of the large size of these blocks several genes may be present within one block. Including dogs from multiple breeds will help narrow the region of interest, but more SNPs will likely be needed as the LD blocks become smaller.

A recent study by Sutter, *et al.* (2007), illustrated how the dog's population structure can be used to identify a genetic variant associated with body size in dogs. First they identified two quantitative trait loci (QTL) associated with size in large and small breed dogs. Then by associating 116 SNPs with skeletal size in 463 Portuguese water dogs, they identified a single peak near the

Insulin growth factor 1 gene (*IGF1*). Further association tests across a wide range of large and small breed dogs identified a single haplotype associated with all small breed dogs that was nearly absent in larger breed dogs. This study shows the use of identifying regions of interest by taking advantage of the long range LD within a breed and then further narrowing that region down to identify key genetic variation associated with a trait by looking across multiple breeds.

Canine Disease Studies

The small effective population size as well as population bottlenecks and founder effects are linked to the increase in frequency of rare "risk" alleles within breeds which results in increased susceptibility to genetic diseases and disorders in certain breeds or lineages of breeds (Ostrander and Kruglyak, 2000). Evidence for this is seen in the increased risk of diseases within certain breeds. Table 1 shows a few of the diseases from the University of Cambridge Canine Inherited Disease database as well as the breeds they are prevalent in (Sargan, 2004).

Pharmacogenetics is a fairly new field that has spun off from genomic studies and emphasizes the identification of genetic variants that influence drug response in an individual. Pharmacogenetic studies have shown promise for predicting response to and dosaging of warfarin, phenytoin and carbamazapine in humans (Tate, *et al*., 2005; Zhu, *et al*. 2007). In the case of warfarin, dosing amounts and drug response has also been linked to age, weight, and ethnicity (Zhu, *et al.,* 2007). Similarly, a study in dogs has identified a polymorphism in the cytochrome P450 gene

CYP1A2 that facilitates the metabolism of xenobiotic drugs in beagles (Tenmizu, *et al.,* 2004). A more recent study showed that a mutation in *MDR1* that was initially linked to drug sensitivity of over 20 drugs in collies, is also shared in nine other breeds that are derived from the collie lineage (Neff, *et al.,* 2004). This last study also highlights the importance of accounting for the genetic relationships among breeds, and it is likely that other anomalies will be shared across dog breeds.

In chapter four, I describe a pharmacogenetic project where I use case control association tests to identify genetic variation associated with drug response to a common anti-epileptic drug, Phenobarbital (PB). Epilepsy is a common neurological disease in dogs and affects a wide range of breeds. Breed specific studies have identified possible genetic links to the disease, however none are shared across multiple breeds and due to the complexity of the disease it is unlikely that there will be common genetic etiology for epilepsy (Patterson, *et al.,* 2003; Patterson, *et al.,* 2005; Jaggy, *et al.,* 1998; Casal, *et al.,* 2006). In contrast, drug response is less complex and is more likely to have a shared mechanism across multiple breeds. Identification of variants associated with drug response could lead to the development of a diagnostic test where owners and veterinarians can quickly assess how the animal will respond to medication. Secondly, breeding practices may be established to remove refractory contributing alleles from the populations. Similar breeding practices are being established in a variety of breeds like Dalmatians and flat coated retrievers where genetic mutations linked to deafness and glaucoma have been identified (Wood, *et al.,* 2004).

Bead-based Genotyping and Gene Expression Profiling

The methods for characterizing DNA polymorphisms and measuring gene expression

profiling and how they affect phenotypic variation have changed rapidly over the last few years. The rapid emergence of many newly sequenced genomes has pushed technology to develop ways to characterize expression and polymorphism across many genes at one time. One such technology is the microarray, which measures gene expression over many genes simultaneously. The first microarrays consisted of a collection of spotted DNAs on filter paper, and have since evolved into several types including PCR amplified cDNA and long and short oligonucleotides printed on glass slides (Schena, *et al.,* 1998). Correlation of gene expression profiles gives us the ability to characterize gene expression changes across multiple genetic and environmental backgrounds and look for patterns of expression among the genes. In the last few years genotyping arrays using the hybridization principles of microarrays have been developed to characterize polymorphism on whole genome scale (Gunderson, *et al.,* 2005; Matsuzaki, *et al.,* 2004). The merger of these two technologies by associating DNA polymorphism and gene expression profiles has uncovered key genetic networks and cis or trans acting factors involved in complex diseases and traits (Cheung, *et al.,* 2005; Schadt, *et al.,* 2008).

The goals of microarray development are accuracy, scalability, and flexibility. Recently, Illumina Technologies has launched a new bead based microarray and genotyping platform where probes are hybridized to 3 µm nano-beads. Approximately 700,000 oligonucleotide probes are covalently attached to each bead (Kuhn, *et al.,* 2004). These beads are then randomly arranged onto etched micro-well slides and have a density of ~40,000 times greater than traditional spotted microarrays (Michael, *et al.,* 1998). This also allows for multiple arrays to be printed on one slide.

Bead positions are determined by a series of complimentary hybridization of fluorescently labeled oligos (Gunderson, *et al.,* 2004). In addition, each beadset is replicated on the array on average 30 times. Currently there are two types of formats used the Sentrix Array Matrix (SAM) and the Sentrix BeadChip. The Illumina Sentrix BeadChips offers between 1 to 16 arrays per slide greatly reducing experimental cost. The hybridization and measurement of transcript abundance methods are very similar to other array platforms and will not be further discussed. The experiments described in chapters 2 and 3 uses the Sentrix BeadChips while the genotyping in chapter 4 uses the SAM platform, which is further characterized below.

The SAMs uses the same beads as the Sentrix BeadChips however the beads are embedded in an 8x12 matrix of fiber optic cables each capable of holding ~50,000 beads. The GoldenGate Genotyping assay which uses the SAM platform allows for the simultaneous genotyping of 384 SNPs across 96 samples (Fan, *et al.,* 2003). Allele specific extension and ligation reaction followed by dye incorporation PCR amplification is used to label DNA. Labeled DNA is then hybridized to the beads and scanning of the array gives an output of dye intensity measurements. The dye channel intensity measurements are clustered then used to determine the genotype (figure 4). The GoldenGate genotyping platform can process up to 300,000 genotype calls in one day (Fan, *et al.,* 2006). Because of the increased accuracy over other sequence based methods and its high throughput, this technology was used to genotype 65% of the SNP loci for the Interantional HapMap Project (Hinds, *et al.,* 2005).

Conclusion

In this thesis, I will use genomic tools to further our understanding about environmental effects, immune response and disease susceptibility in canids. I will demonstrate the potential of these tools as well as problems associated with these experiments. With carefully planned experiments, the use of genomic tools can be beneficial in untangling genetic interactions and responses to the environment. I will also discuss future potential applications of my work and how these projects can be expanded upon.

REFERENCES

Björnerfeldt, S., Webster, M., Vila, C. (2006). Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res*. 16: 990-994.

Case, L. (2008). ASAS Centennial Paper: Perspectives on Domestication: The History of Our Relationship with Man's Best Friend. *J Anim Sci*. *Epub*.

Casal, M, Munuve, R. Janus, M., Warner, P., Henthorn, P. (2006). Epilepsy in Irish Wolfhounds. *J Intern Vet Med.* 20: 131-5.

Cheung, V., Spielman, R., Ewens, K., Weber, T., Morley, M., Burdick, J. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature.* 437: 1365-1369.

Dragoo, J. and Honeycutt, R. (1997). Systematics of mustelid-like carnivores. *J Mamm*. 78: 426-443.

Felsburg, P. (2002). Overview of immune system development in the dog: comparison with humans. *Hum Exp Toxicol*. 21: 487-492.

Fujiwara, R., Geiger, S., Bethony, J., Mendez, S. (2006). Comparative immunology of human and animal models of hookworm infection. *Parasite Immunology*. 28: 285-293.

Gunderson, K., Steemers, F., Lee, G., Mendoza, L., Chee, M. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet*. 37: 549-554.

Holsapple, M., West, L., Landreth, K. (2003). Species comparison of anatomical and functional immune system development. *Birth Defects Research (Part B)*. 68: 321-334.

Idaghdour, Y., Storey, J., Jadallah, S., Gibson, G. (2008). A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan amazighs. *Plos Genetics.* 4: e1000052.

Jaggy, A., Faissler, D., Gaillard, C., Srenk, P., and Graber, H. (1998). Genetic Aspects of Idiopathic Epilepsy in Labrador Retrievers. *J Sm Anim Prac*. 39: 275-80.

Loukas, A. and Provic, P. (2001). Immune Responses in Hookworm Infection. *Clinical Microbiology Review*. 14: 689-703.

Matsuzaki, H., Dong, S., Loi, H., Di., X., Liu, G., *et al.* (2004). Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods.* 1: 109-111.

McGrath, M., Lee, I., Buring, J., Hunter, D., DeVivo, I. (2008). Novel breast cancer risk alleles and endometrial cancer risk. *Int J Cancer.* Epub.

Morey, DF. (1992). Size, Shape and development in the evolution of the domestic dog. *J Archeological Sci*. 33: 158-175.

Neff, M., Robertson, K., Wong, A., Safra, N., Broman, K., *et al.* (2004). Breed distribution and history of canine mdr1-1delta, a pharacogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc Natl Acad Sci USA.* 101: 11725-11130.

Patterson, E., Mickelson, J, Da, Y., Roberts, M., McVey, A., *et al.* (2003). Clinical Characteristics an Inferitance of Idiopathic Epilepsy in Vizlas. *J Am Vet Med Assoc*. 17: 319-25.

Patterson, E., Armstrong, P., O'Brien, D., Roberts, M., Johnson, G., and Mickelson, J. (2005). Clinical Description and Mode of Inheritance of Idiopathic Epilepsy in English Springer Spaniels. *J Am Vet Med Assoc*. 226: 54-58

Parker, H., Kukekova, A., Akey, D., Goldstein, O., Kirkness, E., *et al.* (2007). Breed relationships facilitate fine mapping studies: A 7.8-kb deletion segregates with collie eye anomaly across multiple dog breeds. *Genome Rec*. 17: 1562-1571.

Roy, M., Geffen, S., Smith, D., Ostrander, E., Wayne, R. (1994). Patterns of differentiation and hybeidization in North American wolflike canids, revealed by analysis of microsatellite loci. *Mol. Bio. Evo.* 11: 553-570.

Salvolainen, P, Zhang, Y., Luo, J., Lundeberg, J., Leitner, T. (2002). Genetic Evidence for East Asian origin of domestic dogs. *Science.* 298: 1610-1613.

Sargan, D.R. (2004). *IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. Mamm Genome*. 15: 503-506.

Schadt, E., Molony, C., Chudin, E., Hao, K., Yang, X., *et al.* (2008). Mapping the genetic architecture of gene expression in the liver. *Plos Biology*. 6: e107.

Schena, M., Heller, R., Theriault, T., Konrad, K., Lachenmeier, E., Davis, R. (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Tibtech.* 16: 301-306.

Sutter, N., Bustamante, C., Chase, K., Gray, M., Zhao, K., *et al.* (2007). A single IFG1 Allele is a major determinant of small size in dogs. *Science.* 316: 112-115.

Sutter, N., Eberle, M., Parker, H., Pullar, B., Kirkness, E., *et al.* (2004). Extensive and Breed Specific Linkage Disequilibrium in Canis Familiaris. *Genome Res.* 14: 2388-2396.

Tate, S., Depondt, C., Sisodiya, S., Cavalleri, G., Schorge, S., *et al.* (2005) Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proc Natl Acad Sci USA.* 102: 5507-5512.

Tenmizu, D., Endo, Y., Noguchi, K., and Kamimura, H. (2004). Identification of a novel canine CYP1A2 1117 C>T SNP causing protein deletion. *Xenobiotica.* 34: 835-846.

Trut, L. (1999). Early canid domestication: The farm fox experiment. *Am Sci.* 87: 160-169.

Trut, L., Pliusnina, I., Oskina, I. (2004) An experiment on fox domestication and debatable issues of evolution of the dog. *Genetika*. 40: 794-807.

Vila, C., Maldonaldo, J., Wayne, R. (1999). Phylogenetic Relationships, Evolution and Genetic Diversity of the Domestic Dog. *J of Heredity*. 90: 71-77.

Vila, C., Savolainen, P. Maldonaldo, J., Amorim, I., Rice, J., *et al.* (1997). Multiple and ancient Origins of the domestic dog. *Science*. 276: 1687-1689.

Wayne, R. (1993). Molecular evolution of the dog family. *Trends in Genetics*. 9: 218-224.

Wayne, R. (2001). Consequences of domestication: Morphological diversity of the dog. In . *The genetics of the dog* (ed. A. Ruvinsky and J. Sampson), pp. 43-61. CABI Publishing Wallingford, United Kingdom.

Wayne, R. and Jenks, S. (1991). Mitochondrial DNA analysis implying extensive hybridization of the endandered red wolf *Canis rufus*. *Nature*. 351: 565-568.

Wayne, R., Benveniste, R., Janczweski, D. O'Brien, S. (1989). Molecular and biochemical evolutionof the carnivore. In: Carnivore behavior, ecology and evolution (Gittleman, JL, ed.) Ithaca, NY: Cornell University Press; 465-494.

Wood, J., Lakhani, K., Henley, W. (2004). An epidemiological approach to the prevention and control of three common heritable diseases in canine pedigrees in the United Kingdom. *Vet J*. 168: 14-27.

Zhu, Y., Shennan, M., Reynolds, K., Johnson, N., Herrnberger, M., *et al*. (2007). Estimation of warfarin maintenance dose based on VKORC1 (-1639 G>A) and CYP2C9 genotypes. *Clin Chem.* 53: 1199-1205.

TABLES

Table 1:

Diseases and disorders with a list of some of the breeds in which they are commonly found.

Disease	**Breeds**
Cancer	Boxer, English Setter, German Wirehaired Pointer, Great Dane, Rottweiler
Epilepsy	Belgian Malinois, Boston Terrier, Collies, Dachshund, Dalmatian, Shetland Sheepdog
Atopic Dermatitis	Golden Retriever, Labrador Retriever, West Highland White Terrier, Bull Terrier
Progressive Retinal Atrophy	Akita, English Cocker Spaniel, Maltese, Welsh Pembroke Corgi, Italian Greyhound

FIGURES

Figure 1: Relationship Trees of Dogs (image from Ostrander and Wayne, 2005)

A) Relationship of carnivores based on DNA hybridization data (Wayne, *et al*., 1989). B) Phylogeny of canids based on cytochrome b, cytochrome c oxidase 1, and cytochrome c oxidase II DNA (Wayne *et al*., 1997). C) Neighbor-joining tree of wolf (w) and the 4 main dog clades.

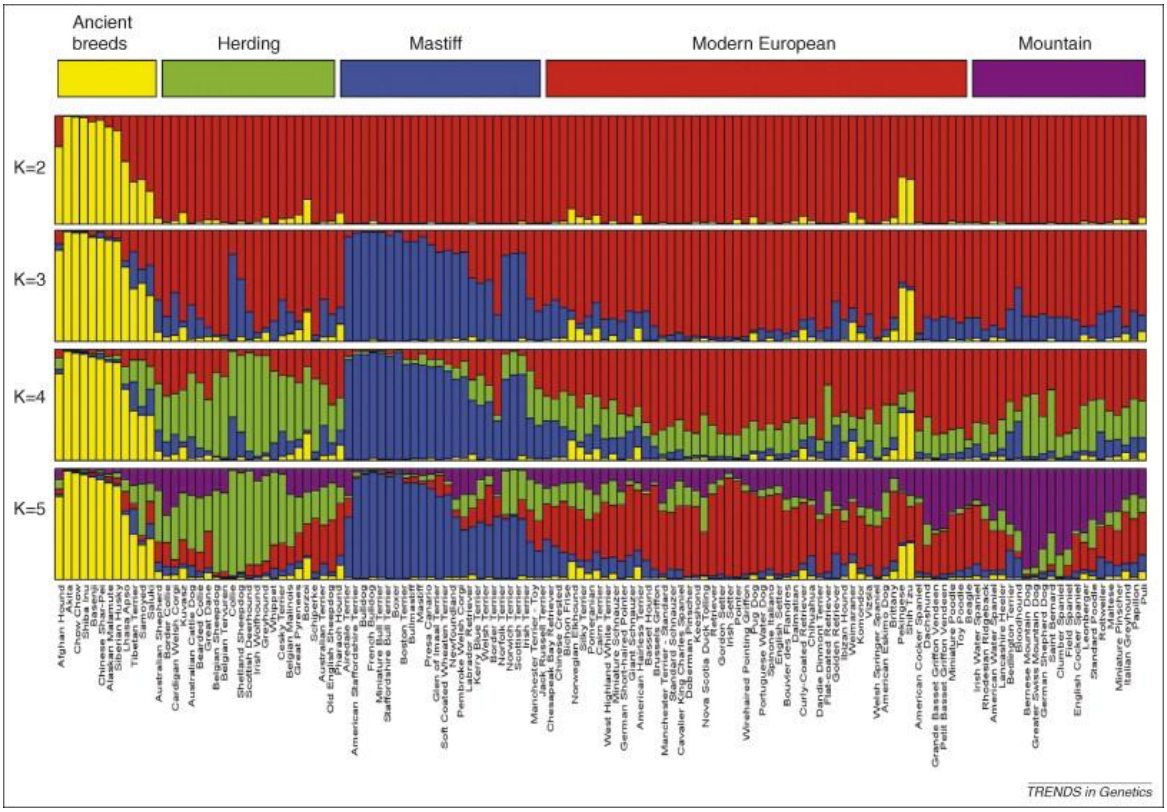

Figure 2:

Structure clusters of 132 breeds described in Parker *et al.,* 2007

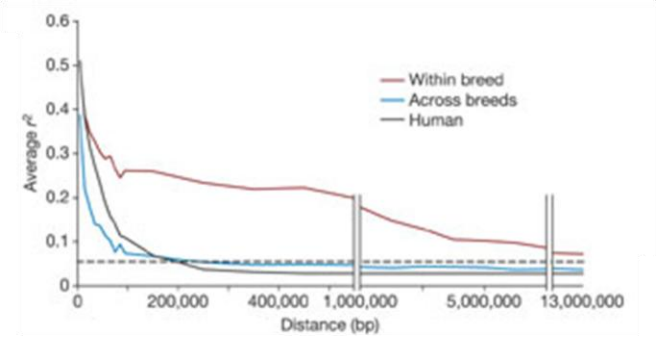

Figure 3:

LD measurements in humans and between and across breeds in dogs (Linbald-Toh, *et al.,* 2005).

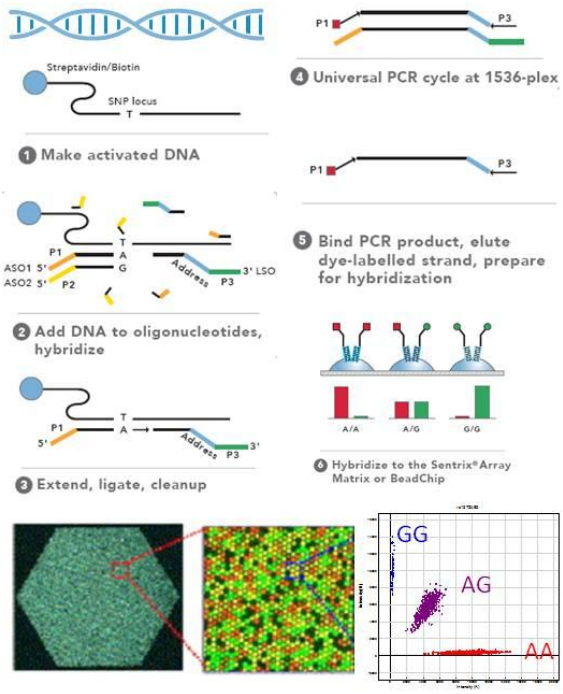

Figure 4:

Illumina throughput for SAM SNP detection (Derived from Fan, *et al.,* 2006).

CHAPTER TWO

A gene expression signature of confinement in peripheral blood of red wolves (*Canis rufus*)

A gene expression signature of confinement in peripheral blood of red wolves (*Canis rufus*)

Erin Kennerly[1], Anne Ballmann[2], Stanton Martin[3], Russ Wolfinger[4], Simon Gregory[5], Michael Stoskopf[2], and Greg Gibson[1*]

1. Department of Genetics, Gardner Hall, North Carolina State University,

   Raleigh NC 27695-7614

2. Environmental Medicine Consortium and Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh NC 27606

3. Information Technology Division, North Carolina State University, Raleigh NC 27695

4. SAS Institute, Cary NC 27513

5. Center for Human Genetics, Duke University Medical Center, Durham NC 27710

*Corresponding Author: Greg Gibson
School of Integrative Biology
University of Queensland, St Lucia Campus
Brisbane, QLD 4072
AUSTRALIA
ggibson.uq@gmail.com
Phone: + 61 (7) 3365 2194 Fax: + 61 (7**)** 3365 1655

REFERENCE:

Kennerly, E., Acton, A., Martin, S., Gregory, S., Wolfinger, R., Stoskopf, M, Gibson, G., A gene expression signature of confinement in peripheral blood of red wolves (canis rufus). *Molecular Ecology*, 2008, Vol. 11 (7). 2782-2791.

ABSTRACT

The stresses that animals experience as a result of modification of their ecological circumstances induce physiological changes that leave a signature in profiles of gene expression. We illustrate this concept in a comparison of free range and confined North American Red Wolves (*Canis rufus*). Transcription profiling of peripheral blood samples from 13 red wolf individuals in the Alligator River region of North Carolina revealed a strong signal of differentiation. 482 out of 2,980 transcripts detected on Illumina HumanRef8 oligonucleotide bead arrays were found to differentiate free range and confined wolves at a false discovery rate of 12.8 percent and $p < 0.05$. Over-representation of genes in focal adhesion, insulin signaling, proteasomal, and tryptophan metabolism pathways suggests the activation of proinflammatory and stress responses in confined animals. Consequently, characterization of differential transcript abundance in an accessible tissue such as peripheral blood identifies biomarkers that could be useful in animal management practices and for evaluating the impact of habitat changes on population health, particularly as attention turns to the impact of climate change on physiology and in turn species distributions.

INTRODUCTION

Genome-wide transcription profiling provides a novel approach to the detection of biomarkers for physiological status in mammals. Transcript abundance markers may be useful, for example, in categorizing levels of immunological suppression or activation in response to parasitization or infection, for identifying individuals that are experiencing unusual levels of stress, and for characterizing the health of populations during habitat disruption. For each of these applications, a relatively non-invasive and benign sampling strategy is needed. Peripheral blood is ideal, both because it is accessible and because the diverse mixture of monocytes are directly involved in immunity.

Biomarker levels in whole organisms have been widely considered for monitoring of community health, most notably with invertebrate and aquatic organisms (Shugart and McCarthy 1990; Attrill and Depledge 1997). Heat shock proteins, specific cytochrome P450s, and metallothionein levels are commonly used as indicators of the presence of environmental pollutants and toxins (Sanders 1993; Hoffmann and Parsons 1993; Monserrat *et al.* 2007), while other genetic markers are proposed to track global climate change (Umina *et al.* 2005). Genome-wide methods for contrasting transcript levels have enormous potential not just for finding novel markers (Snell *et al.* 2003) but also for characterizing the general plasticity of physiological responses (Giger *et al*. 2006; Gibson 2006; Matzkin *et al*. 2006). Before such strategies can be applied generally to vertebrate studies, it is essential to define the range of genetic and environmental effects on peripheral blood leukocyte gene expression (Whitehead and Crawford 2005).

The North American red wolf (*Canis rufus*) once roamed much of eastern North America, but in the early 20th century their numbers declined due to habitat destruction, health decline, and eradication programs (Cohn 1987). Conservation efforts aimed at saving the red wolf began in the 1973, seven years before they were officially considered extinct in the wild (Nowak *et al.* 1995; McCarley 1962; U.S. Fish and Wildlife Service 1994). As part of these efforts, captive breeding programs were set up in various parts of the United States, and reintroduction efforts, including one in the Alligator River region of eastern North Carolina where wolves have been re-introduced to the wild since 1986 (U.S. Fish and Wildlife Service 1986; Parker & Phillips 1991). Recovery efforts for the red wolf include a captive breeding population and extensive health monitoring of the free ranging population.

Previous studies have demonstrated that measuring physiological markers such as cortisol levels can be used to monitor physiological stress, in captive relative to free range African green monkeys, chimpanzees, and cheetahs, as well as functional immunosupression in captive African green monkeys (Suleman *et al.* 1999, 2004; Whitten *et al.* 1998; Terio *et al*. 2004). Here we demonstrate that whole genome expression profiling can be used to provide insight into the physiological differences between confined wolves and free range wolves, and show that alteration of specific stress-response pathways is characteristic of the specific habitats.

