Title

Investigation of Complex Neuropsychiatric Disorders in the Domestic Dog: Genome-Wide Surveys for Loci Underlying Noise Phobia and Adult-Onset Deafness in Purpose-Bred Dogs.

Author

Yokoyama, Jennifer Sachiko

Publication Date

2010

Peer reviewed|Thesis/dissertation

Investigation of Complex Neuropsychiatric Disorders in the Domestic Dog: Genome-Wide Surveys for Loci Underlying Noise Phobia and Adult-Onset Deafness in Purpose-Bred Dogs.

by

Jennifer Sachiko Yokoyama

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedications & Acknowledgements

First, I would like to dedicate **my dissertation in its entirety** to my now-passed and beloved dog, *Molly Yokoyama*. Even though they said you weren't part Border collie we all know you were. You looked like one. You acted like one. You turned on the water faucet when you were thirsty, you ate the donuts when nobody was looking, you were Dad's art critic when nobody else would be. You were a spaz into your ripe old teens. You were awesome. You were my first dog, and you were a great dog. I miss you and your leg crossing. I will always hear your tags rattling when the front door opens. I love you, doggo.

Second, and no less importantly, I would also like to dedicate this **dissertation in its entirety** to my parents, *Leila and William*. Without you, Parents, I would not have made it this far (or made it at all). You raised me to value hard work, and to enjoy the pay off when you know you've earned something for yourself. Therefore, I see nothing more appropriate than this humongous dissertation as an embodiment of that work ethic and perseverance you bestowed upon me, the solitary progeny. Perhaps it's true that I was born of my own diligent spirit, but I as a human would not have been so successful without the love and support you've always provided, unfailingly. You've let me choose my own path (funny I should choose one so parallel to both of yours…), to make my own mistakes, to learn and win and lose at life, and all these things have ultimately made me a winner. I look forward to many more food binges, shopping sprees, scientific and life discoveries that are boundless possibilities for me, the person who is lucky to be your daughter. So, thank you, Parents. For everything.

I would like to acknowledge…

Those who played a role in my formation as a budding scientist – *Louis Pruitt* (Alhambra High School), who made me "queen of the prom" for scoring highest on the hardest exam of the year –

genetics. It was you who made genetics exciting (and who made it make sense), and without your enthusiastic support I would not have chosen this field. *Dr. Bradley Davidson* (UC Berkeley), who convinced me a semester of independent research was more valuable than a lab course in genetics. You were right. *Dr. Edward T. F. Wei* (UC Berkeley), for believing in my ability to thrive in the world of "backstabbing" academic research (no lie). I owe you a soy hot cocoa. *Dr. Thomas Knopfel* (RIKEN Brain Science Institute), for letting the loud American kid take over an independent project. You took a chance hiring me and let me make your lab my second family. I will always appreciate the opportunity you gave me for both scientific and individual growth. *Dr. Junichi Nakai* (RIKEN Brain Science Institute), for mentorship and life support during my starting days and months in Japan. I will always remember sprinting down the hall making competent cells with you, and even though I don't do molecular biology anymore I will always think of myself as a molecular biologist because of your esteemed training.

At UCSF – *Drs. Ying-Hui Fu and Louis Ptacek* for a lab rotation and early support in my graduate career. *Drs. Frank Szoka and Joshua Park* for a lab rotation and outstanding mentorship throughout my time here. Josh, you are my graduate school 오빠. My thesis committee members – *Drs. Laura Bull and Laurence Tecott*—for critical insights and support throughout this epic process. Thank you especially to Laura for reading through my drafts with a fine-toothed comb – wowowowow – both laudable and a super huge help to getting this monster to finality.

The Hamilton Laboratory, a.k.a., The Millenium Falcon: Without you I would have no lab. Literally. I knew from the moment that I rotated that this was a good fit. I'm glad I went with my gut. We've had our ups and downs (in life and numbers) down there in the basement of LPPI, and I wouldn't have had it any other way. We conquered the Halloween contest (well, once), the Benet lab on the wooden alleys (go Rebel Alliance!), and the Deathstar (at least in spirit). Oh to

be the Underdog….I was lucky to be able to be ease in at noon and out before the lights went out at 2am (*Laura* – your late-night camaraderie! Fantastic!), all while wearing my graphic T's and tennies. I will miss you, and you will always be my second 'home' (particularly now that you are reorganized – yay, *Carolyn*!). I have made many lifelong friends during my time here at the HamLab, and I hope we can continue to stay in touch through the years. To highlight a few particularly relevant folks –

To *Dr. Steven P. Hamilton* – it is 100% true when I say that I COULD NOT have asked for a better mentor, period. Aside from those times when you stole all the best Starburst out of my pack, you have ***always*** had my back. I used to wonder where I'd fit in to the pack of graduate trainees that you had – brilliant Peters, prodigious Kraft, thorough Laura. I hope I've been able to become The Worker that you needed to lay the foundation for the canine project, and The Glue to help bolster the lab as a family. Thank you for letting me take on independent projects while simultaneously providing invaluable support as I stumbled along. Thank you for doing your best to meet my self-imposed deadlines. Thank you for letting me be a bit neurotic. And thank you for keeping me company in the corner by the cheese platter at all those awkward canine-feline conferences. You pushed me without being dictatorial, you motivated me with positivity, you taught me through practice – modesty, intellect, insightfulness and a sweet tooth. You have been nothing but supportive, and I hope we can continue to work together as colleagues in the future. I wish you all the success your human capacity deserves. To you I dedicate **Chapter 1**, particularly because you are the only person who will ever read it twice, and that truly means everything.

To *Dr. Melanie Lee Chang* – we had many the arduous struggles and raucous times in the so-called "east wing." I hope I could do your research-baby justice. I think we started something pretty good; here's hoping someone else will come along soon to do it even more better. I dedicate **Chapter 2** to you, because we finally showed that Barbies ≠ Borders. Yee-haw! To

Solo: you are the birth of the noise phobia project, and so to you I dedicate **Chapter 3**. You are crazy and stoic, and I am certain if Melanie had to choose between you and Misha she would pick you. To *Fly*: our eternal positive control, your memory lives on in all three genotyping platforms whose data I would not have been able to merge if not for you. You are crazy in that tongue-wagging goofy way (not that Solo way), and thank goodness or else you wouldn't make a very good control. I dedicate to you my **Exit Seminar**. To *Jett*: thank you for always peeing on my feet when I see you. You are a slender stealthy machine and I wish you lived closer so I could play with you or steal you away from your family (just temporarily). When I look into your eyes I see my Molly [the dog], and sheer bliss. To you I dedicate **Chapter 5**, because you are the future.

To *Dr. Jeffrey B. "Kraftomatic" Kraft, Jr.,* for feeding me gummy worms before going to play basketball, sharing your wife (in a friend sort of way), and for providing a huge dose of fun and sanity in a wealth of anything but. And most of all, thank you for vouching for me when I was long forgotten…I literally would not be here at UCSF without your support and dedication, and for that I will always have a strong love-hate relationship with you. ☺ To *Molly Kraft* – thank you for taking me in to your warm circle of knitting and friendship. You are wonderful, Mama Kraft, and one of these days we will have a cupcake Face-Off.

To *Dr. Stanley I. Shyn* – for keeping me up-to-date on news, politics, scientific gossip, and SyFy. Our pontifications in real life and over the interweb have been invaluable, and I look forward to hearing about how awesome your life is outside of academia. I wish I could've kept you as a bay-mate, work just isn't the same when you can't wheel around to shoot the shoot when you get sick of staring at Manhattan plots or PLINK won't work and Shaun isn't emailing back….

To *Dr. Holly A. Garriock* for being a true womanly role model and inspiration. Alas, I didn't follow your advice and I'm going to do a post-doc, but I will do my best to model other aspects of

my life after your amazingness. If only we'd realized the joy of knitting coffee breaks sooner…instead I will just have to frequently visit your loving abode and offer play breaks from life. To *William* – thank you for teaching me how to not be afraid of babies. You are three-years-old, and soon you will be taller than me. Just remember your roots - crawling on the floor of the HamLab….Say hi to your lil' broski for me. To *Rob* – thank you for always telling it like it is. You are a breath of fresh air.

And yes, *Jody Williams*, it is still my goal to steal you away from Steve. One day.

The Kwok Laboratory, a.k.a., The Deathstar: Thank you for letting me be an honorary member of your noble establishment, and for letting me steal some of your equipment and/or lab members from time to time. In particular – *Ernest Lam*, thank you for taking up the reins of the canine sled with me. WOOF-Seq'ing answers with you has been such a toy joy and novelty. For so long a lone-wolf, when a border-line experiment spitz in your face or beagle-ing [beguiling] findings terrier down your spirit, it's so less shar-peinful having someone else to hound. We've packed a lot into the last few months, chowed down a lot of sequence, and I think we've barked up the right tree and retrieved some true show winners to shepherd in the deafness GWAS. You're in the ring now, and so to [Shi]Tzu I dedicate **Chapter 4**; I hope we find a nice pointed ending to our doggone story. I know you'd setter for nothing less, even in a pinsch.

Thank you to our collaborators outside of UCSF—*Dr. Mark W. Neff* (VARI and TGen) for generously offering to share samples for the noise phobia project, and for providing me the opportunity to work on what has turned into the most exciting chapter of my thesis. You've extended mentorship beyond that of a collaborator, and I've greatly valued our conversations on research, academia, careers, TV, and life. May our next work venture not be in a field full of penetrating dust. Also, *Alison Ruhe* – you are such a trooper. Amazing. *Dr. Hannes Lohi* (Univ.

Helsinki) – your group is astounding at sample collection and collaboration, and I hope you and Steve are able to continue on the quest for loci in canine anxiety. In particular, to *Dr. Katrina Tiira* – the noise phobia project would not be where it is now without your hard work translating and administering our questionnaire in Finnish – thank you! Also, *Dr. Christophe Hitte* (Univ. Renne), for exploring our data with your new computational gadgetry. We are always happy to be your pilot data. And of course, *Dr. Karen Overall* (Univ. Pennsylvania), whose work forms the fundamental basis for utilizing the dog as a model for studying anxiety behaviors analogous to human psychiatric disorders: your work in behavioral veterinary medicine is invaluable.

And finally…

I want to acknowledge all the people who have helped me survive the life experience that is graduate school. In particular, my classmates and close friends – *Dr. Stacy L. Musone, Dr. Melissa A. Calton*, *Howard Horng and Jason Fernandes*. Thank you for providing vent-ilation (ha ha) and much needed breaks from science…buffets, New Kids, and otherwise. I am Happy to have you as my CORE, for now and always. Missy – we did it! Time for a cruise. How – your emails are scary for the work place but always a reminder of your warmth and friendship. And yes, Science Day was our best grad school accomplishment. Jasonation – thanks for all the heart-to-hearts, and for always being super awesome. I look forward to hearing more from the Bell Ringer. A special thanks especially to my *cy for being an amazing roomie and stupendous all-around friend and person. I love that we can be late to doing nothing together, and it's still fantastically fun. You will always be an honorary YokoChild (and you're probably the favorite, too). Also, warm thanks to *Geoffrey Quan* – thank you for making sure I was able to stay safe, sane and survive (in that order). And of course to *William Prendergast* for never-relenting story telling and honesty, and for reminding me that one person can always make a difference.

To *Patricia Seo*. Words cannot even describe how lucky I am to have a BFF as wonderful as you. You are my fighting spirit, my pushy pants, my defender and my Best Friend Forever. You give me the strength to stand up for myself, and to accept the outcome. No other person except Patricia Seo would drive all the way to the middle of Oregon to slog around boxes and dog bloods just to help out her friend, Jen. No other person except Patricia Seo would spend weekends in lab labeling tubes, if only to help check off one item from Jen's headache list. People won't even do these things for money, and yet you do them for nothing but good will and appreciation. You are irreplaceable. I hope one day soon you will truly appreciate yourself as much as I appreciate you as a person on this earth and as a formidable presence in my life. To you, I dedicate **half of my Ph.D.**. Perhaps the P, for Patty.

And lastly and mostly—

To my soon-to-be-husband, *Daniel W. Sirkis*. To my Voice of Reason, without you I certainly would not be here, nor in as happy or healthy a state (well, healthy in spirit). Not only do I suffer less overeating pains these days, but I also think I'm happier for it (most of the time). Nobody but you has ever been able to convince me that I am a person worthy of happiness, and as simple as that sounds, the journey to this realization has been anything but simple or easy. We have a nice and long adventure ahead of us, and I can't wait. You remind me to take a step away from the dessert table as easily as you remind me to take a step back and see things in perspective; and not in My perspective, but in Voice of Reason perspective. How lucky for me that I've found a counterbalance to my ill-enacted wordages and penchant for overdoing it. You know what I'm thinking when I don't, and I appreciate you sharing my thoughts (and yours) with me when I find myself with none. Supplement fromage, you are all the cheddars I will ever need and more. Thank you, d.wes, for being you. To you, I dedicate **half of my Ph.D.**, the D - for Dan. You make me happy (is that the missing 'h' of this equation?) and I love you, and soon you will be stuck with me FOREVER. <3

Original manuscript citations and co-authorship:

Portions of the text of this dissertation are reprints of material as it appears in the following published or submitted manuscripts:

1. (Chapter 2, section 1) Yokoyama JS, Erdman CA, Hamilton SP. "Array-based whole-genome survey of dog saliva DNA yields high quality SNP data." (2010) *PLoS One*. 5(5):e10809.
2. (Chapter 2, section 2) Chang ML, Yokoyama JS, Branson N, Dyer DJ, Hitte C, Overall KL, and Hamilton SP. (2009) "Intrabreed Stratification Related to Divergent Selection Regimes in Purebred Dogs May Affect the Interpretation of Genetic Association Studies." *Journal of Heredity*. 100(Supplement 1):S28-S36.
3. (Chapter 3) Yokoyama JS, Chang ML, Tiira KA, Branson N, Dyer DJ, Juarbe-Diaz S, Ruhe AL, Robertson KR, Neff MW, Lohi H, Overall KL, and Hamilton SP. "Genome-wide association study identifies candidate loci in canine noise phobia." Submitted to *Genes, Brain and Behavior,* November 24, 2010.
4. (Chapter 4) Yokoyama JS, Lam ET, Erdman CA, Corneveaux J, Ruhe AL, Robertson KR, Chang ML, Overall KL, Huentelman MJ, Lohi H, Hamilton SP and Neff MW. "Genome-wide association study of adult-onset deafness in herding Border collies." *(In preparation)*

Statement of work performed by Jennifer S. Yokoyama with regard to material included in this dissertation from the above published/submitted manuscripts by advisor, Dr. Steven P. Hamilton: All of the text included in this dissertation from the above published/submitted works were written and/or based on work primarily performed by Ms. Yokoyama. Specifically: (1) Ms. Yokoyama directed and supervised this study, and was directly responsible for sample recruitment and ascertainment, data generation and QC, and performed all of the analyses associated with this work. She also played a primary role in the preparation of this manuscript, with modifications based on comments from our laboratory research assistant, Carolyn Erdman, and myself. (2) The coauthor, Dr. Melanie L. Chang, was a post-doctoral fellow in my laboratory and directed and supervised the research that forms the basis for this publication. Ms. Yokoyama played a critical role in sample ascertainment, a primary role in data generation, and was solely responsible for the genotype QC and data set creation, clustering analysis, and GWAS simulations. She also prepared the text for the Methods & Results sections for these analyses as

they appear in the published work. (3) Ms. Yokoyama was directly involved in sample and phenotype ascertainment, data generation and genotype QC for this work. She formulated the final study design and carried out all of the statistical analyses for this study. She also prepared the manuscript, and her and I made modifications to her primary drafts based on comments from the co-authors prior to submission. (4) Ms. Yokoyama was involved in sample ascertainment, genotype generation and QC, and played critical roles in study design, statistical analysis for the primary GWAS analysis, and sample preparation for next-generation sequencing. She is currently preparing this manuscript with input from Ernest Lam (graduate student in the Kwok lab) on the next-generation sequencing portions. Further descriptions of the roles played by co-authors for all three studies are included within footnotes for each respective section.

Research funding and support:

A big THANK YOU to the numerous owners and breeders contributing genetic material for this work. J.S.Y was supported by a Predoctoral Training Consortium in Affective Science Research Fellowship (NIMH, T32 MH20006-12). This work was also supported by the National Institute of Mental Health (R21 MH084149), the American Kennel Club (ACORN Award 850-A), the McKnight Foundation, a UCSF bridge award, the Hellman Family Fund (S.P.H.), and the Defense Advanced Research Projects Agency (DARPA FY-06-0028 / 52731-LS-DRP to K.L.O.). Distribution Statement "A" Approved for Public Release, Distribution Unlimited. H.L. was supported by the ERC starting grant (260997), the Sigrid Juselius Foundation, Biocentrum Helsinki, the Academy of Finland, the Jane and Aatos Erkko Foundation, the University of Helsinki Research Funds and the Finnish Cultural Foundation (K.T.). H.L. is also a member of the Nordic Center of Excellence in Disease Genetics. M.W.N. would also like to acknowledge the Veterinary Genetics Laboratory of the University of California at Davis for intramural support. The authors also wish to thank Jonathan Woo, Melanie Dela Cruz and Pamela Doo (Genomics Core Facility, University of California, San Francisco) for genotyping, and Daria Vasiltsov, Cecilia Becerril and Ellen Maclean of DNA Genotek for beta-testing kits and technical support on the Oragene·ANIMAL kits and sample collection.

Foreword

You notice an odd tightening in your chest. You are unable to identify the problem, but you know something is wrong. Eyes widening, you grit your teeth in hardened anticipation of the uncontrollable. Your breath comes in quick gasps as you feel the grip of fear close tightly around you. Faster, faster, your heart attempts to beat itself out of your ribcage, to run away from the terror closing in. You want to run, too, and yet you find yourself frozen, unable to move, unable to breathe, unable to think. Paralyzed by the physiological culmination of your unbridled panic, you crumple into yourself, clutching at your stomach for comfort. Palms moist, you wipe the sweat as it runs into your eyes. You try to open your mouth to scream yet find only faint whimpers pushing to escape from pursed lips, gasping for air – for fortitude – for the end.

Jennifer Sachiko Yokoyama

December 10, 2010

Abstract

Investigation of Complex Neuropsychiatric Disorders in the Domestic Dog: Genome-Wide Surveys for Loci Underlying Noise Phobia and Adult-Onset Deafness in Purpose-Bred Dogs.

Jennifer Sachiko Yokoyama

The domestic dog offers a novel and potentially powerful genetic model for studies of complex neuropsychiatric disease, including maladaptive behavioral conditions analogous to human psychiatric disorders. Dogs provide two critical advantages that facilitate such studies: a) potentially prominent genetic homogeneity due to the foundations of individual breeds; and b) naturally occurring behavioral disorders with clinical features similar to a number of human anxiety disorders (e.g., separation anxiety, obsessive-compulsive disorder, specific phobias). After establishing the validity of the dog as a genetic model for studying neuropsychiatric disorders relevant to human conditions, I will present work conducted in the context of our laboratory's on-going project in canine behavioral genetics. Specifically, we have conducted the first genetic study of the canine anxiety disorder noise phobia, and the first genome-wide association study for adult-onset deafness in dogs. We have identified multiple candidate regions in Border collies that may confer risk for noise phobia, including loci on *Canis familiaris* chromosome (CFA) 5, CFA8 and two loci on CFA10 that all appear to demonstrate epistatic interactions. We have also identified a strong association region on CFA6 for adult-onset deafness in Border collies. Targeted next-generation sequencing of the CFA6 deafness locus identified multiple candidate sequence variants, including non-synonymous SNPs in putative genes *USP31* and *RBBP6*. Together, these findings implicate new potential risk loci for troubling disorders, with implications for further research in larger samples of dogs of different breeds. Additionally, our results highlight the strong potential for studies of complex neuropsychiatric disease in the dog that may be directly relevant for analogous disorders in human populations, providing new opportunities for scientific discovery.

TABLE OF CONTENTS

	PAGE
COPYRIGHT	**ii**
DEDICATIONS & ACKNOWLEDGEMENTS	**iii**
ABSTRACT	**xiii**
TABLE OF CONTENTS	**xiv**
LIST OF TABLES	**xviii**
LIST OF FIGURES	**xx**
CHAPTER 1: INTRODUCTION TO CANINE BEHAVIORAL GENETICS	**1**
MAN'S BEST FRIEND IN GENETICS	2
THE DOG AS A NATURAL MODEL OF BEHAVIOR	7
GENOMICS OF CANIS FAMILIARIS	13
GENE MAPPING IN THE DOG: THEN AND NOW	18
CANINES PROVIDE SANITY IN THE MESS OF HUMAN PSYCHOSIS	26
REFERENCES	29
CHAPTER 2: GENETIC INVESTIGATIONS OF THE DOMESTIC DOG	**41**
2.1. GENOTYPING DOGGIE DROOL: IT MAY NOT BE AS RUFF AS YOU THINK	
2.1.1. Introduction	42
2.1.2. Materials and Methods	
2.1.2.a. Samples	44
2.1.2.b. DNA extraction	45
2.1.2.c. Quantification of DNA	46

2.1.2.d. Genotyping 46

2.1.2.e. CNVs 46

2.1.3. Results

2.1.3.a. Saliva sample collection 47

2.1.3.b. DNA yield 47

2.1.3.c. Genotyping 49

2.1.3.d. CNVs 55

2.1.4. Discussion 56

2.2. WITHIN-BREED SUBSTRUCTURE COMPLICATES GENOME-WIDE ASSOCIATION STUDIES

2.2.1. *Introduction* 61

2.2.2. Materials and Methods

2.2.2.a. Sample recruitment, collection, and data generation 62

2.2.2.b. Cluster and genetic distance analyses 63

2.2.2.c. Pedigree analysis 64

2.2.2.d. Simulation studies 65

2.2.3. Results 65

2.2.3.a. Results of cluster and genetic distance analyses 66

2.2.3.b. Simulations 68

2.2.4. Discussion 70

2.3. LABORATORY METHODOLOGY FOR TWO GENOME-WIDE ASSOCIATION STUDIES 72

2.3.1. Genotype Generation 73

2.3.2. Genotype Quality Control 74

2.4. REFERENCES 75

CHAPTER 3: GENOME-WIDE ASSOCIATION STUDY IDENTIFIES CANDIDATE LOCI FOR THE CANINE ANXIETY PHENOTYPE NOISE PHOBIA

3.1. INTRODUCTION 78

3.2. MATERIALS AND METHODS

3.2.1. Samples 80

3.2.2. Phenotypic assessment 82

3.2.3. Association analysis 83

3.2.4. Multi-breed visualization analysis 87

3.3. RESULTS

3.3.1. Primary analysis in Border collies 87

3.3.2. Secondary analysis in multiple breeds 98

3.4. DISCUSSION 103

3.5. REFERENCES 111

CHAPTER 4: GENOME-WIDE ASSOCIATION STUDY OF ADULT-ONSET DEAFNESS IN HERDING BORDER COLLIES IMPLICATES HUMAN DEAFNESS LOCUS

4.1. INTRODUCTION 116

4.2. MATERIALS AND METHODS

4.2.1. Samples 119

4.2.2. Phenotypes 120

4.2.3. Next-generation Sequencing

4.2.3.a. Sample preparation 121

4.2.3.b. Target-capture 121

4.2.4. Analysis

4.2.4.a. GWAS 122

4.2.4.b. NGS variant calling and assessment 123

4.3. RESULTS

4.3.1. Samples 124

4.3.2. Genome-wide Association Study 124

4.3.3. Fine-mapping 128

4.4. DISCUSSION 138

4.5. REFERENCES 143

CHAPTER 5: CONCLUDING REMARKS AND FUTURE DIRECTIONS

5.1. SUMMARY OF RESEARCH FINDINGS 148

5.2. WHO LET THE DOGS OUT: THE FUTURE OF CANINE GENETICS 149

PUBLISHING AGREEMENT **151**

LIST OF TABLES

PAGE

CHAPTER 2

TABLE 2.1 CONCENTRATION, PURITY AND CONCORDANCE OF SALIVA- VERSUS BLOOD-EXTRACTED DNA SAMPLES. 47

TABLE 2.2 INDIVIDUAL SALIVA AND COMPARISON BLOOD SAMPLE STATISTICS FOR DNA EXTRACTION AND GENOTYPING COMPARED TO PUBLISHED DATA FOR HUMAN SALIVA AND BLOOD, PLUS MANUFACTURER'S REPORT. 49

TABLE 2.3 GENOTYPING STATISTICS FOR SALIVA- AND BLOOD-EXTRACTED DNA SAMPLES BEFORE AND AFTER MARKER QC. 50

TABLE 2.4 MEAN GENOTYPING STATISTICS FOR DISCORDANTLY CALLED MARKERS COMPARED TO FULL MARKER SET. 54

TABLE 2.5 REPLICATE STATISTICS. 59

TABLE 2.6 RESULTS OF 100 SIMULATIONS OF GWAS OF APPROXIMATELY 53,000 AUTOSOMAL SNPS IN A TOTAL OF 48 UNRELATED BOC. 69

CHAPTER 3

TABLE 3.1 SAMPLE DEMOGRAPHICS. 81

TABLE 3.2 TEST FOR BALANCED IBS OF CONTROLS AND CASES IN BOC REPLICATION SAMPLE. 85

TABLE 3.3 SAMPLES AND PSEUDOHERITABILITY ESTIMATES. 86

TABLE 3.4 NOISE PHOBIA IN BOC. 89

TABLE 3.5 GWAS OF NOISE PHOBIA IN BOC WITH STRICTER CONTROL CRITERION. 91

TABLE 3.6 TOP 25 RESULTS FROM FINE-MAPPING/REPLICATION. 92

TABLE 3.7 TOP ASSOCIATIONS FOR EACH REPLICATION/FINE-MAPPING CHROMOSOMAL REGION. 93

TABLE 3.8 TOP 25 HITS FOR 3-SNP HAPLOTYPIC LOGISTIC REGRESSION WITH 1 MDS COVARIATE. 94

TABLE 3.9 FINDINGS FOR REPLICATION/FINE-MAPPING WITH STRICTER CONTROL CRITERION. 98

TABLE 3.10 TOP 25 RESULTS FROM META-ANALYSIS. 101

CHAPTER 4

TABLE 4.1 TOP 25 RANKED FINDINGS FROM ANALYSIS FOR ADULT-ONSET DEAFNESS IN BOC. 126

TABLE 4.2 TOP TEN RANKED FINDINGS FOR ALLELIC ASSOCIATION WITH 1,000,000 PERMUTATIONS IN PLINK. 127

TABLE 4.3 NGS STATISTICS. 132

TABLE 4.4 VARIANT-CALLING STATISTICS. 133

TABLE 4.5 BREAKDOWN OF SNPS HOMOZYGOUS IN CASES AND NOT IN CONTROLS. 134

TABLE 4.6 EXONIC VARIANTS FOR DEAFNESS ON CFA6. 135

TABLE 4.7 *GRIN2A* SSSNP CODON USAGE. 138

LIST OF FIGURES

PAGE

CHAPTER 2

FIGURE 2.1 PLOT OF P10 GENCALL SCORE VERSUS CALL RATE AFTER QC FOR SALIVA- AND BLOOD-EXTRACTED DNA SAMPLES COMPARED TO FULL SET OF GENOTYPED SAMPLES. 52

FIGURE 2.2 CLUSTER PLOTS FOR SELECT SNPS. 53

FIGURE 2.3 ALLELE DISTRIBUTIONS FOR DISCORDANT MARKERS COMPARED TO FULL MARKER SET. 55

FIGURE 2.4 MOLECULAR EVALUATION OF PUTATIVE CNV REGION ON CFA23. 56

FIGURE 2.5 FOUR-BREED *STRUCTURE* RESULTS. 66

FIGURE 2.6 HIERARCHICAL GROUPING BASED ON THE PAIRWISE MATRIX OF GENETIC IDENTITY BY STATE (IBS).

a. Dendrogram of 108 dogs of four breeds constructed by pairwise genetic distance analysis for K = 4. *67*

b. Dendrogram of 108 dogs based on the distance matrix of IBS set to K = 8 hierarchical clustering. *68*

CHAPTER 3

FIGURE 3.1 MDS 1 X MDS 2 PLOTTED FOR ALL UNRELATED BOC. 84

FIGURE 3.2 MANHATTAN PLOT OF BOC NOISE PHOBIA ANALYSIS IN EMMAX. 88

FIGURE 3.3 MANHATTAN PLOTS OF BOC NOISE PHOBIA ANALYSIS WITH STRICTER CONTROL CRITERION. 90

FIGURE 3.4 MANHATTAN PLOT OF THE 58MB REGION ON CFA10 AND THE 22MB REGION ON CFA8 FOR BOC NOISE PHOBIA REPLICATION/FINE-MAPPING. 96

FIGURE 3.5 MDS PLOTS FOR SAMPLES FROM ALL BREEDS ANALYZED FOR NOISE PHOBIA.
MDS1 x MDS2 *99*
MDS3 x MDS4 *100*
FIGURE 3.6 VISUALIZATION OF TOP RESULTS FROM FIVE DIFFERENT BREED GWAS'S FOR NOISE PHOBIA.
a. CFA5 *102*
b. CFA8 *102*
c. Whole genome *103*
FIGURE 3.7 EPISTATIC INTERACTIONS BETWEEN THREE CANINE CHROMOSOMES. 106
FIGURE 3.8 Q-Q PLOTS FOR BOC ANALYSES. 110

CHAPTER 4

FIGURE 4.1 MULTI-DIMENSIONAL SCALING (MDS) VECTOR PLOTS OF BOC USED FOR DEAFNESS ANALYSIS. 122
FIGURE 4.2 Q-Q PLOT OF GWAS ANALYSIS FOR ADULT-ONSET DEAFNESS IN BOC. 123
FIGURE 4.3 MANHATTAN PLOT OF GWAS FOR ADULT-ONSET DEAFNESS. 125
FIGURE 4.4 MANHATTAN PLOT OF P_{GENOME}. 127
FIGURE 4.5 HAPLOTYPES IN CFA6 HIT REGION ~25MB. 129
FIGURE 4.6 HAPLOTYPES IN CFA6 HIT REGION ~35MB. 130

Chapter 1: Introduction to Canine Behavioral Genetics

This dissertation seeks to explore the relevance of using the domestic dog as a model for investigating the genetic basis of mental illness, particularly as it relates to anxiety disorders in human psychiatry. I first examine the validity of the dog as a model for studying behavior, which includes review of recent advances in canine genetics as they relate to behavior, and the implications of these findings for the research field. I next describe our research sample, summarizing the behavioral phenotypes we are studying and the tools used to measure them, and outlining the genotyping methods we use to characterize these samples genetically. I also discuss our examination of within-breed stratification in Border collies and evaluate the use of saliva-extracted DNA samples on the latest array-based high-throughput genotyping technologies. I then describe the strategy and methods used to statistically determine genetic loci underlying noise phobia in herding and working breed dogs via genome-wide association study (GWAS). After the discussion of noise phobia, I also describe similar analytical methods used to determine genetic loci underlying adult-onset deafness in Border collies. The investigation of deafness includes deep sequencing of our top candidate regions for causative mutations. I conclude with further discussion of the clinical relevance, implications, and future directions of this body of work in the context of canine behavioral genetics research, as well as the field of human psychiatric genetics.

In this section, I will outline the development of the dog as a model organism for scientific discovery, evaluating the dog's validity as a model for studying behavior in a genetic manner and briefly summarizing recent advances in canine genetics research, with emphasis placed on studies regarding behavioral pathology.

Then, one day, graduate students will

enthusiastically espouse the "awesome power of dog genetics."

–MW Neff [1]

1.1. Man's Best Friend in Genetics

Canis familiaris has been man's best friend since the domestication of the dog, estimated at 15,000-100,000 years ago [2]. As early as 1872, scientists such as Charles Darwin noted the similarities with which the canine's behavior embodied many of the same attributes as their human companions. In his work, The Expression of the Emotions in Man and Animals [3], Darwin drew parallels between humans and their animal compatriots across many different classes of emotional expression, observations that have lost nothing in their impact and utility after over nearly a century and a half of research. Notably, he remarked this about fear:

> The word 'fear' seems to be derived from what is sudden and dangerous…As fear increases into an agony of terror, we behold, as under all violent emotions, diversified results. The heart beats wildly, or may fail to act and faintness ensue; there is a death-like pallor; the breathing is laboured; the wings of the nostrils are wildly dilated…As fear rises to an extreme pitch, the dreadful scream of terror is heard. [3]

Some 80 years later, John L. Fuller of the Jackson Memorial Laboratory in Bar Harbor, Maine, began publishing articles on the field of "behavior genetics" in animals such as the dog and mouse [4]. The purpose of these studies, performed in the context of the long-range project later known as "Genetics and the Social Behavior of Mammals," was to assess the factor of heredity in behavior [4]. In brief, members from five dog breeds were intensely studied by the group: basenjis, beagles, American cocker spaniels, Shetland sheep dogs, and wire-haired fox terriers. Numerous forms of observational experiments were carried out, including: similarity/difference

observations between breeds, observations of behavioral development, cross-fostered puppies (between different breeds), and home versus kennel rearing. In addition, Fuller and colleagues carried out a cross between two very distinct breeds: basenjis and cocker spaniels. They created a three-generation pedigree including backcrosses to compare inheritance patterns of different physical and behavioral traits, as well as assess variation within-breeds, hybrids and litters [4]. Among many other observations, the researchers concluded:

> In general, the results show that heredity is an important quantitative determiner of behavior in dogs and that genetic differences in behavior can be as reliably measured and analyzed as can hereditary differences in physical size….Furthermore, there are relatively few behavioral traits for which any breed is actually homozygous. Even within the restricted samples chosen for this experiment there was a great deal of individual genetic variability [4].

Similar to how strains of inbred mice could be evaluated for naturally occurring variation between lines (here "naturally" refers to there being no genetic, physical or pharmacologic intervention taken to bring about a particular characteristic or behavior), it makes sense that researchers might also look to other species that represent great phenotypic diversity in the context of methodical breeding practices. The domestic dog fits such a model. The dog is not only used for other laboratory research (and thus bred in a systematic fashion), but in society has been bred very methodically since the creation of pure breeds starting in the Victorian era and continuing over the last 150 years [2, 5]. Registration in breed clubs such as the American Kennel Club requires that a "pure" bred dog have parents who are both registered. Purebred dogs are thus the product of the circular requirement that all dogs' parents must have parents from the same, often closed, breed group [6]. The "popular sire effect" in the dog is very common, where a single stud is used in multiple matings in order to pass on favorable physical and behavioral attributes to future generations of pups. This may be problematic in breeds derived from a small number of

founders, introducing reduced genetic diversity within the breed. This selective breeding within a closed group thus propagates common genetic lines through many generations of dogs. In this context, then, the pure breeds represent something less clonal than a pure genetic mouse strain – where each rodent is an exact genetic match to its strain-mates – but clearly provides a more simplified (homogeneous) genetic background compared to, for example, human populations, on which to study gene contributions to phenotypes of interest.

Early behavior genetics research in the dog during the '60s and '70s focused mainly on the thorough investigation of a line of "nervous" pointers starting in the Dykman laboratory at the University of Arkansas, with research carried out predominantly by Murphree [7-12; reviewed in 20]. In sum, a line of pathologically "nervous" pointer dogs was developed through selective breeding, and compared to a control line of the same breed. Most notably, the nervous dogs demonstrated severe timidity and fearfulness (freezing/immobility bordering on catatonia) towards humans, but not other dogs. Because of this response, Dykman proposed that the nervous pointer line could be a model for anthropophobia (interpersonal relation phobia, or social phobia) [13], though whether or not this was relevant given the distinction between social relationships between dogs and dogs (versus dogs and people) remains in question.

Nevertheless, research on the nervous pointer lines continued over the next couple decades, primarily by Uhde and colleagues at the Unit of Anxiety and Affective Disorders at the National Institute of Mental Health (NIMH), who characterized different biological attributes of these dogs [14-16]. For example, Uhde, *et al.* found that nervous dogs demonstrated lower body weights, lower weight/height body ratios, and lower insulin growth factor-1 (*IGF-1*) serum levels as compared to normal-behaving controls [15]. Interestingly, it was also discovered that approximately 75% of the line of nervous pointers also suffered from bilateral deafness (demonstrated by complete absence of brain stem auditory evoked response), though it appeared

that nervous dogs still responded the same to fear-invoking stimuli (i.e. human interaction), regardless of hearing status [15]. Incidentally, the most recent publication on the nervous pointer line is regarding its use as potential model for studying progressive juvenile hereditary deafness and neuronal retrograde degeneration [19].

In addition to biological characterization of the nervous pointer lines, researchers also attempted pharmacological intervention [17, 18]. Importantly, Tancer and colleagues found that, although three nervous pointers showed marked improvement to short-term treatment with the antipanic medication imipramine HCl and not placebo, chronic administration did not modify any abnormal behavior in the affected dogs [17]. This emphasized the importance of thorough evaluation of potential models of human anxiety disorders and how results from such studies need to be taken in the context of the model under study (see section 1.2 for a discussion of 'validity' in animal models of anxiety). Concurrently, work in the dog was also progressing in another behavioral phenotype, narcolepsy, which will be discussed further in the gene mapping section (1.3) below.

In the 1980s, groups also sought further information on the biological foundation of behavior and its heredity [20]. In their review of "Animal Behavior Genetics," Wimer and Wimer bring up a point that is as relevant now as it was then: "Because of the complex nature of social systems, it has not been feasible to subject them to genetic analysis; instead, the approach has been to manipulate genotype and study the effects on social systems" [20]. With regard to animal models of human disease and behavior disorders, they note that "sometimes major gene effects on behavior are discovered by accident, then subsequently exploited in the search for mechanisms….There are also models in which animals have been selectively bred for extreme behavioral traits, such as…the 'nervous pointer dogs'" [20]. Again, Wimer and Wimer make observations that are still applicable today. Regarding the nervous pointers, the authors make an interesting point, observing that though the model existed, it was still not clear what human

behavioral disorder the animal phenotype parallels [20], though noting the primary investigators' suggestion that the nervous phenotype observed may be similar to phobia in humans [13]. Wimer and Wimer conclude that there is great potential for establishing brain-behavior associations in animal models, and for discovering the impact that genes can have in living systems existing in the real world [20].

While earlier studies focused primarily on understanding behaviors intrinsic to specific lines or breeds of dogs, the year 1990 brought a move into more clinical settings and exploration of treatment options. This was a logical next step in the field given the interest in pathological behavior in the dog—particularly involving fears and phobias—as well as obsessive-compulsive disorder-like behaviors described in the nervous pointers and other dogs [7, 21-24]. For example, the study by Tancer, *et al.* sought to investigate the behavioral effects of chronic treatment of the tricyclic antidepressant (TCA) imipramine in the genetically nervous pointers [17]. In addition, another TCA, clomipramine, was investigated by many groups for its effects on compulsivity and other anxiety behaviors. Importantly, Moon-Fanelli and Dodman found that within 1-12 weeks, 75% of dogs remaining in their study of compulsive tail chasing demonstrated a 75% or greater reduction in tail-chasing behavior [25]. Similar clinical results were also seen by Seksel and Linderman [26] and Overall and Dunham [27]. Interestingly, the use of clomipramine for canine anxiety disorders was found to be efficacious enough to allow its manufacturer, Novartis, to develop a canine-specific formulation, "Clomicalm®," which was approved by the Food and Drug Administration in 1998 (NADA #141-120). The interest in treatment prospects for canines highlights the desire of the veterinary field to develop better treatments for thoroughly characterized anxiety disorders commonly encountered in the clinic [28]. As such, investigations of pharmacological interventions have continued into the 21st century [26, 27, 29, 30].

The ability to use antidepressant medications commonly used in humans for treatment of similar anxiety disorders in dogs supports the fundamental hypothesis that there is at least some underlying neurobiology that is shared between the two species in the systems affected by these disorders. Here we thus see a confluence of scientific motivations towards common ends; by the veterinary community to identify effective pharmacological treatments for anxiety disorders commonly seen in the clinic, and for researchers investigating the neurobiological basis of psychiatric disease in humans. In light of the latter, the NIMH held a special workshop to discuss the development of animal models for anxiety disorders [31]. In addition to evaluating existing and new animal models and study approaches, the workshop also sought to examine how these models relate to clinical anxiety symptoms and syndromes and how they might impact the research field. In summary, they concluded that it is unlikely that researchers will be able to develop a comprehensive animal model that accurately reflects the relative influences of factors contributing to human neuropsychiatric disease. However, ample models exist and can be better developed, defined, and extended to improve research in this field and more effectively model the combinatorial factors that contribute to clinical disease. Of note for our discussion, the group pointed out that the dog may be an important naturalistic model for determining genetic susceptibility to certain discrete anxiety syndromes that demonstrate unique behavioral, epidemiological and treatment response profiles, suggesting different underlying neurobiological etiology [31]. These comments provide a good transition point from which we will further emphasize the strengths of the dog as a model for studying behavior, and in particular in the genetic study of pathological, anxiety-related disorders.

1.2. The Dog as a Natural Model of Behavior

In the article, “Natural animal models of human psychiatric conditions: assessment of mechanism and validity” [32], Karen Overall provides a summary of the classic animal model for human psychiatric conditions—rodents—and notes how the behavior of such prey species may not be an

accurate model for studying anxiety in humans. Specifically, she notes that the vigilance and avoidance demonstrated by rodents is an advantageous behavioral attribute in prey species that must be weary of their environment in order to avoid risky situations that may result in death. However, these normal types of avoidant behaviors would be seen as highly maladaptive in social species like humans. This fundamental difference brings in to question the validity of studying maladaptive anxiety disorders in rodent models – i.e., are behaviors that are adaptive in one species but maladaptive in another representative of the same neurobiological phenomenon?

This point is highlighted by the context within which many researchers obtain their phenotypes; many of the behavioral assessments used to measure behavior in rodent animal models rely on paradigms that incorporate measures of innate or learned responses to aversive situations in normal animals [31]. Though these types of studies can be very useful for pharmacological investigations, it may be inaccurate to extrapolate from this mode of behavioral measurement of *adaptive behavior* in rodents to *maladaptive disorders* in humans. A classic example of this may be seen in the commonly used testing paradigm for anxiety, the open field test (OFT; first described by Hall [33]), in which a rodent is placed in one corner of a novel, enclosed, brightly-lit open field staging area and anxiety-induced locomotor and exploratory behaviors are monitored for five minutes. The extent to which the test subject, for example, avoids the open center of the test arena (versus exploring the novel environment) is evaluated as a measure of the level of anxiety in the rodent. Although the degree to which the test subject will spend time in the brightly-lit open space may be seen as a measure of the level of anxiety possessed by the test subject (where more anxious individuals spend less time there), this is not necessarily an accurate model of maladaptive behavior. Rather, this type of anxiety may be seen as an *adaptive* behavior for survival by the test subject, whose innate fear of bright open spaces is justified given the threat to its safety posed by an environment that would promote exposure to predators. Therefore, although anxiolytics may indeed demonstrate “anti-anxiety” effects in rodents treated with them

by increasing exploratory behavior in the open space, the behavior change in itself may not be wholly demonstrative of a socially relevant behavior modification. Further, the type of generalizable behavior that can be measured in rodents in these testing paradigms may not be analogous to the *discrete* maladaptive behaviors we would like to investigate in the context of human psychiatric disease. The rodent may thus provide a strong genetic model for studying gene function (in the context of transgenic or knockout models) and pharmacological model for studying the effect of compounds on certain exploratory behaviors; however, it is likely not the most ethologically-relevant organism for naturalistic modeling of discrete neurobiological syndromes that comprise a specific suite of physiological behaviors [31].

As an alternative organism of study, Overall presents the domestic dog as a more effective animal model for studying anxiety-related behavior [32]. One specific benefit of studying canines is the spontaneous or endogenous presentation of symptomatology without the need for genetic or neurochemical manipulation. In other words, anxiety-like behaviors are naturally observed in dogs without means of genetic, anatomical, or pharmacologic manipulations that are often required in rodent models. Overall evaluates the dog as a model organism on three types of validity: (1) *face* validity – the extent to which the model organism is phenotypically similar to and representative of the same symptomatology (behavioral and physiological) as the organism it seeks to represent; (2) *predictive* validity – the extent to which the model organism demonstrates the same effect to pharmacological intervention as that seen in the modeled system; and (3) *construct* validity – the extent to which the model organism relies on the same underlying neurobiological mechanism that is responsible for the same condition in the organism it models [32]. In the context of these three forms of validity, the author suggests canine correlates or homologues to different human psychiatric disorders. Although there is limited knowledge yet on underlying neurobiology of anxiety syndromes, behavioral and physiological symptoms suggest that humans and dogs share at least some fundamental underlying systems (face and construct

validity). Moreover, results of clinical studies on the effectiveness of TCAs in canines with obsessional disorders also suggests that dogs demonstrate similar responses to pharmacological interventions used to treat human anxiety disorders (predictive validity). Overall thus concludes that canines may provide a naturally occurring model for studying human psychiatric conditions in a more socially relevant framework as compared to rodents [32].

There are some disadvantages of using canines. First, phenotyping for behavioral traits or disorders often requires either (a) direct laboratory observation, or (b) the use of questionnaires filled out by the dog's owner. Additionally, locus-specific genetic manipulation (transgenic or knockout modeling, informative breeding, etc.) is socially unpalatable in community-based samples and currently not feasible in colony populations. Finally, a limited amount is known about canine psychobiology. However, a critical feature of the dog in terms of genetics is the foundation of pure breeds in the dog. This greatly facilitates gene-mapping efforts as each breed is like a genetically isolated population. Additionally, the canine is similar in physiology to humans, which means any discoveries made in dogs may be directly applicable to further human studies. In terms of phenotype, canines are social animals, constantly encountering both inter- and intra-dog interactions (e.g. owners, other dogs, other household pets, etc.); anxiety-related disorders in this social context, therefore, are maladaptive because they hinder this social interaction and/or bring about adverse consequences. Moreover, such behavior is often the target of negative selection (i.e., removal from breeding).

The use of dogs as models for naturally occurring behavioral disorders ("naturally occurring" again meaning the behavior is expressed without genetic, pharmacologic, or anatomic manipulation) is promising, not only because the presentations and pharmacological treatment responses are often similar to those of humans, but also because of the ability to track particular behaviors within specific pure breeds and/or several generations of pedigreed dogs. This latter

point especially can be key in helping to identify genetic susceptibility factors of discrete behaviors or syndromes. Naturally occurring behavioral disorders, such as those seen in the dog, remove possible confounding factors or epiphenomena such as impaired development due to artificial genetic manipulation. Additionally, the natural occurrence of these behaviors removes the reliance on testing paradigms as a measure of the behavioral phenotype.

Regarding the use of owner-based questionnaires for phenotyping in canine studies, there are two main possibilities for questionnaire development. One method is to adapt human-based questionnaires for their utility in assessing canine behavioral disorders. Vasa, *et al.* [34] attempted to evaluate just that, hypothesizing that owners of dogs experience and observe their dogs every day, similarly to parents with young infants. Therefore, questionnaires designed to survey parents about their infants for phenotypes such as attention and hyperactivity disorder (ADHD) may also be effective for surveying owners of dogs with similar attention deficits and hyperactivity problems. Owners of a total of 220 household pet dogs representing 69 different breeds were administered a questionnaire designed to survey parents about their infants for ADHD. The authors found that the questionnaire had high internal consistency, and all but one item measured the purported subscale. They concluded that human questionnaires administered to dog owners were a valid means of measuring attention deficit and activity in dogs [34]. A second method for phenotyping dogs for behavioral traits is through the use of questionnaires developed specifically for surveying canine behavior in general. Our lab has developed one such questionnaire for evaluating clinical anxiety disorders in dogs, which we used for determining the noise phobia phenotype described in Chapter 3 [35]. The questionnaire utilized by our group specifically aims to objectively quantify observed responses by owners of their dogs' responses to specific, discrete situations or stimuli. In this way, we aim to avoid subjective bias introduced by owners' perception of having a "good" or "bad" dog [35]. Other groups have also developed questionnaires for assessing canine behavior and temperament for use in research and for

evaluating dogs trained for working duties; the most commonly used is the "Canine Behavioral Assessment and Research Questionnaire (CBARQ) developed by Hsu and Serpell [36, 37]. In contrast to the questionnaire developed by our group, CBARQ incorporates measures of "owner impression," where owners are asked to give their opinion of their dog's, for example, 'aggressiveness.' In addition, behavioral categories in CBARQ are also divided into different groups (e.g. "stranger-directed aggression" versus "owner-directed aggression"), a distinction not utilized by the Overall questionnaire [35].

In addition to providing a more accurate representation of socially maladaptive behavioral syndromes, the canine also offers an important genetic context within which to investigate such behaviors as mentioned above. In "A fetching model organism," Neff and Rine [1] expound on the utility of the domestic dog as a model organism for studying evolution, development and behavior. More specifically, the authors suggest the dog to be a valuable model for studying both breed-specific behaviors (such as pointing) in addition to maladaptive behavior within a simplified genetic structure (in the form of pure breeds). The authors emphasize the wealth and diversity—as the dog demonstrates the largest span of phenotypic diversity of any terrestrial species—of naturally occurring phenotypic variation that can now be investigated using new microarray technologies that were enabled by sequencing of the canine genome and cataloguing of millions of single-nucleotide polymorphisms (SNPs) [1]. Spady and Ostrander also highlight the persistence of breed-specific behaviors such as herding, pointing, tracking and hunting in the absence of training or motivation, suggesting that these behaviors are, at least in part, controlled on a genetic level [39]. Further details highlighting the advancement of canine genetics are found in the following section.

1.3. Genomics of *Canis familiaris*

If we are to use the domestic dog as a model for studying the genetic basis of behavior, we must first learn about the genomic structure of the dog. With the creation of "pure breeds" in Europe in the 19th century, concentrated selection for physical attributes and behavioral traits began in the world of dogs. This selection has not only created the largest phenotypic diversity within any single terrestrial species, but has also served as a real-life experimental system in which canine geneticists explore naturally occurring traits in relatively genetically-homogeneous breeds. Pure breeds are akin to population isolates in human genetic studies, and geneticists are ready to exploit this opportunity to forward our knowledge of the genetic contribution to a variety of phenotypes.

The canine genome is comparable in size to most mammalian genomes, and is composed of 38 acrocentric autosomes and two sex chromosomes. The first types of polymorphic variant to be described in the dog for purposes of trait mapping were microsatellite markers (simple sequence repeat polymorphisms). Ostrander and colleagues first characterized dinucleotide repeat (CA)n markers in 1993 [39], and more were discovered in subsequent years [40]. However, studies suggested that dinucleotide (CA) repeat variation may be significantly lower in purebred dogs as compared to mixed-breed dogs, which may make them inadequate for linkage mapping within pure breeds [40, 41]. Thus the discovery of a new class of tetranucleotide repeats in the dog, (GAAA)n, was a critical addition to the canine genomic tool set, particularly given their increased polymorphic rate compared to (CA)n repeats. Additionally, these tetranucleotide repeats demonstrated sufficient distribution throughout the genome for linkage studies and adequate stability for mapping of traits with simple heredity patterns in extended family pedigrees [42].

With polymorphic marker sets in place, the tools were now available for more thorough investigation of the canine genome. In 1996, Fischer, *et al.* applied fluorescence *in situ* hybridization (FISH) techniques for physical mapping of the 78 canine chromosomes, identifying six microsatellite sites, two ribosomal sites and a human chromosome-X paint in short-term peripheral blood cultures [43]. In 1997, construction of a canine linkage map began, first with Lingaas, *et al.*'s establishment of 16 linkage groups [44], then with the publication of a first-generation linkage map by Mellersh, *et al.* linking 139 dinucleotide and tetranucleotide microsatellite markers and identifying 30 linkage groups [45]. Within two years, a second-generation linkage map was published by Neff and colleagues, extending to 276 mapped loci and 39 autosomal linkage groups [46]. To complement the canine linkage map, construction of a whole-genome radiation hybrid (RH) panel was also in development by 1999 [47]. Creation of an integrated linkage-RH map continued [48-50], and by 2004 an integrated FISH/RH map existed for all canine chromosomes except the Y-chromosome [51].