MATERIALS AND METHODS

Microarray Platform

Gene expression profiling was performed using Illumina's HumanRef8 Sentrix bead array platform, which contains 24,354 long (50mer) oligonucleotide probes representing well-annotated human genes. The full content of the platform is accessible under GEO accession: GPL2700. This is a heterologous platform for red wolf comparisons, and given the divergence of red wolf to human we expected and observed a large number of genes to hybridize poorly to the array, limiting the analysis to genes that are more conserved between humans and wolves. We chose the Illumina platform over the canine (*Canis domesticus*) short oligonucleotide array available from Affymetrix (Higgins *et al.* 2003) because the lower cost afforded replication. Since approximately 40 percent of the probes on the Affymetrix array are human derived, there is no guarantee that they accurately represent red wolf (*Canis rufus*) expression. Heterologous hybridization between species as genetically close as human and chimpanzee has been shown to affect inference of differential expression, as the effects of sequence polymorphism are not necessarily linear and so may not be accounted for in commonly used statistical models (Ji *et al*. 2004; Gilad *et al*. 2005).

Sample Collection and Microarray Experiment

Whole blood samples (3 ml) were collected from 13 red wolves (4 females and 9 males) and 1 free range female coyote (Table 1), immediately mixed with RNAlater at a volume of 3 to 1 and then subsequently stored at -20ºC. Total RNA was extracted using Ambion's Ribopure blood kit, yielding on average ~27 µg of total RNA. Seven hundred nanograms of biotin labeled cRNA derived from extracted total RNA using Ambion's Illumina RNA amplification kit (#I1755), was used in four

separate hybridizations onto Illumina HumanRef-8 Sentrix gene chips (#BD-25-201), yielding 4 technical replicates for each sample. Slides were washed and labeled with streptavidin-Cy3, then scanned using an Illumina Beadarray scanner and spot fluorescent intensities were extracted using the Illumina BeadStudio software (version 1.5.0.34). The raw microarray data can be accessed through the Gene Expression Omnibus under accession number GSE8020 at http://www.ncbi.nlm.nih.gov/geo.

Pedigree Analysis

Most animals in the study share a common ancestor within four generations. Pedigree data for all wolves in this study was available from the study records and extends over 8 generations. It was converted into tabular format and imported into the JMP Genomics software package (Cary, NC). SAS PROC INBREED was used to calculate the inbreeding coefficients for each wolf, except for one divergent wolf (identifier11080) and the coyote (identifier 20288). These coefficients were used as a part of the mixed model used to analyze the data in order to account for genetic relatedness, essentially following Yu *et al.* (2006). A kinship matrix of these values was created and used to generate the profile of relatedness in Figure 2, which conforms to the topology inferred from visual inspection of the pedigree.

For the estimate of the contribution of genetic relatedness to overall gene expression variance, a Principal Component analysis was performed on the expression measurements for all expressed genes on all arrays. PC1 through PC4 explain 67.5 % of the variance, and were each subsequently modeled as a function of Habitat, Genetic Relatedness, Sex, Wolf, and Sex-by-Habitat

interaction, where Genetic Relatedness is represented as the 12 columns of Cholesky coefficients from the kinship matrix (see below). The Principal Components Application in JMP Genomics allows estimation of the contribution of each variable to the variance in the gene expression captured by each PC, assuming that each variable is a random effect. No contribution of the Sex-by-Habitat interaction was observed. Figure 3 shows the estimated contributions for each of the first four PC, as well as the average contributions weighted by the percent variance explained by each PC.

Mixed Model Analysis

Statistical data analysis was performed in JMP Genomics software, which makes use of various SAS procedures in specific data steps. Intensity measurements for each gene were derived from (on average) 30 beads per gene. This average exported from BeadStudio software was log base 2 transformed (log_2I) and then median centered by subtracting the median log_2I from the log_2I for each array to yield the relative fluorescent intensity values (log_2RFI). By plotting the average log_2RFI for each gene across all the arrays according to rank of fluorescence intensity, a sigmoidal curve is obtained with a long plateau containing almost 80% of the probes. This level corresponds to background expression. Approximately 12% of the probes (2,980 out of 24,354) were above the inflection point of this curve, namely with an average log_2RFI of 7.732 or greater, and were deemed to be expressed. Only these were included in subsequent analyses. The remainders are presumed to represent non-expressed genes or genes that have diverged in sequence too greatly to cross-hybridize to the human-based probes.

In addition to the median centered normalization reported here, four other normalization

procedures were performed side-by-side, an analysis that is greatly facilitated using the JMP Genomics platform. First, a standard normalization routine was used in which the log2i data was mean centered and the standard deviation was subtracted from the log2i values. This approach yielded differential expression estimates very similar to those obtained with the median centered approach. ANOVA normalization, a partial least squares normalization, and Loess normalization approaches were all deemed unsuitable for this particular experiment due to over-fitting of noise. These methods were applied to both the full data set and the top 12% of probes selected with very comparable results.

Differential expression between confined and free ranging animals was assessed by mixed model analysis of variance. The following model was fit separately to data from each gene:

$$\log_2 RFI_{jklmn} = \mu + Habitat_j + Sex_k + Habitat*Sex_{jk} + (Wolf)_{jkl} + Error_{jklm}$$

Habitat and *Sex* were considered fixed effects with the *j*th treatment (*j*= free range or confined) and the *k*th sex (*k*= male or female), whereas the *l*th wolf is a random effect nested within treatment and sex. The *Error* is assumed to be normally distributed with a mean of zero. We also ran the following model to account for the genetic effects due to relatedness among the individuals:

$$\log_2 RFI_{jklmn} = \mu + Habitat_j + Sex_k + Genetic\ Relatedness_n + Habitat*Genetic\ Relatedness_{jn} + Habitat*Sex_{jk} + Wolf_{jkl} + Error_{jklm}$$

Genetic Relatedness models a polygenic random effect, and is equivalent to the $Z\mathbf{u}$ vector term in the model of Yu *et al* (2006). The variance of this random effect is assumed to be $2KV_g$, where K is the known $n \times n$ matrix of relative kinship coefficients that define the degree of genetic covariance between a pair of individuals (determined from their pedigree using SAS PROC INBREED), and V_g is an unknown variance component estimated from the data. We fit this term by first computing the Cholesky root (a kind of matrix square root) of $2K$ and then using the Cholesky coefficients as the Z matrix, while assuming **u** is a vector of independent normal random effects with mean zero and variance V_g. The lth wolf also has an individual random effect nested within treatment and sex. The residual *Error* is assumed to be normally distributed with a mean of zero.

WebGestalt Analysis

Functional groups and pathways that were enriched for differential expression between confined and free ranging wolves were explored using WebGestalt freeware available from Vanderbilt University at: *http://bioinfo.vanderbilt.edu/webgestalt* (Zhang et al., 2005). WebGestalt acts as a web interface for several publicly available resources including Kyoto Encyclopedia of Genes and Genomes (KEGG) and the database for annotation, visualization, and integrative discovery (DAVID). A Fisher's exact test was used for all analyses, and the Bonferroni cutoff of $P < 0.01$ is exceeded for 4 of the 5

pathways nominally significant at $P < 0.05$ as indicated in Table 3. Since all of the functional categories identified in Table 3 are at similar hierarchical levels, they include completely non-overlapping sets of genes and can be regarded as independent functional categories, despite the fact that they all represent modes of stress response. None of the categories are included as a subset of a higher ontological level, so they are independent in this regard as well.

Realtime Quantitative PCR

Quantitative-PCR was performed on 4 selected genes (3 significant and 1 control) to confirm the direction of their expression between the two treatment groups. Genes were chosen based on their significant differences between the two groups and their involvement in the top pathway hits from the WebGestalt analysis. Since we chose genes that showed relatively high levels of expression compared to all the genes on the microarray, our results may reflect bias towards those genes, but should show no biases against the subset of genes used in the analysis. However, it should be noted that only the two genes that are more highly expressed in free-range wolves were validated with this approach. Primer sets were designed to span an intron on the 3' end of the canine ortholog for the genes in Table 2. Canine orthologs were described in Ensembl database and confirmed by performing a reciprocal BLAST search.

QPCR reactions for a pool of 4 confined samples and a pool of 4 free ranging samples was performed in 6 replicates for each primer set. Two hundred nanograms of each sample were used

for each reverse transcription reaction using Promega's Improm II reverse transcriptase (#A3802). Ten microliter quantitative RT-PCR reactions were run on Applied Biosystems ABI 7900 machine using Applied Biosystems SYBR green PCR master mix (#4309155). A one tailed t-test was used to test the significance of the difference between the absolute values of the log base 2 of the cross over threshold (log_2CT) after they were normalized to the control gene GABRE. A Ct of one equals the cycle number where enough amplicon is present to register fluorescence above background (Hembruff *et al.* 2005).

RESULTS

Whole blood gene expression profiles of 13 wild red wolves and one coyote were examined using Illumina Sentrix HumanRef 8 Bead Arrays. Six animals were born in captivity and confined for up to10 years since birth (with the exception of one individual (10632) which was released to the wild between 4 and 10 months of age) (Table 1), while the remaining eight animals have been free-ranging their whole life. At the time of sample collection, information pertaining to health status, active mange and parasite infection was recorded. A total of 4 technical replicates per animal or 56 microarrays were performed using biotin-labeled cRNA prepared from whole blood that had been stored at -20ºC in RNAlater solution.

After normalization of the expression intensity measurements obtained from an average of 30 beads per transcript, it was estimated that just 12 percent of the long oligonucleotide human probes on the Sentrix arrays (2980 of 24,254), hybridized above background to the red wolf cRNA. This was expected given the sequence divergence between humans and canids (Linblad-Toh, *et al.*

2005), and the remaining probes were simply ignored. Our data thus consists of a small subset of the peripheral blood transcriptome, and it is possible that some measurements represent cross-hybridization to non-orthologous genes. Nevertheless, both hierarchical clustering (Figure 1) and analysis of variance of the data reveals clear differences among wolves in the gene expression profiles.

After two-way hierarchical clustering of transcript abundance measures by gene and wolf, two clearly distinct clusters of wolves are apparent (Figure 1). These are distinguished by the expression of several hundred transcripts. By visual inspection, it was apparent that these two clusters of animals are not distinguished on the basis of sex or parasite load. Rather, the clusters clearly separate wolves that are either confined (C) or free-ranging (F). Individual 20288 is the coyote, and clearly clusters with the free range red wolves, suggesting that DNA sequence polymorphism is unlikely to account for the overall cluster differentiation. It is also noteworthy that three of the animals in the confined cluster indicated by the bracket appear to be much more similar to one another than the others, and these turn out to be siblings. Confined individuals 10632 and 10406 have health problems that might be expected to affect blood expression profiles, but there are too few genes differentiating these from the other confined animals to make any definitive statements in this regard.

To account for expression differences due to relatedness among the wolves, we created a relative kinship matrix (*K*) which assigns each wolf a relatedness score to every other wolf based on methods described in Yu *et al.* (2006). Pedigree data was available for all wolves except animal

11080 and the coyote 20288. The profile of relatedness based on the kinship matrix score shows only superficial correlation with the hierarchical clustering of gene expression among the wolves (Figure 2). Three confined siblings (identifiers 11273, 11274 and 11275) are the most divergent cluster for both gene expression and ancestry, but the other three confined wolves, though related to one another, share common ancestry with four of the free-range wolves. These results suggest that differential gene expression changes are mainly due to habitat but that genetics also makes a contribution.

In order to quantify the proportion of observed variance in gene expression due to Habitat, Sex, Wolf, and Genetic Relatedness, we performed a Principal Component (PC) analysis on the estimated transcript abundance of each of the 2,980 expressed genes in each of the 11 wolves for which relatedness was estimated. Analysis of variance was then performed to estimate the relative contributions of each of these factors. Averaged over the first four PC, and weighting for their contributions, habitat accounts for 25.6%, genetic relatedness 9.7%, individual wolf 17.4%, and sex just 1.5% of the variance in PC values across the sample, with the remainder unexplained. The pie graphs in Figure 3 provide a graphical view of the contributions to each PC separately and clearly imply that PC1 largely captures the effect of habitat, PC2 captures genetic relatedness, PC3 captures differences between individual wolves, and PC4 a mixture of effects.

In the process of data normalization, we also noted a 5.6% decrease in average variance of expression profiles across all expressed transcripts for the confined ($\sigma^2 = 0.78$) relative to the free range wolves ($\sigma^2 = 0.86$). The difference is significant both by *t*-test ($p = 0.04$) and permutation ($p =$

0.04) and is consistent with the hypothesis that the diverse and variable conditions experienced by animals in the wild lead to a wider range of gene expression values overall than confinement. The alternative hypothesis, that perturbation of normal physiology due to a change in the environment increases phenotypic variability (Zhang 2005, Charmantier and Grant 2005), much as the aging process seems to affect the stability of expression profiles in humans (Somel *et al*. 2005), is not supported by these data.

Mixed model analysis of variance was applied to quantify differential gene expression and identify specific genes that distinguish the two groups of wolves. The number of significant differences between the sexes was no greater than expected by chance, but effects of both confinement status (confined versus free-range) and interaction between confinement status and sex were significant at experiment-wide confidence levels for hundreds of genes. The volcano plot in Figure 4 shows the relationship between significance and fold difference between the confined and free ranging expression for each gene. The magnitude of difference between mean expression in the confined and free-ranging samples, in log base 2 units of fluorescence intensity, is on the *x*-axis, while significance is shown as the negative logarithm of the *p*-value (NLP) on the *y*-axis for each probe. While overall there is a relatively symmetric distribution of up- and down-regulation of expression between the two groups of wolves, there is a notable excess of significant and more than 2-fold higher expression in free range wolves (genes up and to the left on the plot).

Genes above the horizontal false discovery rate threshold in Figure 4 are considered significantly differentially expressed between the two confinement status classes. The false

discovery rate procedure of Storey *et al.* (2003) was used to identify a list of 148 genes with *q*-values less than 0.10 ($p < 0.0124$), but because the q-value associated with the nominal test-wise significance value of $p < 0.05$ is just $q = 0.128$ for a total of 482 genes, we adopted this value for comparison of gene categories. At this cutoff, 62 genes are expected to be false positives. Of these 482 genes, 341 genes are expressed at higher levels in free-ranging wolves, while the remaining 141 genes are expressed at higher levels in the confined wolves. Identical analysis using all 24,354 genes revealed comparable numbers of significantly differentially expressed gene. Comparable results were also observed when a genetic relatedness component was incorporated in an analysis, confirming that the expression divergence between free range and confined animals is predominantly environmental in origin. However, the significance of the genetic effect term in the model implies that there is a weak but significant effect of relatedness overall.

The microarray results were validated by quantitative RT-PCR measurement of the expression of three differentially expressed genes shown in Table 2, and one control gene, gamma-aminobutyric acid A receptor epsilon (GABRE). The $2^{-\Delta\Delta Ct}$ method was used to contrast fold differences in expression between free-range and confined animals relative to the control gene (Livak & Schmittgen 2001). A one-tailed t-test confirms mRNA quantities of actin, beta (ACTIN) ($p < 0.0002$) and catenin (cadherin-associated protein), beta 1 (CTNNB1) ($p < 0.03$) are significantly higher in the free ranging wolves compared to the confined wolves, while mitogen-activated protein kinase 1 (MAPK1) shows a slight but non-significant decrease in free-range wolves. Both ACTIN and CTNNB1 are part of the focal adhesion pathway which is described below.

Enrichment of functional pathways for genes that were over- or under-represented relative to all expressed genes on the array was examined using Vanderbilt's WebGestalt. Using a significance threshold of $P < 0.05$ (Fisher's exact test), six pathways were overrepresented in the 341 genes that were up regulated in free ranging wolves. We combined two of these, adherens junctions (6 genes) and tight junctions (7 genes) together in assembling the list of pathways in Table 3. Tryptophan metabolism and cell cycle regulation pathways were significantly ($p<0.05$) overrepresented in the set of 141 genes that are more highly expressed in the confined red wolves. Additional analysis on the 148 genes at $q < 0.1$ gives analogous results for the free ranging wolves as four out of the five aforementioned pathways are significant at $P < 0.005$.

The up regulation of the focal adhesion and proteasomal pathways along with the actin cytoskeleton pathway in free ranging animals is interesting, because they all are linked to responsive states due to dietary changes or exercise stimulation (Reid 2005, Fluck *et al.* 1999, Carson & Wei 2000). The focal adhesion pathway (FAK) and the actin cytoskeleton pathway are also involved in the further regulation of the insulin pathway (Tsakiridis *et al.* 1999, Huang *et al.* 2002). Components of the FAK and actin cytoskeleton pathways are linked to the cellular processes needed for a pro-inflammatory immune response, as well as cytokine and cytokine receptor activation and regulation (Singh *et al.* 1999, Hall 1998, Funakoshi-Tago *et al.* 2003). Down regulation of the proteasomal pathway has been associated with diets that have increased corn gluten as well as muscle turnover associated with increased activity levels (Reid 2005, Wakshlag *et al.* 2003). Starvation also stimulates this pathway because animals must break down muscle to obtain sufficient essential amino acids

(Finn & Dice 2006). The depressed states of these pathways in confined animals could be reflective of the animal's physiological response to confinement caused by the lack of exercise and dietary changes.

DISCUSSION

Use of heterologous microarrays for gene expression profiling

The findings in this study were obtained using a heterologous microarray platform, namely by hybridizing red wolf cRNA to human long oligonucleotide probes. This is clearly not an ideal experimental practice, as it leads to loss of data due to failed hybridization of substantially diverged sequences, and raises the possibility of cross-hybridization, particularly to gene family members, in cases of intermediate sequence divergence. The above-background detection of just 12% of transcripts contrasts with expression of the majority of all genes on Illumina Human bead arrays in human peripheral blood (Göring *et al.* 2007; Idaghdour *et al.* 2008) and confirms the substantial loss of resolution due to heterologous hybridization.

These concerns over artifacts due to mishybridization do not, however, invalidate the general conclusions. First, the clear division between expression in confined and free-range animals was not expected *a priori*. There is no reason why individuals within the species should show such marked differences as a result of cross-hybridization, particularly because the differences fall predominantly into a small number of functional gene ontology categories. As described below, the fact that these categories are readily interpreted in the context of dietary, immunological, and

physiological stress responses, increases our confidence in their significance. Furthermore, two of the three changes in expression that we retested by wolf-specific quantitative RT-PCR were validated, confirming that while some artifacts are present, as a whole, the expression changes are likely real. While inferences about the effect of confinement on any single gene must be treated with caution, there is little doubt that confinement status has a major impact on expression profiles in the red wolf immune system.

Overrepresentation of stress pathways among differentially expressed genes

Each of the pathways listed in Table 4 as overrepresented among differentially expressed genes has been linked to stress, dietary, or immunological responses either through gene expression profiling or other types of studies. It is difficult to disentangle the effects of these environmental factors because they impinge on several of the pathways. For example, increased levels of the pro-inflammatory biomarkers likely indicate an individual's exposure to microbial, viral and macro parasites, but are also known to be involved in mediation of glucose homeostasis and hence dietary response (Grimble 2002, Long & Nanthakumar 2004). In addition, Ohmori *et al.* (2005, see also Morita *et al.* 2005) used gene expression profiling to show that receptors for interleukins are up regulated in peripheral blood by exposure of students to the stress of exam-taking. This is consistent with other findings showing that stress activates an inflammatory immune response in both humans and animals (Nukina *et al.* 2001, Goebel *et al.* 2000).

Similarly, the up regulation of the tryptophan metabolism pathway in confined animals is interesting because the pathway has been linked to stress response and serotonin release (Dunn &

Welch 1991) seen in both dietary and environmental changes. Tryptophan metabolism is up regulated in brains of restrained mice, and this in turn leads to increased serotonin synthesis (Lenard & Dunn 2005). Tryptophan metabolism changes have also been associated with dietary changes such as from a high protein diet to a lower protein diet that will stimulate an uptake of serum tryptophan in the brain to increase serotonin levels (Wakshlag *et al.* 2003, DeMarte & Enesco 1985). When the wolves are confined, they are placed on a commercial dry kibble diet (Hill's Science Diet Active Maintenance), and although this is protein-rich and is occasionally supplemented with deer carcasses, clearly it is not the same as their normal food intake. Simultaneously, the animals experience a marked change in their mobility and daily exercise patterns. Free-range animals may be expected to show increased muscle turnover, consistent with the finding of excess proteasomal pathway activity if our observations on blood also apply to muscle. It is noteworthy in this regard that a recent study using heterologous microarrays to study heat stressed coral reef fish, showed patterns of differential regulation of several gene ontology classes including actin cytoskeleton assembly (Kassahn *et al.* 2007).

Psychoneuroimmunological studies have led to speculation that the brain and immune system interact by sharing common signaling molecules. The immune system has been proposed to act as a "sixth sense", allowing the brain to detect things it cannot otherwise hear, see, taste or feel (Blalock & Smith 2007). Sharing of neurotransmitters, hormones and their respective receptors by the central nervous system and the immune system may enable the mind to influence susceptibility or resistance to disease or stress, and vice versa. In human studies, patients with a high degree of

emotional stress show a significant depression in their immune response, making them more susceptible to diseases like atopic dermatitis or hepatitis C (Raison *et al.* 2005, Hashizume & Takigawa 2006). We emphasize, though, that our sample size is too small and there are too many variables to allow any inferences about the possible impact of psychological effects of confinement on gene expression in the blood.

Peripheral blood biomarker analysis

Biomarkers have been adopted extensively to study physiological status in response to environmental agents in invertebrate and aquatic species (for example: Snell *et al.* 2003; Umina *et al.* 2005; Giger *et al.* 2006), but their use in mammalian ecology remains relatively unexplored. Our study demonstrates that gene expression differentiation due to life history and environmental effects may be at least as strong as genetic differentiation in wolves, implying that environmental and ecological influences can be detected readily. The adoption of microarrays is a particularly promising approach to biomarker discovery in peripheral blood, and applications from conservation genetic animal management to early detection of environmental stress are readily envisaged.

Our data also supports the notion that expression profiling of peripheral blood may be highly informative about the immunological status of an individual animal, and that extensive cross-sectional sampling may support parsing of the effects of such factors as nutrition, exercise, and psychological factors. We have recently observed that human lifestyle affects the expression of at least one third of the human leukocyte transcriptome in a genetically uniform population, the Moroccan Amazigh (Idaghdour *et al.* 2008). Individuals living nomadic, rural, and urban lifestyles

show dramatic differences in expression of suites of genes in specific immunological and disease categories. It thus seems likely that environment-dependent peripheral blood profiles are a common feature of mammalian species, raising the possibility that widespread expression profiling of blood samples from large populations, in combination with relevant data on lifestyle differences, should be considered as a general strategy for the identification of biomarkers for diverse types of ecological stressors. It would not be advisable to use profiles obtained using heterologous platforms to make clinical assessments about a particular individual, so applied conservation genetic work with wild species awaits the development of species-specific arrays.

ACKNOWLEDGEMENTS

We are indebted to the Red Wolf Recovery Implementation Team at the Alligator River Wildlife Refuge in eastern North Carolina, and the U.S. Fish and Wildlife Service Red Wolf Program for access to samples and discussions. We especially thank Will Waddell for pedigree information on the wolves. This work was supported by funds to GG from the North Carolina Agricultural Research Service, whose lab is also supported by NIH grant 2-R01-GM61600. EK is supported by NIH Training Grant GM-08443-13 to the Department of Genetics at North Carolina State University.

REFERENCES

Attrill MJ, Depledge MH (1997) Community and population indicators of ecosystem health: targeting links between levels of biological organization. Aquatic Toxicology, **38**, 183–197.

Blalock JE, Smith EM (2007) Conceptual development of the immune system as a sixth sense. Brain, Behavior, and Immuninty, **21**, 23-33.

Carson J, Wei L (2000) Integrin signaling's potential for mediating gene expression in hypertrophying skeletal muscle. Journal of Applied Physiology, **88**, 337-343.

Charmantier M and Grant D. (2005) Environmental quality and evolutionary potential: lessons from wild populations. Proceedings of the Royal Society: Biological Sciences. **272**, 1415–1425.

Cohn JP (1987) Red Wolf in the Wilderness. Bioscience, **37**, 313-316.

DeMarte ML, Enesco HE (1985) Influence of diet on plasma tryptophan and brain serotonin levels in mice. Experientia, **41**, 48-50.

Dunn AJ, Welch J (1991) Stress- and endotoxin-induced increases in brain tryptophan and serotonin metabolism depend on sympathetic nervous system activity. Journal of Neurochemistry, **57**, 1615-1622.

Finn P, Dice F (2006) Proteolytic and lipolytic responses to starvation. Nutrition, **22**, 830-844.

Fluck M, Carson JA, Gordan SE, Ziemiecki A, Booth FW (1999) Focal adhesion proteins FAK and paxillin increase in hypertrophied skeletal muscle. American Journal of Physiology: Cell Physiology, **277**, C152-C162.

Funakoshi-Tago M, Sonoda Y, Tanaka S, Hashimoto K, Tago K, Tominaga S, et al. (2003) Tumor Necrosis Factor-induced Nuclear Factor κB activation is impaired in Focal Adhesion Kinase-deficient mice fibroblasts. Journal of Biological Chemistry, **278**, 29359-29365.

Gibson G (2006) Evolution: The plastic transcriptome. Current Biology, **16**, R285-R287.

Giger T, Excoffier L, Day PJR, Champigneulle A, Hansen MM, Powell R, et al. (2006) Life history shapes gene expression in salmonids. Current Biology, **16**, R281–R282.

Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP (2005) Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. Genome Research, **15**, 674-680.