Researchers utilized the tools developed for the canine genome for investigations of breed relationships [52] as well as patterns of linkage disequilibrium (LD) between and within pure breeds [53]. These investigations served two purposes, (1) to further describe the genetic structure of pure breeds (i.e., to establish genetically the distinction of pure breeds as genetic 'isolates' and to investigate the genetic clustering of different breeds as it may relate to breed creation), and (2) to assess the canine genome for use in linkage studies. In 2006, Senger and colleagues published a high-resolution 9000-rad RH panel for the dog [54]. Comparative genomics studies continue in the field of canine genetics [55], and tools are becoming available to facilitate multi-species studies [56].

Parker, *et al.* [52] investigated the genetic relationships of different breeds of dogs, hypothesizing that genetic relationships will reflect breed histories and historical accounts for breed creation.

Microsatellite markers were used to characterize 85 different domestic dog breeds. Phylogenetic analysis and clustering were used to separate breeds into different putative groups and genetic clusters of putative shared ancestry. The authors found that phylogenetic analysis separated out several breeds of ancestral origins (Asian, African) from the remaining breeds that have more modern European origins. These groups were also identified via genetic clustering, where first the most ancestral breeds separated from the European breeds, then three subsets of European breeds (mastiffs, herding-type breeds, hunting-type breeds) were also separated. These general breed separations are in line with known breed function, geographic origin and/or morphology, suggesting that the genetic relationships between these dogs reflect the known breed history or purpose of the dogs [52], as hypothesized.

Also published in the same year, Sutter and colleagues characterized the LD patterns in the dog [53]. Their goal was to survey LD in pure breeds of domestic dog and assess haplotype diversity between and within breeds. They theorized that there would be extensive LD in pure breeds, and long common haplotypes within but not between breeds. Five genomic intervals were surveyed for known sequence variants in 20 unrelated dogs each for five different pure breeds (Akita, Bernese mountain dog, golden retriever, Labrador retriever, Pekingese). Sutter, *et al.* found that LD in dogs appears to be ~100x longer than that observed in humans [53]. Additionally, low haplotype diversity is observed within regions of high LD, with small numbers of haplotypes present within breeds and also seen across breeds (though at varying frequencies depending on breed). This highlights the shared ancestry of all domestic dogs as they branched off from wolves evolutionarily, with discrete bottlenecks occurring more recently upon the creation of pure breeds. The extensive LD in the dog also suggests that many fewer markers are required for mapping efforts in the dog compared to human, highlighting the utility of the dog in genetic studies [53].

Although microsatellite markers could be multiplexed to expedite analysis [57], the use of the domestic dog for genetic mapping studies was still laborious. Additionally, the idea of utilizing the canine for genome-wide studies of complex traits – which are predicted to have multiple loci of modest effect – was quite daunting. Thus the publication of the genome of *C. familiaris* was a welcome addition to the tool belt of canine geneticists worldwide. A 1.5X sequence derived from a single male Poodle [58], covering about 78% of the genome was published by the Institute for Genomic Research, and a NHGRI-funded effort to sequence the dog genome resulted in the publication of a 7.5X draft sequence covering approximately 99% of the Boxer genome, excluding highly repetitive regions [2]. As the fifth mammal (human, chimp, mouse, rat), and first animal of the neighboring clade of those mammals, to be sequenced, the dog genome serves as the first mammalian outgroup for comparative genomic searches for conserved elements and function in the human genome. With annotation of SNPs in the canine genome, the tools would also be in place to undergo GWAS to map traits of interest that may be relevant to the study of human disease. Goals of the canine sequencing initiative thus included: sequencing and assembly of a high-quality draft of the domestic dog genome; annotation of SNPs in the dog genome for use in mapping studies; assessment of the genomic structure of the dog, particularly compared to the genomes of the primate and rodent lineages to learn more about genomic evolution and structure; and assessment of genomic structure in limited regions of the genome across multiple breeds.

Whole-genome shotgun sequencing of the boxer was performed and an assembly was created using an improved version of the ARACHNE program. The genome assembly was also aided with information from the previously existing canine radiation hybrid and cytogenetic maps. The "CanFam2.0" version of the genome contains 7.5X sequence coverage of 99% percent of the putative 2.4Gb of canine genome. 2.5 million SNPs are annotated on the genome. Comparative genomic analysis demonstrates extensive synteny between human and dog, and suggests that

~5.3% of the human genome consists of conserved functional elements that are shared between dog, human and mouse. Interestingly, the dog genome demonstrates more nucleotide homology with the human genome than does the mouse, despite the fact that human and mouse are more closely related evolutionarily [59]. However, this is likely due to the estimated 2.5X higher rate of genomic deletion in mouse, which results in human and dog sharing more extant 'ancestral sequence' [59]. LD in the dog extends across several megabases within breeds, but only extends over tens of kilobases when looking across breeds [2]. The approximately 2.5 million SNPs annotated on CanFam2.0 were generated by comparing the boxer genome to the previously existing Poodle genome by Kirkness, *et al.* [58], observation of heterozygous loci in the boxer, and comparison made to shotgun sequence from select chromosomal regions from dogs of nine diverse breeds, four grey wolves and one coyote. The polymorphism rate of the current SNP map is ~1/900bp between breeds, and ~1/1500bp within breeds, which suggests the genotyping of ~10-15 thousand SNPs across the genome will suffice for most studies attempting to identify genetic associations to common DNA variation in population samples. This stands in distinction to human mapping studies, in which 500,000 or more SNPs are commonly used [2].

In addition to the SNP data provided by CanFam2.0 and SNP-discovery efforts, microsatellites also continue to be an important exploratory apparatus for understanding variation in the canine genome [60]. Earlier this year Wong, *et al.* published a comprehensive linkage map of the dog [61]. Utilizing microsatellite data, published genome data and SNP genotyping results the authors created the first complete linkage map, which revealed marked regional and sex-specific differences in recombination rates [61]. With these advances in dog genomics, the most important tools are in place for rigorous genetic analyses of canine behavioral traits [2, 58].

1.4. Gene Mapping in the Dog: Then and Now

Genetics in the dog has been characterized by a long history of traditional mapping approaches, particularly in diseases demonstrating mendelian inheritance, and in fixed traits, with genomic investigations of canine phenotypes and disease appearing as early as the 1960s (e.g. [62]). Causative mutations and quantitative trait loci (QTLs) have been mapped predominantly by investigation within single breeds, or in extended pedigrees of interrelated individuals. QTL mapping in the dog is exemplified by the work done in Portuguese water dogs (PWDs) through the Georgie Project by Lark and colleagues (http://www.georgieproject.com). In 2004, Chase, *et al.* described two side-specific QTLs that regulate laxity in the hip joint (this type of laxity can lead to degenerative joint disease [63]). Another QTL on *Canis familiaris* chromosome (CFA) 3 was also characterized for its role in regulating pathological bone remodeling and osteophyte formation resulting in osteoarthritis [64]. QTLs for osteoarthritis of hip joints were also mapped by Mateescu, *et al.* to four regions, on CFA 5, 18, 23, and 31 [65]. In addition to these explorations of arthritis phenotypes, the PWD was also utilized in mapping of the *IGF1* allele as a major QTL for size variation [66].

One of the early notable discoveries in canine disease gene mapping was the linkage analysis and comparative mapping of canine progressive rod-cone degeneration (PRCD) by Acland and colleagues in 1998 [67]. Acland, *et al.* not only mapped the locus (utilizing a panel of anchor loci and microsatellite markers) but, perhaps more importantly for the field, demonstrated potential locus homology with the human syndrome retinitis pigmentosa: the canine linkage region on CFA9 demonstrated conserved synteny with the retinitis pigmentosa locus on 17q (RP17), for which no gene had yet been identified. In 2006, Zangerl, *et al.* identified a homozygous mutation in the novel gene *PRCD* located in their linkage region on CFA9 that was present in 18 affected dogs of different breeds [68]. Interestingly, they also identified the same mutation in a human

patient with retinitis pigmentosa, demonstrating the potential strength of the canine in mapping disease loci relevant to humans.

Mapping of many other disease loci has also occurred and includes: deletion of exon 2 in the *MURR1* gene via positional cloning in copper toxicosis in Bedlington terriers [69]; linkage mapping of the primary locus for collie eye anomaly to CFA37 [70]; linkage analysis of a susceptibility locus in canine tricuspid valve malformation in Labrador retrievers to CFA9 [71]; mutation in exon 7 of canine Birt-Hogg-Dubé (BHD) segregating with hereditary multifocal renal cystadenocarcinoma and nodular dermotofibrosis in German shepherd dogs (the BHD locus is also implicated in the analogous disease in humans [72]); a missense mutation in *CLN8* in the neurodegenerative disease of Irish setters, neuronal ceroid-lipofuscinosis [73]. In addition to mapping of morphological and disease traits, there has also been interest in mapping genes underlying behavior. This interest was exemplified by the creation of a Newfoundland-Border Collie hybrid colony in the laboratory of Jasper Rine at the University of California at Berkeley in the early 1990s as part of the "Dog Genome Initiative" [74].

Perhaps the strongest example to date of canine research in the behavioral realm is that of the Mignot group at Stanford University on canine narcolepsy. First described in 1973 by Knecht, *et al.* [75] and then in 1974 by Mitler, *et al.* [76], narcolepsy-cataplexy in dogs demonstrates pathology analogous to human narcolepsy via electrograph, with excitation due to presentation of a plaything, food or water most frequently eliciting attacks resulting in atonia [76]. In addition to this face validity, canine cataplexy has also been shown to be responsive to imipramine as seen in human cataplexy [76-80] except in one study [81], demonstrating predictive validity in narcoleptic dogs for this drug. In 1976, Stanford University established a colony of narcoleptic dogs to evaluate the pathophysiology of the disorder [82]. Although narcolepsy observed in their colony dogs had a very similar presentation as that in humans, it did not appear to be associated

with the HLA region as had previously been demonstrated in humans [83]. Thus, the Mignot group hoped to elucidate new candidate genes for narcolepsy research in humans by understanding what causes narcolepsy in their dog colonies.

Over twenty years after the creation of its colony, the Stanford group published discovery of the gene causing the disorder in two large colonies of lab-bred narcoleptic dogs, mapped via positional cloning [84]. Colonies of Doberman pinchers and Labrador retrievers were raised in the laboratory and bred to create large pedigrees of dogs that transmitted narcolepsy. They found heritability in their Doberman pedigree to be autosomal recessive with one allele and full penetrance. The researchers used linkage analysis to localize the causative allele on CFA12. They eventually localized the causative gene to be hypocretin (orexin) receptor 2 (*Hcrtr2*). Hungs and colleagues of the same group later went on to map the causative mutations in the *Hcrtr2* gene that appear to be responsible for narcolepsy in these pedigrees [85]. The hypocretin (also called orexin) family of related proteins are intriguing novel candidates for sleep disorders. The hypocretin excitatory neurotransmitters were only discovered a year prior to their linkage to narcolepsy in dogs, and were previously thought to be involved in appetite regulation. Their role in sleep regulation continues to be investigated in humans, although studies of hypocretin pathway genes have not yet been found to be associated with narcolepsy in humans [86]. Though some might argue that lack of human association brings into question the construct validity of these canine colonies in the context of modeling human narcolepsy, the canine discovery has clearly opened up the field for research in pathways that may otherwise have never been investigated with regard to this disorder. Additionally, it would not be unreasonable that variation in the hypocretin system may only play a small or modifying role (i.e., low penetrance) in human narcolepsy.

Finally, there have also been advances in the relatively new field of pharmacogenetics (where genetic variation is investigated for its effect on drug response, metabolism and toxicity) in canine research. The best example of this is seen in the canine multidrug resistance gene (*MDR1*), which codes the P-glycoprotein (Pgp) transporter that plays a critical role in drug distribution. A mutation in *MDR1* was identified by Mealey and colleagues, and found to be associated with ivermectin sensitivity in collies [87]. The mdr1-1Δ mutation causes toxicity to the common anti-heartworm medication ivermectin, as well as cytotoxicity to several other anticancer agents [88, 89]. Knowledge of this mutation's presence in dogs is critical for two reasons: (a) because of the severity of the toxic side effects seen in dogs carrying the mutation, and (b) because dogs carrying the wildtype alleles are unaffected and can thus be safely dosed with these medications. Although this mutation is most common in dogs of collie lineage, it is also seen in some hounds [88]. Thus Neff, *et al.* sought to characterize this mutation in other breeds, exploiting the presence of the mdr1-1Δ mutation to phylogenetically characterize the collie lineage of dogs and identify other breeds that might be at risk for drug toxicity [90]. Utilizing this highly functionally relevant mutation, they identified the mutation in several additional breeds of herding dogs, plus two breeds of sighthounds not previously thought to share lineages with collie lines, the Longhaired Whippet and the Silken Windhound [90]. Importantly, genetic testing for the *MDR1* mutation is now available to identify dogs that are at risk for toxic side effects, and the list of suggested breeds to test includes the sighthounds identified by Neff and colleagues in 2004 (http://www.vetmed.wsu.edu/depts-vcpl/).

In the last five years, there has been a burgeoning of gene mapping results as a by-product of the publication of the 7.5X canine genome [2]. In addition to microsatellites and RH mapping, there are now over 2 million SNPs published and available for linkage mapping purposes. Further, the development of SNP arrays surveying the genome in a high-throughput manner can now characterize 20,000 - over 170,000 SNPs in a single genotyping assay. Finally, next-generation

sequencing (NGS) is increasing in popularity because of the quantity of data it can provide (currently up to 7Gb of sequence) per single run.

As a proof of principal, Karlsson and colleagues sought to map traits with simple hereditary patterns in a small set of cases and controls using minimal SNP marker coverage. They hypothesized that is possible to perform GWAS in a small number of dogs using a small set of markers to map genetic loci underlying simple traits such as morphological traits demonstrated to have mendelian heredity patterns [91]. The ability to perform GWAS with such small numbers of samples and markers in the domestic dog is due to the breed structure of the dog, with long blocks of LD that allow for single markers to tag very large genomic regions. They therefore investigated two phenotypes: ridging in the Rhodesian ridgeback (in pedigrees shown to be a dominant trait with presumed one mutation causing ridging) and white coloring in the boxer (in pedigrees shown to be a recessive trait with presumed two mutations required to cause full white color, and one copy causing a middle phenotype of white-spotting). About 10 cases and 10 controls of each phenotype were compared at about 27,000 SNPs surveyed via an Affymetrix microarray [91].

For ridging, the authors identified a single region on CFA18 demonstrating the strongest association that withstood permutation to correct for Type I error (raw p-value 9.6×10^{-8}; permuted p-value 0.0014). The causative mutation for ridging was later mapped by Salmon Hillbertz and colleagues to a duplication event involving three fibroblast growth factor genes [92]. For white color/spotting, the authors identified a single region on CFA20 demonstrating strongest association that also withstood permutation (raw p-value 7.1×10^{-10}; permuted p-value $< 3 \times 10^{-5}$). Leegwater and colleagues localized the white color/spotting locus in boxers to a region near the *MITF* gene, which is a strong candidate since the same gene in humans has been found to cause albinism [93].

Array-based GWAS have been successful in identifying a handful of loci playing a role in complex disease. In diseases of simple inheritance, the power of canine genetics is highlighted by Drogemuller, *et al.*'s identification of a missense mutation in the *SERPINH1* gene as a candidate causative mutation for osteogenesis imperfecta (OI) in Dachshunds [94]. Remarkably, only 5 affected dogs were used for first-pass analysis to identify a single extended region of homozygosity on CFA21. Five more obligate carriers were then used to narrow the homozygous region, which was then investigated for functional candidate genes via annotation of the corresponding human interval and linkage analysis in families segregating OI. A novel mapping approach was also taken by Wiik, *et al.* [95] when mapping a deletion in nephronophthisis 4 (*NPHP4*) associated with cone-rod dystrophy in the standard wire-haired dachshund. For the first time, the group used a genome-wide discordant sibling study design, utilizing only 13 discordant sib-pairs in their primary analysis [95]. Notably, Awano, *et al.* discovered a *SOD1* mutation in Pembroke Welsh Corgis that was responsible for canine degenerative myelopathy (DM) in that and five other breeds. DM is a fatal neurodegenerative disease that may be the first naturally occurring animal model for human a myotrophic lateral sclerosis (ALS, [96]). The *SOD1* results are particularly intriguing given the fact that *SOD1* mutations are also found in about 20% of all families with the hereditary form of ALS (http://www.alsa.org/als/genetics.cfm), though it should be noted that only about 10% of all ALS cases are familial.

For disease phenotypes demonstrating more heterogeneity, modest results have also been demonstrated for relatively small sample sizes. Wood and colleagues genotyped 25 golden retrievers with atopic dermatitis and compared them to 23 healthy, matched controls at approximately 22,000 SNPs [97]. The lowest p-value in their initial GWAS was 1.27×10^{-6}, with an odds ratio (OR) = 9.5. They followed up their top 40 hits in 648 additional dogs from eight different breeds, and meta-analysis resulted in two intergenic SNPs on CFA10 and 29 meeting

their criteria for significance, with corrected p-values of 1.4×10^{-4} and 0.0015 and OR = 2 and 0.6, respectively. Of note, neither of the SNPs that demonstrated significance in the eight-breed meta-analysis were the top-ranked hit by p-value in the single breed GWAS [97]. Quite recently, Wilbe, *et al.* [98] conducted a GWAS for a canine systemic lupus erythematosus (SLE)-related disease complex in a primary sample of 81 affected and 57 control Nova Scotia duck tolling retrievers. Their top finding in the initial GWAS for all SLE-related disease was on CFA32 with a raw p-value of 7.9×10^{-6}, and a permuted genomic p-value (100,000 permutations) of 0.06. A sub-analysis of ANA-positive immune-mediated rheumatic disease rendered four hits on four different chromosomes reaching nominal significance, with raw p-values of 2.2×10^{-5} to 1.5×10^{-6} and permuted genomic p-values of 0.18 to 0.02, respectively. Notably, these hit regions included multiple candidate genes involved in immune function and/or human SLE disease [98].

In addition to GWAS with SNP data, other groups have used complimentary approaches that integrate multiple different types of variant information for genome-wide mapping. For example, Werner and colleagues mapped a novel locus on CFA8 for an inherited form of juvenile dilated cardiomyopathy via linkage in 16 PWD families using microsatellite, SNP and short interspersed nuclear element (SINE) markers [99]. A combination of microsatellite and SNP analysis was also used in the discovery of a mutation in the dynamin 1 (*DMN1*) causing exercise-induced collapse in Labrador retrievers [100]. Most recently, Oberbauer and colleagues have investigated idiopathic epilepsy in two large, extended pedigrees of Belgian shepherds using microsatellite markers providing genome-wide coverage. They identified six QTLs on a total of four different chromosomes for further investigation of the disorder in Belgian shepherds, highlighting the complexity of genomic studies of polygenic disorders [101].

Along with mapping for genes related to disease, researchers have also pursued genetic investigations of behavior. Not surprisingly, numerous researchers have investigated the "usual

suspects" – i.e., genes related to neurotransmitters and enzymes involved in their production and metabolism – that have been investigated in humans and rodents with regard to similar behavioral phenotypes. In particular, many studies in the dog have focused on aggression. Studies interrogating polymorphisms in serotonin genes (in particular, serotonin receptor genes *htr1a* and *htr2a*; the serotonin transporter gene *slc6a4*) for their role in aggression in golden retrievers have not rendered any significant results, suggesting these particular variants/genes are unlikely to play a significant role in canine aggression phenotypes [102, 103].

Numerous studies have also been performed examining the role of various dopamine-related genes and receptors with regard to behavior. For example, the variable number tandem repeat (VNTR) present in the dog dopamine D4 receptor gene (*DRD4*) has been associated with ADHD in humans, and was therefore investigated for a role in canine activity-impulsivity. Interestingly, association with a particular allele and increased activity-impulsivity was observed only in police German shepherds, but not pet German shepherds [104, 105]. These results highlight the potential effect within-breed stratification may have on the results of association studies, a point investigated by our group and others, which is discussed in more detail in Chapter 2 [106, 107]. Although the *DRD4* association with activity-impulsivity was not replicated in a group of Belgian tervurens, the researchers did go on to find association between a different variant in *DRD4* and attention-deficit, as well as associations between variants in the dopamine transporter (*DAT*) and dopamine β-hydroxylase (*DBH*) genes and attention-deficit [108].

Most recently, a large study on 62 SNPs in or close to 16 neurotransmitter genes were investigated in 50 aggressive and 81 non-aggressive dogs, but did not identify any haplotypes associated with the phenotype [109]. Takeuchi, *et al.* did find an association between a variant in *SLC1A2* with 'aggression to strangers' in a cohort of Shiba Inu (an indigenous dog breed), though studies in other breeds are required to confirm the generalizability of these results to other breeds

[110]. These results further highlight the genetic and phenotypic complexity of behavioral phenotypes such as aggression, and strongly suggest that methods beyond assessment of hypothetical candidate genes are required when investigating their genetic basis, primarily due to the limiting requirement of preexisting knowledge of underlying neurobiology.

With regard to GWAS of behavioral phenotypes, one letter has been published in *Molecular Psychiatry* [111] detailing the results of a GWAS in Doberman pinschers for canine compulsive disorder (CCD). Ninety-two cases and 68 controls were analyzed across 14,700 SNPs, with 3 SNPs on CFA7 withstanding permutation for significance. The most significant SNP had a raw p-value of 7.6×10^{-7} and a permuted genomic p-value of 0.013 [111]. The most associated SNP after fine-mapping with 84 SNPs in the 1.7Mb region surrounding their associated SNP is located within the cadherin 2 gene. Interestingly, the cadherin family has also been implicated in human autism spectrum disorder, a disorder that often includes aspects of repetitive and compulsive behaviors [112]. This study highlights the feasibility of identifying genetic risk loci for complex behavioral phenotypes.

Now I will conclude this chapter with a brief overview of the current status of genetic research for human psychiatric conditions relevant to the current work.

1.5. Canines Provide Sanity in the Mess of Human Psychosis

A PubMed search for "psychiatric genetics" leads to over 10,000 references for studies starting in the 1950s through today. From twin studies to candidate gene sequencing to linkage analysis in families, the breadth of psychiatric genetics research is far and wide, ranging from topics such as substance abuse to sleep disorders. However, for the purposes of this introduction, I will only briefly review genetic findings for two main areas of psychiatric research related to the topic of this dissertation, panic disorder and specific phobia, with an emphasis on GWAS if applicable.

Panic disorder (PD) as well as phobic disorders have been genetically investigated more thoroughly than other, broader disorders such as generalized anxiety disorder (GAD) [113]. Numerous linkage analyses for PD, as well as for specific phobia and social anxiety (phobia) have been reported, though research investigating putative linkage regions has made little progress, and linkage results across groups have not converged on common genomic regions, nor replicated in independent samples [113, 114]. Though numerous candidate genes have been studied for their involvement in anxiety, the most consistent positive finding to date is catechol-O-methyltransferase (*COMT*), though the actual associated allele has yet to be determined due to locus heterogeneity across populations [114]. In addition to linkage studies, a GWAS of PD has also been conducted in a Japanese sample [115]. Though the same group was not able to replicate their top findings in an independent sample cohort at statistical significance for a multiple-testing correction, they found two nominally associated SNPs located in the *APOL3* and *CLU* genes, respectively [116]. Most recently, a German group reported an association for SNPs in TMEM132D with PD [117]. These results were demonstrated in three independent PD sample cohorts, and expression studies in a mouse model for extreme trait anxiety also suggest a role of TMEM132D in PD and possibly also other anxiety disorders as well [117].

Finally, a GWAS of neuroticism (a personality trait thought to confer risk toward anxiety and depression disorders) implicated the novel gene candidate *MAMDC1* in both American and German samples [118]. However, the same group was not able to replicate their own findings in a larger, independent American cohort, despite sufficient statistical power to do so [119]. Most recently, a GWAS for the five major dimensions of personality resulted in a suggested association signal for neuroticism at *SNAP25* in a Sardinian population isolate, though this finding was not replicated in subsequent follow-up in an independent replication cohort [120]. These results have

been somewhat disappointing since they involve a dimensional quantitative phenotype that is thought to provide a potential endophenotype for human mood and anxiety disorders.

Despite relatively large sample sizes, genetic heterogeneity of human populations and inconsistency in disease phenotyping of psychiatric disorders have likely decreased the ability to find genetic risk loci for complex behavioral diseases related to anxiety. Although recent GWAS studies have suggested new candidate genes that may contribute risk towards these disorders, few of these candidates were previously thought to play a role based on known etiology, and would demonstrate novel roles in disease should they be replicated and shown to contribute risk to the disorder. The canine offers the possibility of overcoming some of the obstacles facing human genetic studies in mood and anxiety disorders by allowing discovery of genetic variants that may influence predisposition towards anxiety-related disorders, contributing vital knowledge to an important field of research.

In the next chapter, I will outline the genetic methods utilized for conducting such an investigation of the anxiety disorder noise phobia in the domestic dog. The intricacies of investigating this complex behavioral trait will also be paired with the genetic investigation of another complex disorder, adult-onset deafness, which demonstrates a case of simplified genetics in herding Border collie dogs. Chapter 3 will describe the samples, analytical methods and results of our genome-wide survey for loci underlying noise phobia in Border collies and other purpose-bred dogs. Chapter 4 follows with a description of our investigation of adult-onset deafness, which includes next-generation sequencing to follow up a strong candidate region implicated by our GWAS in herding Border collies. Finally, I will briefly summarize these results, highlighting the implications for genetic research of complex neuropsychiatric traits in the dog and elaborating on future directions for this work in the context of our laboratory's Canine Behavioral Genetics Project.