Goebel M, Mills P, Irwin M, Ziegler M (2000) Interleukin-6 and Tumor Necrosis Factor-α production after acute physiological stress, exercise and infused isoproteronol: Differential effects and pathways. Psychosomatic Medicine, **62**, 591-598.

Göring HH, Curran JE, Johnson MP, Dyer TD, Charleworth J, et al. (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nature Genetics, **39**, 1208-1216.

Grimble RF (2002) Inflammatory status and insulin resistance. Current Opinion in Clinical Nutrition and Metabolic Care, **5**, 551-559.

Hall A (1998) Rho GTPases and actin cytoskeleton. Science, **279**, 509-514.

Hashizume H, Takigawa M (2006) Anxiety in allergy and atopic dermatitis. Current Opinion in Allergy Clinical Immunology, **6**, 335-339.

Hembruff S, Villeneuve D, Parissenti A (2005) The optimization of quantitative reverse transcription PCR for verification of cDNA microarray data. *Analytical Biochemistry*, **345,** 237-249.

Higgins MA, Berridge BR, Mills BJ, Schultz AE, Gao H, Searfross GH, et al.(2003) Gene expression analysis of the acute phase response using a canine microarray. Toxicological Science, **74**, 470-484.

Hoffmann AA, Parsons PA (1993) Evolutionary Genetics and Environmental Stress. Oxford University Press, New York.

Huang D, Khoe M, Illic D, Bryer-Ash M (2002) Reduced Expression of focal adhesion kinase disrupts insulin action in skeletal muscle. Endocrinology, **147**, 3333-3343.

Idaghdour, Y, Storey, J, Jadallah, S, Gibson, G (2008) A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan amazighs. Plos Genetics, **4**, e1000052.

Ji W, Zhou W, Gregg K, Yu N, Davis S (2004) A method for cross-species gene expression analysis with high-density oligonucleotide arrays. Nucleic Acids Research, **32**, e93.

Kassahn K, Caley J, Ward A, Connolly A, Stone G, Crozier R (2007) Heterologous microarray experiments used to identify the early gene response to heat stress in a coral reef fish. Molecular Ecology, **16**, 1749-1763.

Lenard N, Dunn A (2005) Mechanisms and Significance of the Increased Brain Uptake of Tryptophan. Neurochemical Research, **30**, 1543-1548.

Linbald-Toh K, Wade C, Mikkelsen T, *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438 (7069)**, 745-746.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, **25**, 402-408.

Long K, Nanthakumar N (2004) Energetic and Nutritional Regulation of the adaptive Immune Response and Trade-Offs in Ecological Immunology. American Journal of Human Biology, **16**, 499-507.

Matzkin LM, Watts TD, Bitler BG, Machado CA, Markow TA (2006) Functional genomics of cactus host shifts in Drosophila mojavensis. Molecular Ecology, **15**, 4635-4643.

McCarley H (1962) The taxonomic status of wild Canis (Canidae) in south central United States. Southwestern Naturalist, **7**, 227–235.

Monserrat JM, Martinez PE, Geracitano LA, Amado LL, Martins CM, Pinho GL, Chaves IS, Ferreira-Cravo M, Ventura-Lima J, Bianchini A (2007) Pollution biomarkers in estuarine animals: critical review and new perspectives. Comparative Biochemistry and Physiology C Toxicology and Pharmacology, **146**, 221-234.

Morita K, Saito T, Ohta M, Ohmori T, Kawai K, Teshima-Kondo S, et al. (2005) Expression analysis of physiological stress-associated genes in peripheral blood leukocytes. Neuroscience Letters, **381**, 57-62.

Nowak RM, Phillips MK, Henry VG, Hunter WC, Smith R (1995) The origin and fate of the red wolf. In: Ecology and Conservation of Wolves in a Changing World. (ed Carlyn LN, Fritts SH, Seip DR), pp. 409–415. Edmonton, Alberta: Canadian Circumpolar Institute Occasional Publication **35**.

Nukina H, Sudo N, Aiba Y, Oyama N, Koga Y, Kubo C (2001) Restraint stress elevates the plasma interleukin-6 levels in germ-free mice. Journal of Neuroimmunology, **115**, 46-52.

Ohmori T, Morita K, Saito T, Ohta M, Ueno S, Rokutan K (2005) Assessment of human stress and depression by DNA microarray analysis. Journal of Medical Investigation, Supp **52**, 266-271.

Parker WT, Phillips MK (1991) Application of the experimental population designation to recovery of endangered red wolves. Wildlife Society Bulletin, **19**, 73-79.

Raison CL, Broadwell SD, Borisov AS, Manatunga AK, Capuron L, Woolwine BJ, et al. (2005) Depressive symptoms and viral clearance in patients receiving interferon-alpha and ribavirin for hepatitis C. Brain, Behavior, and Immunity, **19**, 23-27.

Reid MB (2005) Response of the ubiquitin-proteasome pathway to changes in muscle activity. American Journal of Physiology: Regulatory, Integrative, and Comparative Physiology, **288**, 1423-1431.

Sanders BM (1993) Stress proteins in aquatic organisms: an environmental perspective. CRC Critical Reiews in Toxicology, **23**, 49–75.

Shugart L, McCarthy J (1990) Biomarkers of Environmental Contamination. Lewis Publishers, Boca Raton FL.

Singh R, Wang B, Shirvaikar A, Khan S, Kamat S, Schelling JR, et al. (1999) The IL-1 receptor and Rho directly associate to drive cell activation in inflammation. Journal of Clinical Investigation, **103**, 1561-1570.

Snell TW, Brogdon SE, Morgan MB (2003) Gene expression profiling in ecotoxicology. *Ecotoxicology*, **12**, 475-483.

Somel M, Khaitovich P, Bahn S, Pääbo S, Lachmann M (2006) Gene expression becomes heterogeneous with age. Current Biology, **16**, R359-R360.

Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. Proceedings of the National Academy of Science, **100**,9440-9445.

Suleman MA, Yole D, Wango E, Sapolsky R, Kithome K, Carlsson HE, Hau J (1999) Peripheral blood lymphocyte immunocompetence in wild African green monkeys (*Cercopithecus aethiops*) and the effects of capture and confinement. In Vivo, **13**, 25-27.

Suleman M, Wango E, Sapolsky R, Odongo H, Hau J (2004) Physiologic manifestations of stress from capture and restraint of free-ranging male African Green Monkeys (*Cercopithecus Aethiops*). Journal of Zoo and Wildlife Medicine. **35**, 20-24.

Terio KA, Marker L, Munson L (2004) Evidence for chronic stress in captive but not free-ranging cheetahs (*Acinonyx jubatus*) based on adrenal morphology and function. Journal of Wildlife Disease, **40**, 259-266.

Tsakiridis T, Tong P, Matthews B, Tsiani E, Bilan P, Klip A, Downey G (1999) Role of the Actin Cytoskeleton in Insulin Action. Microscopy Research and Technique, 47, 79-92.

Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005) A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. Science, **308**, 691-693.

U.S. Fish and Wildlife Service (1986a) Red wolf proposed for reintroduction. Endangered Species Technical Bulletin, XI, **3**.

U.S. Fish and Wildlife Service (1994) Untitled. Endangered Species Technical Bulletin, XIX, **7**.

Wakshlag J, Barr S, Ordway G, Kallfelz F, Flaherty C, Christensen B, et al. (2003) Effect of dietary protein on lean body wasting in dogs: correlation between loss of lean mass and markers of proteasome-dependent proteolysis. Journal of Animal Physiology: Animal Nutrition, **87**, 408-420.

Whitehead A, Crawford DL (2005) Variation in tissue-specific gene expression among natural populations. Genome Biology, **6**, R13.

Whitten P, Stavisky R, Aurell F, Russell E (1998) Response of fecal cortisol to stress in captive chimpanzees (Pan troglodytes). American Journal of Primatology, **44**, 57-69.

Yu J, Pressior G, Briggs W, Bi I, Yamasaki M, Doebley J, et al. (2006) A Unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, **38**, 203-208.

Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Research, **33**, W741-W748.

Zhang S-U (2005) Evolution and maintenance of the environmental component of the phenotypic variance: Benefit of plastic traits under changing environments. The American Naturalist, **166**, 569-58.

TABLES

Table 1:

Biological information for each wolf including current status and average variance of the four technical replicates

Wolf ID	Sex	Age	Birth site	Current Status	Average Variance[1]
10537	M	12.5	Confined	Confined since 1994, unknown before	0.84
11273	M	1.5	Confined	permanent confinement	0.81
11274	F	1.5	Confined	permanent confinement	0.72
11275	F	1.5	Confined	Permanent confinement	0.89
10632	M	11.5	Confined	Confined, Free-ranging from 4-10mos of age	0.69
10406	M	14.5	Confined	permanent captivity	0.72
11136	M	4.5	Wild	Free-ranging	0.58
11105	M	6	Wild	Free-ranging	0.97
11148	F	4.5	Wild	Free-ranging	0.71
11080	M	6	Wild	Free-ranging	0.97
11270	F	5	Wild	Free-ranging	0.90
11206	M	3	Wild	Free-ranging	0.94
11310	M	1	Wild	Free-ranging	0.91
20288	F	Unknown	Wild	Coyote Free-ranging	0.97

1. Average variance of expression profile for all transcripts in the four technical replicates of each wolf.

Table 2:

Accession numbers for Canine Orthologues used for Quantitative PCR

Gene Name	Symbol	Illumina Target ID	Human Genbank Accession	Canine Genbank Accession
catenin (cadherin-associated protein), beta 1	CTNNB1	GI_40254459-S	NM_001904	XM_855875
mitogen-activated protein kinase 1	MAPK1	GI_20986528-I	NM_002745	XM_534770
actin, beta	ACTIN	GI_5016088-S	NM_001101	XM_536230
gamma-aminobutyric acid A receptor, epsilon (GABRE)	GABRE	GI_12707557-I	NM_021990	XM_549340

Table 3:

Pathways identified by Gene Ontology analysis

Pathway	# of Genes	Significance
Focal Adhesion pathway	13	**
Regulation of Actin Cytoskeleton	10	*
Insulin Signaling Pathway	12	***
Tight and Adherens Junction Pathway	13	**
Proteasome	6	**

* for $0.01 < p < 0.05$; ** for $0.001 < p < 0.01$; *** for $p < 0.001$.

FIGURES

Figure 1: Two-way hierarchical clustering of differentially expressed genes.

Each row represents the expression signature of one of the 13 wolf samples and a single coyote sample (20288), each column one measure of transcript abundance with red values high expression and blue low. The 6 samples names beginning with a C are confined animals while the remaining 7 samples beginning with an F are free-ranging animals, whereas red colored samples are female and males are blue: these fall into distinct clusters. Three siblings are the three confined samples indicated by the bracket. This figure shows the analysis based on all genes that are significantly differentially expressed between free-range and confined animals. A plot based on all 2,980 expressed genes has a very similar structure, except that the second (11105) and fourth (11148) free-range wolves above cluster with the confined animals.

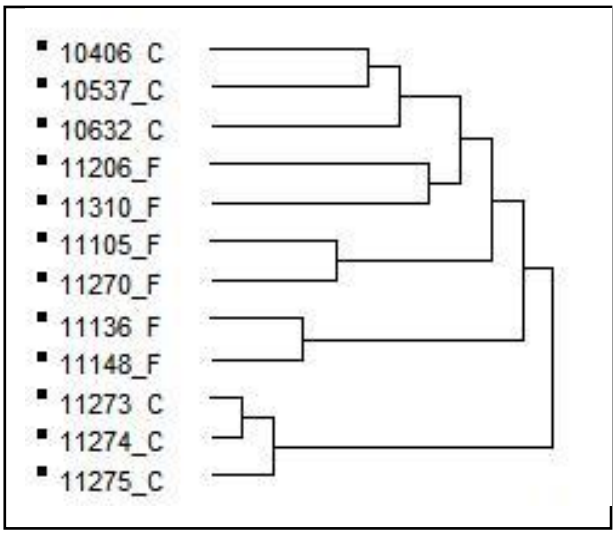

Figure 2: Pedigree clustering based on kinship matrix scores.

To determine the degree of gene expression clustering due to relatedness, wolves were clustered using their kinship matrix scores. Wolves 11273, 11274, and 11275 are all siblings as well as wolves 11148 and 11136. With exception to the 11273-11275 sib pairs there were no correlations between the gene expression clustering and the pedigree clustering.

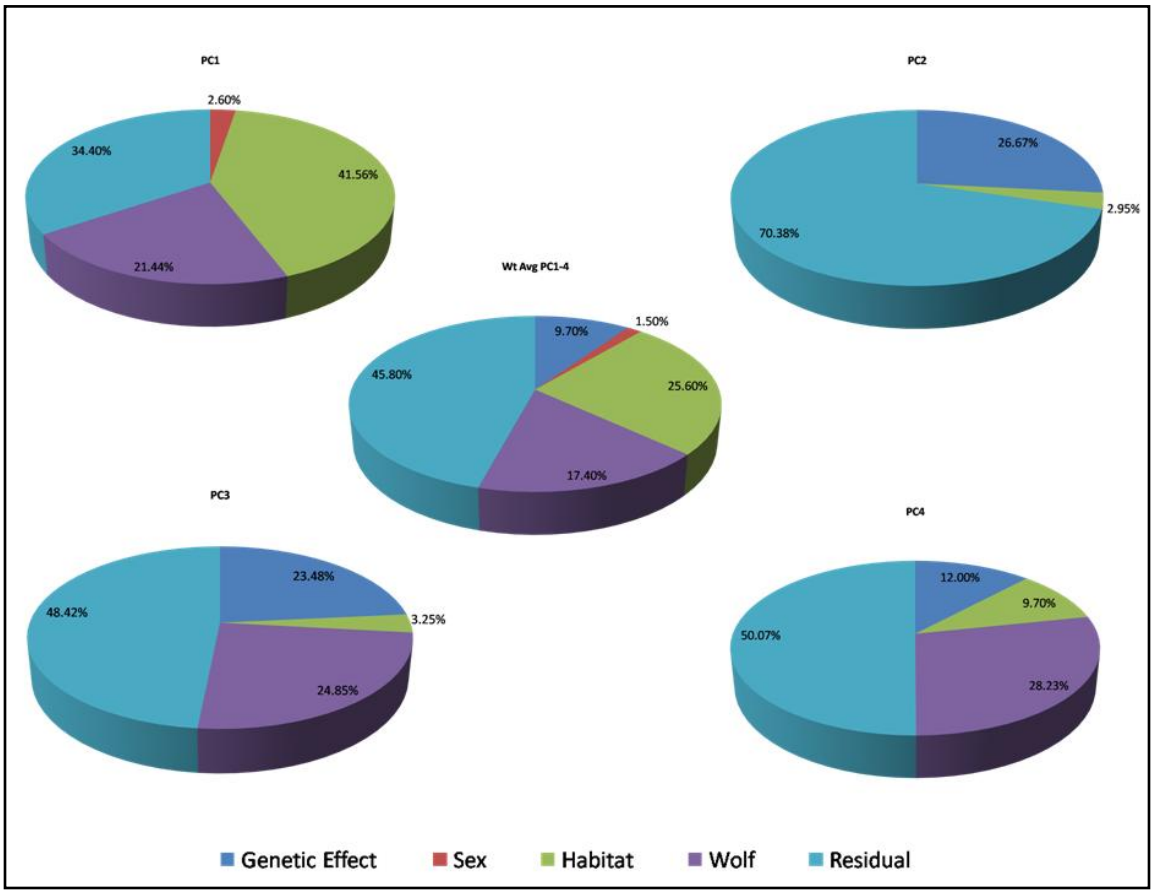

Figure 3: Contributions of habitat and genetic relatedness to expression variation.

Each pie chart illustrates the proportion of the variance for the indicated Principal Component that is explained by the indicated factors (Habitat, Genetic Relatedness, Sex, or Wolf) or left unexplained (Residual). The central pie chart shows the weighted average contribution of each effect in proportion to the variance explained by each of the four PCs (PC1: 38.2%; PC2: 15.8%; PC3: 7.6%; PC4: 5.9%).

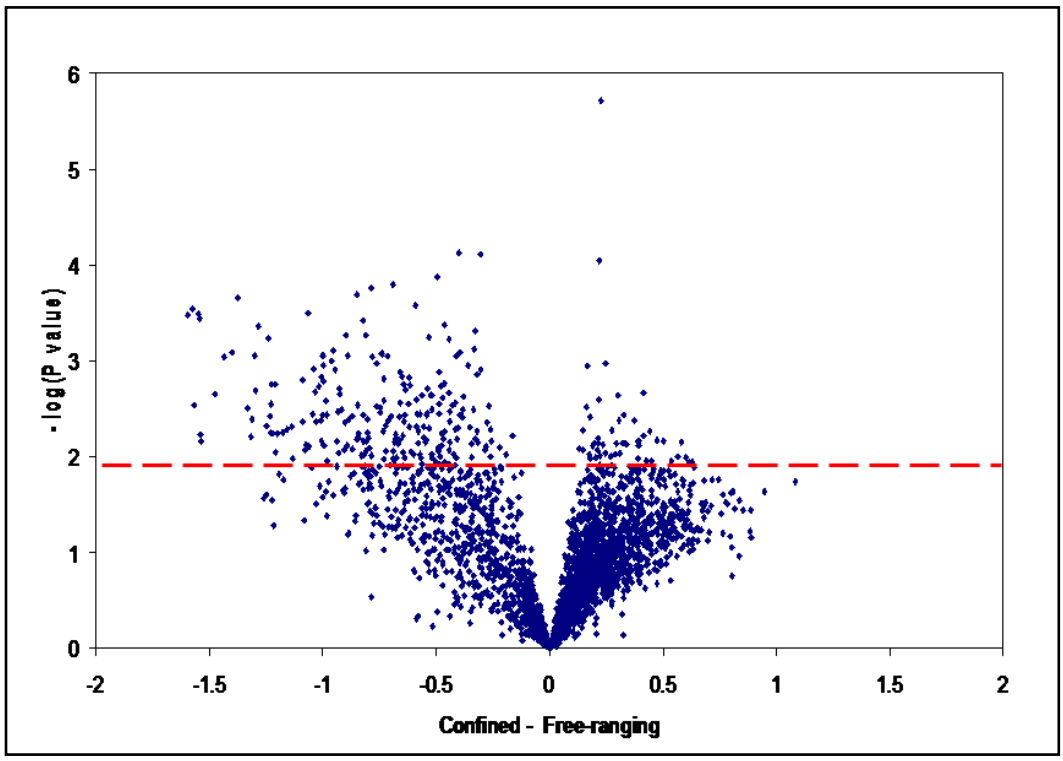

Figure 4: Volcano plot of significance against magnitude of effect for each gene.

The difference between mean expression in confined and free-range animals (excluding the coyote) in $\log_2$ units is plotted on the *x*-axis, and significance on the y-axis was determined by mixed model analysis of variance performed using the PROC MIXED procedure in SAS. The red dashed horizontal line indicates an FDR *q*-value less than 0.10 and corresponds to $p < 0.0124$. The 132 genes above the line to the left of zero are higher in the free-ranging wolves and the 16 genes above the line to the right of zero are higher in the confined wolves.

CHAPTER THREE

Gene expression profiling during acute phase infection of dogs with *Ancylostoma caninum* intestinal hookworms

ABSTRACT

The World Health Organization (WHO) estimates that each year approximately 2 billion people are infected with soil transmitted nematode infections. *Ancylostoma caninum*, a common canine intestinal hookworm causes weight loss, malnutrition, intestinal blood loss and anemia in dogs, especially in naïve pups. *A. Caninum* is the most characterized hookworm and serves as a model for its close relatives the human hookworms *Necatur americanus* and *Ancylstoma duodenale.* Unlike humans, dogs are able to mount an effective immune response and acquire long term protective immunity against future hookworm infection making them an ideal model to understand the mechanism involved the immune response to infection. In this study I have used both a canine specific and human microarrays to examine gene expression changes associated with hookworm infection in beagle pups. I have identified several pathways which show differential expression through the course of the infection as well as evaluated the use of heterologous microarray platform for immune studies in canines.

Introduction

Hookworms are the most common soil-transmitted parasite in the world, infecting more than 700 million people each year worldwide (Bethony, *et al.,* 2006). The blood feeding human intestinal hookworms *Ancylostoma duodenale* and *Necator americanus* cause anemia, weight loss, and malnutrition in their hosts. Humans can also be infected by the most prevalent hookworm in the world, the canine hookworm *Ancylostoma caninum*, but these do not sexually mature in humans (Croese, *et al.,* 1994; Fujiwara, *et al.,* 2006a). In recent years, efforts to understand the immunobiology of hookworm infection and to develop vaccines have been performed with several animal models including chickens, rabbits, hamsters, and dogs (Fujiwara, *et al.,* 2006a).

The *A. caninum* lifecycle is the best characterized of all hookworms. First stage L1 larvae hatch from eggs deposited on the ground in fecal matter, then pass through 2 molts to form the L3 stage. The infective L3 larvae enter their host through oral ingestion or penetration of the skin (Loukas and Provic, 2001). Once inside the host, serum stimulation occurs and the L3 larvae will either continue maturation to sexually mature L5 blood feeding worms whose eggs are shed in the feces or will arrest their development until more favorable conditions are present (Hawdon and Schad, 1990). The ability of the L3 to arrest development and maintain a dauer state for extended periods reduces the effectiveness of traditional anti-helminthics (Frietas and Arasu, 2005; Moser, *et al.,* 2005, Datu, *et al.,* 2008). Upon activation in the intestine of its host, *A. caninum* elicit a combination of a cellular mediated response (Th1) and humoural (antibody) mediated response

(Th2). In most hosts, invocation of a Th2 response is enough to overcome the infection and maintain long-term protective immunity; however in humans this is not the case (Else and Finkelman, 1998; Loukas, 2005). Humans infected with hookworms often develop a mixed cytokine response and are unable to confer long-term protective immunity seen in other animals like the dog (Pit, *et al.,* 2000; Geiger, *et al.,* 2004; Pit, *et al.* 2001).

Typical infections in dogs are associated with anemia and increased eosinophil counts as well as increased titers for Th2 controlled antibodies: IgM, IgG, and IgE (Loukas, *et al.* 1996). Hookworm vaccination with irradiated *A. caninum* larvae in dogs has been shown to increase IgG1 production as well as the ratio of IL-4 to IFN-γ (Fujiwara, 2006b). Similar studies in mice with other helminth infections have shown that induction of a Th2 response is imperative to overcome an infection and induce resistance to future infections. Demonstration that the increased IL-4 levels are detected in whole blood supernatant and not in CD4+ T cells, implies that the source of the IL-4 production is from mast cells, basophils and eosinophils (Fujiwara, *et al.* 2006b). Computational analysis of the hookworm genome sequence also suggests that hookworm secreted proteins can suppress the host's inflammation immune response and increase Th1 cytokines such as interferon gamma (IFN-γ) to provide a more favorable environment for the worm and delay detection from the host's immune system (Abubucker, *et al.* 2008; Dondji, *et al.* 2008). The result of this immunomodulation is a mixed pool of cytokines for both Th1 and Th2 response, which compete and allow for long-term survival of the worm (Hsieh, *et al.* 2004).

The sequencing of the canine genome provided new genomic tools that can be used to

study the complex immune response to hookworm infections in dogs (Lindblad-Toh *et al*., 2002). Traditional studies focused on mouse models because of the readily available resources, however unlike in mice, infective hookworm larvae are able to develop and mature to blood-feeding parasites, making dogs an ideal model to study the immune system's response to hookworm infection. Dogs also develop protective immunity after repeated exposure to hookworm; understanding the mechanisms behind this protective immunity should provide insight and direction for future vaccine development (Loukas and Prociv, 2001). To further characterize the immune response that is mounted by dogs infected with *A. caninum* and to determine if a Th1 or Th2 response is evoked, I have examined gene expression changes in whole blood during the 12 weeks following *A. caninum* infection in three male beagle pups. Correlation of these changes with other markers of infection including immunoglobulin levels (Ig) and fecal egg counts provides insight to key steps of the immune response, and identifies potential vaccine targets.

Materials and Methods

Dogs/Parasites/Samples

Three 8-10 week beagle pups were obtained from Marshall Farms, North Rose, NY and maintained in accordance with North Carolina State University Animal Care and Use Committee guidelines. Pups were infected with 120 infective larvae from a naturally occurring North Carolina (Wake County) strain of *A. caninum* as previously described (Arasu and Kwak, 1999). Blood samples were collected pre-infection, and at week 1, week 4, and week 12 post-infection. Whole blood was

collected in vacuum containers containing EDTA and a 750 µl aliquot was immediately mixed with RNA STAT LS-50 to preserve RNA. RNA was extracted from the RNA-Stat LS 50 mixture according to the manufacturer's protocol. For each time point a full CBC panel and ELISAs were performed to quantify Ig levels as well as quantify white blood cell, eosinophils and red blood cell levels. Fecal egg counts were done using sucrose fecal floats as described in Nolan *et al*., 1994.

Arrays/Hybridizations

Both the Affymetrix Canine GeneChips (#90 0725) and Illumina Human Ref8 Sentrix bead array platform (#BD-25-201) were used to characterize gene expression in whole blood. Affymetrix arrays consist of an average of 14 perfect and mismatched 25 mer probes per gene, the majority of which were designed to match canine gene sequences, although some probes were designed from human sequences. Probe sequence and information for the 21,700 genes can be found at NCBI's GEO under accession GPL3979. Illumina's Ref8 bead array consists of 24,354 oligonucleotides that are complementary to human transcript sequences, attached to beads which are held into etched wells on glass slides by electrostatic forces. Each probe is replicated on average 30 times at random locations throughout the array and for each slide there are 8 arrays. Full probe information for the Ref8 array can be found under GEO accession number, GPL2700.

Total RNA extracted from whole blood was aliquoted into two groups; half to be used for the Affymetrix arrays and the second half for the Illumina arrays. RNA from whole blood collected from weeks 0, 1, and 4 was prepared for hybridization to Affymetrix gene chip using Agilent Low RNA Input Flourescent Linear Amplification Kit (product # 5184-3523). A single round of linear

amplification was performed on 500ng of total RNA to yield complementary strand RNA (cRNA). Six hundred nanograms of complementary RNA was amplified and biotin labeled using Affymetrix two cycle cDNA synthesis reaction kit (#900494). Twenty micrograms of biotin-labeled cRNA was fragmented and added to 400ul of the hybridization cocktail per Affymetrix protocol. Hybridizations and scans on a GeneChip Scanner 3000 were performed by Expression Analysis (Durham, NC) using Affymetrix Genechip standard protocols. GeneChip Operating Software (GCOS) was used to extract feature data from scanned images.

Total RNA from whole blood collected at pre-infection and weeks 1, 4, and 12 post-infection from three beagles infected with hookworm was used for hybridization to Illumina's Ref8 genechip. Seven hundred nanograms of biotin labeled cRNA derived from Ambion's Illumina RNA amplification kit (#I1755) was hybridized onto Human Ref-8 Sentrix arrays. Slides were stained with streptavidin-Cy3, washed and then scanned using an Illumina BeadArray Scanner at Duke University's Center for Genome Science. Average fluorescent spot intensities for each gene were extracted using Illumina BeadStudio Software (version 1.5.0.34).

Data Analysis

Perfect match (PM) probe intensities were extracted from the Affymetrix dataset, log base two transformed and subsequently median centered using JMP Genomics software (SAS, Cary NC). For the Illumina dataset raw probe intensities were derived from an average of 30 beads per gene.

The average raw intensity was transformed on the log base 2 scale and median centered for each array to yield a relative fluorescent intensity value (log_2RFI) for each probe. The average log_2RFI for each gene across all the arrays was plotted in rank order, yielding a sigmoidal curve. All points below the point of inflection (8.10) for this curve were considered below background and removed from further analysis.

A repeated measures model that examines response trends over time was used to analyze the data, because for each individual there were multiple collection times over the course of infection. In this case measurements across time on an individual and measurements taken closer together in time will be correlated, violating assumptions of independence of variance that underlie standard analysis of variance models (Littell, *et al.,* 1993). Because the data is from repeated measures within individuals that have unequal spacing between collection points, a spatial power law repeated measure model was used to fit the data. For both the Affymetrix and Illumina data the following repeated measures model was used to fit the data:

$$\log_2(PM_{jl}) = \mu + T_j + D_l + \varepsilon_{jl}$$

Where *T* represents the *j*th time (0, week 1, week 4 or week 12* (*Illumina only), and *D* for the individual dog (*l* = 1, 2, 3). Dog is considered a random effect and time is a fixed effect. In this case the error (ε_{jl}) from one individual will be correlated but not between individuals and is therefore divided into two components $Cov(y_{t1}, y_{t2})$ and ε_{jl}, where $Cov(y_{t1}, y_{t2})$ is equal to $\sigma^2\rho^{|t1-t2|}$ and is the

measurement of the variance associated with differences between time points within an individual while the remaining error ε is the variance between individuals. This model will account for the individual effect and helps distinguish the gene expression over the course of infection. Data analysis for the Illumina data included a replicate term in the model to account for variation between the two replicate arrays, but was not significant and was subsequently left out of further analyses. In addition to the repeated measures model, the above model was run as mixed model ANOVA to look for differences associated with time.

Database for Annotation, Visualization, and Integrated Discovery (DAVID)

Genes above a false discovery rate (FDR) of 0.05 were selected for further investigation using DAVID (Dennis, *et al.,* 2003). At this cutoff the Affymetrix dataset consisted of 85 genes while the Illumina dataset consisted of 1129 genes that showed differential expression throughout the course of the infection. A more stringent Bonferroni cut off for each dataset yielded 22 and 305 genes for the Affymetrix and Illumina datasets respectively (Table 2). In both cases the hierarchal clustering showed the same patterns of gene expression. Using the FDR gene list for each experiment, genes were divided into correlated gene expression groups based on the hierarchal cluster analysis. For the Affymetrix experiment there were 4 groups (A, B, C, D) and the Illumina dataset consisted of 6 groups (A, B, C, D, E, F) (Figure 4a & b). In addition to using DAVID to identify pathways specific to small group gene clusters, I also examined combinations of gene clusters that showed similar trends of expression through the course of infection (i.e. genes showing high to low

expression through the course of infection). Since the gene clusters were derived from the hierarchal cluster of significant genes they represent non-overlapping genes, however it is possible that genes will be represented multiple times in pathways identified through DAVID. In this case it is difficult to determine appropriate significant cutoffs for pathways which share multiple genes. The p-values reported in Table 4 are derived using Fisher's exact test where a Bonferroni significance cut off is $p<0.01$ and pathways at $p<0.05$ are only marginally significant.

Real-Time Quantitative PCR (RT-PCR)

RNA extracted from whole blood using RNA STAT LS 50 was used for further verification of gene transcript levels by RT-PCR. First strand cDNA synthesis was created from 10 ug of total RNA, using Superscript II Reverse Transcriptase (Invitrogen). RT-minus cDNA reactions as well as intron spanning primers were used to confirm the absence of genomic DNA. RT-PCR was performed on a Bio-Rad iCycler Q Detection System using Bio-Rad SYBR Green Supermix for the primer sets listed in Table 1. Twenty five microliter RT-PCR reactions as well as relative transcript level calculations were done using the methods outlined in Freitas and Arasu, 2006.

Results

Previous studies in dogs have shown that pups start to shed hookworm eggs as early as 14 days post infection (Loukas and Prociv, 2001). As a means to track parasite infection, fecal egg count measurements were taken at 14, 28, 49 and 96 days (figure 1). The average egg per gram of feces

was 117, 972, 1285, and 447, for 14, 28, 49, and 96 days post infection, respectively. The decline in egg voiding after 49 days post infection suggest that the dog is overcoming the hookworm infection. The egg counts drop consistently between 49 days and 96 days and is correlated with the increase in both IgE and IgA levels between weeks 4 and 12 (figures 2 & 3). This data is suggestive of a Th2 driven response as both IgE and IgA are under the control of Th2 cytokines

IgG and IgE levels

Immunoglobulin levels have traditionally been used to measure the degree of activation of an immune response as well as to determine whether a Th2 or a Th1 response is occurring. A typical *A. caninum* infection is characterized by increased levels of hookworm specific IgE, IgG1, and IgG4 (Loukas, *et al.* 1996; Loukas, *et al.,* 1994). These immunoglobulins are under the control of Th2 cytokines like IL-4 which also stimulate eosinophil production. Increased levels of IgE is important in protecting the body against pathogens that cross the epithelial boundaries because they stimulate an inflammatory response by mast cells, basophils and eosinophils, while IgG immunoglobulin chains are recognized by phagocytic cells like macrophages and neutrophils (Pit, *et al.*, 2000, 2001). While IgE appears to be more specific to *A. caninum* antigens, IgG is one of the few immunoglobulins that will pass across the placenta into the intrauterine bloodstream and establish sufficient levels in the gut of the fetus to confer protection from pathogens. Both IgG and IgE invoke a humoral immune response which leads to an active antibody specific defense against an antigen, however IgG can differentiate further into subsets that are also associated with a Th1 response.

As seen in figures 2 & 3, showing data generated in the laboratory of Dr Prema Arasu, the

dogs used in my study exhibited a classical response to hookworm infection by showing an increase in both IgE and IgG levels by 2 weeks post infection as well as increased eosinophil levels (data not shown). This data is consistent with previous studies of canine hookworm infection in dogs which found a marked increase in IgG, 2 to 4 weeks post infection (Carroll and Grove, 1984). It is noteworthy that humans infected with *A. caninum* have detectable levels of hookworm specific IgE, but not IgG during early stages of infection, implying a difference in the immune response to the canine parasite(Loukas, *et al.* 1994). From this study, the slight increase of IgG levels from weeks 4 to 12 is only marginally significant at $p<0.06$ compared to the increase in IgE levels during the same time point (*p-value* = 0.01). IgE levels show a remarkable increase from weeks 1 to 12 (*p-value* 0.002). This data provides supportive evidence that IgE levels may be a good indicator of hookworm infection in dog, similar to what is seen in humans.

Microarray Analysis

To test the feasibility of using gene expression profiling in detecting a signature of immune response during a hookworm infection, I utilized two relatively untested microarray platforms, namely the canine specific Affymetrix whole genome microarray and a human specific Illumina Sentrix Ref8 BeadArray whole genome gene chip. The cost effectiveness of the Illumina platform allowed us to add an additional time point (week 12) as well as second replicate for each time point, but this advantage was offset against the heterologous nature of the human-derived probes. For the Illumina dataset, only 8% of the probe set was expressed above background levels, and the remaining 92% of the probes were excluded from further analysis. Nevertheless, I was able to

detect genes which showed a significant signature of differential gene expression through the course of infection with both platforms (Figure 4 a & b).

Comparison of the number of genes differentially expressed at an FDR of 0.05 showed 85 genes in the Affymetrix and 1129 genes for the Illumina platform. To determine if the additional time point or the replicate used in the Illumina experiment accounts for the increased number of genes detected, additional analysis was done using the Illumina dataset where a single replicate was randomly removed for each time point and/or the week 12 data was left out (Table 2). With the inclusion of week 12 expression data an additional 343 genes were detected; however, removal of the array replicates reduced the number of detected expression changes by 78%. Removal of both the replicate and week 12 data reduced the gene set to only 91 genes detected at an FDR of 0.05, and no genes at Bonferroni significance levels, comparable to the Affymetrix results.

In both the Affymetrix and Illumina experiments trends in gene expression can be divided into 4 or 6 main groups respectively based on the changes in gene expression throughout the course of the experiment (Figure 4 a & b). Group A in both datasets consist of genes that are down regulated after week 1 or 4 post infection. Genes in group B are genes that are down regulated between pre-infection and at week 1 post infection. Group C are genes which are up regulated at 4 (and 12) weeks post infection, while group D are genes which are up regulated between pre-infection and 1 week post infection. The Illumina dataset has two additional groups to account for the addition of the week 12 time point. Group E are those genes that are up regulated by week 12 post infection, while group F are those genes that are down regulated at week 12 post infection (Figure 4b).

Functional Pathway Analysis

DAVID analysis was performed for each dataset separately, first as a complete dataset of all differentially expressed genes, then in the separate groups identified determined from Figure 4. Because of the lower number of genes for groups A and B from the Affymetrix and group F from the Illumina data, they were not analyzed in DAVID. Instead they were contained in large gene groups which were aimed at capturing overall trends of pathway regulation during the course of the hookworm infection (i.e. genes expression high to low). There was no significant gene set enrichment from the Affymetrix data, except for a possible down regulation of ribosomal pathway ($p \sim 0.002$). Table 3 lists the pathways which had at least 10 genes represented and a *p-value* threshold of $p<0.01$. Groups not listed in Table 3 did not yield pathways above the significance threshold. DAVID provides a starting point of which pathways maybe involved in the immune response but without further characterizations of these pathways, it is difficult to determine their exact role during the hookworm infection.

The ribosomal pathway is the only pathway that was significant from the Affymetrix experiment. The genes present in this pathway showed decreased expression over the course of the infection. The ribosomal pathway was also present in the Illumina data and showed the same signature of expression. The Illumina data showed several other pathways that were associated with altered gene expression due to infection. The significance of the gene set enrichment is debatable, since some of the pathways are represented by fewer than 10 genes, and p-values are only marginal after adjustment for multiple comparisons. The analysis provides suggestive evidence for the

involvement of some of these pathways, which are interesting a priori because they may provide further evidence of immunomodulation by hookworm secreted proteins, and insight into how the dog's immune system overcomes the infection. Hookworms are known to secrete or excrete several proteins including Neutrophil Inhibitory Factors, C-Type Lectins, anticoagulant peptides, proteases and protease inhibitors to alter the host response to an infection (reviewed in Loukas and Provic, 2001; Allen and MacDonald, 1998). Functional pathways like proteasome, TCA cycle, and actin cytoskeleton pathways are all linked to cell turnover, cellular proliferation and inflammation response and are down regulated early on during the infection, which may be evidence of a response to hookworm secretions. In addition to these pathways the antigen processing and presentation present in group B cluster shows a drastic decrease between pre-infection and post-infection, whose decreased activity has also been linked to protease inhibitors secreted by the hookworm (Milestone, *et al.,* 2000). In further support of this theory, hamsters infected with *A. ceylanicum* as close relative of *A. caninum*, showed decreased activity of antigen presenting cells as well as decreased lymphocyte proliferation (Dondji, *et al.,* 2008).

Because of the low gene numbers (16 genes) I was unable to perform a functional pathway analysis through DAVID for groups C and D of the Affymetrix data, however the Illumina groups comparable to these groups showed several pathways that are upregulated over the course of the infection. Among these pathways are cytokine-cytokine receptor interactions and neuroactive ligand receptor interactions. To date little is known about the specific cytokine interaction profiles involved in an *A. caninum* infection, thus research aimed at identifying specific cytokines acting during the

hookworm infection would be useful to further characterize immune response. The neuroactive ligand receptor interaction pathway is a novel pathway associated with the infection but may be linked to the host's method of worm expulsion from the gut. Previous studies have linked Th2 cytokines as well as secretions from nematodes to nerve dependent increases in gut muscle contractility (Goldhill, *et al.,* 1997; Zhao, *et al.,* 2003). Th2 cytokines IL-4 and IL-13 have been shown to activate enteric nerve activity in intestinal muscles which increase muscle contractility and secretions used to expel the worms from the gut. There is also evidence that entero-endocrine cells are also stimulated by cytokines in the intestine and may act as important sensory cells and activator cells of the enteric nervous system (Zhao, *et al.,* 2003; Khan and Collins, 2004). The upregulation of genes involved in these pathways may be a reflection of a cytokine activated expulsion of worms from the gut caused by increased contractility and mucous production.

To further evaluate the usefulness of using a heterologous microarray, I compared genes that showed altered expression over the course of infection between the Affymetrix and Illumina platforms (Table 4). It is understood that the only genes that will hybridize above background levels are those that have retained sufficient sequence similarity to support robust hybridization. I also note that I may not be able to distinguish genes within the same gene family; however from the genes that do hybridize I can identify pathways that may be involved with an environmental response.

Over all there were 19 genes which were found to be in both significant gene lists which also showed the same gene expression patterns. These genes were selected strictly by the commonality

of their names; further studies into sequence homology may reveal more consistencies between the two datasets. The same Illumina Ref8 platform was previously used to examine the effects of gene expression due to habitat differences between captive and free ranging red wolves (Kennerly, *et al.,* 2008). In addition to identifying a genetic component to gene expression signatures of the red wolves, I found several pathways linked to stress and physiological changes due to the environment. Interestingly in this study, both datasets as well as RT PCR failed to characterize previously known contributors to hookworm infection like IL-4 and other TH2 cytokines. This may be that whole blood is not ideal tissue to measure cytokine mRNA levels and isolating T cells would be better. It is also reasonable to predict that increasing the number of dogs and the time points sampled may improve the effectiveness of the Affymetrix platform. Because the Affymetrix platform is canine specific it should be more capable of detecting gene expression changes in canine immune molecules which have significant sequence divergence from human sequence.

RT-PCR Analysis

RT- PCR was used to verify microarray results as well as look at levels for specific Th1 and Th2 markers. IL-4 is a classical Th2 response mediator and has been shown to be a mediator of other Th2 cytokines as well as to be important in mounting an effective immune response to nematode infections. However the activation of the Th2 cytokine IL-4 is not required and each species of nematodes may require a unique combination of cytokines to initiate expulsion (reviewed in Else and Finkelman, 1998). In my results IL-4 mRNA levels were not significantly different over the course of infection. This may not be surprising as IL-4 independent effector mechanisms like IL-13 have

been identified with other nematode infections like *Necator brasiliensis* (canine helminth parasite) (Finkelman, *et al.* 1997). IL-4 secretions in humans infected with the human hookworm *N.americanus* also showed no remarkable differences between infected and non-infected individuals, even though other biomarkers clearly indicated a Th2 polarized response (Geiger, *et al.,* 2007). RNA levels for Th1 and T cell regulator markers, IFN-γ and FoxP3 respectively, showed no remarkable changes over the course of infection indicating a lack of a fully mounted Th1 response. Tumor Necrosis factor Alpha (TNF-α) showed a slight but not significant decrease at weeks 2 and 3 with a slight increase at week 12 post infection. This same trend was also seen in TNF-α levels in humans infected with adult larvae of *N. americanus*, suggesting a possible effect of immune modulation from antigens secreted from adult worms (Geiger, *et al.,* 2007). Although not significant, the slight decrease in mRNA levels of TNF-α at week 2 post-infection coincides with the earliest time interval that an *A. caninum* larva can mature into an adult hookworm. FK506BP has not been previously linked to nematode infections but showed significant down regulation of expression over the course of the hookworm infection. This gene was also found to show consistent expression between the two platforms, therefore I chose this gene to verify microarray gene expression. Derivatives of this gene have been developed into drugs used in organ transplant patients as a way to suppress the immune system. In my microarray results this gene was significantly down regulated in both platforms, which was further validated by RT-PCR results and may be seen as a marker for increased immune response. Overall the RT-PCR results do not show a clear Th1 or Th2 polarized response which is consistent with other studies of hookworm infections, although the statistical

power of a study involving just three animals is limited. Increasing the number of dogs used in the experiment, or examining other cytokines more specific to *A. caninum* infection may demonstrate a clearer Th1 or Th2 driven response.

Discussion

Hookworm infection is one of the most common parasite infections in the world, infecting billions of people each year. The canine hookworm, *A. caninum* is the most commonly used model for studying human hookworm infection. However due to the limited genomic resources available to study infections in dogs, most studies investigating hookworm infections are performed in mice. Here I report the first whole genome expression analysis during an acute phase hookworm infection in its natural host, the dog. My data is consistent with previous studies, and also highlights new potential areas for research into our understanding of hookworm infection. Dogs make an ideal model to study hookworm infection as the hookworms are able to complete their lifecycle and repeated exposure in dogs leads to acquired immune protection. Humans are unable to obtain long standing protection from infection; therefore identifying key steps involved in the dog's development of protective immunity may help further development of a vaccine against hookworm infection. In addition to the host response to hookworm infection *A.caninum* and the human hookworm *A.duodenale* share a recent ancestry making them a good model to study parasite-host interactions. The second most common human hookworm *N. americanus* which mainly resides in the lungs as opposed to *A. caninum* which resides in the gut, share many characteristics indicating that *A. caninum* may be a good model for this helminth as well.

Use of Heterologous Microarrays

In this experiment, naïve beagle pups with serum stimulated L3 *A. caninum* larvae and recorded changes in their fecal egg counts, eosinophil counts, and hookworm antigen specific Ig levels. I also tracked gene expression changes associated with the course of infection over a 12 week period on both a canine specific Affymetrix microarray platform and an Illumina human specific microarray platform. Previously I have shown that by filtering out probes that did not hybridized above background levels, Illumina's human Ref8 platform is capable of distinguishing changes in gene expression due to habitat (Kennerly, *et al.,* 2008). In this study only 8% of the original 24,254 probes hybridized above background levels, which is consistent with 12% hybridization from the earlier study. The reduction of hybridization can be explained by loss of hybridization due to sequence divergence between humans and dogs. Even with a reduced gene set, a clear signature of gene expression associated with time during a hookworm infection is detected. I also noted similar trends in expression patterns between the Affymetrix platform and Illumina platforms, which is further evident by overlapping functional pathways and genes within expression groups. I also tested a previously unassociated gene FK506BP using RT PCR to validate expression in both platforms.

While the Illumina microarray was able to determine specific gene expression changes associated with time through the course of infection that was consistent with the Affymetrix data, it was fairly ineffective at identifying changes in immune specific genes which have known involvement in hookworm infections. However this may be an artifact of looking at whole blood

instead of another tissue like T-Cells, which will show the most immune specific response. Also when the array replicate and the week 12 data were removed, the Illumina dataset the arrays performed at more comparable levels to that of the Affymetrix gene chips (Table 2). Although the trends established from RT-PCR data and other response indicators like fecal egg counts were consist between the dogs, there appeared to be variation between the dogs' response. The variation of response between the three dogs may potentially make it more difficult to establish a clear Th1 or Th2 response from the expression arrays, increasing the number of dogs used will likely help. While the Illumina platform may provide a potential cost effective alternative to the Affymetrix gene chip, serious consideration of experimental expectations need to be considered. To detect canine specific gene expression changes it may be more useful to use the Affymetrix gene chip with increased the numbers of time points, dogs or replicates or using RNA from other tissue types.

Infection/Immune Response Assays

During the course of infection I observed several markers at weekly intervals. Using fecal egg count to track the hookworm infection in the dog, I noted a sharp increase in egg counts between weeks 2 and 7 (figure 1). This is consistent with previous data that shows the first mature hookworm eggs are shed at week 2 post infection. I also note a decrease in egg shedding between weeks 7 and 13 which is consistent with Ig levels showing an active immune response to hookworm specific antigens by week 12 (figures 2 & 3). The microarray results also show an increased immune response around weeks 4 and 12 that is associated with the cytokine-cytokine interaction pathway as well as neuroactive pathway receptors. Both of these pathways may be linked to worm expulsion due to active Th2 immune response in the gut as discussed below.

I also noticed increased eosinophil counts over the course of infection. Eosinophil counts have been directly linked to leukocyte response to adult hookworms as well as a reflection of intestinal worm loads (Pritchard, *et al.,* 1990; White, *et al.,* 1986). Increase of eosinophil cell counts, or eosinophilia is linked to expression of Th2 cytokine IL-5 and granulocyte-macrophage colony stimulating factors (GM-CSF) (Nishinakamura, *et al.*, 1996). From the expression results there was a significant increase in mRNA levels over the course of infection for GM-CSF receptor and CCL14, both of which are receptors for IL-5 and critical in productions of eosinophils (Nishinakamura, *et al.,* 1996).

The relationship between hookworm infection and Ig and eosinophil levels may provide valuable insight into thelink between geohelminth infection and allergies. As reviewed in Cooper, 2004 there is an inverse relationship between parasitic infection (including hookworms) and susceptibility to allergic diseases like asthma, suggesting that parasite infections may provide protective immunity against these diseases. Both allergies and parasite infections are marked by increased IgE levels and eosinophilia as a response to Th2 cytokines. One hypothesis is that early exposure to hookworms can prime the immune system especially the T regulatory cells to limit strong pro-inflammatory response thus reduce the response to allergic inflammatory response later in life (Maizels and Yazdanbakhsh, 2003; 2008). Dogs may provide a useful model system to study this interaction. The Th2 allergic response to environmental factors such as food and mites in dog is exhibited as atopic dermatitis and the cytokines and Ig markers mimic those found in asthmatic humans. Study results comparing the number and severity of atopic dermatitis allergic outbreaks in

dogs which are exposed to hookworm infections as puppies may parallel those trends found in humans.

Functional Pathway Analysis

Several functional pathways were identified from the Illumina dataset as being up-regulated or down-regulated over the course of the infection. Among the more interesting pathways is the down-regulation of the antigen processing and presentation pathway. Several studies have highlighted the immune-suppressive capabilities of *A. caninum* secretions including neutrophil inhibitory factors, anticoagulants, proteases, and protease inhibitors (Moyle, *et al.,* 1994; Cappello, *et al.,* 1995; Hsieh, *et al.,* 2004). The down regulation of antigen presentation and processing supports the claim of hookworm stimulated immune suppression early on during the infection (i.e. between pre-infection to 1 week post infection) which may be an artifact of secretions from the hookworm. However over the duration of the infection other pathways that are indicative of an active immune response like cytokine-cytokine interactions are present in genes that are up-regulated as time since infection increases. This may be evidence of a mounting Th2 immune response that is initially stimulated by hookworm induction of the Th1 response by hookworm-secreted antigens. Further studies would need to be done to highlight the exact targets of the hookworm secreted antigens and their effect on long term gene expression.

Recent studies on helminth expulsion have investigated an immune system activation of the nervous system to induce muscle contractility and globlet cell secretions. The activation of these

pathways by the immune system has been implicated in not only hookworm infection but other autoimmune disorders of the gut like inflammatory bowel disease (Khan and Collins, 2004). Further understanding of the mechanism that ignites the altered physiology of the gut will help identify immunological targets for other intestinal inflammatory states in addition to those associated with hookworm infection.