1.6. References

1. Neff MW, Rine J. (2006) A fetching model organism. Cell. 124(2):229-31.
2. Lindblad-Toh et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438(7069):803-19
3. Darwin, Charles. 1872. The expression of the emotions in man and animals. London: John Murray.
4. Scott JP and Fuller JL. 1965. Genetics and the Social Behavior of the Dog: The classic study. The University of Chicago Press, Ltd., London.
5. Vonholdt BM *et al.* (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature. 464(7290):898-902.
6. http://www.akc.org/reg/registeradog.cfm
7. Murphree OD, Dykman RA. (1965) Litter patterns in the offspring of nervous and stable dogs. I. Behavioral tests. J Nerv Ment Dis. 141(3):321-32.
8. Murphree OD, Peters JE, Dykman RA. (1969) Behavioral comparisons of nervous, stable and crossbred pointers at ages 2, 3, 6, 9, and 12 months. Cond Reflex. 4(1):20-3.
9. Murphree OD and Newton JE. (1971) Crossbreeding and special handling of genetically nervous dogs. Cond Reflex. 6(3):129-36.
10. Murphree OD, Angel C, DeLuca DC. (1974) Limits of therapeutic change: specificity of behavior modification in genetically nervous dogs. Biol Psychiatry. 9(1):99-101.
11. Murphree OD, Angel C, DeLuca DC, Newton JE. (1977) Longitudinal studies of genetically nervous dogs. Biol Psychiatry. 12(4):573-6.
12. Dykman RA, Murphree OD, Peters JE. (1969) Like begets like: behavioral tests, classical autonomic and motor conditioning, and operant conditioning in two strains of pointer dogs. Ann N Y Acad Sci. 159(3):976-1007.
13. Dykman RA, Murphree OD, Reese WG. (1979) Familial anthropophobia in pointer dogs? Arch Gen Psychiatry. 36:988-93.
14. Uhde TW, Malloy LC, Slate SO. (1992) Fearful behavior, body size, and serum IGF-1 levels in nervous and normal pointer dogs. Pharmacol Biochem Behav. 43(1):263-9.

15. Klein E, Steinberg SA, Weiss SR, Matthews DM, Uhde TW. (1988) The relationship between genetic deafness and fear-related behaviors in nervous pointer dogs. Physiol Behav. 43(3):307-12.

16. George MS, Malloy LC, Slate SO, Uhde TW. (1994-5) Pilot MRI study of brain size in nervous pointer dogs. 1(3):129-33.

17. Tancer ME, Stein MB, Bessette BB, Uhde TW (1990). Behavioral effects of chronic imipramine treatment in genetically nervous pointer dogs. Physiol Behav. 48(1):179-181.

18. Uhde TW, Malloy LC, Benson BB. (1994) Growth hormone response to clonidine in the nervous pointer dog model of anxiety. Anxiety. 1(2):45-9.

19. Coppens AG, Gilbert-Gregory S, Steinberg SA, Heizmann C, Poncelet L. (2005) Inner ear histopathology in the "nervous Pointer dogs" with severe hearing loss. Hear Res. 200(1-2):51-62.

20. Wimer RE and Wimer CC. (1985) Animal Behavior Genetics: A Search for the Biological Foundations of Behavior. Ann Rev Psychol. 36:171-218.

21. Tuber DS, Hothersall D, Peters MF. (1982) Treatment of fears and phobias in dogs. Vet Clin North Am Small Anim Pract. 12(4):607-23.

22. Luescher UA, McKeown DB, Halip J. (1991) Stereotypic or obsessive-compulsive disorders in dogs and cats. Vet Clin North Am Small Anim Pract. 21(2):401-13.

23. Dehasse J. (1994) Pathological anticipatory defence behavior in dogs. Vet Q. 16 Suppl 1:49S.

24. Overall KL. (1998) Concerns regarding study of obsessive-compulsive disorder in dogs. J Am Vet Med Assoc. 213(2):198-9.

25. Moon-Fanelli AA, Dodman NH. (1998) Description and development of compulsive tail chasing in terriers and response to clomipramine treatment. J Am Vet Med Assoc. 212(8):1252-7.

26. Seksel K and Linderman MJ. (2001) Use of clomipramine in treatment of obsessive-compulsive disorder, separation anxiety and noise phobia in dogs: a preliminary, clinical study. Aust Vet J. 79(4):252-6.

27. Overall KL and Dunham AE. (2002) Clinical features and outcomes in dogs and cats with obsessive-compulsive disorder: 126 cases (1989-2000). J Am Vet Med Assoc. 221(10):1445-52.

28. Overall,K.L. (1997). Clinical Behavioral Medicine for Small Animals. St. Louis: Mosby

29. Crowell-Davis SL, Seibert LM, Sung W, Parthasarathy V, Curtis TM. (2003) Use of comipramine, alprazolam, and behavior modification for treatment of storm phobia in dogs. J Am Vet Med Assoc. 222(6):744-8.

30. Gruen ME, Sherman BL. (2008) Use of trazodone as an adjunctive agent in the treatment of canine anxiety disorders: 56 cases (1995-2007). J Am Vet Med Assoc. 233(12):1902-7.

31. Shekhar A, McCann UD, Meaney MJ, Blanchard DC, Davis M, Frey KA, Liberzon I, Overall KL, Shear MK, Tecott LH, Winsky L. (2001) Summary of a National Institute of Mental Health workshop: developing animal models of anxiety disorders. Psychopharmacology (Berl). 157(4):327-39.

32. Overall KL. (2000) Natural animal models of human psychiatric conditions: assessment of mechanism and validity. Prog Neuropsychopharmacol Biol Psychiatry. 24(5):727-76.

33. Hall, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. J. Comp. Psychol. 18:385–403.

34. Judit Vasa, József Topálb, Éva Pécha and Ádám Miklósia (2007) Measuring attention deficit and activity in dogs: A new application and validation of a human ADHD questionnaire. Applied Animal Behaviour Science 103(1-2):105-117.

35. Overall 2006 Behavioral questionnaire: Overall KL, Hamilton SP, Chang ML (2006) J. Vet Behav: Clin Appl Res. 1(3): 124-41.

36. Serpell JA, Hsu Y. (2001) Development and validation of a novel method for evaluating behavior and temperament in guide dogs. Appl Anim Behav Sci. 72(4):347-364.

37. Hsu Y, Serpell JA (2003) Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. J Am Vet Med Assoc. 223(9):1293-300.

38. Spady TC, Ostrander EA. (2008) Canine behavioral genetics: pointing out the phenotypes and herding up the genes. Am J Hum Genet. 82(1):10-8.

39. Ostrander EA, Sprague GF Jr, Rine J. (1993) Identification and characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. Genomics. 16(1):207-13.

40. Ostrander EA, Mapa FA, Yee M, Rine J. (1995) One hundred and one new simple sequence repeat-based markers for the canine genome. Mamm Genome. 6(3):192-5.

41. Zajc, I., Mellersh, C. S., and Sampson, J. (1997). Variability of canine microsatellites within and between different dog breeds. Mamm. Genome 8: 182 – 185.

42. Francisco, L. V., Langston, A. A., Mellersh, C. S., Neal, C. L., and Ostrander, E. A. (1996). A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mamm. Genome 7: 359 – 362.

43. Fischer PE, Holmes NG, Dickens HF, Thomas R, Binns MM, Nacheva EP. (1996) The application of FISH techniques for physical mapping in the dog (Canis familiaris). Mamm Genome. 7(1):37-41.

44. Lingaas F, Sorensen A, Juneja RK, Johansson S, Fredholm M, Wintero AK, Sampson J, Mellersh C, Curzon A, Holmes NG, Binns MM, Dickens HF, Ryder EJ, Gerlach J, Bäumle E, Dolf G. (1997) Towards construction of a canine linkage map: establishment of 16 linkage groups. Mamm Genome. 8(3):218-21.

45. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, Ostrander EA. (1997) A linkage map of the canine genome. Genomics. 46(3):326-36.

46. Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, Aguirre GD, Ziegle JS, Ostrander EA, Rine J. (1999) A second-generation genetic linkage map of the domestic dog, Canis familiaris. Genetics. 151(2):803-20.

47. Vignaux F, Hitte C, Priat C, Chuat JC, Andre C, Galibert F. (1999) Construction and optimization of a dog whole-genome radiation hybrid panel. Mamm Genome. 10(9):888-94.

48. Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, Jouquand S, Werner P, André C, DeRose S, Patterson DF, Ostrander EA, Galibert F. (2000) An integrated linkage-radiation hybrid map of the canine genome. Mamm Genome. 11(2):120-30.

49. Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C, Holmes NG, Chéron A, Suter N, Vignaux F, Bristow AE, Priat C, McCann E, André C, Boundy S, Gitsham P, Thomas R, Bridge WL, Spriggs HF, Ryder EJ, Curson A, Sampson J, Ostrander EA, Binns MM, Galibert F. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. Genome Res. 11(10):1784-95.

50. Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, Quignon P, Lowe JK, Renier C, Gelfenbeyn B, Vignaux F, DeFrance HB, Gloux S, Mahairas GG, André C, Galibert F, Ostrander EA. (2003) A 1-Mb resolution radiation hybrid map of the canine genome. Proc Natl Acad Sci USA. 100(9):5296-301.

51. Breen M, Hitte C, Lorentzen TD, Thomas R, Cadieu E, Sabacan L, Scott A, Evanno G, Parker HG, Kirkness EF, Hudson R, Guyon R, Mahairas GG, Gelfenbeyn B, Fraser CM, André C, Galibert F, Ostrander EA. (2004) An integrated 4249 marker FISH/RH map of the canine genome. BMC Genomics. 5(1):65.

52. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L. (2004) Genetic structure of the purebred domestic dog. Science. 304(5674):1160-4.

53. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA. (2004) Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res. 14(12):2388-96.

54. Senger F, Cadieu E, Evanno G, Hitte C, Berkova N, Priat C, André C, Galibert F. (2006) Construction and characterization of a high-resolution, 9000-rad canine radiation hybrid panel. Anim Genet. 37(5):527.

55. Faraut T, de Givry S, Hitte C, Lahbib-Mansais Y, Morisson M, Milan D, Schiex T, Servin B, Vignal A, Galibert F, Yerle M. (2009) Contribution of radiation hybrids to genome mapping in domestic animals. Cytogenet Genome Res. 126(1-2):21-33.

56. Derrien T, André C, Galibert F, Hitte C. (2007) AutoGRAPH: an interactive web server for automating and visualizing comparative genome maps. Bioinformatics. 23(4):498-9.

57. Eggleston ML, Irion DN, Schaffer AL, Hughes SS, Draper JE, Robertson KR, Millon LV, Pedersen NC. (2002) PCR multiplexed microsatellite panels to expedite canine genetic disease linkage analysis. Anim Biotechnol. 13(2):223-35.

58. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, et al. (2003) The dog genome: survey sequencing and comparative analysis. Science 301: 1898-1903.

59. Springer MS, Stanhope MJ, Madsen O, de Jong WW. (2004) Molecules consolidate the placental mammal tree. Trends Ecol Evol. 19(8):430-8.

60. Sargan DR, Aguirre-Hernandez J, Galibert F, Ostrander EA. (2007) An extended microsatellite set for linkage mapping in the domestic dog. J Hered. 98(3):221-31.

61. Wong AK, Ruhe AL, Dumont BL, Robertson KR, Guerrero G, Shull SM, Ziegle JS, Millon LV, Broman KW, Payseur BA, Neff MW. (2010) A comprehensive linkage map of the dog genome. Genetics.1 84(2):595-605.

62. Roberts SR. (1967) Color dilution and hereditary defects in collie dogs. Am J Ophthalmol. 63(6):1762-75.

63. Chase K, Lawler DF, Adler FR, Ostrander EA, Lark KG (2004) Bilaterally asymmetric effects of quantitative trait loci (QTLs): QTLs that affect laxity in the right versus left coxofemoral (hip) joints of the dog (Canis familiaris). Am J Med Genet A. 124A(3):239-47.

64. Chase K, Lawler DF, Carrier DR, Lark KG. (2005) Genetic regulation of osteoarthritis: A QTL regulating cranial and caudal acetabular osteophyte formation in the hip joint of the dog (Canis familiaris). Am J Med Genet A. 135(3):334-5.

65. Mateescu RG, Burton-Wurster NI, Tsai K, Phavaphutanon J, Zhang Z, Murphy KE, Lust G, Todhunter RJ. (2008) Identification of quantitative trait loci for osteoarthritis of hip joints in dogs. Am J Vet Res. 69(10):1294-300.

66. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, Quignon P, Johnson GS, Parker HG, Fretwell N, Mosher DS, Lawler DF, Satyaraj E, Nordborg M, Lark KG, Wayne RK, Ostrander EA. (2007) A single IGF1 allele is a major determinant of small size in dogs. Science. 6;316(5821):112-5. Erratum in: Science. 2007 Jun 1;316(5829):1284.

67. Acland GM, Ray K, Mellersh CS, Gu W, Langston AA, Rine J, Ostrander EA, Aguirre GD. (1998) Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. Proc Natl Acad Sci USA. 95(6):3048-53.

68. Zangerl B, Goldstein O, Philp AR, Lindauer SJ, Pearce-Kelling SE, Mullins RF, Graphodatsky AS, Ripoll D, Felix JS, Stone EM, Acland GM, Aguirre GD. (2006) Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. Genomics. 88(5):551-63.

69. van De Sluis B, Rothuizen J, Pearson PL, van Oost BA, Wijmenga C. (2002) Identification of a new copper metabolism gene by positional cloning in a purebred dog population. Hum Mol Genet. 11(2):165-73.

70. Lowe JK, Kukekova AV, Kirkness EF, Langlois MC, Aguirre GD, Acland GM, Ostrander EA. (2003) Linkage mapping of the primary disease locus for collie eye anomaly. Genomics. 82(1):86-95.

71. Andelfinger G, Wright KN, Lee HS, Siemens LM, Benson DW. (2003) Canine tricuspid valve malformation, a model of human Ebstein anomaly, maps to dog chromosome 9. J Med Genet. 40(5):320-4.

72. Lingaas F, Comstock KE, Kirkness EF, Sørensen A, Aarskaug T, Hitte C, Nickerson ML, Moe L, Schmidt LS, Thomas R, Breen M, Galibert F, Zbar B, Ostrander EA (2003) A mutation in the canine BHD gene is associated with hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog. Hum Mol Genet. 12(23):3043-53.

73. Katz ML, Khan S, Awano T, Shahid SA, Siakotos AN, Johnson GS. (2005) A mutation in the CLN8 gene in English Setter dogs with neuronal ceroid-lipofuscinosis. Biochem Biophys Res Commun. 327(2):541-7.

74. Sargan DR (1995) Research in canine and human genetic disease. J Med Genet. 32:751-4.

75. Knecht CD, Oliver JE, Redding R, Selcer R, Johnson G. (1973) Narcolepsy in a dog and a cat. J. Am. Vet. Med. Assoc., 162(12): 1052-3.

76. Mitler MM, Boysen BG, Campbell L, Dement WC. (1974) Narcolepsy-cataplexy in a female dog. Exp. Neurol., 45(2): 332-40.

77. Mitler MM, Soave O, Dement WC. (1976) Narcolepsy in seven dogs. J Am Vet Med Assoc. 168(11):1036-8.

78. Babcock DA, Narver EL, Dement WC, Mitler MM. (1976) Effects of imipramine, chlorimipramine, and fluoxetine on cataplexy in dogs. Pharmacol Biochem Behav. 5(6):599-602.

79. Hendricks JC and Hughes C. (1989) Treatment of cataplexy in a dog with narcolepsy. J Am Vet Med Assoc. 194(6):791-2.

80. Tonokura M, Fujita K, Morozumi M, Yoshida Y, Kanbayashi T, Nishino S. (2003) Narcolepsy in a hypocretin/orexin-deficient Chihuahua. Vet Rec. 152(25):776-9.

81. Van Heerden J and Eckersley GN. (1989) Narcolepsy in a long-haired dachschund. J S Afr Vet Assoc. 60(3):151-3.

82. Cederberg R, Nishino S, Dement WC and Mignot E. (1998) Breeding history of the Stanford colony of narcoleptic dogs. Vet Rec. 142(2):31-6.

83. Mignot E, Tafti M, Dement WC and Grumet FC. (1995) Narcolepsy and immunity. Adv Neuroimmunol. 5(1):23-37.

84. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qui X, de Jong PJ, Nishino S, Mignot E. (1999) The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. Cell. 98(3):365-76.

85. Hungs M, Fan J, Lin L, Lin X, Maki RA, Mignot E. (2001) Identification and functional analysis of mutations in the hypocretin (orexin) genes of narcoleptic canines. Genome Res. 11(4):531-9.

86. Olafsdottir BR, Rye DB, Scammell TE, Matheson JK, Stefansson K, Gulcher JR. (2001) Polymorphisms in hypocretin/orexin pathway genes and narcolepsy. Neurology. 57(10):1896-9.

87. Mealey KL, Bentjen SA, Gay JM, et al. (2001) Ivermectin sensitivity in Collies is associated with a deletion mutation of the MDR-1 gene. *Pharmacogenetics.* 11(8):727-33.

88. Mealey KL, Northrup NC, Bentjen SA. (2003) Increased toxicity of P-glycoprotein-substrate chemotherapeutic agents in a dog with the MDR1 deletion mutation associated with ivermectin sensitivity. J Am Vet Med Assoc. 223(10):1453-5, 1434.

89. Sartor LL, Bentjen SA, Trepanier L, Mealey KL. (2004) Loperamide toxicity after therapeutic doses in a Collie with the MDR1 mutation associated with ivermectin sensitivity. Journal of Veterinary Internal Medicine. 18:117-118.

90. Neff MW, Robertson KR, Wong AK, Safra N, Broman KW, Slatkin M, Mealey KL, Pedersen NC. (2004) Breed distribution and history of canine mdr1-1Delta, a pharmacogenetic mutation that marks the emergence of breeds from the collie lineage. Proc Natl Acad Sci USA. 101(32):11725-30.

91. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ 3rd, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kämpe O, Hedhammar A, Lander ES, Andersson G, Andersson L, Lindblad-Toh K. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet. 39(11):1321-8.

92. Salmon Hillbertz NH, Isaksson M, Karlsson EK, Hellmén E, Pielberg GR, Savolainen P, Wade CM, von Euler H, Gustafson U, Hedhammar A, Nilsson M, Lindblad-Toh K, Andersson L, Andersson G. (2007) Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat Genet. 39(11):1318-20.

93. Leegwater PA, van Hagen MA, van Oost BA. (2007) Localization of white spotting locus in Boxer dogs on CFA20 by genome-wide linkage analysis with 1500 SNPs. J Hered. 98(5):549-52.

94. Drögemüller C, Becker D, Brunner A, Haase B, Kircher P, Seeliger F, Fehr M, Baumann U, Lindblad-Toh K, Leeb T. (2009) A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta. PLoS Genet. 5(7):e1000579.

95. Wiik AC, Wade C, Biagi T, Ropstad EO, Bjerkås E, Lindblad-Toh K, Lingaas F. (2008) A deletion in nephronophthisis 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund. Genome Res. 18(9):1415-21.

96. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC, Taylor JF, Perloski M, Biagi T, Baranowska I, Long S, March PA, Olby NJ, Shelton GD, Khan S, O'Brien DP, Lindblad-Toh K, Coates JR. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. Proc Natl Acad Sci USA. 106(8):2794-9.

97. Wood SH, Ke X, Nuttall T, McEwan N, Ollier WE, Carter SD. (2009) Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. Immunogenetics. 61(11-12):765-72.

98. Wilbe M, Jokinen P, Truvé K, Seppala EH, Karlsson EK, Biagi T, Hughes A, Bannasch D, Andersson G, Hansson-Hamlin H, Lohi H, Lindblad-Toh K. (2010) Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet. 42(3):250-4.

99. Werner P, Raducha MG, Prociuk U, Sleeper MM, Van Winkle TJ, Henthorn PS. (2008) A novel locus for dilated cardiomyopathy maps to canine chromosome 8. Genomics. 91(6):517-21.

100. Patterson EE, Minor KM, Tchernatynskaia AV, Taylor SM, Shelton GD, Ekenstedt KJ, Mickelson JR. (2008) A canine DNM1 mutation is highly associated with the syndrome of exercise-induced collapse. Nat Genet. 40(10):1235-9.

101. Oberbauer AM, Belanger JM, Grossman DI, Regan KR, Famula TR. (2010) Genome-wide linkage scan for loci associated with epilepsy in Belgian shepherd dogs. BMC Genet. 11:35.

102. van den Berg L, Kwant L, Hestand MS, van Oost BA, Leegwater PA. (2005) Structure and variation of three canine genes involved in serotonin binding and transport: the serotonin receptor 1A gene (htr1A), serotonin receptor 2A gene (htr2A), and serotonin transporter gene (slc6a4). J Hered 96(7):786-796

103. van den Berg L, Vos-Loohuis M, Schilder MB, van Oost BA, Hazewinkel HA, Wade CM, Karlsson EK, Lindblad-Toh K, Liinamo AE, Leegwater PA. (2008) Evaluation of the serotonergic genes htr1A, htr1B, htr2A, and slc6a4 in aggressive behavior of golden retriever dogs. Behav Genet. 38(1):55-66.

104. Hejjas K, Vas J, Topal J, Szantai E, Ronai Z, Szekely A, Kubinyi E, Horvath Z, Sasvari-Szekely M, Miklosi A. (2007) Association of polymorphisms in the dopamine D4 receptor gene and the activity-impulsivity endophenotype in dogs. Anim Genet. 38(6):629-33.

105. Hejjas K, Kubinyi E, Ronai Z, Szekely A, Vas J, Miklósi A, Sasvari-Szekely M, Kereszturi E. (2009) Molecular and behavioral analysis of the intron 2 repeat polymorphism in the canine dopamine D4 receptor gene. Genes Brain Behav. 8(3):330-6.

106. Quignon P, Herbin L, Cadieu E, Kirkness EF, Hédan B, Mosher DS, Galibert F, André C, Ostrander EA, Hitte C. (2007) Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. PLoS ONE. 2(12):e1324.

107. Chang ML, Yokoyama JS, Branson N, Dyer DJ, Hitte C, Overall KL, and Hamilton SP. (2009) Intrabreed Stratification Related to Divergent Selection Regimes in Purebred Dogs May Affect the Interpretation of Genetic Association Studies. Journal of Heredity 100(Supplement 1):S28-S36.

108. Hejjas K, Vas J, Kubinyi E, Sasvari-Szekely M, Miklosi A, Ronai Z. (2007) Novel repeat polymorphisms of the dopaminergic neurotransmitter genes among dogs and wolves. Mamm Genome. 18(12):871-9.

109. Våge J, Wade C, Biagi T, Fatjó J, Amat M, Lindblad-Toh K, Lingaas F. (2010) Association of dopamine- and serotonin-related genes with canine aggression. Genes Brain Behav. 9(4):372-8.

110. Takeuchi Y, Kaneko F, Hashizume C, Masuda K, Ogata N, Maki T, Inoue-Murayama M, Hart BL, Mori Y. (2009) Association analysis between canine behavioural traits and genetic polymorphisms in the Shiba Inu breed. Anim Genet. 40(5):616-22.

111. Dodman NH, Karlsson EK, Moon-Fanelli A, Galdzicka M, Perloski M, et al. (2010) A canine chromosome 7 locus confers compulsive disorder susceptibility. Mol Psychiatry 15: 8-10.

112. Wang et al., (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528-533.

113. Smoller JW, Gardner-Schuster E, Covino J. (2008) The genetic basis of panic and phobic anxiety disorders. Am J Med Genet C Semin Med Genet. 148C(2):118-26.

114. Hamilton SP (2009) Linkage and association studies of anxiety disorders. Depress Anxiety. 26(11):976-83.

115. Otowa T, Yoshida E, Sugaya N, Yasuda S, Nishimura Y, Inoue K, Tochigi M, Umekage T, Miyagawa T, Nishida N, Tokunaga K, Tanii H, Sasaki T, Kaiya H, Okazaki Y. (2009) Genome-wide association study of panic disorder in the Japanese population. J Hum Genet. 54(2):122-6.

116. Otowa T, Tanii H, Sugaya N, Yoshida E, Inoue K, Yasuda S, Shimada T, Kawamura Y, Tochigi M, Minato T, Umekage T, Miyagawa T, Nishida N, Tokunaga K, Okazaki Y, Kaiya H, Sasaki T. (2010) Replication of a genome-wide association study of panic disorder in a Japanese population. J Hum Genet. 55(2):91-6.

117. Erhardt A, Czibere L, Roeske D, Lucae S, Unschuld PG, Ripke S, Specht M, Kohli MA, Kloiber S, Ising M, Heck A, Pfister H, Zimmermann P, Lieb R, Pütz B, Uhr M, Weber P, Deussing JM, Gonik M, Bunck M, Keßler MS, Frank E, Hohoff C, Domschke K, Krakowitzky P, Maier W, Bandelow B, Jacob C, Deckert J, Schreiber S, Strohmaier J, Nöthen M, Cichon S, Rietschel M, Bettecken T, Keck ME, Landgraf R, Müller-Myhsok B, Holsboer F, Binder EB. (2010) TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. Mol Psychiatry. 2010 Apr 6 [Epub ahead of print].

118. van den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Möller HJ, Hettema JM, Giegling I, Bukszár J, Rujescu D. (2008) Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Arch Gen Psychiatry. 65(9):1062-71.

119. Hettema JM, van den Oord EJ, An SS, Kendler KS, Chen X. (2009) Follow-up association study of novel neuroticism gene MAMDC1. Psychiatr Genet. 19(4):213-4.

120. Terracciano A, Sanna S, Uda M, Deiana B, Usala G, Busonero F, Maschio A, Scally M, Patriciu N, Chen WM, Distel MA, Slagboom EP, Boomsma DI, Villafuerte S, Sliwerska E, Burmeister M, Amin N, Janssens AC, van Duijn CM, Schlessinger D, Abecasis GR, Costa PT Jr. (2010) Genome-wide association scan for five major dimensions of personality. Mol Psychiatry. 15(6):647-56.