Conclusions

My data highlights the potential benefits of using a heterologous array platform to examine peripheral blood gene expression over the course of an acute *A. caninum* infection. I have shown correlations between indicators of hookworm infection (i.e. fecal egg counts and Ig levels) and gene expression patterns which are consistent with previous studies of hookworm infection. I have also demonstrated the use of DAVID to identify new pathways which show altered gene expression patterns associated with different time points post infection. Further investigation of genes specific to these pathways may reveal key mechanism in parasite host interactions, host defense and worm expulsion, which can be used to develop new vaccine targets for treatment of hookworm infection.

Acknowledgements

I would like to thank Dr. Prema Arasu and her lab for their assistance with this project. Her lab oversaw the care and infection protocols of the dogs used in this study as well as the collection of the samples used in this study.

REFERENCES

Abubucker, S., Martin, J., Yin, Y., Fulton, L., Yang, S-P., *et al.* (2008). The canine hookworm genome: Analysis and classification of *Ancylostoma caninum* survey sequences. *Molecular and Biochemical Parasitology*. 157: 187-192.

Allen, J., and MacDonald, A. (1998). Profound suppression of cellular proliferation mediated by the secretions of nematodes. *Parasite Immunology*. 20: 241-247.

Arasu, P. and Kwak, D. (1999). Developmental arrest and pregnancy-induced transmammary transmission of Ancylostoma caninum larvae in the murine model. *J Parasitol.* 85: 779-784.

Bancroft, A. and Grencis, R. (1998). Th1 and Th2 cells and immunity to intestinal helminthes. *Chemical Immunology*. 71: 192-208.

Bethony, J, Brooker, S, Albonico, M, Geiger, S, Loukas, A, *et al.* (2006). Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet*. 367: 1521-1532.

Cappello, M., Valsuk, G., Bergum, P., Huang, S., Hotez, P. (1995). Ancylostoma caninum anticoagulant peptide: a hookworm-derived inhibitor of human coagulation factor Xa. *Proc Natl Acad Sci.* 92: 6152-6156.

Carroll, S. and Grove, D. (1984). Parasitological, hematologic, and immunologic responses in acute and chronic infections of dogs with *Ancylostoma ceylanicum*: a model of human hookworm infection. *Journal of Infectious Diseases*. 150: 284-294.

Cooper, P. (2004). Intestinal worms and human allergy. *Parasite Immunology*. 26: 455-467.

Croese, J., Loukas, A., Opdebeeck, J., Fairley, S., Prociv, P. (1994). Human Enteric Infection with Canine Hookworms. *Annals of Internal Medicine*. 120: 369-374.

Datu, B., Gasser, R., Nagaraj, S., Ong, E., O'Donoghuo, P, McInnes *et al.* (2008). Transcriptional Changes in the Hookworm, Ancylostoma caninum, during the Transition from a Free-Living to a Parasitic Larva. *Plos: Neglected Tropical Disease*. 2: e130.

Dennis, G., Sherman, B., Hosak, D., Yang, J., Gao, W., *et al*. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 4: R60.

Dondji, B., Bungiro, R., Harrison, L., Vermeire, J., Bifulco, C., *et al.* (2008). Role for Nitric Oxide in hookworm-associated immune response. *Infection and Immunity*. 76: 2560-2567.

Else, K. and Finkelman, F. (1998). Intestinal nematode parasites, cytokines, and effector mechanisms. *International Journal for Parasitology*. 28: 1145-1158.

Finkelman, F, Shea-Donahue, T., Goldhill, J., Sullivan, C., Morris, S., *et al.* (1997). Cytokine regulation of host defense against parasitic gastrointestinal nematodes: lessons from studies with rodent models. *Annual Review of Immunology*. 15: 505-534.

Freitas, T. and Arasu, P. (2005). Cloning and characterisation of genes encoding two transforming growth factor-β-like ligands from the hookworm, *Ancylostoma caninum. International Journal of Parasitology*. 35: 1477-1487.

Fujiwara, R., Geiger, S., Bethony, J., Mendez, S. (2006a). Comparative immunology of human and animal models of hookworm infection. *Parasite Immunology*. 28: 285-293.

Fujiwara, R., Loukas, A., Mendez, S., Williamson, A., Bueno, L., *et al.* (2006b). Vaccination with irradiated Ancylostoma caninum third stage larvae induces a Th2 protective response in dogs. *Vaccine*. 24: 51-509.

Geiger, S., Caldas, I., McGlone, B., Campi-Azevedo, A., DeOliveira, L., *et al.* (2007). Stage-Specific Immune Responses in Human *Necator americanus* infection. *Parasite Immunology*. 29: 347-358.

Geiger, S., Massara, C., Bethony, J., Soboslay, P., Correa-Oliveira, R. (2004). Cellular Response and cytokine production in post-treatment hookworm patients from an endemic area in Brazil. *Clinical Experimental Immunology*. 136: 334-340.

Goldhill, J., Morris, S., Maliszewski, C., Urban, J., Funk, C., *et al.* (1997). Interleukin-4 modulates cholinergic neural control of mouse small intestinal longitudinal muscle. *American Journal of Physiology*. 272: G1135.

Hawdon, J., Schad, G. (1990). Serum-stimulated feeding in vitro by third-stage infective larvae of the canine hookworm *Ancylostoma caninum*. J Parasitol. **76**: 394–398.

Hsieh, G., Loukas, A., Wahl, A., Bhatia, M., Wang, Y., *et al.* (2004). A secreted protein from the human hookworm Necator Americanus binds selectively to NK cells and induces IFN-γ production. *Journal of Immunology*. 173: 2699-2704.

Khan, W., and Collins, S. (2004). Immune-mediated alteration in gut physiology and its role in hoste defence in nematode infection. *Parasite Immunology*. 26: 319-326.

Littell, R., Henry, P., Ammerman, C. (1993). Statistical analysis of repeated measures data using SAS procedures. *J Anim Sci*. 76: 1216-1231.

Loukas A, Opdebeeck J, Croese J, Prociv P. (1996). Immunoglobulin G subclass antibodies against excretory/secretory antigens of *Ancylostoma caninum* in human enteric infections. Am J Trop Med Hyg. **54**: 672–676.

Loukas, A, Opdebeeck, J., Croese, J., and Prociv, P. (1994). Immunologic incrimination of *Ancylostoma caninum* as a human enteric pathogen. *American Journal of Tropical Medical Hygiene*. 50: 69-77.

Loukas, A. and Prociv, P. (2001). Immune Responses in Hookworm Infection. Clinical *Microbiology Review*. 14: 689-703.

Maizels, R. and Yazdanbakhsh, M. (2003). Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol.* 3: 733-744.

Maizels, R. and Yazdanbakhsh, M. (2008). T-cell regulation in helminth parasite infections: implications for inflammatory diseases. *Chem Immunol Allergy.* 94: 112-123.

Maizels, R., and Yazdanbakhsh, M. (2003). Immune Regulation by helminth parasites: cellular and molecular mechanisms. *Nature Review Immunology*. 3: 733-744.

Milestone, A, Harrison, L, Bungiro, R, Kuzmic, P., Capello, M. (2000). A broad spectrum Kunitz-type serine protease inhibitor secreted by hookworm, Ancylostoma ceylanicum. *Journal of Biology and Chemistry*. 275: 29391-29399.

Moser, J., Freitas, T., Arasu, P., Gibson, G. (2005). Gene expression profiles associated with the transition to parasitism in *Ancylostoma caninum* larvae. *Molecular and Biochemical Parasitology*. 143: 39-48.

Moyle, M., Foster, D., McGrath, D., Brown, S., Laroche, Y., *et al.* (1994). A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *Journal of Biochemisty*. 269: 10008-10015.

Nishinakamura, R., Miyajima, A., Mee, P., Tybulewicz, V., and Murray, R. (1996). Hematopoiesis in mice lacking the entire granulocyte-macrophage colony-stimulating factor/interleukin-3/interleukin-5 functions. *Blood*. 88: 2458-2464.

Nolan, T., Bhopale, V., Megyeri, Z.,. Schad, G. (1994). Cyroperservation of human hookworms. *J Parasitol.* 80: 648-650.

Pit, D., Polderman, A., Baeta, S., Shulz-Key, H. and Soboslay, P. (2001). Parasite-specific antibody and cellular immune responses in humans infected with *Necator Americanus* and *Oesophagostomum bifurcum. Parasitology Research*. 87: 722-729.

Pit, D., Polderman, A., Schulz-Key, H., and Soloslay, P. (2000). Prenatal immune priming with helminth infections: parasite-specific cellular reactivity and Th1 and Th2 cytokine responses in neonates. *Allergy*. 55: 732-739.

Pritchard, D., Quinell, R., Slater, A., McKean, P., Dale, D., *et al.* (1990). Epidemiology and immunology of *Necator americanus* infection from excretory-secretory and cuticular collagen antigens. *Parasitology*. 100: 317-326.

White, G., Maxwell, C., and Gallin, J. (1986). Changes in structural and functional properties of human eosinophils during experimental hookworm infection. *Journal of Infectious Disease*. 154: 778-783.

Zhao, A., McDermott, J., Urban, J., Gause, W., Madden, K., *et al.* (2003). Dependence of IL-4, IL-3, and nematode-induced Alterations in Murine Small Intestinal Smooth Muscle Contractility on Stat6 and Enteric Nerves. *The Journal of Immunology*. 171: 948-954.

TABLES

Table 1: RT-PCR gene accession numbers and primer sets.

Gene	accession #	Forward primer	Reverse primer
60S ribosomal protein	AF044496	catctccccttctcctttg	gcgagaatgcagagtttcct
IL-4	AF187322	tgattccaactctggtctgct	tttctcgctgtgaggatgtt
IFN-γ	AF327901	ttaagctgattcaaattcctgtg	cgcttccttaggttggatctt
TNF-α	AF327899	tgttgtagcaaaccccgaag	aggacctgggagtagatgagg
FKBP	XM_534923	gaaccatgaaggtgggagaa	ctcccctgaactcgaacaac

Table 2: Significant gene totals from the original analysis versus the removal of the week 12 time point and the second replicate from the Illumina dataset.

	Bonferroni	**FDR**
Affymetrix	22	85
Illumina	305	1129
No Week 12	57	786
No Replicate	5	247
Both Removed	0	91

Table 3: DAVID Analysis of Illumina Data with pathways which have at least 10 genes represented or a p-value less than 0.01.

Groups (total # genes)	Pathway	Genes in Pathway	Significance
A, B, F (614)	Ribosome	17	8.14 e -4
	Proteasome	8	2.7 e- 3
A (215)	Regulation of Actin cytoskeleton**	12	3.3 e -3
B (328)	Citrate Cycle (TCA)**	5	1.4 e -2
	Antigen Processing and Presentation**	4	5.1 e -2
C, D, E (486)	Cytokine-Cytokine Receptor Interaction	11	7.2 e -4
	Arachidonic Acid	5	8.5 e -3
E (207	Neuroactive Ligand Recpter Interaction	6	1.9 e -2

**These pathways failed to meet inclusion criteria, but are listed for biological relevance.

Table 5: Comparison of Between Similar Gene Groups from the Affymetrix and Illumina Platforms

Genes found in both Platforms for Groups A, B, & E
Ribosomal Protein L18
Ribosomal Protein L13a
Retinoblastoma Associate Protein
DNAJ (HSP40) homolog
Peptidylproyl Isomerase A (cyclophilin A)
RAS-GTPase Activating Protein
Translation Initiation Factor 3
Translation Elongation Factor
NADH Dehydrogenase (Ubiquinone 1)
Splicing Factor Arginine/Serine Rich
RAB1A
Proteasome 26s
Interferon Regulatory Factor 3
Ubiquinol Cytochrome C Reductase
FK506 BP
Acyl-CoEnzyme A
Vasolin Containing Protein

Genes Found in both Platforms for Groups C, D & E
Finkel-Diskis-Rielly Murin Sarcoma Virus
Peptidyl Isomerase

FIGURES

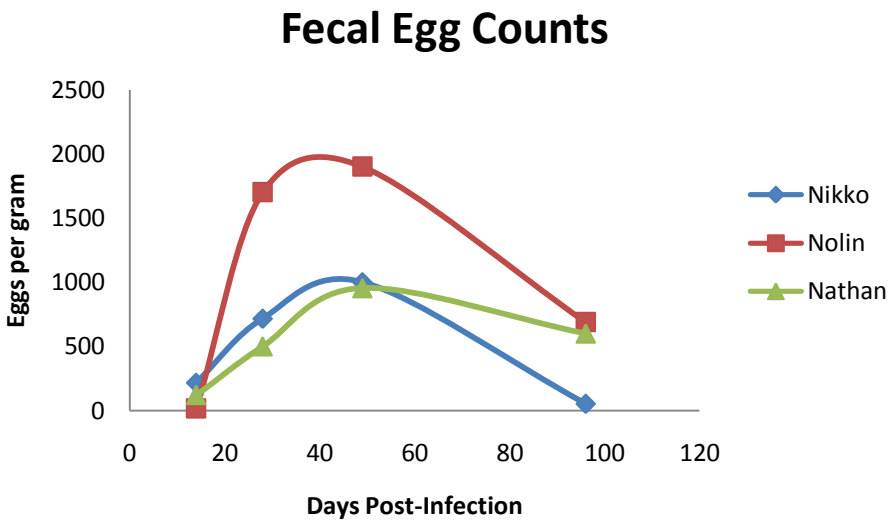

Figure 1: Eggs per gram of fecal matter, as seen during the course of infection.

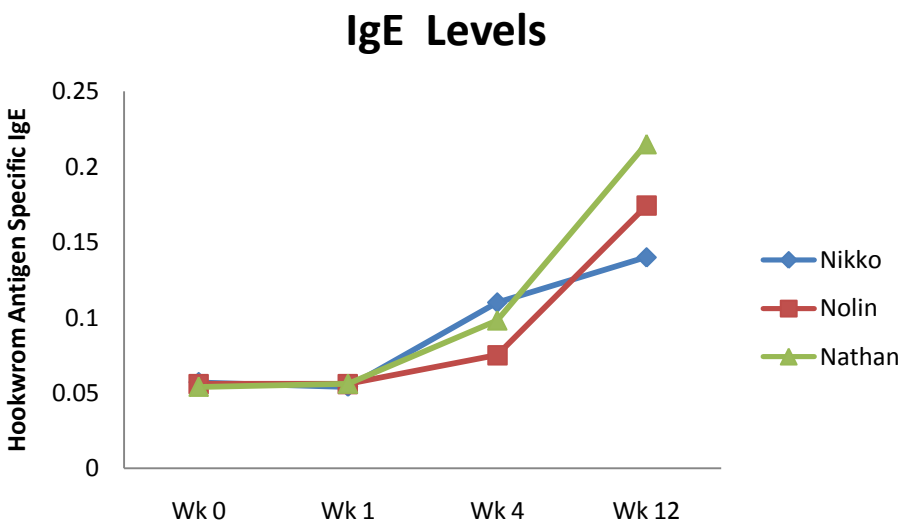

Figure 2: Hookworm antigen specific IgE levels over 12 week course of infection.

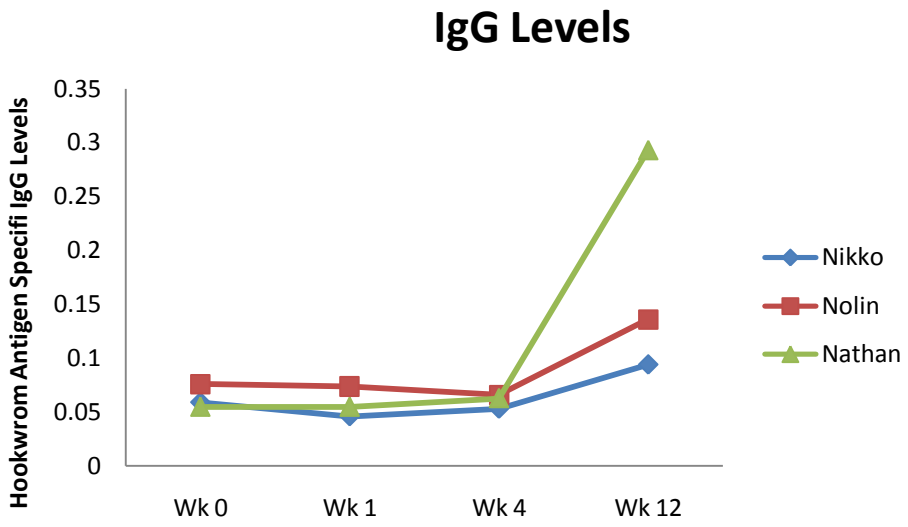

Figure 3: Hookworm antigen specific IgG levels over 12 week course of infection.

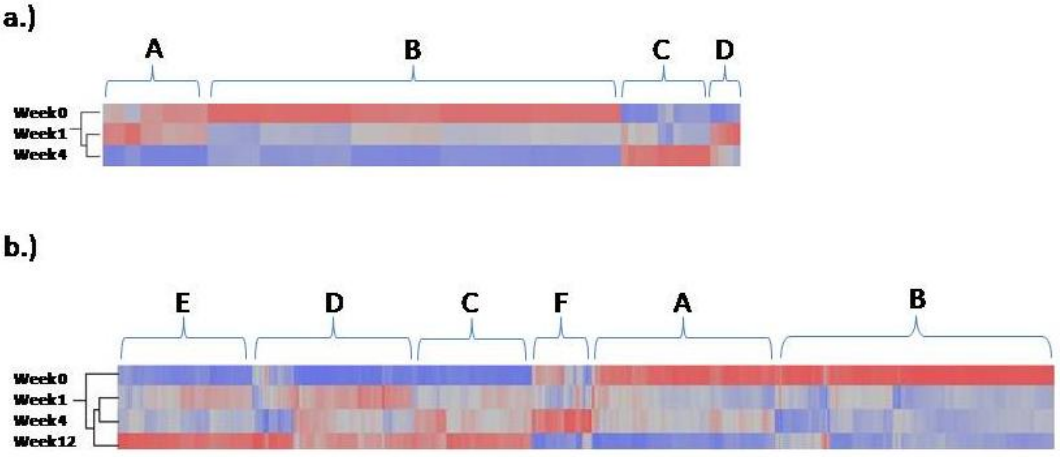

Figure 4: Hierarchal cluster of differentially expressed genes.

Red is high expression while blue is low expression. A) Cluster of 85 FDR significant genes from the Affymetrix dataset. B) Cluster of 1129 FDR significant genes from the Illumina dataset.

CHAPTER FOUR

Pharmacogenetic Case Control Association Study of Phenobarbital Drug Response in Epileptic Dogs

ABSTRACT

Epilepsy is the most common neurological disorder in dogs with a prevalence as high as 6%. Although several anti-epileptic drugs (AED) are readily available, adequate seizure control is not achieved with a single medication requiring combinatorial drug treatments in about third of epileptic dogs. Depending on the breed, refractoriness to Phenobarbital, a commonly prescribed AED, is between 10-20 percent of epileptic dogs. The genetic mechanisms involved in drug response have yet to be determined and need to be explored to further our knowledge and provide better treatment options for epileptic patients. The long term objective of this research program is to establish dogs as a model organism to study drug response in epilepsy, and to use natural genetic variation in dogs to explore the genetic mechanisms involved in drug response. We have recently designed a custom Illumina BeadChip which allows for high throughput genotyping of 384 single nucleotide polymorphisms (SNPs) in 30 genes involved in drug metabolism, drug targeting, and drug transport. A case control association study on 125 epileptic dogs identified 5 genes with an association to drug response.

Introduction

Idiopathic epilepsy is characterized by reoccurring seizures whose onset has no known underlying cause (O'Brien, 2003). It is the most common neurological disorder in dogs, with prevalence estimated between 0.5% and 5.7%. However this is thought to be an underestimate due to lack of recognition by owners of atypical seizure behaviors (Licht, *et al.,* 2002). There are several anti-epileptic drugs (AEDs) approved for use in dogs and humans, but owing to a short pharmacokinetic half-life in dogs it has proven difficult to maintain therapeutic serum levels of the drugs in epileptic dogs and consequently there are fewer treatment options available (Chandler, 2006, Benet, 1996). Nevertheless, treatment with more common AEDs, like Phenobarbital (PB), can reduce or stop the occurrence of seizure activity in 60-80% of dogs (Dowling, 1994; Schwartz-Porsche, *et al.,* 1985). Dogs that do not respond to Phenobarbital are considered refractory and are often treated with increased dosages of PB or supplemented with other AEDs like potassium bromide (KBr). The success of combination therapies like PB and KBr is as high as 90% in epileptic dogs (Poddell and Fenner, 1993).

The ability to quickly assess epileptic dogs that are likely to respond to PB could reduce both the financial and emotional burden placed on owners, and would potentially reduce the harmful side-effects associated with ineffective AED usage. High dosages of drugs like PB expose dogs to an elevated risk of unwanted side effects, including liver toxicity and blood dyscrasia (Gaskill, *et al.,* 2005; Jacobs, *et al.,* 1998). The average cost of treating an epileptic dog including routine blood work and monthly medication is estimated to be around $800 a year when the dog is responsive to

medication. However, if the animal is unresponsive or refractory to medication, the cost can exceed $500 a month. This does not include costs incurred during trips to the emergency clinic or those associated with long term drug effects like liver failure.

Pedigree studies of epileptic dogs show high heritability within breeds, as well as variation in prevalence among affected breeds. To date, no genetic markers have been identified that may explain the heritable component, though research is ongoing (Patterson, *et al.,* 2003; Patterson, *et al.,* 2005; Jaggy, *et al.,* 1998; Casal, *et al.,* 2006). Due to the complexity of the disease, there may not be a shared mechanism found across multiple breeds and identification of contributing loci is not expected to lead to immediate treatment options. By contrast, the mechanism of drug response is likely to be simpler and may be more consistent across breeds. Targeting of genes associated with drug response is a more direct approach to effective design of interventions that benefit dogs and their owners.

The concept of individualized medicine, namely using an individual's genotype to guide administration of drug regimens, has shown promise in relation to human epilepsy. A recent study by Tate, *et al.* (2005) identified polymorphisms in the cytochrome P450 gene, *CYP2C9*, and a sodium channel gene, *SCN1A*, which are associated with the effective dosage of the AEDs carbamazepine and phenytoin. Similarly, a common mutation in the multi-drug resistance gene, MDR1-1Δ, in dog breeds from the collie lineage explains hypersensitivity to Ivemectrin and 20 other drugs (Neff, *et al.,* 2005). Here I have taken a candidate gene approach to survey whether variation in PB transporters and metabolizing enzymes, or ion channels thought to mediate the response to PB, might explain variation for refractoriness to PB in epileptic dogs.

Taking a pharmacogenetic approach, Youssef Idaghdour and I designed a custom Illumina Beadarray platform to examine 384 single nucleotide polymorphisms (SNPs) across 30 candidate genes with a potential role in Phenobarbital pharmacology, in 125 responsive and non-responsive epileptic dogs. We selected an average of 12 SNPs spread over a 1MB region of each gene, expecting these to capture much of the haplotype structure at each locus (Lindblad-Toh, *et al.* 2005). Case control associations were done with two independent replicates as well as a combined data analysis to reveal several strong associations with drug response. After adjustment for multiple comparisons, five genes were found to show suggestive evidence for association to drug response, namely a potassium channel gene (*KCNQ3*), a sodium channel gene (*SCN2A2*), a GABA receptor gene (GABRA2), which are all potential targets of Phenobarbital, a ABC transporter gene (ABCC4)and a relatively rare variant of *epoxide hydrolase* (*EPOX HYD*) (Meldrum and Rogawski, 2007). Further investigation of the effect of genetic variation in these drug target genes in relation to Phenobarbital drug response is warranted, as our data strongly support the development of a diagnostic test that may be used to recommend selection of an alternative treatment regimen in epileptic dogs.

Methods

Criteria for Selection of Dogs

Epileptic dogs were recruited through local/regional veterinarians, the North Carolina State University Veterinary Teaching Hospital, Breed Clubs, and the project website (www.carolinacanineepilepsy.org). Dogs were included in the study if they had been treated with Phenobarbital for at least one year, had up to date blood work, seizure logs, and showed a normal

physical examination prior to initiation of seizure activity. Refractory dogs were those dogs that had increased incidence of seizures (>1/month)) while maintaining a therapeutic PB blood level of >30 ug/ml. We also accepted epileptic dogs that were supplemented with other AEDs like KBr, as long as the owners provided evidence of PB as the initial method of treatment. Responsive participants were those dogs that had less than one seizure per month while on PB in the absence of other AEDs. Non-epileptic purebred dogs were also accepted into the study for use in population structure analysis. All samples were collected under North Carolina State University Institutional Animal Care and Use Committee protocol #05-125-B.

Samples and Breed Representation

A total of 205 blood samples were collected representing 45 breeds. Genotyping was attempted for a total of 161 samples on 2 plates (75 for plate 1 and 86 for plate 2). Replicates were added to fill each plate to the complete 96 samples. Calling of genotypes was successful on 152 samples from both plates including 75 male and 77 female dogs (Table 1). There were 27 non-epileptic, 40 PB responsive, and 85 refractory dogs. Sample collection took two years, so the second plate was genotyped 12 months after the first, but overall genotype frequencies were similar between the two plates.

SNP Selection

A custom Illumina Sentrix Bead genotyping array was designed for 384 single nucleotide polymorphisms (SNPs), across 30 candidate genes. The thirty candidate genes were chosen for their

involvement in human epilepsy or as known targets of Phenobarbital, drug metabolizing enzymes, or drug transporters (Table 2). For each gene, 11 to 18 SNPs were identified using the online SNP database at the Canine Genome Project at the Broad Institute. SNPs were spaced on average at 100kb intervals within the coding regions and spanned out to 500kb intervals on the flanking regions for each gene. We also assessed known allele frequencies for each SNP, requiring that each had a minor allele frequency (MAF) of greater than 5% and was polymorphic across multiple breeds (Table 2).

DNA Extraction and Illumina Sentrix BeadArray

Owners were asked to submit a 3-5 ml blood sample in an EDTA tube. DNA was extracted from 200 µl of whole blood to a minimum concentration of 75ng/µl using a Qiagen QIAamp DNA Extraction Kit (#51304). All samples were quality controlled by quantification using a nanodrop analyzer, and 25 randomly selected samples were also assessed for integrity on an agarose gel. Samples were considered of good quality if they showed a concentration above 75ng/µl with a 260/280 ratio of approximately 1.8.

The Illumina Sentrix BeadArray platform uses an oligo-specific ligation reaction to amplify regions of interest in genomic DNA surrounding each SNP site in one highly parallel reaction. Allele-specific dye labeled amplified product is hybridized to beads that are held electrostatically into etched grooves on the surface of fiber optic bundles. The platform allows genotyping of 384 SNPs for 96 samples simultaneously, yielding almost 37,000 SNP calls per run.

Of the 205 blood samples collected, 161 samples were hybridized to the Illumina BeadArray Matrix across 2 separate plates. Successful genotype calls from plate 1 were obtained for 75 samples including 20 non-epileptic and 55 epileptic dogs (35 refractory and 20 responsive). Plate 2 generated calls for 7 non-epileptic and 71 epileptic dogs (24 responsive and 47 refractory). A total of 8 samples (5 refractory, 3 responsive) or ~5% of the samples did not generate genotype calls and could not be used for the analysis. Both plates also contained a minimum of 9 replicates per plate and one sample from plate 1 was repeated on plate 2. In addition to in-plate replicates, Youssef Idaghdour independently verified Illumina SNP calls in 6 of my samples for the *MDR1* gene using derived cleaved amplified polymorphic sequence (dCAPS). I also sequenced a region in *CYP2A1* in 7 of the samples to further verify allele calls. In all cases replicates had a 100% concordance with each other.

Of the 384 SNPs, 353 had a successful call rate in 70% of the samples across both plates. For both plates an expected ~ 5% of the SNP markers failed Illumina's clustering algorithm and were not called (Table 3). An additional 24 SNPs were homozygous for all the samples across the combined plate data. Since allele frequencies differ among breeds that show extensive inbreeding, Hardy-Weinberg equilibrium was not adopted as a criterion for inclusion, but only 58 of the 353 SNPs showed significant departure from HWE.

Case/Control Association Test

Case control association tests were performed using the JMP genomics package (Cary, NC), which implements the PROC CASECONTROL procedure for SAS/Genetics. Genotype marker

frequencies were compared between refractory (case) and responsive (control) dogs; non-epileptic dogs were not included in the analyses described here, though these samples were included in the analysis of breed structure and in genotype calling algorithm. Three chi square based tests were performed: an allele case-control test, a genotype case-control test, and the Armitage (1955) trend test (SAS Manual). For each test a genotype marker table is created for each binary or biallelic marker (Table 4).

Both the allele and trend tests identify additive allele effects; however the trend test is the most appropriate test when the Hardy-Weinberg Equilibrium assumption has been violated. The statistics for both tests are given below (Sasieni, 1997):

$$X_T^2 = \frac{N[N(r_1+2r_2)-R(n_1+2n_2)]^2}{R(N-R)[N(n_1+4n_2)-(n_1+2n_2)^2]}$$

$$X_A^2 = \frac{2N[2N(r_1+2r_2)-2R(n_1+2n_2)]^2}{(2R)2(N-R)[2N(n_1+2n_2)-(n_1+2n_2)^2]}$$

The genotype test is best when a dominance effect is suspected. The statistics for this test, which includes components for additive and dominance effects, is (Nielson and Weir, 1999):

$$X_G^2 = \sum_{i=0}^{2}\left[\frac{(Nr_i-Rn_i)^2}{NRn_i}+\frac{(Ns_i-Sn_i)^2}{NSn_i}\right]$$

Chi-square values were then compared to the X^2 values for the appropriate degrees of freedom of 1 (allele/trend tests) or 2 (genotype test) to determine the p-values, which are reported as negative logarithms on the base 10 scale (NLP values).

Due to linkage disequilibrium among markers, which we confirmed extends over 100s of kilobases in most of our loci, there are fewer than 353 independent tests. Since most of the strongest associations were detected at a lower significance level in at least two other SNPs per gene, a reasonable adjustment for multiple comparisons assumes 200 independent evaluations, and an experiment-wide significance threshold of 0.05/200, or $p<0.00025$ (NLP > 3.6).

Population Structure Estimates

In order to assess the extent of haplotype diversity and its potential effect on association testing, the level of population structure among breeds was estimated. I used the program *STRUCTURE*, which adopts a Bayesian modeling algorithm to identify *K* sub-populations that are defined by allele frequencies across the loci (Pritchard, *et al.,* 2000). The admixture model was employed as this assumes that a portion of each individual's ancestry has been inherited from each sub-population. This model does not assume that all markers are unlinked (Falush, *et al.,* 2003). Parameters for *STRUCTURE* were set at 30,000 burnin, with 1,000,000 Markov Chain Monte Carlo reps after burn-in. There are 44 breeds represented in the sample, but setting k=44 would clearly over fit the data since most breeds are represented by just a few dogs. Parker *et al.* (2004) have previously suggested on the basis of microsatellite analyses that there are four ancestral groups of dogs, and setting k=4 recapitulated these groups (data not shown). I also fit a compromise model with k=10 to see how much substructure can be explained with the data, however higher k values were unable to further separate sub-populations. To account for possible spurious associations due to population structure, the *STRUCTURE* output was run through *STRAT,* which is a case control

association program that looks at individuals within subpopulations and identifies markers that exhibit significant associations within subpopulations (Pritchard and Rosenberg, 1999). Population structure in cases of significance was also considered by manual inspection.

Power analysis

To estimate the effect size of the alleles implicated in this study I first estimated the fold increase in responsiveness that would account for the observed difference in genotypes in the cases and controls assuming an overall refractoriness of 20% in dogs. For example, a common allele with both homozygote classes 25% and an incidence of refractoriness of 8% in dogs homozygous for the responsive allele, 16% in heterozygotes (2-fold increase) and 40% in refractory-allele homozygotes (5-fold increase) would be expected to have 50% of the refractory dogs homozygotes for the refractory allele, and 40% heterozygotes. These values are similar to allele frequencies observed at *SCN2A2*. Similar estimates suggest that effect sizes in the range of 3 to 5-fold feasibly explain the observed association at *KCNQ3* as well. Subsequently, the Genetic Power Calculator at http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html was used to determine that this study, with approximately 80 refractory cases and 40 responsive controls, had 80% power to detect common alleles with effect sizes of 5-fold increased partially dominant refractoriness at p=0.01, and 3.5-fold increased completely dominant refractoriness at 0=0.05 (Table 5) (Purcell, *et al.,* 2003). LD between the causal and tagging SNP as well as departures from Hardy-Weinberg equilibrium due to inbreeding in breeds will tend to decrease the power further.

Results

Genotype statistics

In total 353 SNPs were genotyped in 161 samples, with just 8 samples that failed to provide data. The average MAF was 0.23 and the heterozygosity was 0.28 while the allelic diversity was 0.35.

Linkage disequilibrium (LD) is the non random association between loci. Previous studies in dogs have shown LD to extend beyond 1Mb blocks within major breeds, however across multiple breeds LD can break down beyond 200kb (Sutter, *et al.,* 2004; Linblad-Toh, *et al.,* 2005). To ensure the ability to capture the majority of the haplotype diversity, SNP coverage was designed to extend over 1Mb regions encompassing each gene of interest. Consistent with these expectations, within the 125 epileptic dogs (representing 44 breeds), LD typically decayed to background levels (D' ~0.2) over a distance of 700kb, although for some pairs of SNPS, LD extended beyond 1Mb region (Figure 1).

A corollary of this analysis is that LD was not observed above background for markers between genes, even those on the same chromosome. Figure 2 illustrates this trend for chromosome 6, which has 4 genes separated by the vertical dotted lines. Within each gene there are haplotype blocks defined by significant LD between 2 to 5 adjacent SNPS, but D' is always less than 0.2 between genes on the same chromosome. Within genes, LD typically ranges between 0.3 and 0.8, but in some cases there is only low-level LD even between adjacent SNPs. These results indicate that our SNP selection strategy does guarantee some level of association between tagging and

causal SNPs. Considerably more fine scale typing would be required to guarantee that polymorphisms that affect drug response are tagged effectively enough to detect associations given the statistical power calculations above.

Breed Structure Analysis

Currently, there are over 400 known dog breeds, with 152 of those officially recognized by the American Kennel Club (AKC) (Parker, *et al.,* 2004). These breeds normally derive from strict regimented breeding, making the genetic diversity within a breed low. This also creates well defined population or breed structure where dogs within a breed will be more genetically related than dogs between breeds. With intense selection for behavior and morphology it is not surprising that over 350 inherited disorders have been characterized in dogs (Patterson, *et al.,* 1988). Because association statistics can produce false positives when both the case prevalence and allele frequencies differ between sub-populations, as is the case in dogs, care must be taken in drawing conclusions drawn from any associations which have not accounted for breed structure. Recently, using 96 microsatellite markers Parker *et al.* (2004) were able to successfully assign 410 of 414 dogs to the correct breed. Application of *STRUCTURE* using the allele frequencies of the 353 SNPs in all 152 dogs, similarly successful parsed out breeds (Pritchard, *et al.,* 2000). I first ran *STRUCTURE* with the inclusion of the boxer breed, but was concerned that the analysis may be biased by the fact that the SNP database was derived from the sequence of a boxer, so reran the analysis excluding the boxer breed. Figure 3A (left hand plots; each color represents a sub-population) shows the output for just the epileptic dogs from *STRUCTURE* for k=2 through k=4 for the dataset which included the

boxers. The analysis that excluded boxers (Figure 3B) generally recognizes similar breed groups, but individual breeds were separated out as early as k=4 where Labradors were separated out from the border collies and golden retriever subgroup. Every additional level of k after k= 4 was successful at pulling out a breed up through k=9. Clearly from the *STRUCTURE* data there is a significant breed structure among our data that implies caution in interpretation of the association statistics.

Association with Phenobarbital Response

Case control association tests were performed for each plate separately and as a combined dataset. From the first plate which contained 55 epileptic dogs, 23 markers for the allele test, 16 markers for the genotype test and 11 markers for the trend test were significant at an unadjusted p< 0.05. For the second plate, 28 markers were significant for the allele test, 16 markers for the genotype test, and 20 markers for the trend test. The corresponding numbers for the combined dataset were 28 SNPs for the allele test, 17 SNPs for the trend test, and 15 SNPs for the genotype test. Among the 30 genes, 10 genes had no SNPs that showed a p-value below 0.05 for any of the three association tests. Those genes were *5HT2A*, *ABCC1*, *ABCC3*, *CACNA1H*, *CAT-O-METH*, *DPYD*, *MDR1*, *SCN1B*, *SLC4A3*, and *SLC6A4*. There were several genes which had suggestive evidence for at least one SNP, as well as markers that were significant for only one test or did not replicate between plates. Among those genes are *ABCC11*, *ABCC2*, *CYP3A12*, *Dopamine Receptor D2*, *Dopamine Receptor D4*, *Thiop-S-Meth*, *CYP2A13*, *CYP1A2*, and *KCNQ2*. The four most interesting genes are those that had 2 or more significant SNPs, with increased significance in the combined analyses: these were *ABCC4*, *GABRA2*, *KCNQ3*, and *SCN2A2*. Both *ABCC4* and *GABRA2* had 2 significant SNPs,

while *KCNQ3* and *SCN2A2* had 4 and 3 significant SNPs, respectively. Finally, *EPOX HYD* had 2 adjacent SNPs, one of which showed the most significant association in the combined dataset, but these two SNPs individually showed significance on plates 1 and 2 separately.

In order to evaluate the experiment-wide significance of these associations I considered the allele test significance values, as these consistently gave the strongest evidence for association. There was a slight excess of significant tests at $p<0.05$ and $p<0.01$ (26 and 8 tests respectively, compared with expectations of 17 and 3), but the false discovery rate does not approach acceptable levels. The most significant combined association was for SNP BICF229J54854 in *EPOX HYD*, p = 0.00055, which falls just short of the Bonferroni-adjusted significance level with alpha equals 0.05 and assuming 200 independent tests. The strongest associations in *KCNQ3* (BICFG630J249879, p = 0.0015), *SNC2A2* (BICF233J22762, p = 0.0049) and *GABRA2* (BICF230J51677, p = 0.0130) correspondingly fail to meet experiment-wide significance thresholds. None of the above SNPs show a Bonferroni significance level departure from HWE. Consequently, the genetic data can only be regarded as suggestive that variation in these genes affects responsiveness to Phenobarbital. To adjust for potential biases due to the wide range of allele frequencies as well as the unequal number of cases and controls, I permutated the trait values in the dataset 100,000 times to yield derived chi-square distributions, but observed almost identical p-values as those documented in Table 6.

Regarding *EPOX HYD*, as noted, the strongest association was only seen in the second replicate, whereas a weak association was observed for another site in the first replicate. These two sites are in apparent complete LD (D' = 1; see Table 7), so the inconsistency is likely due to the

relatively small sample size of the control (responsive) dataset and low frequency of the minor (refractory) allele. Examination of the locus-specific LD plot in Figure 4 indicates that there is generally low LD across the gene. Consequently, more extensive genotyping would be required to establish whether other markers at this locus may provide stronger evidence for association with refractoriness to PB. Examination of the histograms in Figure 5 showing the counts of each genotype in responsive (red) and refractory (blue) dogs strongly suggests that heterozygosity or homozygosity for the minor allele (BICF229J54854 freq = 0.173) is predictive of refractoriness. Almost all of the responsive dogs are homozygous for the major allele, but typing of a much larger sample of responsive dogs will be required to confirm this association.

Markers in *KCNQ3* and *SNC2A2* show consistent evidence for association of the major allele with refractoriness to PB. The three most strongly associated markers in *KCNQ3* (BICFG630J249879, BICFG630J249554 and BICFG630J249887) are in strong LD as part of a haplotype block that has a minor allele frequency of approximately 0.5 in responsive dogs, but less than 0.3 in refractory ones. Similarly, the two strongest associations in *SNC2A2* (BICF233J22762 and BICFPJ1467385) involve alleles that are in strong LD and at similar frequencies in responsive dogs, but have a clear excess of homozygotes for one allele in refractory dogs. In both cases, the most associated SNPs are significant at the nominal $p<0.05$ on both plates and the association is strengthened in the combined analysis.

Prediction of Phenobarbital Response

I also ran multi-locus nominal logistic regression to test for interactions between SNPs from

different genes. No evidence for departure from additive effects on PB responsiveness was detected. Nevertheless, figure 6 below shows that the combination of homozygotes for the major alleles at *KCN3* and *SCN2A2* is highly predictive of lack of response to Phenobarbital, and the addition of a third genotype at *ABCC4* defines a completely refractory class. Two-way analysis suggest that *SCN2A* 2762 GG paired with the *KCNQ3* 887 AA, essentially only occurs in dogs that are refractory to PB, while with the addition of a third marker *ABCC4* 550 GG ensures that the dog is refractory in my samples.

Receiver Operating Characteristics Analysis (ROC)

The ultimate goal of this project is to identify testable genetic variation that is associated with drug response for application in a clinical setting. The receiver operating characteristic curve (ROC, Figure 7) is a common method for assessing predictive power: it contrasts sensitivity (the fraction of true positives that are identified by the predictive genotype) against 1 minus specificity (the fraction of true negatives that are called positive) (Fischer, *et al.,* 2003). There are two ways to assess the reliability of the diagnostic test; one is to identify the point closest to the upper left hand corner, and the second is to measure the area under the curve (AUC). The upper leftmost point is the point of best discrimination on the graph, as this is where the selectivity is optimized relative to the specificity, while the AUC value represents a summary statistic for overall diagnostic performance (Greiner, *et al.,* 2000). The discriminatory point serves as a reference for physicians as moving up the curve increases selectivity at the expense of increased inclusion of false positives, and vice versa. AUCs of 0.5 to 0.7 are considered of little utility, 0.7 to 0.9 are considered moderately

accurate and 0.9 to 1 are considered highly accurate. The AUC value derived from a diagnostic test for refractoriness using the most strongly associated SNPs in *SCN2A2* (762), *KCNQ3* (887) and *ABCC4* (550) is 0.74. By this measurement, using these three SNPs for a diagnostic test would be moderately sensitive, but would still misdiagnose too many responsive dogs as being refractory to be used as a sole criterion for exclusion from Phenobarbital administration. Additionally, the ROC curve AUC value for *KCNQ3* (887), *SCN2A2* (762), and *EPOX HYD* (854) was .76, indicating that this combination also has potential for a diagnostic test. Further characterizations of these SNPs in a larger study, may help increase the model fit and indicate whether a better predictive value can be obtained for specific breeds. In addition parsing the data and looking at individual breed groups may improve the fit of the model.

Discussion

Interpretation of the Associations

Pharmacogenetic studies direct attention away from discovering the complex underlying mechanism of a disease towards being able to understand how a patient will respond to treatments. The objective is to provide insight into common mechanisms underlying response to drugs, in this case the AED Phenobarbital, and how genotypic differences affect variability among individuals in their responsiveness to the drug. Such information may support clinical applications including the provision of predictive tools for drug response or dose management.

Although I identified several strong associations in 5 candidate genes, none were above the experiment-wide significant levels. The data consequently only provide suggestive evidence that variation in the genes affects responsiveness to PB. The differences in allele frequency of the cases and controls for *KCNQ3* and *SCN2A2* suggest an approximately five-fold increase in refractoriness of homozygotes for the protective allele relative to the responsive genotypes. Power calculations indicate that a study with 110 individuals is sufficient to detect an effect of this magnitude at p = 0.001, and the replication on two separate plates with half this size also supports the contention that the strongest associations are real. To obtain definitive experiment-wise significance a sample ten times larger would be required. However retesting of just the SNPs I have highlighted, in another study with 100 cases and controls would greatly increase confidence in the involvement of the genes.

Two other factors may contribute to the failure to obtain more significant associations. Our LD measurements within a gene showed high local LD usually extending across several SNPs, but often showed low level LD between adjacent SNPs. This suggests that causal SNPs may not have been sufficiently tagged to detect associations that exceed the experiment-wide significance cutoff. More fine scale mapping of these genes will help to identify more appropriate tagging SNPs and increase the likelihood of tagging SNPs that are causally related to the drug response. Another possibility is that refractoriness is a polygenic trait and that individual loci increase the likelihood of refractoriness less than 3-fold. Even under the best-case scenario of complete LD with a common SNP, this study would only have detected such loci with 80% power at p=0.05, well below the

experiment-wide threshold. It is of course also possible that the major genes that contribute to responsiveness were not included in the set of candidate genes included on the custom genotyping chip.

Potential mechanisms of refractoriness to Phenobarbital

There are two main hypothesis used to explain drug refractoriness; the target hypothesis and the multidrug transporter hypothesis (Schmidt and Loscher, 2005). The target hypothesis argues that unresponsiveness may be attributed to variability in the drug targets, while the transporter hypothesis implicates failure of the drug to reach its target because it does not cross the blood brain barrier.

Assuming for the purposes of this discussion that the associations with *KCNQ3*, *SCN2A2*, *ABBC4*, and *GABRA2* are true positives, both of these hypotheses are supported. *SCN2A2*, *GABRA2* and *KCNQ3* are all ion channel genes and play important roles in maintaining balance in the nervous system and are often targets for AEDs (Armijo, *et al.,* 2005). Voltage gated sodium channel genes and GABA receptor genes are specific targets of Phenobarbital, as well as other ion channels (Armijo *et al.,* 2005). Voltage gated sodium channel genes such as *SCN2A2* are responsible for maintaining the balance of neuronal Na+ ion currents. Mis-regulation of channel activity due to mutations in Na+ voltage channel genes or due to neurotoxins leads to hyper excitability of neurons. Phenobarbital binds to the Na+ channel to block the sodium current (Armijo, *et al.*, 2005; Meldrum and Rogawski, 2007). Several of the voltage-gated sodium channel genes, including SCN2A2, have been linked to various forms of epilepsy in humans (Wallace, *et al.,* 1998; Sugawara, *et al.,* 2001). Another gene in

the same family, *SCN1A*, is known to affect carbamazepine and phenytoin drug dosage in epileptic human patients and could potentially have the same implications in dogs (Tate, *et al*., 2005). Voltage gated potassium channels genes like *KCNQ3* have been implicated in several forms of epilepsy and are important for post excitatory membrane repolarization. Inhibition of K+ channel genes leads to hyperexcitability of nerve terminals (NeuBauer, *et al.,* 2008; Armijo*, et al.,* 2005; Meldrum and Rogawski, 2007). This is the first study to suggest an association between polymorphism in *KCNQ3* and response to PB.

GABA receptor A genes, like *GABRA2* are the main targets of Phenobarbital, which serves to enhance their function to prevent excitable nerve discharge (Armijo, *et al.,* 2005). GABA is the main inhibitory neurotransmitter, and acts upon three receptors which work to remove GABA from the synaptic cleft into nerve terminals and glial cells. Activation of both GABA receptors A and B inhibits nerve potential and prevents nerve discharge. Mutations in the GABA receptors are linked to several forms of human epilepsy as well as Phenobarbital and phenytoin resistance in a rat model of temporal lobe epilepsy (NeuBauer, *et al.,* 2008; Armijo, *et al.,* 2005; Volk, *et al.,* 2006).

The potential association of *ABBC4* with AED refractoriness provides some support for the multidrug transporter hypothesis, since *ABBC4* encodes a multidrug resistance gene. Generally, overexpression of these genes prevents AEDs from penetrating the brain. Phenobarbital resistant humans as well as rat temporal lobe epilepsy models show increased levels of a multidrug resistance protein ABCB1 in the brain (Volk and Loscher, 2005). Interestingly, in PB resistant rats treated with tariquidar, a multidrug resistant inhibitor, Phenobarbital's anticonvulsant activities were completely restored (Brandt, *et al.,* 2006).

Finally, dogs are known to metabolize drugs much faster than humans, which make the possible *EPOX HYD* connection to drug response interesting, as it is likely involved in the metabolism of PB (Chandler, 2006; Farnback, 1984; Knowles, 1998). This leads to a third hypothesis to explain refractoriness, which is that some dogs metabolize the drug more quickly than others, preventing them from maintaining therapeutic levels. Human studies have identified alleles in CYP genes that are linked to poor and extensive metabolizers of phenytoin and carbamazepine (Klotz, 2007). Other pharmacogenetic studies in dogs identified poor and fast metabolizers of a xenobiotic drug, showing a link to drug metabolism rates and perception of responsiveness (Tenmizu, *et al.,* 2004). *EPOX HYD* functions in detoxification of epoxides that are the results of the metabolism of aromatic compounds, like PB. *EPOX HYD* is also involved in the conversion of vitamin K epoxide to usable vitamin K_1, which is required for proper cytochrome P450 function (Waxman and Azaroff, 1992). Studies looking at vitamin K levels in mice treated with PB show a significant decrease in vitamin K levels compared to non-PB treated mice (Mochizuki, *et al.,* 2008). In humans, vitamin K is often given as a supplement to pregnant epileptic women to prevent neonatal hemorrhaging due to decreased vitamin K levels caused by CYP activity induced by AEDs (Mochizuki, *et al.,* 2008). In addition, *EPOX HYD* has been implicated in warfarin pharmacogenetic dosaging studies, as patients with a combination of *EPOX HYD* mutations and the *CYP2C9* mutation require maximum dosages of warfarin (Loebstein, *et al.,* 2005). *CYP2C9* and *CYP2B* are the main cytochrome P450s involved in Phenobarbital metabolism in humans and mutations in *CYP2C9* have been linked to Phenobarbital clearance times in humans (Anderson, 2004; Goto, *et al.,* 2007). *CYP2C9* has not been identified in dogs, instead *CYP1A1* and *CYP2B11* are considered the main metabolizers of PB in dogs (Graham, *et*

al., 2006). Although, more research needs to be done to show a definite link between *EPOX HYD* and Phenobarbital metabolism in dogs, it may be possible that supplementing vitamin K into the diet of epileptic dogs will be effective in improving seizure control in refractory epileptic dogs.