Chapter 2: Genetic Investigations of the Domestic Dog

This chapter represents work that sets the stage for the laboratory and analytical work described in later chapters. Given the relative novelty of the methods we have used in this work, a series of preliminary studies were carried out to determine suitability of various SNP genotyping technologies for the experiments comprising the main research goals. Similarly, we carried out our own assessments of the nature of genetic stratification in the canine genome, which would help guide study design and analysis for my thesis work. Although the studies composing the body of my dissertation work consist only of DNA samples obtained through whole blood, I first present work by our group describing the validation of DNA samples obtained from saliva for use on whole-genome genotyping arrays. The purpose of this study is two-fold; first, to demonstrate that high-quality genotyping data can be obtained through use of saliva samples, and that these genotypes are of high fidelity when compared to genotype data from blood of the same animal; second, and more broadly, to validate the use of Illumina arrays for high-throughput genotyping in the dog. The second section of this chapter follows similar themes, with a description of our work investigating within-breed stratification in a subset of our sample dogs and the impact this type of breed substructure can have on genome-wide association studies (GWAS). We demonstrate validation of yet another genotyping platform – the canine array produced by Affymetrix – as well as describing apparent substructure in our sample of Border collies. This section also explores the consequences of including stratified samples in a GWAS, as well as means of statistically correcting for population substructure to reduce false positives.

Our successful experiments on samples of even questionable provenance on multiple new forms of array technology highlight the production of high-quality, reliable genotyping data that can then be used for genome-wide surveys of genetic loci underlying complex traits. Additionally, a thorough understanding of our samples and the substructure that exists within them due to

sampling at different canine community events as well as across international borders allows us to take the necessary precautions to minimize confounding due to breed stratification. With this information in hand, I will then conclude Chapter 2 with a description of the materials and methods used for conducting both genome-wide surveys described in Chapters 3 and 4.

2.1 Genotyping Doggie Drool: It May Not Be as Ruff as You Think*

2.1.1. Introduction

Assemblies of the *Canis familiaris* genome [1,2] have facilitated genomic research in the domestic dog, fostering discovery of genetic loci influencing a range of canine traits and diseases. Though targeted gene-mapping efforts using microsatellite markers and resequencing of candidate genes have resulted in discoveries for traits with simple hereditary patterns, the study of complex disease and behavioral phenotypes has proven to be very challenging. However, with over 2.5 million single-nucleotide polymorphisms (SNPs) annotated on the canine genome, the potential for performing unbiased surveys for genetic loci underlying traits via GWAS has become a practical tool for canine geneticists, leading to compelling association signals for traits with reduced genetic complexity [3–9]. Even a GWAS performed for presumably more complex phenotypes such as canine compulsive disorder (potentially analogous to human obsessive-compulsive disorder) has rendered promising results in a single genomic region [10], as summarized in the previous chapter. Array-based genotyping platforms are now available and provide data for tens- to hundreds- of thousands of SNPs across the dog genome in a single genotyping assay.

*Portions of this section have been published in the manuscript: "Array-based whole-genome survey of dog saliva DNA yields high quality SNP data." Yokoyama JS, Erdman CA, Hamilton SP. *PLoS One*. (2010) May 25;5(5):e10809. **JSY** – study design, sample recruitment and ascertainment, data QC, statistical analysis, manuscript preparation; CAE – sample handling, DNA extractions, contributions to extraction Methods.

Because array-based genome-wide genotyping platforms require large quantities of high quality genomic starting material, DNA for such studies has traditionally been obtained from whole blood. However, with increasing demands for large sample sizes to ensure statistical power to detect multiple signals of modest effect as is expected for complex phenotypes, obtaining whole blood samples from large numbers of dogs becomes challenging. In fact, sampling can even become the limiting factor when studying behavioral traits such as severe anxiety disorders where handling by a clinician in itself causes great duress to the animal, and is often only possible with sedation.

The utility of dog DNA obtained from buccal sampling is well established for microsatellite marker typing, targeted SNP genotyping and limited resequencing. We have found that use of whole-genome amplification (WGA) provided sufficient quantities of genomic material for use in higher throughput multiplex genotyping assays surveying up to several hundred SNPs [11]. Although WGA of canine buccal DNA produces reasonable (~3 μg) quantities of total DNA [12], previous studies by our group suggest that only 3-15% (90-450 ng) of this total WGA sample actually represents canine DNA [11]. Use of WGA buccal DNA from dogs on genome-wide arrays—which require 250-500 ng of genomic DNA input—presents a challenge given the level of microbial DNA contamination. Preliminary studies by other groups have found performance of buccal swab DNA on Illumina's Infinium canine array to be modest, suggesting the total amount of canine DNA present in WGA buccal samples is insufficient for high-quality data production for use in GWAS (MW Neff, personal communication).

Another mode of DNA sampling that has gained increasing utilization is saliva collection, from which DNA has been shown to be of equivalent quality as blood-extracted DNA [13]. The most notable strengths of saliva collection involve convenience: 1) samples can be collected at home by users themselves; 2) once saliva is mixed with stabilization buffer samples are stable for

several months at room temperature; and 3) saliva can be sent through postal mail and across international borders without infringement of shipping laws or animal welfare restrictions. Saliva collection has a higher sample return rate than blood in human subjects [13,14]. Additionally, bacterial DNA content has been reported to compose only 16.1% of the total DNA obtained from canine saliva samples [15]. Perhaps most importantly, saliva collection provides a painless, non-invasive alternative to venous draws—one of the main reasons many researchers have switched to saliva collection for research in infants and children.

Saliva-extracted DNA has been demonstrated to be of equivalent quality as blood-extracted DNA in humans [13]. Very recently, Mitsouras and Faulhaber [17] also demonstrated high yields of high quality DNA from canine saliva, sufficient for PCR-restriction fragment length polymorphism (RFLP) genotyping. We therefore proposed saliva collection as an alternative to blood draws for obtaining DNA samples from dogs in a minimally invasive fashion for use on genome-wide genotyping platforms to yield high-quality data for use in GWAS. We describe here our verification of DNA yield and quality, genotyping performance, copy number variant (CNV) calling, and data quality via comparison with blood-extracted DNA samples. We also report owner feedback from kit usage and highlight the utility of saliva collection for future studies in canine genetics.

2.1.2. Materials and Methods

2.1.2.a. Samples. Saliva and blood samples were collected from four bearded collies (BEC) and one Border collie (BOC) in the context of our ongoing genetic studies of canine behavior. Saliva only was also obtained from six additional BEC recruited for the same study. Saliva samples were collected by owners using the Oragene·ANIMAL (OA-400 Tube Format, DNA Genotek, Ontario, Canada) kit as per manufacturer's instructions. Briefly, saliva was collected from dog's mouth using 2-3 absorbent sponges (http://www.dnagenotek.com/DNA_Genotek_

Support_Lit_UI_ANIMAL.html). After sample collection, DNA was preserved by placing the sponges in Oragene·ANIMAL stabilization solution, labeled, and then sent to our laboratory by mail. All saliva samples were stored at room temperature before and after shipping. Blood samples were obtained by 3-5 ml blood draw. All animal work was approved by the local review committee.

2.1.2.b. DNA extraction. Extraction of dog DNA was performed as suggested by manufacturer's instructions except as noted based on our lab's extensive experience with human saliva DNA. Samples were incubated for two hours in water at 50°C. Swabs absorbed the full volume of stabilization buffer in addition to saliva, and thus required manual extraction ('squeezing' with sterile tweezers) to remove solution for use in extraction. The solution was collected in original holding container, and then 500 µl was aliquoted via pipette into a 1.5 ml microcentrifuge tube. Absorbent sponges were kept in remaining stabilization solution in the event that additional extractions were required. Because twice the amount of solution was aliquoted for extractions, 20 µl of Purifier Solution was used, and the use of glycogen was omitted. The Oragene Animal protocol contains a NaCl step to ensure efficient recovery of DNA; as this step is not in the Human protocol and was added between the two versions of the Animal protocol that we performed (beta testing kit courtesy of DNA Genotek vs. published version PD-PR-095 Issue 2.1), extractions were carried out both with and without the use of NaCl on the same sample (which was not used for the reported genotyping). The single BOC sample was extracted without the NaCl step via the beta kit instructions, whereas the ten BEC samples were all extracted using the NaCl step from the updated protocol. For the final hydration step, 100 µl of Hydration Buffer from the Qiagen kits used for blood extractions (Qiagen Inc., Valencia CA) was used to rehydrate DNA, and samples were incubated for at least 24 hours at room temperature prior to final storage at 4°C. Blood sample DNA was extracted in-house using standard methods with the Puregene Blood Kit (Qiagen Inc.). All animal work was approved by Institutional Animal Care and Use

Program at the University of California, San Francisco (AN079848-02). All dogs were recruited from private owners, who consented to use of de-identified data for research purposes.

2.1.2.c. Quantification of DNA. Quantification of all extracted DNA samples was performed on a NanoDrop (ND-1000 v3.3.0) spectrophotometer (Thermo Fisher Scientific Inc., Wilmington DE). Quantification of saliva-extracted samples was not corrected as per notes suggested by DNA Genotek (Laboratory Protocol PD-PR-095 Issue 2.1), but rather were reported as calculated by the NanoDrop for direct comparison with results published by other groups for human saliva and blood. For more details, please see **(2.1.4) Discussion**.

2.1.2.d. Genotyping. Samples were genotyped on the Infinium Canine SNP20 BeadChip (Illumina Inc., San Diego CA) by the Genomics Core Facility at the University of California, San Francisco. Genotypes were called and quality control (QC) was conducted in-house using the GenomeStudio Data Analysis Software package (1.0.2.20706, Illumina Inc.). Clusters of all samples with GenTrain Scores (a measure of reliable SNP detection) <0.60 were visually assessed for quality and either manually reclustered or zeroed due to poor performance (i.e., excluded from the data set). Further exclusion criteria removed SNPs with call rates <95% or minor allele frequency (MAF) <0.02. Genotyping was also performed on a subset of saliva-blood replicate samples using the next-generation Infinium canine array, CanineHD. Genotyping was also performed by the Genomics Core Facility and genotypes were called and QC'ed as described above.

2.1.2.e. CNVs. Copy number variation was evaluated *in silico* with the GenomeStudio software (cnvPartition v2.4.4, Illumina Inc.) using default criteria. One predicted CNV locus was also evaluated by direct PCR of two genomic segments within the putative deletion region using the following primer pairs: (PLSCR1exon amplicon) forward 5'-

TCTAAACCCAGGATTAGCAAGAA-3', reverse 5'-CCATGTAATTTTGATAGGGTATTTCA-3' and (CFA23CNV44Mb amplicon) forward 5'-TGTAAACCTCATTTCACTTACATGG-3', reverse 5'-GGTCCATGGAGGACTCTCTCT-3'. Platinum-Taq was used to amplify segments with a 58ºC touchdown protocol in presence of 0.4μM primer, 100μM dNTPs, 2.5mM Mg and 1mM Betaine.

2.1.3. Results

2.1.3.a. Saliva sample collection. Twelve sample kits were sent to six BEC owners who previously consented via written communication to participate in the sample collection. Of those, four owners representing 10 kits (dogs) returned samples to our laboratory, representing a 67% by-owner and 83% by-dog return rate. Surveys sent out with the beta testing version of the Oragene·ANIMAL saliva collection kits reported that owners found the collection to be very easy overall. For all owners, sample collection was successful and took less than 10 minutes.

Source	n	Concentration (ng/ul)	260/280	Concordance (%)
Saliva	11	125.5 (46.9-212.4)	1.67 (1.39-1.86)	-
Blood	5	384.4 (317-521.2)	1.96 (1.84-2.24)	-
Saliva vs. Blood *(SNP20)*	4	-	-	99.9 (99.9-100)
Saliva vs. Blood *(HD)*	11	-	-	100 (99.9-100)

Table 2.1: Concentration, purity and concordance of saliva- versus blood-extracted DNA samples. Mean values plus ranges for DNA concentration and 260/280 ratios (as a measure of purity) as calculated by NanoDrop spectrophotometer for saliva and blood samples, and mean genotype concordance for paired samples from the same individual. Concordance is the proportion of agreeing genotype calls over total genotypes that were called for both samples (saliva and blood). Source – tissue source of DNA extraction; n – number of dogs or pairs (for concordance).

2.1.3.b. DNA yield. DNA extraction was successful for all saliva samples received from owners. For each sample, 500 μl of saliva-buffer solution was easily extractable from the swabs, with

additional volume remaining after the liquid transfer step to the 1.5 ml tube, thus allowing for another extraction for more DNA if necessary. The extraction protocol was very straightforward, and DNA fibers were visible for all extractions performed. When quantified via NanoDrop spectrophotometer, saliva-extracted samples (n = 11) had a mean concentration of 125.5 ng/µl (Table 2.1), for an average yield of 12.6 µg of total DNA. This compares to the mean concentration of 384.4 ng/µl for our comparison blood-extracted samples (n = 5). The 260/280 mean for all saliva-extracted samples was 1.67, as compared to blood samples that had a mean of 1.96 (Table 2.1). However, the 260/230 mean for saliva-extracted samples was much lower than that of comparison blood samples, with an average of 0.53 for saliva versus 1.61 for blood (Table 2.2). Low 260/230 ratios suggest presence of contaminants, which absorb at 230nm.

	Saliva Samples					Blood Samples				
	ng/ul	260/280	260/230	Call Rate	p10 GC	ng/ul	260/280	260/230	Call Rate	p10 GC
Laboratory Control	46.91	1.79	0.44	0.999	0.805	317.00	2.35	0.47	0.999	0.806
Sample 1[a]	71.21	1.50	0.34	0.990	0.792	521.22	1.86	1.93	0.942	0.762
Sample 2	127.05	1.75	0.63	0.999	0.805	345.50	1.88	1.89	1.000	0.805
Sample 3	171.16	1.80	0.76	0.997	0.804	369.31	1.86	1.95	1.000	0.806
Sample 4	212.43	1.84	0.79	0.995	0.799	369.05	1.84	1.80	1.000	0.806
Sample 5	83.93	1.56	0.37	1.000	0.806	*NA*	*NA*	*NA*	*NA*	*NA*
Sample 6	100.38	1.80	0.61	0.997	0.802	*NA*	*NA*	*NA*	*NA*	*NA*
Sample 7	92.71	1.86	0.71	0.996	0.800	*NA*	*NA*	*NA*	*NA*	*NA*
Sample 8	146.41	1.58	0.44	0.986	0.786	*NA*	*NA*	*NA*	*NA*	*NA*
Sample 9	161.95	1.39	0.33	0.999	0.804	*NA*	*NA*	*NA*	*NA*	*NA*
Sample 10	165.97	1.45	0.39	1.000	0.806	*NA*	*NA*	*NA*	*NA*	*NA*
Mean	**125.46**	**1.67**	**0.53**	**0.996**	**0.801**	**384.42**	**1.96**	**1.61**	**0.988**	**0.797**
Mean (no outlier[a])	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1.000**	**0.806**
Hansen *et al.* (2007 - humans)	108.00	1.63	0.80	*NA*	*NA*	56.80	1.79	1.44	*NA*	*NA*
Iwasiow *et al.* (2009 - dogs)	232.00	1.60	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

[a] *Sample 1 removed as blood sample genotyping performance outlier*

Table 2.2: Individual saliva and comparison blood sample statistics for DNA extraction and genotyping compared to published data for human saliva and blood, plus manufacturer's report.

Individual statistics are given for each saliva and comparison blood sample for DNA concentration (ng/ul), DNA purity (260/280), contamination (260/230), post-QC genotype call rate and post-QC p10 GenCall score (p10 GC). Saliva vs. blood sample concordance rates are also given for every individual represented by both tissue types. Mean values as reported in the main text are provided, as well as the published values for human saliva and blood samples as reported by Hansen et al. [13] and dog saliva statistics as reported by researchers from the manufacturer of the Oragene·ANIMAL collection kit, DNA Genotek [15].

2.1.3.c. Genotyping. Illumina's Infinium Canine SNP20 genotyping array was developed by Illumina to survey the canine genome at sufficient coverage for use in GWAS as suggested by Lindblad-Toh *et al.* [2]. The array contains 22,362 SNPs with a median of 565 markers per chromosome (mean 573.4, maximum 1,146, minimum 267). The average intermarker spacing is

103.6kb, with median intermarker spacing of 67.8kb. Several very large gaps inflate this mean, with the largest gap at 5.6Mb (on the X chromosome). There are 3, 4 and 27 gaps that are >3Mb, >2Mb and >1Mb in size, respectively, and 292 SNPs with gaps >500kb. The average call rate for the 22,362 SNPs surveyed by the Infinium CanineSNP20 array before QC of SNP data was 99.2% for saliva samples (n = 11) and 98.5% for comparison blood samples (n = 5; Table 2.3). This compared to an average call rate of 99.4% for all blood samples genotyped on this platform by our group (n = 192, data not shown). The mean genotyping statistics for the five comparison blood samples included in this report are lower than the overall average we saw in our total samples because of one poor-performing sample (see Table 2.2). When this poorly performing sample was removed, the mean call rate was 99.6% (Table 2.3). After QC, 20,753 SNPs remained, with average call rates of 99.6% for saliva samples and 98.8% for all comparison blood samples (Table 2.3).

Source	n	# SNPs	Call Rate (%)
Pre-QC			
Saliva	11	22,362	**99.2** (98.5-99.6)
Blood	5	22,362	**98.5** (94.2-99.6)
Blood (no outlier)	4	22,362	**99.6** (99.5-99.6)
Saliva	7	173,662	**99.5** (99.2-99.8)
Blood	8	173,662	**99.7** (99.6-99.8)
Post-QC			
Saliva	11	20,753	**99.6** (98.6-100)
Blood	5	20,753	**98.8** (94.2-100)
Blood (no outlier)	4	20,753	**100** (99.9-100)
Saliva	7	169,361	**100** (99.6-100)
Blood	8	169,361	**100** (99.9-100)

Table 2.3: Genotyping statistics for saliva- and blood-extracted DNA samples before and after marker QC.

Mean values plus ranges for the call rates for each sample type before and after marker QC. Source – tissue source of DNA extraction; n – number of dogs; # SNPs – number of total markers used for calculating statistics across samples; Call Rate – in percent, with range in parentheses.

Another useful metric for evaluating sample quality and performance is the Illumina GenCall score (GenCall Version 6.3.0), which is calculated for each genotype. GenCall scores range from 0 to 1, with smaller values representing data points that fall further from the center of the genotype call cluster with which the sample is associated. Genotypes with a GenCall score ≤ 0.15 received no call. Post-QC, average Illumina 10% GenCall scores—the 10^{th} percentile (p10) of the range of GenCall scores across all genotypes called for the individual—were 0.801 and 0.797 for saliva and comparison blood, respectively (Table 2.3; the average for all blood samples genotyped by our group on this platform was 0.803, data not shown). Plotting call rate versus p10 GenCall scores demonstrated that all saliva samples performed equally well as comparison blood samples after QC (Figures 2.1-2.2). Mean genotyping statistics excluding the performance outlier are also provided in Table 2.3, and genotyping statistics after QC for each sample are given in Table 2.2. We have also recently observed high concordance rates (99.9999%) between blood and saliva replicate pairs on the next generation Illumina canine array with 170,403 QC-filtered SNPs, with a mean call rate for saliva samples of 99.78% (n = 3, Table 2.1).

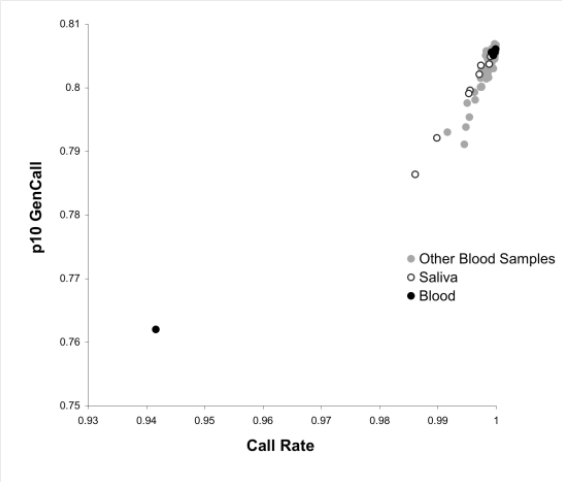


Figure 2.1: Plot of p10 GenCall score versus call rate after QC for saliva- and blood-extracted DNA samples compared to full set of genotyped samples.

Sample 10% (p10) GenCall score is plotted against sample call rate as a means of visualizing overall sample performance. Each sample is represented by one data point, with saliva samples represented by open circles, comparison blood samples as filled circles and remaining blood samples also genotyped by our group as grey circles (n = 192 – see section 3 of this chapter). Overall, saliva samples performed in the same range as all blood samples genotyped. (Please note: axes do not start at the origin.)

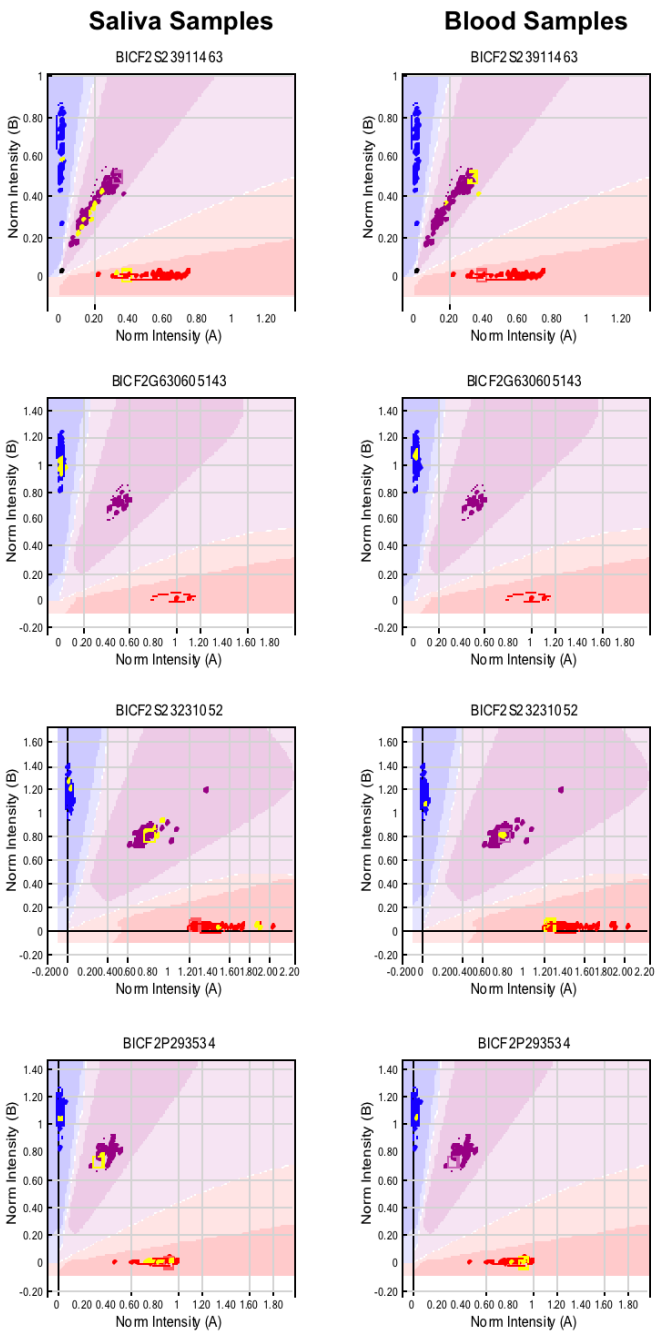


Figure 2.2: Cluster plots for select SNPs.

Saliva- versus blood-extracted DNA samples (by columns) are highlighted in yellow in cluster plots of genotyped samples (n = 192 – see section 3) from GenomeStudio. Each sample is represented by one dot, and clusters represent genotype (AA in red, AB in purple, BB in blue).

Of the 11 saliva samples genotyped, five dogs were also represented by blood samples. For four of the replicate samples, mean concordance of called genotypes in both samples (saliva and blood) was 99.98% (Table 2.1). One replicate sample was dropped from our analysis due to low concordance that suggested within-breed sample mixing (Table 2.2). We examined the characteristics of the SNPs responsible for sample discordance in the remaining four samples to see if particular marker characteristics may predict discordance. However, we found that only one out of 28 discordantly called SNPs had >1 discordant call, whereas the majority were discordant singletons (Table 2.4). Binning markers by MAF suggested a trend towards higher frequencies in discordantly called SNPs (Figure 2.3), though the mean MAF for the 28 discordant SNPs was very similar to that of the entire marker set (Table 2.4). It also appears that the discordant SNPs had lower performance than the full marker set; however, the averages between the two sets were not markedly different (Table 2.4).

		Mean Call Statistics				
SNP set	**# SNPs**	**Call Rate (%)**	**Rep Errors**	**MAF**	**GenTrain Score**	**p10 GC**
Full Marker Set	22362	99.4	-	0.248	0.854	0.871
Discordant SNPs	28	98.7	1.036	0.241	-	0.763
>1 Discordant calls	1	96.9	2	0.449	-	0.541

Table 2.4: Mean genotyping statistics for discordantly called markers compared to full marker set. Mean statistics are provided for the full marker set (before QC) as well as for SNPs whose genotypes were called discordantly between saliva versus blood replicate samples. # SNPs – number of total markers used for calculating statistics across samples; Rep Errors – number of replicate errors (discordant genotype calls); MAF – minor allele frequency; p10 GC – 10^{th} percentile of GenCall score range.

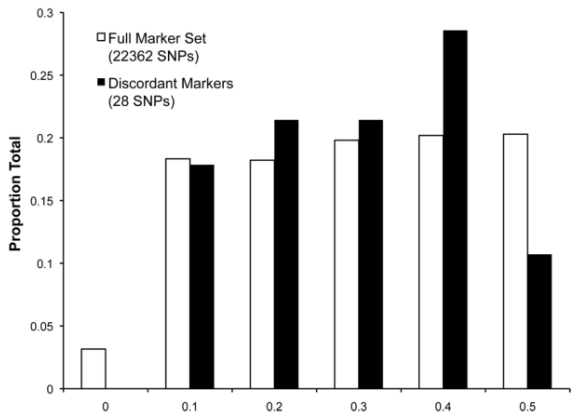


Figure 2.3: Allele distributions for discordant markers compared to full marker set.

Histogram of allele frequencies for the full marker set (open bars) and discordant SNPs (solid bars). Discordant markers appear to trend towards larger allele frequencies. X-axis – allele frequency upper bound for bin; Y-axis – proportion of total (respective) marker set.

2.1.3.d. CNVs. Copy number variation can be readily evaluated with SNP data within the GenomeStudio software package. CNVs called *in silico* were evaluated in all genotyped samples and those appearing to specifically include a subset of the saliva-extracted samples were further assessed for validity via manual inspection of genotype data. A region on *Canis familiaris* chromosome (CFA) 23 had copy number losses predicted for Sample 2 (homozygous loss) and Sample 5 (heterozygous loss), as well as predicted homozygous loss in three other BEC blood samples (data not shown). To validate these calls, we investigated this region via direct PCR of genomic samples for two amplicons located within the putative deletion region: PLSCR1exon amplicon designed to span the 8th exon of the *PLSCR1* gene, and CFA23CNV44Mb amplicon designed to span a predicted conserved region (annotated in the UCSC Genome Browser) in the middle of a hypothesized minimally deleted region based on no-call genotypes in the three samples predicted to have homozygous deletions (Figure 2.4a). PCR results confirmed deletion of the hypothesized minimally deleted region in all samples with homozygous deletion calls but

presence of the PLSCR1exon region as expected from present genotype calls in homozygous loss samples (Figure 2.4b).

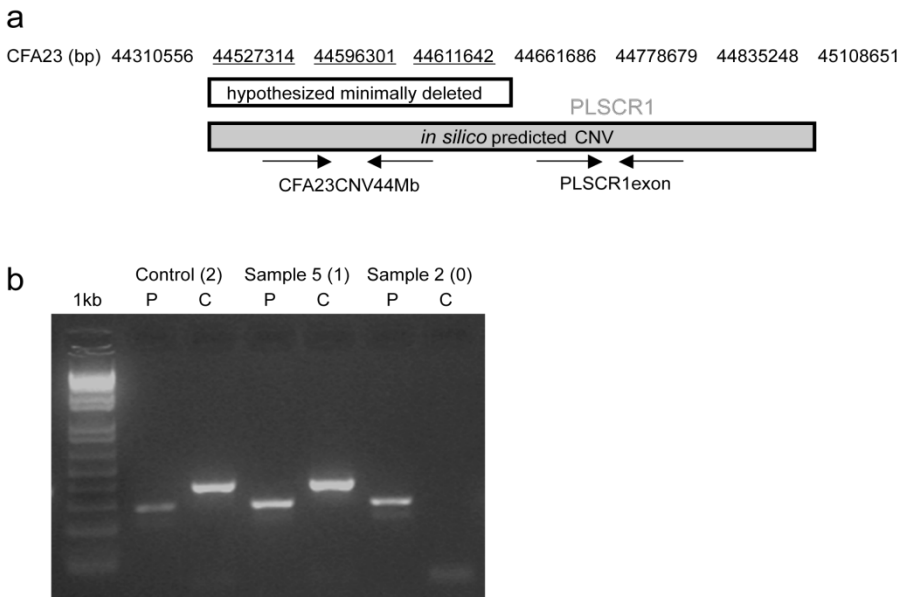


Figure 2.4: Molecular evaluation of putative CNV region on CFA23.

(a) Two amplicons within the *in silico* predicted copy variable region on CFA23 were evaluated for presence/absence in saliva samples that had called copy loss. Base position (bp) on CFA23 is given at top of diagram (not to scale). PLSCR1exon - amplicon is an exonic region of the *PLSCR1* gene; CFA23CNV44Mb - amplicon is a predicted conserved region in the middle of a hypothesized minimally deleted block based on no-call genotypes for three SNPs spanning this region (indicated by underlined base positions). (b) PCR amplicons visualized by UV on 2% agarose gel with 1kb DNA ladder for reference. Predicted size is 293 and 389 bases for the PLSCR1exon ("P") and CFA23CNV44Mb ("C") amplicons, respectively. Sample identities are provided with predicted copies present in parenthesis; Control - Lab Control blood sample predicted to have no loss (i.e. 2 copies present).

2.1.4. Discussion

Our results demonstrate that saliva collection from dogs is facile, convenient, and yields large amounts of high-quality DNA that provide excellent performance on high-throughput whole

genome arrays. Overall, the DNA yield our group obtained was similar to that found in a previous study examining human saliva specimens [13]. Our mean yield was higher than another research group's [17] but lower than the reported yield by the kit manufacturer for canine samples, although it should be noted that Iwasiow, *et al.* [15] report corrected ratios that were adjusted for presence of turbid material that absorb at 320nm, a step that we elected not to perform. The DNA purity we obtained (as measured by 260/280 ratios) was, however, similar to reported values for both human and canine saliva samples (Table 2.2). Our results also suggest that the extraction method used for saliva samples is important, and that subtle differences in extraction protocols may produce differences in DNA purity and/or introduce contaminants, though in our case this did not appear to alter genotyping performance. Whether or not NaCl was used in extractions appeared to produce slight differences in yield, but did not appear to alter DNA purity or level of contamination as measured by 260/280 or 260/230 ratios, respectively (Table 2.2). Other aspects of quality to be explored in the future include measuring levels of contaminating RNA and microbial DNA load, determining if DNA is of high molecular weight, and investigating long-term stability post-extraction.

Genotype concordance between blood and saliva samples from the same individual was equally high as that seen between replicate blood samples from the same individual (data not shown), which further demonstrates the high fidelity of genotypes obtained from saliva samples. However, because DNA from different tissues may produce source-specific profiles with regard to probe fluorescence (Figure 2.2)—which may ultimately affect genotype calling—it is prudent that samples for association studies have balanced representation of cases and controls from each DNA source to reduce spurious associations due strictly to tissue type, sample provenance, and genotyping batch effects [16]. Excluding the outlier sample (which was suspected of sample mixing) and examining the SNPs that were called discordantly between highly concordant replicates, we found only one marker that was called discordant >1 time. This suggests the

discordance is random, and that saliva does not lead to differential discordance when compared to blood.

Because they provide high-fidelity SNP genotypes, it appears that saliva-extracted samples can also be used for successful CNV calling *in silico*. Calls are made for putative CNVs and regions of homozygosity based on genotypes across multiple markers. It thus follows that the size of putative copy variable regions relies on the density of the SNP data, and that the size of a reported CNV may be artifactually large due to the requirements of the calling algorithm. Because of the large inter-SNP distances in our data set, direct assessment of genotype calls was therefore also used to hypothesize a *minimally* deleted region in our samples (where no-calls suggest absence of region), which we verified by direct amplification of genomic DNA. Our results demonstrate that saliva samples can also be used reliably in copy variation analysis, although similar requirements for case-control tissue sample consistency still apply.

We found one saliva sample that demonstrated low concordance (82.6%) with its replicate blood sample. This was likely due to switching samples of dogs within the same breed, as pair-wise concordance between known but different dogs of the same breed was similar to that seen in our low-concordance sample, whereas concordance between known dogs of different breeds was much lower (Table 2.5). Further, our calculations of concordance between known related versus known unrelated dogs suggested that sample switching likely occurred between related dogs. Our analysis also suggested the switching was specifically in the saliva sample, as concordance between the saliva sample and the dog's sire was lower than that of the blood sample and sire. The concordance we saw between the saliva versus blood samples was similar to that of distantly-related dogs, which suggests the sample switching could have resulted from mislabeling or sampling the wrong dog from a household with multiple related dogs (Table 2.5). Additionally, chimeric samples due to dogs licking each other or sharing water bowls could produce

heterogeneous genotyping results and warrants further investigation, although this would very likely generate heterozygosity outliers.

Sample A	Tissue A	Breed A	Origin	Sex A	Sample B	Tissue B	Breed B	Origin	Sex B	# Correct	# Errors	Total	Rep Freq	Relation	comments
Lab Control	Saliva	BOC	US	F	Lab Control	Blood	BOC	US	F	22246	0	22246	1.000		
Sample 1	Saliva	BEC	US	M	Sample 1	Blood	BEC	US	M	20839	24	20863	0.999		
Sample 2	Saliva	BEC	US	M	Sample 2	Blood	BEC	US	M	15187	7051	22238	0.826		***suspected different samples***
Sample 3	Saliva	BEC	US	M	Sample 3	Blood	BEC	US	M	22208	0	22208	1.000		
Sample 4	Saliva	BEC	US	M	Sample 4	Blood	BEC	US	M	22157	5	22162	1.000		
Sample 2	Blood	BEC	US	M	Sample 6	Saliva	BEC	US	M	17257	4941	22198	0.882	father-son	*high concordance suggestive of true relationship*
Sample 2	Saliva	BEC	US	M	Sample 6	Saliva	BEC	US	M	15061	7130	22191	0.824	father-son	***lower concordance suggests saliva sample is switched***
Sample 5	Saliva	BEC	US	F	Sample 6	Saliva	BEC	US	M	15313	6889	22202	0.830	1 shared grandparent	*known different samples, distantly related*
Sample 1	Saliva	BEC	US	M	Sample 3	Saliva	BEC	US	M	14279	7706	21985	0.806	unrelated	*known different samples, US vs. US*
Sample 11	Blood	BEC	foreign	M	Sample 12	Blood	BEC	foreign	M	14511	7749	22260	0.807	unrelated	*known different samples, foreign vs. foreign*
Sample 11	Blood	BEC	foreign	M	Sample 6	Blood	BEC	US	M	14524	7682	22206	0.809	unrelated	*known different samples, US vs. foreign*
Sample 6	Saliva	BEC	US	M	Lab Control	Saliva	BOC	US	F	11810	10385	22195	0.729	unrelated	*known different samples, different sex and breed*
Sample 11	Blood	BEC	foreign	M	Lab Control	Blood	BOC	US	F	11752	10500	22252	0.727	unrelated	*known different samples, different origin, sex, breed*

Table 2.5: Replicate statistics.

Sample identification, tissue source, breed, geographic origin (US vs. foreign) and gender are given for samples (A versus B) that were compared for replicate (concordance) statistics. Sample 2 is suspected to be a switched sample, and demonstrates similar concordance rates as distantly related dogs of the same breed. Samples 11 & 12 are dogs from a geographically distinct population (see section 3). BEC - bearded collie; BOC - Border collie. # Correct – total concordant genotype calls; # Errors – total discordant genotype calls; Total – total number of markers with genotype calls in both samples; Rep Freq – replicate frequency (concordance rate); Relation – unrelated refers to dogs that share no grandparents.

One caveat of this work is that the blood samples were genotyped on a separate run several months earlier than the saliva samples, which could introduce artifacts when comparing genotypes and statistics for samples representing the same individual. However, it is more likely that these artifacts would introduce inconsistencies between duplicated samples; this would result in an underestimation of the total concordance seen between duplicates in our study. As our concordances are already greater than 99% (excluding the suspected wrong sample pairing), this suggests that even higher fidelity in genotype calls between blood- versus saliva-extracted DNA samples may be possible if all samples are run in the same genotyping batch. The similarity of clustering data also suggests that samples of diverse provenance can be clustered together using Infinium data. We also observed high concordance rates between blood and saliva replicate pairs

on the next generation Illumina 170K CanineHD arrays, further demonstrating very strong performance by saliva-extracted DNA samples on even the most recent genotyping technologies.

One limitation to this study is the ascertainment bias introduced by our study design. Because we required prior written assent from owners to participate in saliva sample collection prior to kits being sent out, it is likely that our return rates are overestimates of the population at large. However, because saliva collection is so simple and non-invasive, it is probable that return rates would be quite significant, and likely higher than the rate of blood sample collection. Another limitation of this study is the small number of duplicated samples. This limitation highlights the need for replication with larger numbers of dogs from different breeds (large and small), and on different genotyping platforms by other groups for further validation of the performance of saliva-extracted DNA for high-throughput assays.

In summary, we demonstrate for the first time that saliva sample collection in dogs is a noninvasive means of obtaining high quality DNA for successful use with genome-wide array genotyping, with little danger of loss of information due to the source of data. The dual conveniences of owner sampling in the home and ease of shipping provide alternative means of obtaining samples from rural locales or foreign countries where collection of blood samples may be difficult or impossible. Additionally, ease of sampling allows for collection of large numbers of samples with minimal investment of time and manpower, creating potential for collecting an entire study cohort at a small number of targeted sampling events. Finally, the non-invasive nature of saliva collection makes it particularly appealing when studying dogs whose conditions may otherwise prevent blood collection, such as high levels of anxiety or repeated use of veins for other medical purposes related to disease status. In sum, these factors will lead to increased sample return rates which will increase study sizes and ultimately enhance the ability for

geneticists to detect novel genetic loci underlying disease and behavioral traits in a GWAS framework.

2.2 Within-Breed Substructure Complicates Genome-Wide Association Studies[**]

2.2.1. Introduction

As mentioned in Chapter 1, the domestic dog has become an increasingly popular genetic model for use in studies of disease and behavior. However, several assumptions are made in most genetic association studies using purebred dogs. First, it is assumed that purebred dogs constitute separate, closed, inbred populations exhibiting an intense founder effect. We therefore expect that there will be limited phenotypic and genetic variation within breeds, but broad variation between breeds. These assumptions are probably safe given registration practices for purebred dogs that dictate a dog cannot be considered purebred unless both of its parents are registered purebreds. However, the common expectation, which is that pure breeds make up homogeneous populations, may be problematic.

The present section addresses the phenomenon of "breed splits" and its possible consequences for genetic association studies. Although it is clear from previous evidence, as well as a point of common knowledge among purebred dog owners, trainers, and handlers, that population structure exists within breeds, this structure has not been systematically characterized. Previous studies of population structure in dogs have focused primarily on the relationships between breeds, incorporating relatively small samples of a large number of breeds, and using clustering methods to compare overall degrees of similarity between samples characterized either by microsatellite

[**]Portions of this section have been published in the manuscript: "Intrabreed Stratification Related to Divergent Selection Regimes in Purebred Dogs May Affect the Interpretation of Genetic Association Studies." *Journal of Heredity* (2009) 100(Supplement 1):S28-S36. Chang ML, Yokoyama JS, Branson N, Dyer DJ, Hitte C, Overall KL, and Hamilton SP. MLC – phylogenetic analysis (not described here), pedigree analysis, manuscript preparation for Introduction, Methods & Results of these analyses, Discussion; **JSY** – genotyping and data QC, clustering analysis, GWAS simulations, manuscript preparation for Methods & Results of these analyses; CH – genetic distance analysis. All other co-authors – sample ascertainment, funding.

markers or small numbers of single nucleotide polymorphisms localized to a limited sampling of the genome (e.g. [18]). Existing assessments of within-breed population structure are characterized by restricted genomic coverage [19, 20] or were accomplished via pedigree analysis [21].

We sought to determine if stratification may be predicted by knowledge of sample origin, geography, or selection regime. We incorporated autosomal SNP genotype data with broad genomic coverage, taking advantage of sizable, well-characterized samples in four breeds of interest. We interpreted our findings in the context of owner-reported demographic and pedigree information, in an effort to understand how we may identify probable stratification within samples for future genetic analyses. Finally, we conducted simulations to explore the effects of such stratification on GWAS, and explore strategies for minimizing the risks of false positive results.

2.2.2. Materials and Methods

2.2.2.a. Sample recruitment, collection, and data generation. We recruited and collected samples of four pure dog breeds, for investigation of within-breed stratification in the context of on-going genetic studies of complex behavior and disease. Owners of participating dogs were recruited at dog shows and working competitions (sheepdog trials), and through direct mail, email lists, breed clubs, and training organizations. We drew samples from dogs on site, or asked owners to send blood samples to our laboratory using a standardized protocol. We also collected pedigrees, demographic data, and a detailed behavioral questionnaire [22] for each dog.

Our sample included three herding breeds of interest for a project exploring the genetic background of canine noise phobia, a discrete behavioral phenotype with a probable genetic component discussed in greater detail in Chapter 3: Border collies (BOC, n = 76 total genotyped),

Australian shepherds (AUS, n = 49), and German shepherd dogs (GSD, n = 17). We also included the Portuguese water dog (PWD, n = 17), characterized by a breed community that is enthusiastically supportive of canine genetic studies. Our sample included unrelated dogs selected for GWAS, extended pedigrees segregating noise phobia in two breeds (BOC and AUS), and five small family groups (BOC and AUS, one trio and four quartets) included for assessment of mendelization errors.

We collected whole blood samples of approximately 5 ml from each dog and extracted genomic DNA from each sample using the Gentra Puregene Blood Kit (Qiagen, Valencia, CA). We surveyed approximately 127,000 SNPs per dog using Affymetrix's Canine v2.0 SNP array and called genotypes using the BRLMM-P algorithm. We dropped X-chromosome markers due to poor quality, and filtered the remaining markers for call rate, concordance for a single dog between multiple (four) genotyping runs, significant deviations from Hardy-Weinberg equilibrium, mendelization errors and MAF < 0.02, which resulted in a final dataset of approximately 53,000 SNPs. We also generated marker subsets consisting of 2,100 and 21,000 SNPs that were spaced evenly across the genome for use in the clustering analysis. Multiple marker sets were used to address computational limitations associated with some analyses, and to test the consistency of different-sized marker sets.

2.2.2.b. Cluster and genetic distance analyses. Preliminary cluster analyses were conducted because their use in previous studies [18-20] would afford us comparable assessment of our results. We conducted a preliminary clustering analysis of 2,100 high quality SNPs (100% call rate, median inter-SNP distance 850kb) from all 38 autosomes using *structure* [23-25]. This data set included 48 BOC, 27 AUS, 17 PWD, and 16 GSD, for a total of 108 unrelated dogs and was subjected to 30 iterations of K = 1 through K = 8, where the user-assigned value for 'K' is the number of putative population groups predicted to be present in the given sample. We then used

methods outlined by Evanno, *et al.* [26] to determine the "best fit" or number of population groups predicted given our data set.

Genetic distance analyses were then performed for comparison to previous analyses of the same type that used limited-coverage SNP data [19]. Average genome-wide proportions of alleles sharing identity-by-state (IBS) were calculated pair-wise for 108 dogs from four breeds (48 BOC, 27 AUS, 17 PWD, 16 GSD) across 21,000 uncorrelated SNPs covering all 38 canine autosomes. These were used to create a distance matrix (1-IBS) of 108 x 108 individuals with PLINK v1.02 [27]. The distance matrix was visualized in R, and the number of optimal clusters "K" was calculated with a hierarchical agglomerative clustering method ("agnes" as implemented in R). The optimal value of clusters was calculated to be K = 8. Cluster stability was then assessed for K = 8, as well as for K = 4 (which corresponds to the number of breeds evaluated) via bootstrapping including outliers.

2.2.2.c. Pedigree analysis. As part of data collection for our larger study examining the genetic basis of canine noise phobia, we collected pedigrees and questionnaire information from owners [22], and used this information to characterize the dogs in our sample. Data that proved particularly informative for this purpose were: pedigree information about show ring performance of ancestors (such as show championships) and geographical origin of dogs; type of event at which the sample was collected (i.e., working trial vs. dog show); organization with which the dog is registered; and owner-reported information about titles achieved or activities regularly engaged in with their dogs, as well as the type of breeder from which the dog was obtained (show dog breeder or working sheepdog breeder). Using this information allowed us to describe the individuals of one breed in particular, BOC, as "show dogs" or "working dogs." These types were correlated with different geographical origins, with show dogs tracing back to Australia/New Zealand, and working dogs tracing back to the United Kingdom.

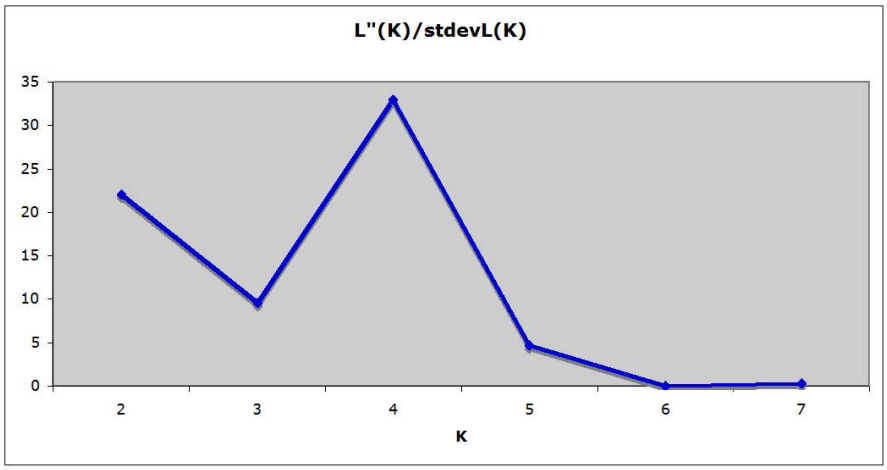
2.2.2.d. Simulation studies. We conducted a simulated case-control association study (100 iterations) using observed genotypes and information about population substructure in BOC revealed by the cluster, distance, phylogenetic (not presented in this chapter; for more information please see published manuscript [28]), and pedigree analyses. Our sample of unrelated BOC split into two groups across all analyses: a larger group of 43 dogs, and a small group of five that were consistently differentiated (see Results (2.2.3), below). We randomly assigned case-control status to the large group of 43 BOC. We then assigned case-control status to the smaller group of five BOC as follows: *split sample* (two randomly assigned as cases and three randomly assigned as controls), *all cases* (with the balance of the 43 randomly assigned case or control status), or *all controls* (with the balance of the 43 randomly assigned case or control status). We performed a genome-wide allelic association analysis on approximately 53,000 SNPs using all 48 unrelated BOCs, using the adjusted p-value calculation to obtain the average chi-squared value and genomic inflation factor based on median chi-squared (PLINK v1.04), and evaluated our simulated results for significantly inflated false positive association rates. Principal components were calculated using Eigenstrat [29], and logistic regression with covariates was implemented in PLINK v1.04.

2.2.3. Results

To summarize our results, individual dogs were correctly assigned to their respective breeds using all methods. Related dogs that were included in some analyses consistently grouped together, supporting the credibility of the results of the analyses. A group of unrelated BOC formed a separate, well-supported subgroup across analyses. These five dogs are distinguished by the type of purpose for which they were bred (show vs. working) and by geographical origin either of themselves or close ancestors. In simulated GWAS, this stratification led to significantly inflated false positive association rates.

2.2.3.a. Results of cluster and genetic distance analyses. Cluster analysis of unrelated individuals in four breeds (BOC, AUS, PWD, GSD) identified four clusters corresponding to breed in the data (Figure 2.5a) and correctly assigned all dogs to the four reported breeds. These results suggest some degree of heterogeneity within BOC when K = 4. At user-assigned values of K > 4, five BOC become distinct from the rest of their breed across runs (Figure 2.5b). We also identified clustering that suggests some proportion of AUS ancestry is shared with the BOC, a result that would be predicted given the history of these breeds.

a



b

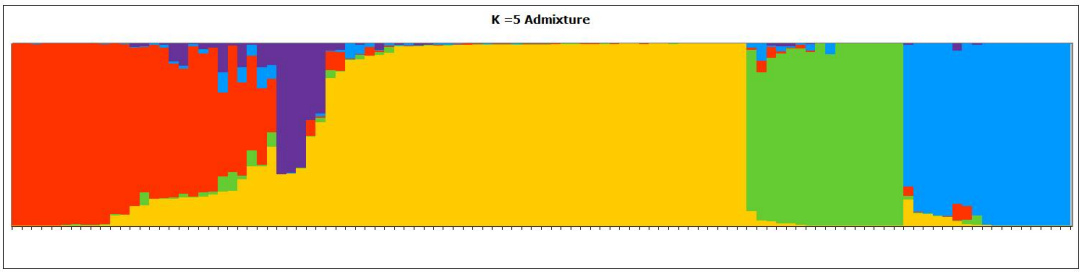
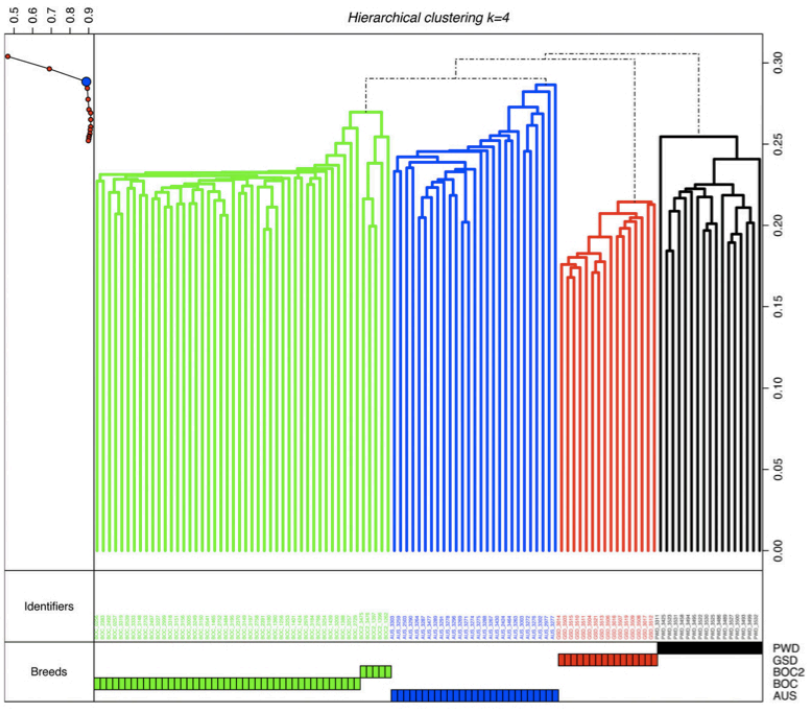


Figure 2.5: Four-breed *structure* results.