Conclusion

Owners of epileptic dogs rank "quality of life" followed by "seizure frequency" as the top factors considered when treating their epileptic dog (Chang, *et al.,* 2006). A diagnostic test that predicts the usefulness of Phenobarbital for seizure control would improve both of these factors. Based on the results from this study, the concept of individualized medicine in canines has potential, but larger scale studies and possibly even focused studies within breeds are essential before clinical usage of these associations is warranted. If the *KCNQ3* and *SCN2A2* effects are replicated in further studies in dogs, then it is feasible that a diagnostic test could be developed. ROC analysis indicates that our model is moderately effective at predicting drug response with SNPs from *KCNQ3*, *SCN2A2*, and *ABCC4*, but not ideal for clinical applications. Ideally further sequence analysis would also identify the causal polymorphisms, suggesting a mechanism of function and ensuring that the accuracy of the test is not compromised by imperfect linkage disequilibrium. This test would be useful for veterinarians and owners of epileptic dogs to establish an appropriate treatment plan more quickly after the diagnosis of canine epilepsy.

Acknowledgements

I would like to thank the Epil-K9 Foundation for their support in advertising the project, the recruitment efforts and the funding of this project. Dr. Natasha Olby for her help in project planning

and with recruitment of normal and epileptic dogs as well as the members of her lab for their assistance with blood draws. Also thanks to Julie Osborne and Dr. Karen Munana for their assistance with the IACUC procedures and recruitment efforts. Dr. Breen's lab at the NCSU vet school also provided non-epileptic blood samples from pure bred dogs. Youssef Idaghdour was instrumental in selection of genes and design of the Illumina SNP chip.

References:

Anderson, G. (2004). Pharmacogenetics and enzyme induction/inhibition properties of antiepileptic drugs. *Neurology.* 63: S3-S8.

Armijo, J., Shushtarian, M., Valdizan, E., Cuadrado, A., de las Cuevas, I., Adin, J. (2005). Ion Channels and Epilepsy. *Current Pharmaceutical Design.* 11: 1975-2003.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics.* 11: 375-386.

Brandt C, Bethmann K, Gasten A, Loscher W. (2006) The multidrug transporter hypothesis of drug resistance in epilepsy: Proof of principle in a rat model of temporal lobe epilepsy. *Neurobiol Dis* . 24: 202-211.

Benet, L., Kroetz, D., Sheiner, L. (1996). Pharamacokinetics: Dynamics of Drug absorption, distribution, and elimination. *Pharmacology Basis Therapy*. 9: 3-28.

Casal, M., Munuve, R., Janus, M., Warner, P., Henthorn, P. (2006). Epilepsy in Irish Wolfhounds. *J Intern Vet Med*. 20: 131-5.

Chandler, K. (2006). Canine Epilepsy: what we can learn from human seizure disorders?. *The Veterinary Journal*. 172: 207-217.

Chang, Y., Mellor, D., Anderson, T. (2006). Idiopathic epilepsy in dogs: owners' perspectives on management with phenobarbitone and/or potassium bromide. *J Small Animal Practice*. 47: 574-581.

Dowling, P. (1994). Management of canine epilepsy with Phenobarbital and potassium bromide. *Canadian Veterinarian Journal*. 35: 724-725.

Faulsh, D., Stephens, M., Pritchard, J. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* 164: 1567-1587.

Farnback, G. (1984). Serum concentratins and efficacy of phenytoin, Phenobarbital, and primidone in canine epilepsy. *J Am Vet Med Assoc*. 184: 1117-1120.

Fischer, J., Bachmann, L., Jaeschke, R. (2003). A reader's guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med*. 29: 1043-1051.

Gaskill, C., Miller, L, Mattoon, S., Hoffmann, W., Burton, S., *et al*. (2005). Liver histopathology and liver and serum alanine aminotransferase and alkaline phosphatase activities in epileptic dogs receiving Phenobarbital. *Veterinary Pathology*. 42: 147-160.

Goto, S., Seo, T., Murata, T., Nakada, N., Ueda, N., *et al*. (2007). Population estimation of the effects of cytochrome P450 2C9 and 2C19 polymorphisms on phenobarbital clearance in Japanese. *Ther Drug Monit*. 29: 118-21.

Graham, R., Tyler, L., Krol, W., Silver, I., Webster, L., *et al.* (2006). Temporal Kinetics and Concentration–Response Relationships for Induction of CYP1A, CYP2B, and CYP3A in Primary Cultures of Beagle Dog Hepatocytes. *J Biochem Molecular Toxicology.* 20: 69-78.

Greiner, M., Pfeiffer, D., Smith, R. (2000). Principles and pratical application of the receiver-operating characteristics analysis for diagnostic test. *Preventive Veterinary Medicine.* 45: 23-41.

Jacobs, G., Calvert, C., Kaufman, A. (1998). Neutropenia and thrombocytopenia in three dogs treated with anticonvulsants. *Journal of American Medical Association*. 212: 681-684.

Jaggy, A., Faissler, D., Gaillard, C., Srenk, P., and Graber, H. (1998). Genetic Aspects of Idiopathic Epilepsy in Labrador Retrievers. *J Sm Anim Prac*. 39: 275-80.

Klutz, U. (2007). The role of pharmacogenetics in the metabolism of antiepileptic drugs: pharmacokinetic and therapeutic implications. *Clin Pharamcokinet.* 46: 271-279.

Knowles, K. (1998). Idiopathic Epilepsy. *Clinical Techniques in Small Animal Practice*. 13: 144-151.

Licht, B., Licht, M., Harper, K., Lin, S., Curtin, J., Hyson, L., Willard, K. (2002). Clinical presentations of naturally occurring canine seizures: similarities to human seizures. *Epilepsy and Behavior*. 3: 460-470.

Lindblad-Toh, K., Wade, C., Mikkelsen, T., Karlsson, E., Jaffe, D., *et al*. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature.* 438: 803-819.

Loebstein, R., Vecsler, M., Kurnik, D., Austerweil, N., Gak, E., Halkin, H., Almog, S. (2005). Common genetic variants of microsomal epoxide hydrolase affect warfarin dose requirements beyond the effect of cytochrome p450 2C9. *Clin Pharmacol Ther.* 77: 365-72.

Meldrum, B. and Rogawski, M. (2007). Molecular targets for antiepileptic drug development. *Neurotherapuetics.* 4: 18-61.

Mochizuki, M., Shimizu, S., Kitazawa, T., Umeshita, K., Goto, K., *et al.* (2008). Blood coagulation-related parameter changes in Sprague-Dawley (SD) rats treated with Phenobarbital (PB) and PB plus vitamin K. *J Toxicol Sci.* 33: 307-314.

Neff, M., Robertson, K., Wong, A., Safra, N., Broman, K., *et al.* (2004). Breed distribution and history of canine mdr1-1delta, a pharamcogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc Natl Acad Sci USA.* 101: 11725-11130.

NeuBauer, B., Waldegger, S., Heinzinger, J., Hahn, A., Kurlemann, G., Fiedler, B., *et al*. (2008). KCNQ2 and KCNQ3 mutations contribute to different idiopathic epilepsy syndromes. *Neurology*. 72: 177-83.

Nielsen, D. and Weir, B. (1999). A classical setting for associations between markers and loci affecting quantitative traits. *Genet. Res.* 74**:** 271–277

O'Brein, D. (2003) Pathogenesis of idiopathic epilepsy, in *Proceedings. American College of Veterinary Internal Medicine Annual Forum*. 404-406.

Parker, H., Kim, L., Sutter, N., Carlson, S., Lorentzen, T., *et al.* (2004). Genetic Structure of the purebred domestic dog. *Science.* 304: 1160-1164.

Patterson, D., Haskins, M., Jezyk, P., Giger, U., Meyers-Wallen, V., *et al.* (1988). Research on genetic diseases: reciprocal benefits to animals and man. *J. Am. Vet. Med. Assoc.* 193**:** 1131–1144.

Patterson, E., Armstrong, P., O'Brien, D., Roberts, M., Johnson, G., Mickelson, J. (2005). Clinical Description and Mode of Inheritance of Idiopathic Epilepsy in English Springer Spaniels. *J Am Vet Med Assoc*. 226: 54-58.

Patterson, E., Mickelson, J, Da, Y., Roberts, M., McVey, A., O'Brien, D., Johnson, G., and Armstrong, P. (2003). Clinical Characteristics an Inferitance of Idiopathic Epilepsy in Vizlas. *J Am Vet Med Assoc*. 17: 319-325

Poddell, M., Fenner, W. (1993). Bromide Therapy in refractory canine idiopathic epilepsy. *Journal of Veterinary Internal Medicine*. 7: 318-327.

Pritchard, J. K. and Rosenberg, N. A. (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum.Gen*. 65: 220–228.

Pritchard, J., Stephens, M., Rosenberg, N., and Donnelly, P. (2000). Association Mapping in Structured Populations. *Am J Hum Genet.* 67: 170-181.

Purcell S, Cherny SS, Sham PC. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics.* 19:149-150.

Sasieni, P. (1997). From genotype to genes: doubling the sample size. *Biometrics***.** 53: 1253–1261.

Schmidt, D. and Loscher, W. (2005). Drug resistance in epilepsy: putative neurobiologic and clinical mechanisms. *Epilepsia.* 46: 858-877.

Schwartz-Porsche, D., Loscher, W., Frey, H. (1985). Therapeutic efficacy of Phenobarbital and primidone in canine epilepsy: a comparison. *Journal of Veterinary Pharmacology and Therapeutics*. 8: 113-119.

Sugawara, T., Mazaki-Miyazaki, E., Ito, M., Nagafuji, H., Fukuma, G., Mitsudome, A., *et al.* (2001). Na-v-1.1 mutations cause febrile seizures associated with afebrile partial seizures. *Neurology.* 57: 703-705.

Sutter, N., Eberle, M., Parker, H., Pullar, B., Kirkness, E., *et al.* (2004). Extensive and breed-specific linkage disequilibirium in canis familiaris. *Genome Research.* 14: 2388-2396.

Tate, S., Depondt, C., Sisodiya, S., Cavalleri, G., Schorge, S., *et al*. (2005) Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proc Natl Acad Sci USA*. 102: 5507-5512.

Tenmizu, D., Endo, Y., Noguchi, K., Kamimura, H. (2004). Identification of the novel canine CYP1A2 1117 C>T SNP causing protein deletion. *Xenobiotica.* 34: 835-846.

Volk, H. and Loscher, W. (2005). Multidrug resistance in epilepsy: rats with drug resistant seizures exhibit enhanced brain expression of P-glycoprotein compared with rats with drug-responsive seizures. *Brain.* 128: 1358-1368.

Volk, H., Arabadzisz, D., Fritschy, J., Brandt, C., Bethmann, K., Loscher, W. (2006). Anti-epileptic drug-resistant rats differ from drug-responsive rats in hippocampal neurodegeneration and $GABA_A$ receptor ligand binding in a model of temporal lobe epilepsy. *Neurobiology of Disease*. 21: 633-646.

Wallace, R., Wang, D., Singh, R., Scheffer, I., George, A., Phillips, H. *et al.* (1998). Febrile seizures and generalized epilepsy associated with a mutation in the Na(+)-channel beta-1 subunit gene SCN1B. *Nature Genet.* 19: 366-370.

Waxman, D. and Azaroff, L. (1992). Phenobarbital induction of cytochrome P-450 gene expression. *Biochem J.* 281: 577-592.

TABLES

Table 1: Synopsis of breed, sex, and epileptic status of the 152 samples selected for genotyping.

Breed	Male	Female	Normal	Refractory	Responsive	Total
Amstaff	4	1	5	0	0	5
Australian Shepherd	1	1	0	1	1	2
Basset Hound	1	1	1	1	0	2
Beagle	2	3	0	1	4	5
Belgian Malinois	0	1	0	1	0	1
Labrador Retriever	4	9	1	7	5	13
Collie	1	1	0	2	0	2
Border Collie	5	4	1	5	3	9
Boxer	1	4	2	2	1	5
Chihuahua	1	0	0	1	0	1
Cocker Spaniel	0	2	1	0	1	2
Welsh Pembroke Corgi	2	0	0	1	1	2
Dachshund	0	1	1	0	0	1
Dalmatian	2	4	1	3	2	6
Doberman Pinscher	3	1	3	0	1	4
German Shepherd	3	3	1	5	0	6
German Shorthair Pointer	1	1	0	2	0	2
Golden Retriever	4	2	0	4	2	6
Greyhound	1	1	1	1	0	2
Irish Setter	1	1	0	2	0	2
Keeshond	1	1	0	1	1	2
Lhasa Apso	1	1	0	0	2	2
Mastiff	1	2	3	0	0	3
Old English Sheepdog	1	0	1	0	0	1
Pug	1	0	0	1	0	1
Rat Terrier	1	0	1	0	0	1
Rhodesian Ridgeback	1	1	1	0	1	2
Rottweiler	1	1	0	1	1	2

Table 1: (continued)

Shetland Sheepdog	1	2	1	2	0	3
Shih Tzu	1	0	0	0	1	1
Saint Bernard	1	1	0	2	0	2
Standard Poodle	2	0	0	1	1	2
Weimaraner	2	0	0	1	1	2
West Highland Terrier	0	1	1	0	0	1
Whippet	1	1	0	0	2	2
Mixed Breed	13	17	1	22	7	30
Yorkshire Terrier	1	1	0	1	1	2
Misc. Terrier	1	3	0	2	2	4
Misc. Hounds	4	1	0	4	1	5
Bulldog	1	1	0	2	0	2
American Eskimo	0	1	0	0	1	1
English Field Setter	0	1	0	1	0	1
Newfoundland	1	0	0	1	0	1
Cavalier King Charles Spaniel	1	0	0	1	0	1
Totals	75	77	27	82	43	152

Table 2: Table shows the gene symbols, the number of SNPs used and average minor allele frequency for the 30 genes examined in this study.

Gene	# of SNPs	Average MAF
ABCC1	13	0.253
ABCC11	12	0.213
ABCC2	13	0.258
ABCC3	13	0.203
ABCC4	13	0.218
ABCC5	12	0.238
ABCG2	12	0.224
MDR1	18	0.262
5HT2A	12	0.246
DRD2	12	0.254
DRD3	13	0.229
DRD4	12	0.244
DPYD	13	0.103
GABRA2	13	0.180
SLC4A3	13	0.275
CAT O-METH	12	0.214
CYP1A2	14	0.199
CYP2A13	14	0.207
CYP2C21	13	0.271
CYP3A12	13	0.233
EPOX HYD	11	0.234
THIO-S-METH	13	0.232
CACNA1H	12	0.111
KCNQ2	12	0.290
KCNQ3	13	0.216
SCN1B	12	0.240
SCN2A2	13	0.233
RALBP1	12	0.272
SLC6A4	12	0.167
TPH1	12	0.213

Table 3: Synopsis of SNP calls from the BeadArray Genoptyping Chip

Total SNPs	Failed Plates 1 & 2	Failed Plate 1	Failed Plate 2	Total SNPs
384	7	10	14	353

Table 4: Genotype distribution for cases and control samples for the marker.

	0	**1**	**2**	**Total**
Case	r_0	r_1	r_2	R
Control	s_0	s_1	s_2	S
Total	n_0	n_1	n_2	N

Table 5: Genetic Power Calculator Results showing the required number of cases needed to detect association for each alpha level. Disease allele frequency equals 0.3.

# Cases	Ratio (case/control)	Relative fold increase	Number of cases needed for 80% at each alpha level		
			0.05	0.01	0.001
80	0.5	3.5	68	102	150
80	0.5	5	50	75	110
100	1	3.5	44	66	96
100	1	5	32	48	70

Table 6: Genes of interest with the Genotype/Allele/Trend association test results (negative log P-values) as well as their respective minor allele frequencies (MAF). Alleles calls are given with the minor allele in parentheses. Stars (*) are when the minor allele is associated with refractoriness

Gene	SNP	Genotype	Allele	Trend	MAF	Alleles (minor)
EPOX HYD	BICF229J54854	2.21	3.26	2.82	0.173	A(G)*
	BICFPJ702064	1.22	1.86	1.73	0.184	A(C)
KCNQ3	BICFG630J249879	1.56	2.82	1.97	0.359	A(G)
	BICFG630J249554	1.82	2.81	2.34	0.324	A(G)
	BICFG630J249887	1.99	2.56	2.12	0.44	A(G)
	BICF232J60180	1.29	0.51	0.44	0.444	A(G)
SNC2A2	BICF233J22762	1.72	2.31	2.14	0.348	G(A)
	BICFPJ1467385	1.26	1.9	1.74	0.416	A(G)
	BICF234J60448	0.87	1.45	1.25	0.28	C(A)*
ABCC4	BICFG630J443807	1	1.65	1.49	0.336	A(C)
	BICFG630J443550	1.23	2.04	1.76	0.328	G(A)*
GABRA2	BICF230J51677	1.14	1.89	1.65	0.116	G(A)*
	BICF233J15851	0.85	1.38	1.13	0.496	G(A)*

Table 7: LD measurements between individual SNPs within a gene.

Gene	Locus 1	Locus 2	Distance	Absolute D'	P-Value
ABCC4	BICFG630J443807	BICFG630J443550	237819	0.03	0.74
EPOX HYD	BICFPJ702064	BICF229J54854	121669	1	2.0E-2
GABRA2	BICF233J15851	BICF230J51677	482574	0.32	0.19
KCNQ3	BICFG630J249554	BICFG630J249879	313235	1	4.49E-07
KCNQ3	BICFG630J249554	BICFG630J249887	388349	0.29	1.00E-2
KCNQ3	BICFG630J249554	BICF232J60180	1080221	0.22	6.0E-2
KCNQ3	BICFG630J249879	BICFG630J249887	75114	1	3.00E-3
KCNQ3	BICFG630J249879	BICF232J60180	766986	1	8.00E-2
KCNQ3	BICFG630J249887	BICF232J60180	691872	0.13	0.16
SNC2A2	BICF233J22762	BICFPJ1467385	76070	0.64	5.05E-10
SNC2A2	BICF233J22762	BICF234J60448	375677	0.64	1.00E-3
SNC2A2	BICFPJ1467385	BICF234J60448	299607	0.69	5.01E-05

FIGURES

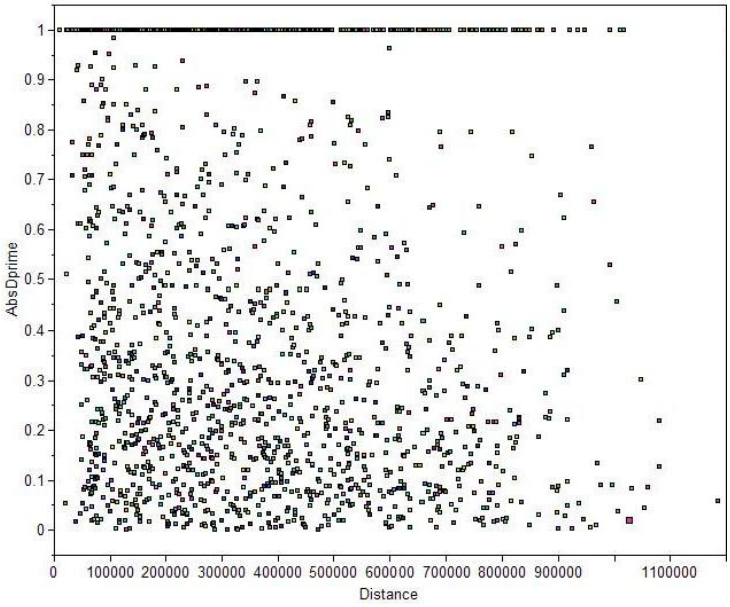

Figure 1: Rate of Decay of LD is plotted as the absolute value of D prime against pairwise distance measurements between marker sets within a gene. Across the dataset LD decays to less than 0.2 after 700kb.

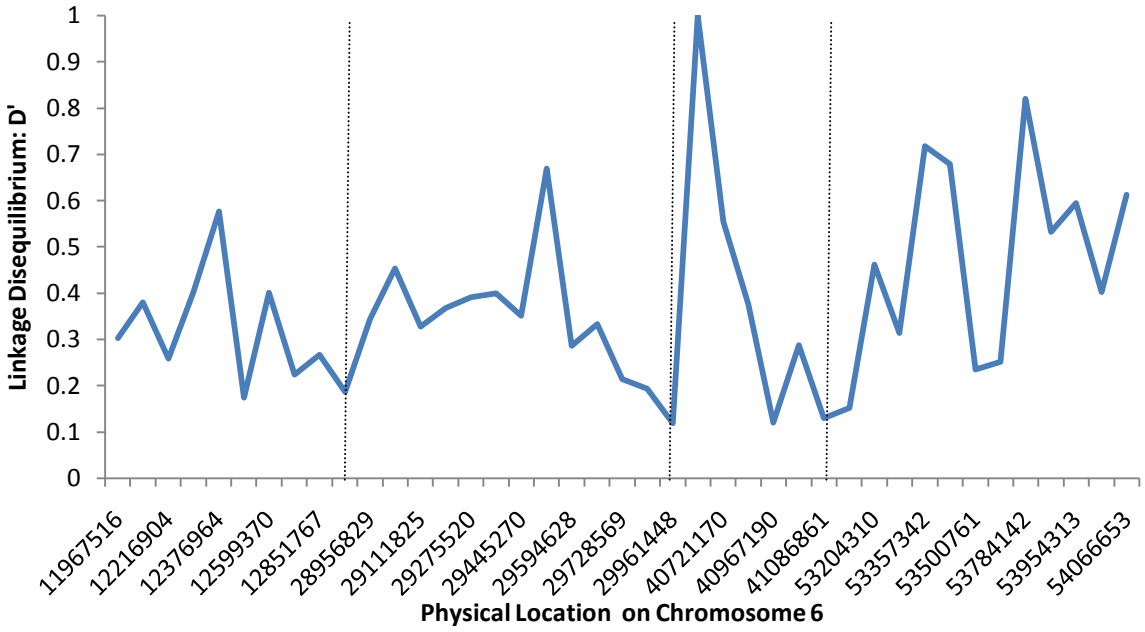

Figure 2: Sliding window along chromosome 6, where each dashed line are the boundaries for gene regions along the chromosome. LD declines to 0.2 or below between genes. *Note X-axis is not to scale.

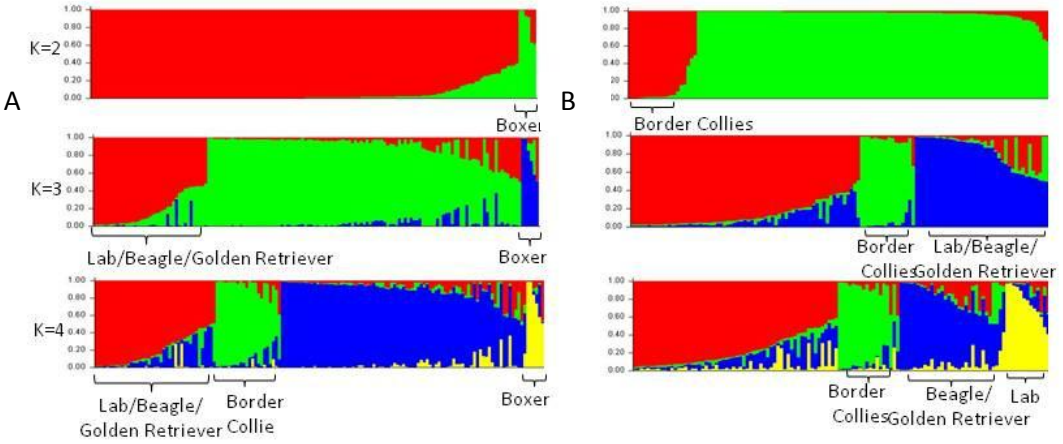

Figure 3: (A) *STRUCTURE* output including the boxers; (B) Structure output with Boxers excluded. Each dog is represented by a column with the proportion of its ancestry derived from k=2, 3, or 4 populations, each represented by a color, on the y-axis.

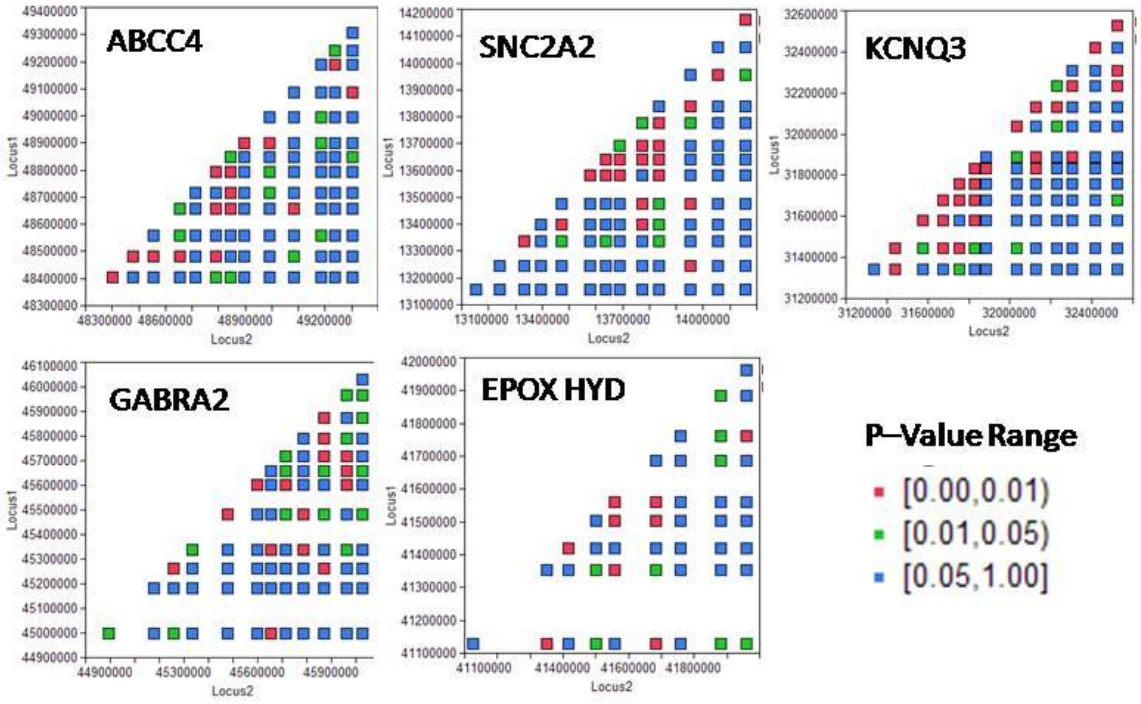

Figure 4: LD P-value plots for top five genes. Each colored square represents the p-value associated with the D' measure of LD between two SNPs at the positions indicated along the locus. Red blocks mark haplotype blocks consisting of two or more SNPs in strong LD.