(a) Predicted number of clusters in a sample of 108 dogs (48 BOC, 27 AUS, 16 GSD, and 17 PWD) after Evanno *et al.* [26], based on 2,100 SNPs. Based on this method, K = 4 is the most appropriate for this sample. (b) *structure* results using 2,100 SNPs and K = 5 suggest within-breed stratification for BOC (yellow), with five show dogs (purple and yellow) appearing distinct. These five samples, from dogs that are unrelated at the grandparent level, were collected at conformation shows, or were sent to us by owners

who participated in AKC-sponsored conformation events with their dogs. The majority of our BOC sample, by contrast, was collected at working sheepdog trials. Other breeds cluster within appropriate breed: AUS (red), GSD (green) and PWD (blue).

The hierarchical grouping via genetic distance analysis for K = 4 demonstrated perfect stability, with all dogs falling into their respective breed clusters (Figure 2.6a). Hierarchical grouping for K = 8, calculated to be the "best fit" for this data set, demonstrated correct separation of dogs into four breeds, and intra-breed stratification of AUS and BOC (Figure 2.6b). The AUS were broken into four separate clusters, though two of those were made up of singletons. The same five BOC that were differentiated in the cluster analyses grouped together, and were distinct from the rest of the breed sample, forming a separate branch from the rest of the breed in the K = 8 dendrogram. A suggestive clustering of AUS was also detectable in the K = 8 dendrogram, though the clusters were composed of too few individuals to withstand rigorous stability testing.



a

b

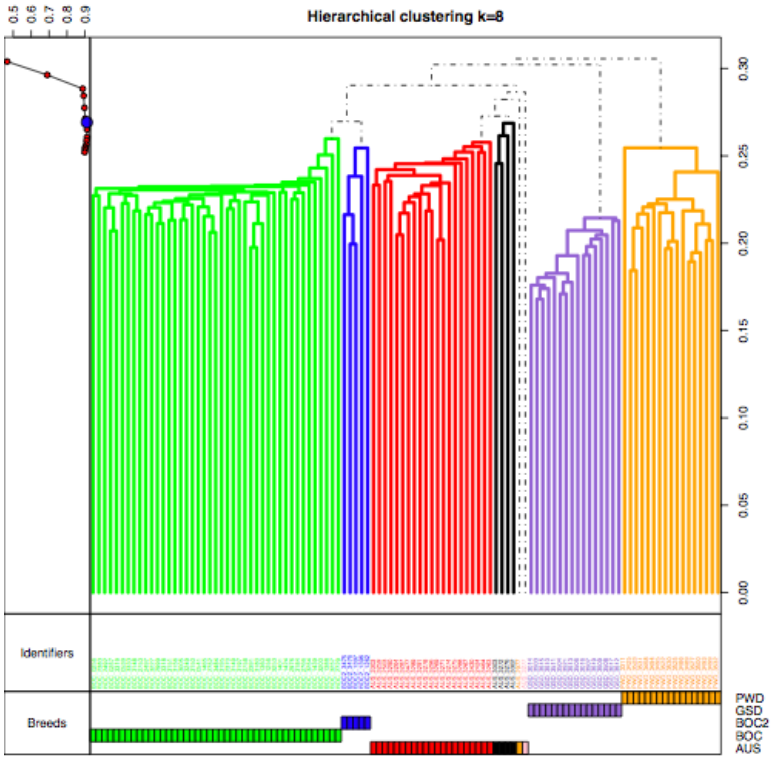


Figure 2.6: Hierarchical grouping based on the pairwise matrix of genetic identity by state (IBS). Analyses performed by and courtesy of Dr. C. Hitte. (a) Dendrogram of 108 dogs of four breeds constructed by pairwise genetic distance analysis for K = 4. Each dog is plotted on the x-axis with the distance of IBS given on the y-axis. Clusters are represented by different colors, with breeds indicated in the bottom-most panel of the x-axis. Breeds included: BOC (green), AUS (blue), PWD (black), GSD (red). (b) Dendrogram of 108 dogs based on the distance matrix of IBS set to K = 8 hierarchical clustering. Each dog is plotted on the x-axis with the distance of IBS given on the y-axis. The panel on the left shows the score for different values of K (dots), with the blue dot at K = 8 demonstrating the optimal score. Clusters are represented by different colors, with breeds indicated in the bottom-most panel of the x-axis (note - colors are different from *a*): working BOC (green), show BOC (blue), AUS (red/black/yellow/pink), GSD (purple) and PWD (orange).

2.2.3.b. Simulations. We sought to determine if intra-breed stratification would confound GWAS by carrying out simulations using our observed genotypes. As described above, we identified a group of BOC divergent from the larger group of BOC samples. We randomly assigned the 43

unrelated dogs of this latter group to case or control status. When the five distinct but unrelated BOC were split between case-control status, we obtained a near-null distribution with an average chi-squared statistic of 1.005 and a genomic inflation factor (lambda) of 1.112 (Table 2.6). The genomic inflation factor is expected to be 1 if there is no stratification occurring, with average chi-squared statistics ≤ 1. However, when all five of these outlier BOCs were assigned to either a case or control group, average chi-squared statistics were 1.180 or 1.168, respectively, with genomic inflation factors of 1.358 and 1.343, respectively, demonstrating significantly increased false positive rates secondary to stratification artifact (Table 2.6).

Subpopulation assignment	Test performed	Genomic inflation factor	Average chi-square
Australasian—split case/control[a]	Allelic association	1.112 ± 0.051	1.005 ± 0.030
Australasian—all cases[b]	Allelic association	1.358 ± 0.060	1.180 ± 0.042
Australasian—all cases[c]	Logistic regression w/3 covariates	1 ± 0	0.829 ± 0.023
Australasian—all controls[c]	Allelic association	1.343 ± 0.055	1.168 ± 0.042
Australasian—all controls[c]	Logistic regression w/3 covariates	1 ± 0	0.831 ± 0.024

Five show BOC of Australasian descent were split (2:3) between cases : controls or assigned all to cases or all to controls as noted above, with the balance of dogs randomly assigned either case or control status. Genomic inflation factors and average chi-square values were calculated for all simulations, and descriptive statistics of each are given for allelic association tests or logistic regression using 3 covariates to account for population structure. GWAS with all 5 divergent BOC assigned either case or control status demonstrated inflated false-positives. However, this inflation can be reduced to null by using covariates that account for population substructure.

[a] 22–26 Cases assigned randomly.

[b] 24–28 Cases assigned randomly.

[c] 25–28 Controls assigned randomly.

Table 2.6: Results of 100 simulations of GWAS of approximately 53,000 autosomal SNPs in a total of 48 unrelated BOC. (see table for caption)

Principal components analyses (PCA) were carried out using the uncorrelated set of 21K genome-wide markers, and the positions on each of the first three eigenvectors were used as covariates in the simulated GWAS in the BOC samples in a logistic regression framework. This led to genomic inflation factors of 1.0 regardless of whether the five divergent BOCs were all assigned to either case or control status, effectively correcting for the observed stratification (Table 2.6).

2.2.4. Discussion

The Border collie, the primary breed of interest for our studies of noise phobia and adult-onset deafness, has a long history of selection as a working sheepdog. Breeders of working BOC have historically followed a selection regime that prioritizes behavioral traits considered desirable for herding. But since the 1960s, the breed has also been developed as a show dog. Conformation breeders select for appearance, and evaluate the breedworthiness of their dogs on the basis of success in the show ring. The full breed standard published by the Australian National Kennel Council illustrates idealized physical attributes [30]; selection regimes based on this have resulted in dogs of extreme homogeneity in appearance, both in Australasia and America. In general, these dogs exhibit few or none of the behavioral characteristics desired in working sheepdogs.

Using pedigree, registration, and other demographic information, we were able to determine that the five BOC that consistently formed a separate, well supported subgroup were distinguished from the rest of our sample, because either they or their ancestors were successful show dogs, and all five traced back to show champions from Australasia, either directly or within less than four generations (see also [28]). These five samples were all collected at conformation shows, or were sent to us by owners who participated in AKC-sponsored conformation events with their dogs. The majority of our BOC sample, by contrast, was collected at working sheepdog trials, traces back to British ancestors, and came from owners who use working farm dogs, or breed and train dogs for sheepdog trial competitions, or both. Some suggestive population structure was also found within our AUS sample, but the variation within this breed is not as straightforward to characterize. Australian Shepherds are characterized by a long history of "dual purpose" breeding, and the heterogeneity we found within this sample probably reflects this fact.

It has previously been suggested that differences in geographic origin in case versus control samples may confound genome-wide association results [19]. We suggest that differing selection regimes may exacerbate the situation. Our results are consistent with the results of studies using pedigree analysis or smaller marker sets to identify population substructure within single dog breeds [20, 21], further emphasizing the importance of understanding the geographic origin and functional context within which samples are collected for large-scale studies.

Results of previous analyses of smaller marker sets, or those sampling only a portion of the genome are concurrent with our results using extensive genome-wide coverage. However, the samples and methods we used in our study, utilizing dense SNP data sampling of all 38 canine autosomes, allow us to assess relationships both between and within breeds with much finer resolution than previous studies. Awareness of sample origin helps explain the patterns of population substructure that were revealed through our analyses, and should allow other researchers to avoid introducing stratification into future analyses by constructing study samples in ways that reduce this confounding effect.

For practical reasons, it may not always be the case that balanced study samples can be obtained. Rather than limit a study's sample size, it may be desirable to explore and implement other means to statistically account for population substructure. In addition to the methods outlined here, intra-breed stratification can also be detected by multi-dimensional scaling (MDS). Covariates from either MDS or PCA can then be used in GWAS to statistically correct for substructure, a practice used in human studies to correct for population stratification [29]. For example, the inflated genomic inflation factors resulting from our simulation studies were reduced to null when the complete BOC sample including working and show dogs was instead analyzed by logistic regression using the first three principal component vectors as covariates. Similarly, analyses utilizing mixed model approaches can account for population stratification due to cryptic

relatedness, which may be of particular importance in studies within single breeds where individuals are likely to be distant relatives.

These results have important implications for genetic association studies in dogs. Contrary to common assumptions, within-breed population structure can be significant in some breeds, and this stratification may be explained by geographical origin, by artificial selection criteria used by dog breeders, or both. Demographic and pedigree information should be used to guide the collection of study samples that are free of significant within-breed population structure when possible, and/or to inform study design when stratified samples are to be used. In addition, genetic data gathered in the performance of genome-wide association studies can be used to statistically measure and, when required, account for breed substructure. This last point is particularly relevant when samples are obtained internationally, or utilize samples for which pedigree information is not available (e.g., the noise phobia GWAS described in Chapter 3). In these ways, genome-wide data is invaluable for characterizing cohort attributes to inform appropriate study design.

2.3. Laboratory Methodology for Two Genome-Wide Association Studies[***]

The sections above describe the use and validation of three distinct genotyping platforms—the Infinium Canine SNP20 and Infinium CanineHD in section 2.1 and the Affymetrix Canine v2.0 SNP array in section 2.2—for investigating the utility of saliva samples for whole-genome SNP surveys, and investigating the role intra-breed stratification may play in GWAS and how best to detect and take it into consideration. We have established the validity of these new array-based technologies for surveying several tens of thousands of SNPs across the canine genome. We have

[***]Portions of this section have been submitted for publication: Yokoyama JS, Chang ML, Tiira KA, Branson N, Dyer DJ, Juarbe-Diaz S, Ruhe AL, Robertson KR, Neff MW, Lohi H, Overall KL, and Hamilton SP. "Genome-wide association study identifies candidate loci in canine noise phobia" (submitted to *Genes, Brain and Behavior* on November 24, 2010).

also examined within-breed stratification in our primary breed of study, the Border collie, and demonstrated the importance of taking this into consideration when conducting GWAS to reduce the chance of obtaining false-positive results due strictly to sample stratification. These preliminary studies provide the backdrop upon which I will now describe the main analyses comprising my dissertation. In total, our group has genotyped over 600 different samples collected in the context of our Canine Behavioral Genetics Project (www.k9behavioralgenetics.com) to survey loci underlying complex disease and behavioral phenotypes as well as breed diversity. In the last section of this chapter I will describe the genotyping and quality control used for the genome-wide association studies for loci underlying noise phobia in herding breed dogs as described in Chapter 3, and loci underlying adult-onset deafness in Border Collies as described in Chapter 4.

2.3.1. Genotype Generation

Genotyping of SNPs providing genome-wide coverage of the dog was performed on three different array platforms: Affymetrix Custom Canine Array v2.0 (v2.0; Affymetrix, Santa Clara CA), or on the Illumina Infinium CanineSNP20 BeadChip (SNP20) and the Illumina Infinium CanineHD BeadChip (HD; Illumina Inc., San Diego CA). All Infinium (SNP20 and HD) genotyping was performed by the Genomics Core Facility at the University of California, San Francisco as per manufacturer's instructions. The Affymetrix genotyping was performed in-house utilizing equipment available for use by the Institute for Human Genetics (University of California, San Francisco) using a modified protocol based on the *Sty* fraction of the Affymetrix Human 500K protocol, as recommended by the Broad Institute who developed the v2.0 canine array. Genotypes were called using the BRLMM-P algorithm in Affymetrix Power Tools (apt-1.12.0) on all Affymetrix samples or with Illumina's GenomeStudio Data Analysis Software Package (1.0.2.20706, Illumina Inc.) for SNP20 and HD samples.

2.3.2. Genotype Quality Control

Genotype QC was first implemented on each platform's dataset using the same criteria for each dataset. Exclusion criteria were: call rates by marker and by individual < 95%, concordance of replicate control sample genotypes across all genotyping runs < 100%, X-chromosome markers due to poor quality data, Hardy-Weinberg equilibrium $p < 0.001$ (for the previously unvalidated Affymetrix platform – as a means of filtering out SNPs with high heterozygosities), MAF < 0.02, mendelization errors of >5% per SNP and >10% per family. This resulted in datasets of about 40,000 SNPs for Affymetrix, 22,000 SNPs for SNP20, and 125,000 SNPs for HD. After the single-platform SNP data was cleaned, two merged datasets were created. Affymetrix v2.0 and SNP20 datasets were merged to create a dataset with ~12,000 SNPs common to both platforms after QC. The three cleaned datasets were also merged across ~9,000 SNPs that were common to all three platforms after QC to create the final dataset. Strand and reference alleles were matched using SNP annotation data, genotype information from replicate samples genotyped across all platforms, and breed-specific allele frequency data. Further QC also required all SNPs to demonstrate 100% concordance for duplicate sample genotypes across all platforms included in the combined datasets. QC for SNP20 and HD samples was performed using GenomeStudio software. QC for Affymetrix samples, plus data merging for v2.0/SNP20 and v2.0/SNP20/HD datasets was performed using Stata10.1/MP (StataCorp LP, College Station TX) and PLINK (v1.06-1.07 [27]).

Now that I have described the genotyping and QC methods used for both of the association studies presented in this dissertation, I will move on to describe our studies of noise phobia (Chapter 3) and adult-onset deafness (Chapter 4).

2.4. References

1. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, et al. (2003) The dog genome: survey sequencing and comparative analysis. Science 301: 1898-1903.
2. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803-819.
3. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet 39: 1321-1328.
4. Drögemüller C, Karlsson EK, Hytonen MK, Perloski M, Dolf G, et al. (2008) A mutation in hairless dogs implicates FOXI3 in ectodermal development. Science 321: 1462.
5. Awano T, Johnson GS, Wade CM, Katz ML, Johnson GC, et al. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. PNAS 106: 2794-2799.
6. Drögemüller C, Becker D, Brunner A, Haase B, Kircher P, et al. (2009) A Missense Mutation in the SERPINH1 Gene in Dachshunds with Osteogenesis Imperfecta. PLoS Genet 5: e1000579.
7. Wiik AC, Thoresen SI, Wade C, Lindblad-Toh K, Lingaas F (2009) A population study of a mutation allele associated with cone-rod dystrophy in the standard wire-haired dachshund. Anim Genet 40: 572-574.
8. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, et al. (2009) Coat Variation in the Domestic Dog Is Governed by Variants in Three Genes. Science 326: 150-153.
9. Parker HG, Vonholdt BM, Quignon P, Margulies EH, Shao S, et al. (2009) An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science 325: 995-998.
10. Dodman NH, Karlsson EK, Moon-Fanelli A, Galdzicka M, Perloski M, et al. (2010) A canine chromosome 7 locus confers compulsive disorder susceptibility. Mol Psychiatry 15: 8-10.
11. Chang ML, Terrill RL, Bautista MM, Carlson EJ, Dyer DJ, et al. (2007) Large-Scale SNP Genotyping with Canine Buccal Swab DNA. J Hered 98: 428-437.

12. Short AD, Kennedy LJ, Forman O, Barnes A, Fretwell N, et al. (2005) Canine DNA Subjected to Whole Genome Amplification is Suitable for a Wide Range of Molecular Applications. J Hered 96: 829-835.
13. Hansen TV, Simonsen MK, Nielsen FC, Hundrup YA (2007) Collection of blood, saliva, and buccal cell samples in a pilot study on the Danish nurse cohort: comparison of the response rate and quality of genomic DNA. Cancer Epidemiol Biomarkers Prev 16: 2072-2076.
14. Nishita D, Jack L, McElroy M, McClure J, Richards J, et al. (2009) Clinical trial participant characteristics and saliva and DNA metrics. BMC Medical Research Methodology 9: 71.
15. Iwasiow RM, James CMP, Niles J (2009) Quality, Yield, Performance and Bacterial Content of DNA from Canine Saliva Collected and Purified using Oragene®•ANIMAL. PD-WP-014 Issue 1.0, www.dnagenotek.com:
16. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, et al. (2005) Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 37: 1243-1246.
17. Mitsouras K, Faulhaber EA. (2009) Saliva as an alternative source of high yield canine genomic DNA for genotyping studies. BMC Research Notes 2:219.
18. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L, 2004. Genetic Structure of the Purebred Domestic Dog. Science 304: 1160-1164.
19. Quignon P, Herbin L, Cadieu E, Kirkness EF, Hédan B, Mosher DS, Galibert F, André C, Ostrander EA, Hitte C, 2007. Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. PLoS ONE. 2(12):e1324.
20. Björnerfeldt S, Hailer F, Nord M, Vilà C, 2008. Assortative mating and fragmentation within dog breeds. BMC Evolutionary Biology 8(1).
21. Calboli FC, Sampson J, Fretwell N, Balding DJ, 2008. Population structure and inbreeding from pedigree analysis of purebred dogs. Genetics. 179(1):593-601.
22. Overall KL, Hamilton SP, Chang ML, 2006. Understanding the genetic basis

of canine anxiety: phenotyping dogs for behavioral, neurochemical, and genetic assessment. Journal of Veterinary Behavior: Clinical Applications and Research, 1(3): 124-141.

23. Pritchard JK, Stephens M, Donnelly P, 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945-959.
24. Falush D, Stephens M, Pritchard JK, 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164: 1567-1587.
25. Falush D, Stephens M, Pritchard JK, 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7(4):574-578.
26. Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14(8): 2611-2620.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC, 2007. pLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81(3):559-75.
28. Chang ML, Yokoyama JS, Branson N, Dyer DJ, Hitte C, Overall KL, and Hamilton SP, 2009. Intrabreed Stratification Related to Divergent Selection Regimes in Purebred Dogs May Affect the Interpretation of Genetic Association Studies. Journal of Heredity 100(Supplement 1):S28-S36.
29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D, 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38(8):904-9.
30. Australian National Kennel Council (ANKC), 2005. The Extended Breed Standard of the Border Collie. The Border Collie Club of Victoria and the Australian National Kennel Council.

Chapter 3: Genome-wide Association Study Identifies Candidate Loci for the Canine Anxiety Phenotype Noise Phobia[***]

Chapter 3 describes our genome-wide association study (GWAS) for the anxiety disorder noise phobia in Border collies and other breeds of dog. After providing background on current noise phobia research, I describe the results of our primary GWAS, in addition to a fine-mapping/replication performed in an independent sample of Border collies. I also detail our investigation of noise phobia in multiple breeds, and highlight the challenges we've faced with between-breed heterogeneity. Finally, I discuss the relevance of these findings for future studies of complex behavioral traits in the domestic dog.

3.1. Introduction

The domestic dog offers a novel and potentially powerful genetic model for studies of behavior, including complex suites of selected behaviors as well as maladaptive behavioral conditions analogous to human psychiatric disorders [1]. Dogs provide two critical advantages that facilitate such studies: a) potentially prominent genetic homogeneity due to the foundations of individual breeds; and b) naturally occurring behavioral disorders with clinical features similar to a number of human anxiety disorders (e.g., separation anxiety, obsessive-compulsive disorder, specific phobias). As mentioned in previous chapters, Dodman and colleagues recently demonstrated the potential of studies of behavioral disorders in dogs for allowing identification of possible risk alleles as well as facilitating insights into human psychiatric illnesses [2]. In a genetic investigation of canine compulsive disorder in Doberman pinschers, they examined 92 cases and

[***]Portions of this section have been submitted for publication: Yokoyama JS, Chang ML, Tiira KA, Branson N, Dyer DJ, Juarbe-Diaz S, Ruhe AL, Robertson KR, Neff MW, Lohi H, Overall KL, and Hamilton SP. "Genome-wide association study identifies candidate loci in canine noise phobia" (submitted to *Genes, Brain and Behavior* on November 24, 2010). **JSY** – sample and phenotype ascertainment, data generation and QC, final study design, all statistical analyses, manuscript preparation; MLC – conception of study idea and basic study design with SPH; MLC and all other co-authors – sample and phenotype ascertainment, technical support, funding.

68 controls across 14,700 SNPs, and demonstrated a putative risk locus on canine chromosome 7, with fine-mapping confirming localization to the canine homologue of the *CDH2* gene, which encodes neuronal cadherin, a calcium dependent cell-cell adhesion glycoprotein. Based on the role of this gene in neuronal functioning [3], canine gene-mapping has potentially provided access to the molecular mechanisms of compulsivity.

The focus of our study is noise phobia, a canine anxiety disorder characterized by excessive fear or panic responses to sound stimuli such as thunder, fireworks and gunshots, resulting in attempts to avoid or escape from the sound [4]. It has been estimated that 40-50% of dogs show signs of fear to noises, and noise phobia has thus been suggested to be a major welfare problem in dogs [5]. This disorder can be severe enough to adversely impact the working utility or quality of life of dogs affected by it, and can sometimes even be the cause of euthanasia. Noise phobia is strikingly analogous to specific phobia in humans, where marked and persistent fear is invoked by a specific stimulus or situation resulting in avoidance [6] and genetic predisposition plays a measurable role [7].

Although noise phobia occurs across breeds of dogs, there is some suggestion of enrichment in herding breeds [8, 9]. Apart from genetic causes, noise phobia may also develop from environmental factors such as trauma, lack of habituation, stress-induced dishabituation, sensitization and social transmission [5]. Some studies of noise phobia have focused on characterization of clinical manifestations and treatment [10, 11]. For example, Seksel and Lindeman [11] found that use of the psychiatric medication clomipramine improved anxiety symptoms in dogs diagnosed with one or more of three anxiety disorders when used in conjunction with a behavioral modification program. In addition to these clinic-based studies, Branson and Rogers investigated the role of handedness in dogs as a potential predictor of noise phobic predisposition, finding that dogs with decreased lateralization (i.e. ambilateral paw

preference) demonstrated more reactivity to the noises of thunderstorms and fireworks than dogs with either left- or right-paw preferences [12]. The results were consistent across repeated sound stimuli exposures. These findings are in line with previous research in humans suggesting that reduced lateralization may be associated with anxiety, depression and psychosis [13, 14]. Importantly, Branson and Rogers [12] also found that owner questionnaire responses regarding their dogs' noise reactivity were highly correlated with experimenter-observed reactivity in a controlled setting. This suggests that the familiarity of owners with the behavior of their dogs can be reliably captured by questionnaire.

Although noise phobia is suggested to have increased prevalence in certain breeds of dogs, and anecdotal and pedigree evidence suggests that it segregates in families, no formal genetic epidemiological investigations have yet been undertaken. We hypothesized a genetic risk and conducted a genome-wide survey for associated loci. Because it has been suggested that noise phobia is enriched in herding breeds, we performed our primary studies in Border collies collected mostly from working bloodlines, with additional samples collected from related herding and utility breeds to afford multi-breed analyses.

3.2. Materials and Methods

3.2.1. Samples. A total of 211 dogs, unrelated at the grandparental level had noise phobia phenotypes and were of sufficient numbers to include in the primary mapping analysis (≥5 individuals in each category of case/control per breed). Samples from the following breeds were used for primary and exploratory secondary analyses, where n is the size of the sample with noise phobia phenotypes: Australian shepherd (AUS, n = 17), Border collie (BOC, n = 84), bearded collie (BEC, n = 26), Belgian shepherd (BES, n =16), Belgian tervuren (BGT, n = 16), great Dane (GRD, n = 22), and German shepherd dog (GSD, n = 30). Gender and mean ages for cases and controls are provided in Table 3.1.

			cases				controls				
breed	**dataset**	**N**	**n_{case}**	**female**	**male**	**avg age (yr)**	**$n_{control}$**	**female**	**male**	**$n_{<2yo}$**	**avg age (yr)**
AUS	v2.0/SNP20	17	6	2	4	9.6	11	6	5	3	5.9
AUS	*strict controls*	*13*	*6*	*2*	*4*	*9.6*	*7*	*4*	*3*	*0*	*7.2*
BEC	SNP20	26	15	7	8	5.5	11	3	8	2	6.1
BEC	*strict controls*	*22*	*15*	*7*	*8*	*5.5*	*7*	*2*	*5*	*0*	*8.3*
BES	HD	16	7	7	0	9.8	9	4	5	0	4
BGT	HD	16	11	6	5	6.5	5	5	0	0	6.3
BOC	v2.0/SNP20	84	47	24	23	8	37	22	15	3	4.7
BOC	*strict controls*	*76*	*47*	*24*	*23*	*8*	*29*	*17*	*12*	*0*	*5.2*
BOC	HD	50	31	18	13	6.8	19	12	7	1	4.7
BOC	*strict controls*	*45*	*31*	*18*	*13*	*6.8*	*14*	*9*	*5*	*0*	*4.9*
GRD	HD	22	8	5	3	4.1	14	7	7	3	4.2
GRD	*strict controls*	*18*	*8*	*5*	*3*	*4.1*	*10*	*5*	*5*	*0*	*4.9*
GSD	v2.0/SNP20/HD	37	7	5	2	5.7	30	16	14	8	3.4
GSD	*strict controls*	*28*	*7*	*5*	*2*	*5.7*	*21*	*10*	*11*	*0*	*4.1*

Table 3.1: Sample demographics.