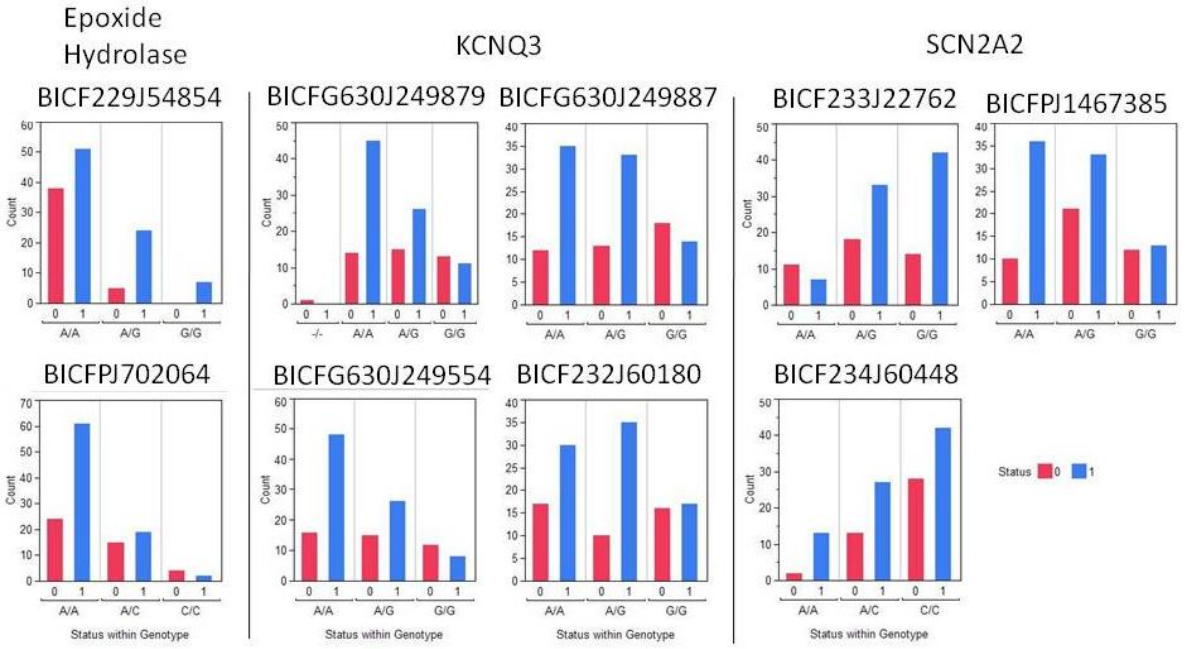

Figure 5: Genotype counts for responsive and refractory dogs for SNPs within *EPOX HYD*, *KCNQ3* and *SCN2A2* which showed significant association for drug response. Red bars are responsive dogs while blue bars are the refractory dogs.

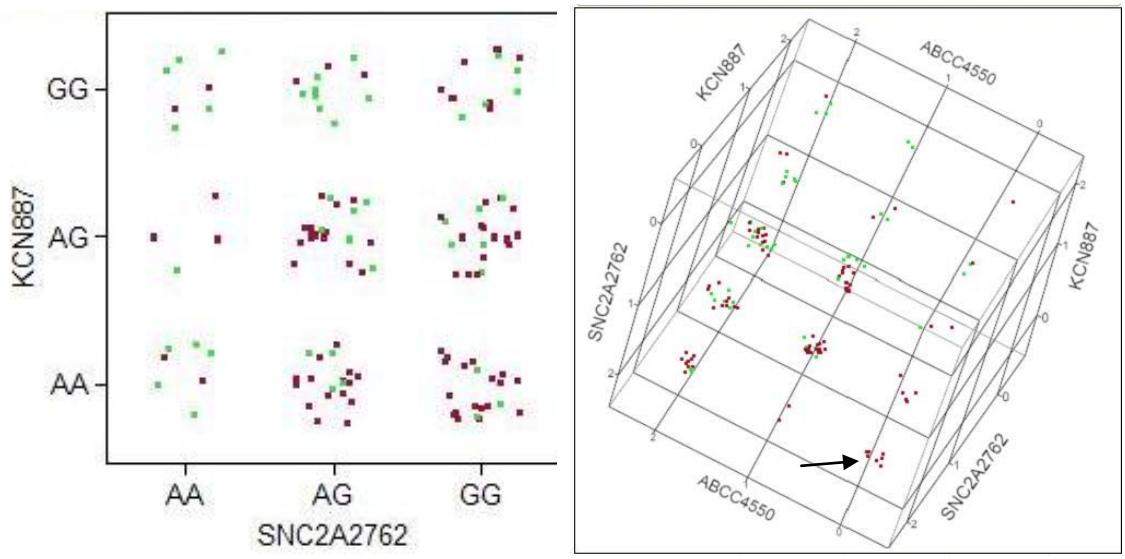

Figure 6. Two and Three-locus scatterplots of responsiveness to PB. (A) The combination of major alleles for *KCNQ3* 887 (AA) and *SNC2A2* 2762 (GG) defines animals that are highly unlikely to be responsive to the drug (green dots, responsive; red dots, refractory). (B) Addition of the *ABCC4* SNP 4550 genotype further differentiates these dogs such that three-way homozygotes are all refractory to PB (Arrow).

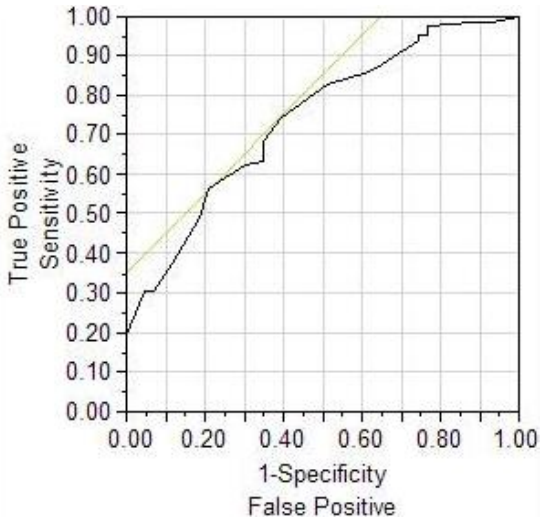

Figure 7: ROC curve for refractoriness as a function of *SCN2A2* (762), *KCNQ3* (887), and *ABCC4* (550).

CHAPTER FIVE

Conclusion and Future Directions

CONCLUSIONS AND FUTURE DIRECTIONS

The emergence of new genomic tools has enhanced our ability to study how genes and the environment interact to promote complex diseases, both in humans and emerging model systems, like the domestic dog. Canines shares a history with man that extends back 14,000 to 15,000 years ago (Vila, *et al*., 1997; Morey, 1992). Today, about 36% of households in the United States own at least one dog and ~55% of those are purebred dogs (Sutter and Ostrander, 2004). Because of breeding practices and rigorous selection on dogs, they have become an ideal system to study monogenic and polygenic diseases. Susceptibility to diseases and disorders such as various cancers and congenital deafness and hip dysplasia is common in some breeds and almost non-existent in others. In this thesis I have demonstrated how genomic tools can be applied in the canine system to address the effects of habitat change in the red wolf, the immune response to hookworm infections, as well as the pharmacogenetic response to epileptic drugs.

I performed gene expression profiling of 13 red wolves and 1 coyote from the NC Alligator River refuge using Illumina's Human Sentrix Ref8 BeadArrays in order to identify potential biomarkers for physiological status. The use of biomarkers has been widely explored in invertebrates and aquatic species, but its use as a conservation tool in mammals has yet to be explored (Snell, *et al.,* 2003; Umina, *et al.,* 2005; Giger, *et al.,* 2006). By incorporating pedigree information, I was able to account for gene expression changes associated with genetic differentiation, which were almost as strong as the environmental response. With the use of the heterologous array, only 12% of the probes hybridized above background levels. Despite the

decreased hybridization, I was able to identify several differentially expressed pathways which had been previously linked to stress, dietary, or immunological responses. The identification of these pathways provide a starting point for large scale studies which can look at specific effects of life style in wolves. This study may also serve as a guide for future conservation studies which aim to monitor and potentially improve the status of captive animals by investigating the effects of nutrition, exercise and other physiological factors associated with captivity.

The development of new tools raises issues with regard to the practicality of their application for certain studies. Chapter 2 shows the potential benefit of using microarrays to evaluate potential differences between wolves in a captive and free range environment. This is a relatively unexplored method for identification of physiological differences between the two environments. Although it is important to capture these differences, it should be understood that microarrays are subject to experimental biases and hypotheses derived from their results should be further validated. While it can be argued that microarrays are an unnecessarily complicated tool for some purposes, they do provide a more global view of physiological changes that accompany for example pathogen exposure in wild animals, than immune and other physiological markers can provide. If measuring the amount of stress captivity places on an animal, evaluation of cortisol levels to measure stress levels in animals would validate and enhance the microarray approach. Interestingly, cortisol levels can even be evaluated from fecal samples and therefore can be collected non-invasively from wild and captive animals (Chelini, *et al.,* 2006). In my study, unfortunately we did not have cortisol levels on these animals, but they would provide secondary

measurement of the physiological state of these animals, although I question whether we would see a difference in levels between the two groups as they are probably well adjusted to their environmental states. A study in wild African green monkeys which measured cortisol and its counter agent prolactin, found a sharp increase in cortisol levels one day after capture, which dropped by the second day of captivity while prolactin levels increased continuously over a 7 month period (Suleman, *et al.,* 2004). Other studies have addressed specific effects of captivity on wild animals and have found items like altered social behavior in Rhesus monkeys, differential tooth wear and longevity in ruminants, and altered body fat make up in rhinoceroses (Clauss, *et al.*, 2008; Jurado, *et al.,* 2008; Judge and De Waal, 1997). In these cases good veterinarians and conservations biologist were effective for identifying differences between captive and free ranging animals, and it remains to be seen whether expression profiling will be a truly beneficial diagnostic tool.

As with a lot of science, research has social implications, and in this case demonstration of a systemic stress response in captive wolves raises arguments about the mental state of captive animals. Even the use of the term "captive" was controversial for some of my collaborators, and the word "confined" was substituted instead. The bigger question is "Are we serving a greater purpose by maintaining animals in captivity in order to preserve a species?" My study was not designed to address this question, but rather to address some of the potential consequences of holding animals in captivity that cannot and should not be avoided. In regards to the mental state of the animals in this study, it should be pointed out that these animals have been in captivity for most if not all of their lives. The gene expression differences may be simply a reflection of the dietary or exercise

differences between the animals in the two environments, and have no reflection on their mental state. Releasing animals which have lived in captivity all of their lives into the wild may be more detrimental than maintaining the status quo.

Microarray technology can also be used to gain a better understanding of genes involved in disease and health. Using microarrays, I investigated transcript changes over a 12 week period in 3 beagle pups that were infected with parasitic hookworms. This was the first study to characterize the genome wide response to infection, and adds considerably to the list of biomarkers that may be used to measure the perturbed activity of the immune system in infected animals. This type of study is important as it gives a broader understanding of how the immune system fights off infections and suggests possible connections between the immune system and other bodily functions. An example is the role of cytokines in stimulation of specific cells in the gut to increase secretions and muscle contractions that lead to expulsion of worms. This study was originally performed using Affymetrix's canine whole genome microarrays, however the cost associated with using the platform prevented adequate technical replication. Hoping to achieve the same success as the red wolf project, the experiment was replicated using Illumina's Sentrix Ref8 human whole genome beadarray. The cost associated with Illumina's platform allowed the inclusion of a fourth time point (week 12) as well as a replicate for each array. However, for this experiment only about 8% of the original 24,354 probes showed hybridization above background levels. Using a repeated measures model, I was able to identify genes from both platforms that showed altered expression through the course of in the infection.

Although several pathways were identified which suggested broad activation of an immune response, neither the Illumina nor the Affymetrix platforms clearly distinguished a Th1 or Th2 immune response and it appears instead that the canine hookworm elicits a hybrid response that presumably relates to the ability of the parasite to mount a chronic and relatively benign infection. With experimental cost decreasing as technology advances, increasing the number of dogs used as well as the number of time points on a canine specific platform may help define which aspects of the response allow for the balance of mutualistic and parasitic interactions. The annotated sequence of the hookworm genome identifies several transposons and excretion/secretion proteins that may serve to suppress the immune response in the host (Moyle, *et al.,* 1994; Abubucker, *et al.,* 2008). Dogs are a good model to study the host-parasite interaction as they are able to mount an immune response which overcomes these immunomodulaton cues and they develop protective immunity against further infection. More focused cellular immunology study may be the best way to understand these interactions which may explain differences between human and canine response to infection.

In line with the "hygiene hypothesis", studies have identified a negative correlation between the prevalence of an atopic dermatitis or asthmatic allergic response and nematode infections (Wordemann, *et al.*, 2008). This link was first discovered in the 1970s from surveys in countries with high incidences of nematode infections but low occurrence of asthmatic reactions (Turner, 1978; Klein, *et al.,* 1971). Conversely, people who have increased hypersensitive atopic dermatitis or asthma show an increased resistance to hookworm infections (Grove and Forbes, 1975). The

hypothesis is that the immunoregulatory mechanisms of the hookworm inhibit IgE, which is the main immunoglobulin responsible for the Th2 type allergic reaction. Therapeutic hookworm infections are currently being studied as an alternative method to treat asthma (Mortimer, *et al.,* 2006). Another study showed that the protection provided from nematode infection was age and parasite specific (Wordemann, *et al.,* 2008). A further advantage of dogs as a model is that they illicit a similar IgE driven immune response to allergies and may provide insight into parasite and age specific immune protection.

The recent identification of a gene for body size which distinguishes large and small breed dogs, is an example of how the population dynamics of dogs can be a useful tool in understanding the genetics of complex traits (Sutter, *et al*., 2007). Similarly, we can use dogs as a model in biomedical studies to characterize genetic variation's role in diseases and drug response. Recent studies in both humans and dogs have demonstrated how genetic variation can influence drug response (Tate, *et al.,* 2005; Neff, *et al.*, 2004). As a tool to identify genetic variation which influences drug response in epileptic dogs, we developed a canine specific Illumina SNP chip which was used to characterize 354 SNPs spread across 30 genes which had previously been linked to drug response or epilepsy disorders in humans and mice. A case control association study between 125 refractory and responsive epileptic dogs identified associations to drug response in 5 candidate genes; however a larger scale study is needed to increase confidence that these genes are involved in drug response. Current pharmacogenetic studies focus on two main hypotheses to explain drug response, the drug target and multidrug transport hypotheses. With the exception of the *ABCC4*

transporter, which only showed a weak association, my data is more consistent with the target theory as the major explanation for variation in anti-epileptic drug response in dogs. In mouse knockout studies the P-glycoprotein, *MDR1,* which is responsible for drug transport showed promise in explaining pharamacoresistent epilepsy, but has failed to show an association in human pharmacoresistent epilepsy. Conversely another study showed how genetic variation in *SCN1A* a drug target of both phenytoin and carbamazepine was associated with dosage amount in humans.

Results from studies like this one can be used to develop genetic test which can help doctors predict drug response as well as dosage amounts in epileptic patients. ROC analysis from my data gave a moderate predictive value of .74 or 0.76 depending on the exact SNP combination, which is not yet high enough to justify a clinically predictive test. Using the *SCN2A2* (762), *KCNQ3* (887) and *ABCC4* (550) SNPs as predictors of drug response will predict refractory dogs 74% of the time, but misdiagnose 40% of the responsive dogs as refractory. Currently, the benefit of using Phenobarbital over other available drugs is the decreased cost associated with treatment. As more drugs become available for canine patients, an ROC value of 0.74 may be an acceptable level of specificity as the cost of alternative drugs are roughly equal to that of Phenobarbital.

One of the genes identified (*EPOX HYD*) also implicates a third explanation, that is not unique to dogs, which is that genetic variation in drug metabolism genes may lead to faster or slower metabolism of a drug. In this case, dogs may not be resistant to the drug, but are unable to maintain sufficient levels of the drug in their bodies. Evidence for a genetic contribution to drug metabolism has been identified in both canine and human cytochrome p450 genes (Tenmizu, *et al.,*

2004; Lamba, *et al.,* 2002). Further verification of the *EPOX HYD* SNP association with Phenobarbital response in epileptic dogs would prompt research into the role this gene plays in the metabolism of Phenobarbital. *Epoxide Hydrolase 1* has been linked to metabolism of another AED, carbamazepine, in Japanese epileptic patients (Nakajima, *et al.,* 2005). A possible starting point for this research would be to compare drug plasma level clearance times against genetic variation within this gene. Increased drug clearance levels associated with the genetic variation would indicate a link to *EPOX HYD* and drug metabolism. The inverse may also be possible in that *EPOX HYD* is unable to convert vitamin K epoxide into usable Vitamin K_1, which is required for proper CYP function. This would imply that PB may be poorly metabolized due to lack of usable vitamin K, and supplementing the diet with additional vitamin K may be enough to increase the effectiveness of Phenobarbital.

The use of diet was first recognized as an effective tool to treat refractory epilepsy in the 1920s, but gained popularity in the 1990's after a true life movie about a boy whose epilepsy was effectively controlled by diet alone (Bailey, *et al.,* 2005). Children who did not gain adequate seizure control with treatment of two AEDs, but who were fed a ketogenic diet which is high in fat and low in protein and carbohydrates showed a significant decrease in seizure activity over a 3 month period (Neal, *et al.,* 2008). The exact mechanism behind this effect is not understood although some have suggested that diet may affect hepatic metabolism and microsomal enzyme activity (Anderson and Kappas, 1991; Pantuck, *et al.,* 1999). A study looking at the effects of diet on the pharmacokinetics of Phenobarbital in dogs showed a low protein and fat diet actually increased the metabolism of Phenobarbital compared to a low protein or maintenance diet (Maguire, *et al.,* 2000). This study

however did not investigate a high fat diet nor did they limit food intake in the low protein and fat dieted dogs where they reported increased food consumption, so the effects of diet in epileptic dogs is still unclear. However, a Google search for canine epilepsy diets yields several websites with testimonials as well as dietary regimes which have worked in reducing seizure activity in dogs. Future work into the mechanism behind dietary treatment of epilepsy may reveal interesting results and assist in the treatment of epilepsy.

The aim of this thesis project was to show the application of genomic tools in dogs, which will serve to stimulate larger scale studies. From these studies I have demonstrated that future studies will give a better understanding of the relationship between genes and phenotypes and the possible application of those results in a medical setting. However it is clear from these studies that steps must be taken in planning of these experiments which account for power issues as well as population structure for them to be successful. In the case of association studies, it may also be useful to begin studies focusing on one breed, and then to further narrow down regions of interest by examining multiple breeds.

References:

Abubucker, S., Martin, J., Yin, Y., Fulton, L., Yang, S-P., *et al.* (2008). The canine hookworm genome: Analysis and classification of *Ancylostoma caninum* survey sequences. *Molecular and Biochemical Parasitology*. 157: 187-192.

Anderson, K., Kappas, A. (1991). Dietary regulation of cytochrome P450. *Annu Rev Nutr.* 11: 141-167.

Bailey, E., Pfeifer, H., Theile, E. (2005). The use of diet in the treatment of epilepsy. *Epilepsy & Behavior.* 6: 4-8.

Chelini, M., Souza, N., Cortopassi, S., Felippe, E., Oliveira, C. (2006). Assessment of the physiologic stress response by quantification of fecal corticosteroids. *J Am Assoc Lab Anim Sci*. 45: 8-11.

Clauss, M., Dierenfeld, E., Bigley, K., Wang, Y., Ghebremeskel, K., et al. (2008). Fatty acid status in captive and free-ranging black rhinoceroses (Diceros bicornis). *J Anim Physiol Anim Nutr*. 92: 231-241.

Giger, T., Excoffier, L., Day, P., Champigneulle, A., Hansen, M., Powell, R., et al. (2006) Life history shapes gene expression in salmonids. *Current Biology.* 16: R281–R282.

Grove, D., Forbes, I. (1975). Increased resistance to helminth infestation in an atopic population. *Med J Aust.* 1: 336-338.

Judge, P., De waal, F. (1997). Rhesus monkey behavior under diverse population densities: coping with long-term crowding. *Anim Behav.* 54: 643-662.

Jurado, O., Clauss, M., Streich, W., Hatt, J. (2008). Irregular toothwear and longevity in captive wild ruminants: a pilot survey of necropsy reports. *J Zoo Wildl Med.* 39: 69-75.

Klein, A., Kaufman, H., Most, H. (1971). Intestinal parasites and asthma. *N Engl J Med*. 285: 179.

Lamba, J., Lin, Y., Schuetz, E., Thummel, K. (2002). Genetic contribution to variable human CYP3A mediated metabolism. *Adv Drug Deliv Rev*. 54: 1271-1294.

Maguire, P., Fettman, M., Smith, M., Greco, D., Turner, A., *et al.* (2000). Effects of diet on pharmacokinetics of Phenobarbital in healthy dogs. *JAVMA*. 217: 847-852.

Morey, DF. (1992). Size, Shape and development in the evolution of the domestic dog. *J Archeological Sci*. 33: 158-175.

Mortimer, K., Brown, A., Feary, J., Jagger, C., Lewis, S. (2006). Dose-ranging study for trials of therapeutic infection with *Necator Americanus* in humans. *Am J Trop Med Hyg.* 75: 914-920.

Moyle, M., Foster, D., McGrath, D., Brown, S., Laroche, Y., *et al.* (1994). A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *Journal of Biochemisty*. 269: 10008-10015.

Nakajima, Y., Saito, Y., Shiseki, K., Fukushima-Uesaka, H., Hasegawa, R., *et al.* (2005). Haplotype structures of EPHX1 and their effects on the metabolism of carbamazepine-10, 11-epoxide in Japanese, epileptic patients. *Eur J Clin Pharmacol.* 61: 25-34.

Neal, E., Chaffe, H., Schwartz, R., Lawson, M., Edwards, N., *et al.* (2008). The ketogenic diet for the treatment of childhood epilepsy: a randomized controlled trial. *Lancet Neurol.* 7: 500-506.

Neff, M., Robertson, K., Wong, A., Safra, N., Broman, K., *et al.* (2004). Breed distribution and history of canine mdr1-1delta, a pharacogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc Natl Acad Sci USA.* 101: 11725-11130.

Pantuck, E., Pantuck, C., Kappas, A. ….(1999). Effects of protein and carbohydrate content of diet on drug conjugation. *Clin Pharmacol Ther.* 50: 254-258.

Snell, T., Brogdon, S., Morgan, M. (2003) Gene expression profiling in ecotoxicology. *Ecotoxicology.* 12: 475-483.

Suleman, M., Wango, E., Sapolsky, R., Odongo, H., Hau, J. (2004). Physiological manifestations of stress from capture and restraint of free-ranging male African green monkeys (Cercopithecus aethiops). *J Zoo Wildl Med.* 35: 20-24.

Sutter, N., Eberle, M., Parker, H., Pullar, B., Kirkness, E., *et al.* (2004). Extensive and breed-specific linkage disequilibirium in canis familiaris. *Genome Research.* 14: 2388-2396.

Tate, S., Depondt, C., Sisodiya, S., Cavalleri, G., Schorge, S., *et al*. (2005) Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proc Natl Acad Sci USA*. 102: 5507-5512.

Turner, K. (1978). The conflicting role of parasitic infections in modulating the prevalence of asthma. *P N G Med J.* 21: 86-104.

Umina, P., Weeks, A., Kearney, M, McKechnie, S., Hoffmann, A. (2005) A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science*. 308: 691-693.

Vila, C., Savolainen, P. Maldonaldo, J., Amorim, I., Rice, J., *et al.* (1997). Multiple and ancient Origins of the domestic dog. *Science*. 276: 1687-1689.

Wordemann, M., Diaz, R., Heredia, L, Madurga, A., Espinosa, A., et al. (2008). Association of Atopy, Asthma, allergic rhinoconjuctivitis, atopic dermatitis and intestinal helminth infections in Cuban children. *Trop Med Int Health*. 13: 180-186.