A detailed breakdown of the sample demographics is given for each breed analyzed, including sex distribution and mean age of sample group (divided by case/control status). In addition to the control criterion described in section 3.2.1, stricter control criterion (0% frequency response to thunder, fireworks and gunshots and >2 years old) were also assessed *post hoc* to test whether utilizing strict control samples would enhance association signal strength in the BOC primary analysis and fine-mapping/ replication samples. In grey are the numbers for the stricter control criterion for AUS, BEC, GRD and GSD samples, although these were *not* assessed in the current study due to sample size limitations. Numbers are not provided for BES and BGT because all control samples in these groups already met the stricter criterion in the main analysis. Breed: AUS – Australian shepherd, BEC – Bearded collie, BES – Belgian shepherd, BGT – Belgian tervuren, BOC – Border collie, GRD – great Dane, GSD – German shepherd dog. Dataset – array type on which samples were genotyped, N – total number of samples (case+control), n_{case} – total number of case samples, female – number of females in given group (case or control), male – number of males in given group (case or control), avg age (yrs) – mean age of cases or controls, $n_{control}$ – total number of control samples, given in years, $n_{<2yo}$ – number of control samples under the age of 2 years at the time of phenotypic assessment.

An additional 50 purebred BOC collected from an American sheepdog herding trial and an American breed specialty were also used for a replication/fine-mapping analysis. Complete pedigrees were not available for all dogs, so we also addressed relatedness by analysis of genotype data. Two individuals—one from each of two pairs—were removed based on a genome-wide estimate of identity-by-state (IBS $\geq$ 0.50), which suggested each pair was closely related. Whole blood samples (3-8 ml blood draw by veterinarians or licensed veterinary technicians) were collected at trialing events, conformation events and breed specialties, or were sent directly to the laboratory by owners and breeders in the context of ongoing genetic studies of canine behavior and complex disease. DNA was extracted from all blood samples using standard protocols. All procedures were approved by local institutional review boards.

3.2.2. Phenotypic assessment. We ascertained which dogs were affected with the noise phobia phenotype using a published behavioral questionnaire developed by our group [4]. ***Cases*** were defined as dogs that responded >60% of the time to at least one of three noises (thunder, fireworks, gunshots) with one or more observable reaction(s): salivate, defecate, urinate, destroy, hide, tremble, vocalize, pace, escape, freeze, or pant. ***Controls*** were defined as dogs that responded <60% of the time to all three problem noises. Controls were vetted to ensure that they had experienced problem noises regularly. All dogs studied had normal hearing. Most controls had a 0% frequency response to thunder, fireworks and gunshots (Table 3.1). As the mean age of onset of noise phobia has been reported as 2 years of age [15], the majority of controls were >2 years old at the time of sampling. Based on owner response, we omitted cases where there was the possibility of noise trauma at an early age. Finnish samples were collected using the same, translated questionnaire (the noise phobia section), however in addition to reactions: salivate, defecate, urinate, destroy, hide, tremble, vocalize, pace, escape, freeze, and pant, "tail between the legs" was also added to reactions that qualified cases. Only one dog (Belgian tervuren) had "tail between the legs" scored as the *only* reaction to loud noises. All other Finnish dog owners

that had the “tail between the legs” reaction indicated also had indicated other behavioral reactions on the questionnaire. For samples collected in Finland, owners delivered the blood samples taken by a veterinarian or a veterinary professional to the Finnish research group.

3.2.3. Association analysis. Genotypes for all samples were attained by the methods described in Chapter 2, Section 2.3.1. After marker quality control (QC, described in Chapter 2, Section 2.3.2), primary GWAS for noise phobia was performed on unrelated BOC only using the v2.0/SNP20 combined dataset of ~12,000 SNPs. Efficient mixed-model association (EMMA) analysis of the 84 unrelated BOC with noise phobia phenotypes was performed using the program EMMA eXpedited (EMMAX). EMMAX uses a variance component approach that accounts for population substructure and cryptic relatedness by utilizing empirically estimated pair wise relatedness in a computationally efficient manner [16]. Correction for population stratification *within* the BOC sample was performed given evidence for within-breed substructure (Figure 3.1a).

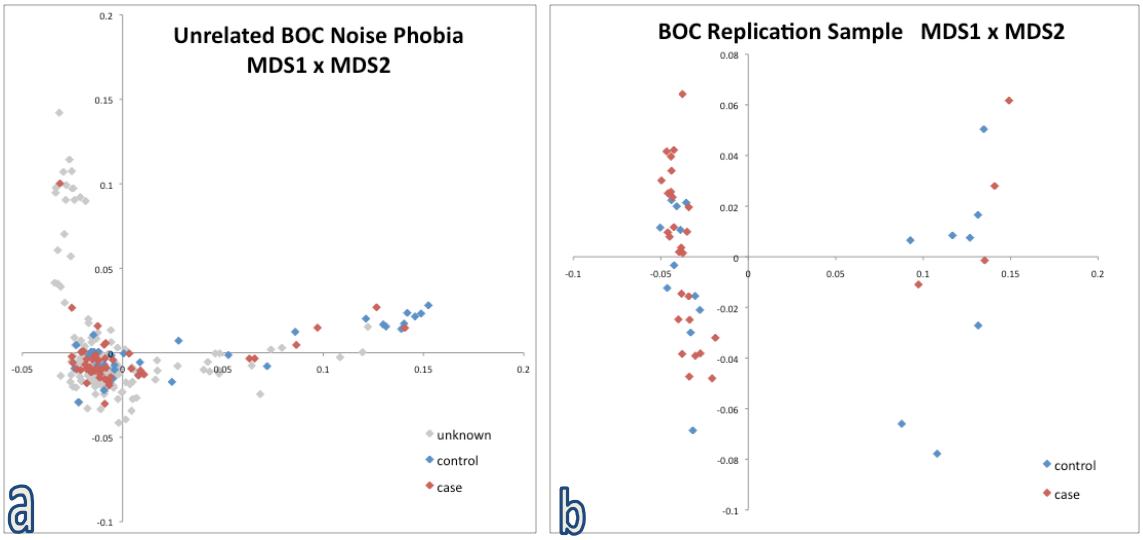


Figure 3.1: MDS 1 x MDS 2 plotted for all unrelated BOC.

(a) Multi-dimensional scaling (MDS) covariates were calculated using data from all unrelated BOC genotyped for the 12K 2-platform dataset. Those for which noise phobia phenotypes were available at the time of analysis are indicated as control (blue) or case (red). The two 'arms' extending from the central cluster of samples at the origin suggest within-breed stratification. Samples in gray were not utilized in this study. (b) BOC replication sample. MDS covariates were calculated using an LD-trimmed whole-genome dataset of approximately 83K SNPs. There is clear stratification within this cohort, warranting use of covariate 1 for the fine-mapping logistic regression analysis.

EMMAX was used to estimate heritability, or "pseudoheritability", which is calculated as the fraction of phenotypic variance explained by the relatedness matrix that was empirically estimated by the provided genotypic data [16]. To assess empirical significance, permutations were performed in PLINK (v1.07) under a logistic regression framework, utilizing one multi-dimensional scaling (MDS) vector as a covariate to correct for within-breed stratification. Replication/fine-mapping was performed using 3,260 SNPs from the CanineHD array corresponding to four genomic regions demonstrating the greatest statistical support from among the top 100 ranked p-values (i.e., the top 100 associated hits were tallied for chromosome representation, and the chromosomes with the most representation were utilized for

replication/fine-mapping). The windows of SNPs that were investigated for replication were demarcated by the outer limits of the largest cluster of top-ranked findings for that chromosome in the primary GWAS. These SNPs were then evaluated for association in an additional 50 BOC of equal relatedness in PLINK (mean pi-hat = 0.028 ± 0.04 for 1,225 pairs). There was no significant difference in relatedness among controls versus cases (Table 3.2).

Between-group IBS (mean, SD) = 0.73, 0.02
In-group (2) IBS (mean, SD) = 0.73, 0.02
In-group (1) IBS (mean, SD) = 0.73, 0.02
Approximate proportion of variance between group = 0.008
IBS group-difference empirical p-values:

T1: Case/control less similar	**p = 0.03**
T2: Case/control more similar	p = 0.97
T3: Case/case less similar than control/control	p = 0.91
T4: Case/case more similar than control/control	p = 0.09
T5: Case/case less similar	p = 0.94
T6: Case/case more similar	p = 0.06
T7: Control/control less similar	p = 0.21
T8: Control/control more similar	p = 0.79
T9: Case/case less similar than case/control	p = 0.95
T10: Case/case more similar than case/control	**p = 0.05**
T11: Control/control less similar than case/control	p = 0.91
T12: Control/control more similar than case/control	p = 0.09

Table 3.2: Test for balanced IBS of controls and cases in BOC replication sample.

In PLINK the *--ibs-test* command was used to determine whether there was a significant difference in relatedness in the controls versus the cases in the BOC replication sample. Modest differences ($p \sim 0.05$) are highlighted in bold. We found a modest difference in relatedness between cases and controls, with cases more similar to cases than controls are to controls. However, this is not unexpected if there is a true association for noise phobia, as one would expect cases to be more similar to one another given shared risk loci.

In the replication set, one MDS vector was associated with the phenotype. The vector was used in a logistic regression to correct for population stratification (Figure 3.1b). In addition to the logistic regression analysis, set-based permutation, haplotypic logistic regression (with one MDS

covariate) and epistasis analyses were all tested using PLINK. All MDS covariates were calculated from sample data for which the specific analysis was performed using a subset of genome-wide, unlinked ($r^2 < 0.8$) markers (i.e., the covariate used in the replication was independent of that calculated for the primary association analysis).

Analyses among other (non-BOC) breeds were also performed with EMMAX with SNP data from each platform (BEC – SNP20; BES, BGT and GRD – HD; AUS and GSD – v2.0/SNP20/HD combined dataset, 9,000 SNPs). Meta-analysis was performed in PLINK (v1.07) with EMMAX beta values and the "no-allele" option since information about reference alleles is not provided in the EMMAX output. Only markers represented in three or more breed analyses are reported. The numbers of cases and controls for each analysis, as well as the array type used for genotyping are presented in Table 3.3.

Breed	# SNPs	Array	n	Cases	Controls	Pseudoheritability
BOC	12,208	v2.0/SNP20	84	47	37	0.76
BOC	12,208	v2.0/SNP20	76	47	29	0.82
BOC	3,260	HD	50	31	19	n/a
BOC	3,260	HD	45	31	14	n/a
AUS	9,152	v2.0/SNP20/HD	17	6	11	0.000012
BEC	21,797	SNP20	26	15	11	0.99
BES	141,443	HD	16	7	9	0.99
BGT	141,443	HD	16	11	5	0.27
GRD	141,443	HD	22	8	14	0.99
GSD	9,152	v2.0/SNP20/HD	37	7	30	0.000010

Table 3.3: Samples and pseudoheritability estimates.

Number of samples, markers and genotyping platform for each breed analyzed, in addition to the pseudoheritability for noise phobia in that breed as calculated by EMMAX. Using stricter control criterion removed 8 control samples from the primary GWAS sample and slightly increased the pseudoheritability for noise phobia (for more details, see text and Table 3.1). Utilizing the same criterion removes 5 controls from the replication analysis. AUS and GSD were dropped from the study because the pseudoheritability ~ 0). BOC - Border collie, AUS - Australian shepherd, BEC - bearded collie, BES - Belgian shepherd, BGT -

Belgian tervuren, GRD - great Dane, GSD - German shepherd dog; v2.0 - Affymetrix Custom Canine Array v2.0, SNP20 - Illumina Infinium CanineSNP20 BeadChip, HD - Illumina Infinium CanineHD BeadChip.

3.2.4. Multi-breed visualization analysis. For the noise phobia analysis, results from breed analyses were compared to each other by scoring the 300 top-ranked findings (where the most associated finding was scored as 300, and the 300th finding was scored as 1) and visualizing those rankings with the Integrative Genomics Viewer (IGV v1.5, http://www.broadinstitute.org/igv/).

3.3. Results

3.3.1. Primary analysis in Border collies. We first sought to calculate a "pseudoheritability" value for noise phobia in BOC. This is not a true heritability calculation since the estimated pair wise relatedness does not correspond exactly to the kinship coefficient [16]. Calculations of pseudoheritability are also found in the livestock literature, for example when estimating heritability for quantitative traits using a Poisson distribution [17, 18]. In our dataset, the pseudoheritability for noise phobia in BOC was estimated by EMMAX to be 0.76 (Table 3.3). Interestingly, using the stricter criterion for controls, the pseudoheritability increased to 0.82 (Table 3.3).

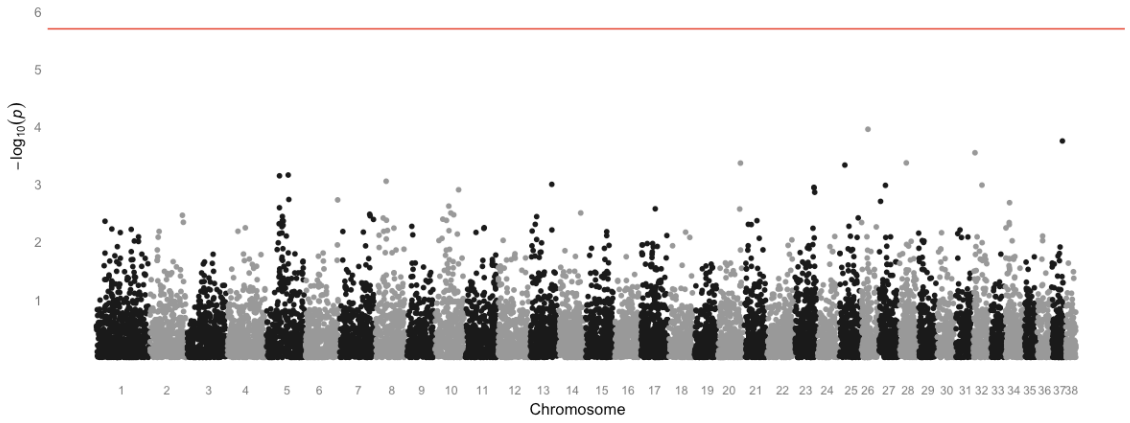


Figure 3.2: Manhattan plot of BOC noise phobia analysis in EMMAX.

Each SNP is represented by a dot, with markers plotted along the x-axis in order by chromosome. The y-axis is the $-\log_{10}$(p-value). Red line indicates genome-wide significance at a Bonferroni threshold.

Association analysis of 84 BOC, including 47 cases and 37 controls for the noise phobia phenotype, demonstrated no results that reached a level of genome-wide significance at a Bonferroni correction for 12,208 tests ($p < 4.1 \times 10^{-6}$). A plot of the genome-wide results is shown in Figure 3.2, and the top 25 findings ranked by p-value are given in Table 3.4. The strongest association in BOC was at SNP chr26.18210457 with $p = 1.10 \times 10^{-4}$. Additionally, two markers on *Canis familiaris* chromosome (CFA) 5 represented the only multiple findings for a single chromosome within the ten highest ranked findings (chr5.54587936 and chr5.33995804 were ranked 7th and 8th, respectively). There was modest regional support for association on CFA8 and CFA10, though only one SNP from CFA8 was among the ten best findings (Table 3.4). Allele frequencies in cases and controls are shown in Table 3.4. The allele frequency differences range from 0.14 - 0.32 between cases and controls, suggestive of a strong effect by the respective top associated loci (odds ratios (OR) are not calculated in EMMAX).

CFA	BP	SNP	P	A1	CASE-freq	CTRL-freq
26	18210457	chr26.18210457	1.10E-04	A	0.37	0.69
37	28614773	chr37.28614773	1.76E-04	C	0.21	0.45
32	6940786	chr32.6940786	2.82E-04	T	0.30	0.53
28	18779050	chr28.18779050	4.23E-04	A	0.29	0.57
20	55413165	chr20.55413165	4.27E-04	C	0.22	0.53
25	16005438	chr25.16005438	4.62E-04	T	0.33	0.64
5	54587936	chr5.54587936	6.87E-04	C	0.22	0.06
5	33995804	chr5.33995804	7.09E-04	T	0.31	0.07
8	28181714	chr8.28181714	8.81E-04	G	0.04	0.24
13	53367679	chr13.53367679	9.98E-04	A	0.49	0.23
32	23294467	chr32.23294467	0.0010	G	0.22	0.41
27	17997320	chr27.17997320	0.0010	T	0.45	0.73
23	48778821	chr23.48778821	0.0011	C	0.50	0.77
23	49142150	chr23.49142150	0.0011	T	0.26	0.57
10	59670840	chr10.59670840	0.0012	T	0.09	0.32
23	50635945	chr23.50635945	0.0014	G	0.11	0.31
5	55987596	chr5.55987596	0.0018	C	0.16	0.38
6	78683285	chr6.78683285	0.0019	T	0.49	0.26
27	6570682	chr27.6570682	0.0020	C	0.30	0.54
34	12715014	chr34.12715014	0.0021	A	0.10	0.34
10	36535452	chr10.36535452	0.0024	T	0.02	0.24
5	33902404	chr5.33902404	0.0025	A	0.33	0.12
17	37690609	chr17.37690609	0.0026	G	0.39	0.70
20	53644425	chr20.53644425	0.0027	A	0.00	0.18
14	55161655	chr14.55161655	0.0031	A	0.03	0.18

Table 3.4: Noise phobia in BOC.

Top 25-ranked findings for noise phobia analysis performed in EMMAX. P-values are given with each SNP in addition to genomic location information. Allele frequencies are provided for cases and controls with respect to a reference allele. Large differences in allele frequencies between cases and controls suggest a strong genetic effect. CFA – canine chromosome; BP – base pair, SNP – marker; p – p-value from EMMAX analysis; A1 – reference allele; CASE-freq – allele frequency for A1 in noise phobia cases; CTRL-freq – allele frequency for A1 in controls.

We performed 100,000 permutations under a logistic regression framework with one MDS covariate to correct for stratification; no signals were significant hits at a threshold of p_{genome} <

0.05 (data not shown). We sought to address the possibility that our controls were enriched with cases since we included dogs with some noise response, although below our chosen threshold. When we removed these dogs from the analysis (n = 8), the results remained largely unchanged (Figure 3.3 and Table 3.5), although the signal on CFA5 was more apparent in the analysis utilizing the stricter control criterion.

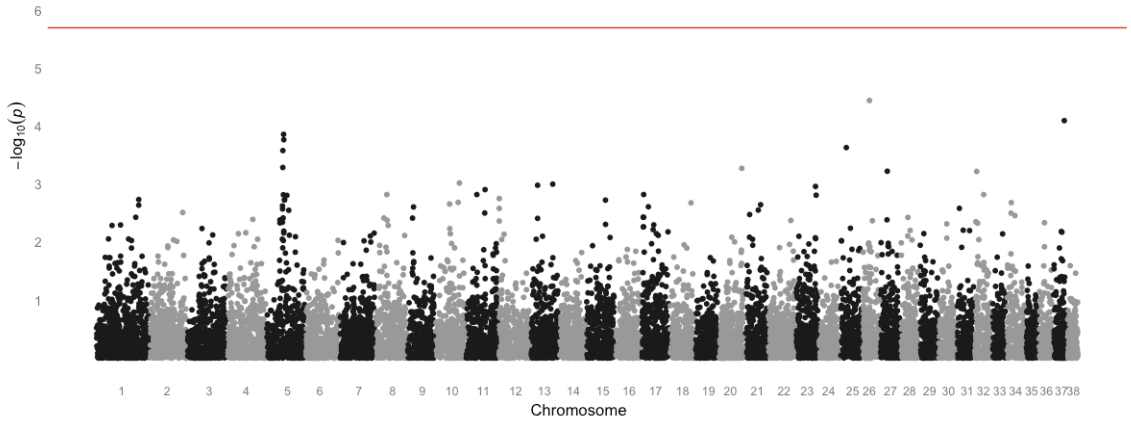


Figure 3.3: Manhattan plots of BOC noise phobia analysis with stricter control criterion.

Red line indicates genome-wide significance at a Bonferroni threshold. Findings on CFA5 are stronger in the strict-control phenotype sample.

CFA	BP	SNP	P_{strict}	A1	CASE-freq	CTRL-$freq_{strict}$
26	18210457	chr26.18210457	3.60E-05	A	0.3696	0.7241
37	28614773	chr37.28614773	8.02E-05	C	0.2128	0.5
5	42453935	chr5.42453935	1.38E-04	A	0.2979	0.5862
5	42976185	chr5.42976185	1.71E-04	T	0.4891	0.1724
25	16005438	chr25.16005438	2.34E-04	T	0.3298	0.6607
5	40827188	chr5.40827188	2.63E-04	T	0.4149	0.1379
5	40640574	chr5.40640574	5.16E-04	A	0.4787	0.7586
20	55413165	chr20.55413165	5.34E-04	C	0.2234	0.5345
27	18708815	chr27.18708815	5.98E-04	C	0.2553	0.4655
32	6940786	chr32.6940786	6.05E-04	T	0.3043	0.5345
10	59670840	chr10.59670840	9.56E-04	T	0.08511	0.3448
13	53367679	chr13.53367679	0.0010	A	0.4894	0.2069
13	17819909	chr13.17819909	0.0010	G	0.4022	0.7069
23	49142150	chr23.49142150	0.0011	T	0.2553	0.5862
11	47448928	chr11.47448928	0.0012	A	0.2979	0.06897
17	7827093	chr17.7827093	0.0015	G	0.3191	0.08621
32	23294467	chr32.23294467	0.0015	G	0.2234	0.4138
11	28185376	chr11.28185376	0.0015	A	0.04255	0.1897
8	28181714	chr8.28181714	0.0015	G	0.04255	0.2414
5	41978918	chr5.41978918	0.0015	C	0.4149	0.7241
23	50635945	chr23.50635945	0.0016	G	0.1064	0.3276
5	50218801	chr5.50218801	0.0016	A	0.09783	0.3448
12	3285545	chr12.3285545	0.0018	T	0.4681	0.7241
1	102282421	chr1.102282421	0.0018	T	0.2128	0.4655
5	44606853	chr5.44606853	0.0019	A	0.2553	0.06897

Table 3.5: GWAS of Noise phobia in BOC with stricter control criterion.

Top 25-ranked findings for noise phobia analysis performed in EMMAX using stricter control criterion. P-values are given with each SNP in addition to genomic location information. Allele frequencies are provided for cases and controls with respect to a reference allele. Stronger associations are observed in the sample set utilizing the stricter control criterion. CFA – canine chromosome; BP – base pair, SNP – marker; p_{strict} – p-value from EMMAX analysis; A1 – reference allele; CASE-freq – allele frequency for A1 in noise phobia cases; CTRL-$freq_{strict}$ – allele frequency for A1 in strict criterion controls.

We interrogated regions of suggestive association to test for regional support of association with greater power and resolution. Secondary association was assessed with a higher density SNP panel and an independent sample of herding BOC. A total of 3,260 SNPs from the regions on CFA 5, 8 and 10 showing the strongest regional evidence for association were analyzed. The top 25 ranked findings are presented in Table 3.6.

CFA	SNP	BP	A1	OR	L95	U95	P	P_{strict}	in 12k?
10	chr10.31603131	31603131	A	0.0075	0.0006469	0.08694	9.09E-05	2.67E-04	no
10	chr10.57383571	57383571	A	0.04572	0.006877	0.304	0.001414	0.003187	no
10	chr10.58630360	58630360	G	0.04624	0.006971	0.3067	0.001451	0.003269	no
10	chr10.36054599	36054599	G	0.08014	0.01497	0.4289	0.003186	0.009063	no
10	chr10.58018214	58018214	A	0.06519	0.01	0.4248	0.0043	0.01214	0.76
10	chr10.58041409	58041409	C	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58071141	58071141	A	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58088621	58088621	G	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58110655	58110655	T	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58250922	58250922	G	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58635986	58635986	A	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58667130	58667130	A	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58673713	58673713	A	0.06519	0.01	0.4248	0.0043	0.01214	no
10	chr10.58680393	58680393	A	0.06519	0.01	0.4248	0.0043	0.01214	no
8	chr8.22135168	22135168	A	6.29	1.645	24.05	0.007197	0.01046	no
8	chr8.21861892	21861892	G	0.2542	0.08997	0.718	0.009728	0.004291	no
10	chr10.56874515	56874515	A	0.08913	0.01402	0.5664	0.01039	0.03269	no
5	chr5.54582216	54582216	A	4.6	1.418	14.92	0.01101	0.02542	no
8	chr8.21835629	21835629	A	3.774	1.35	10.55	0.01136	0.008071	no
8	chr8.21842264	21842264	A	3.774	1.35	10.55	0.01136	0.008071	no
8	chr8.21875109	21875109	A	3.774	1.35	10.55	0.01136	0.008071	no
10	chr10.48582562	48582562	A	0.2035	0.0587	0.7055	0.01208	0.00803	no
10	chr10.49715500	49715500	G	0.1512	0.03446	0.6631	0.01226	0.008735	no
10	chr10.56982895	56982895	A	0.09255	0.01415	0.6052	0.01299	0.04695	no
10	chr10.57950063	57950063	A	0.09255	0.01415	0.6052	0.01299	0.04695	no

Table 3.6: Top 25 results from fine-mapping/replication.

Results have been combined for all the chromosomes included in fine-mapping, with indications of whether the SNP was present in the original GWAS (12K) dataset. If the SNP was in the original GWAS, the p-value for association at that marker in the original sample is provided (n = 1 SNP). P-value only is also given for marker association signals in the analysis utilizing the stricter control criterion sample set. CHR - chromosome, SNP - snp name, BP - base pair, OR - odds ratio, SE - standard error of the odds ratio, L95 - lower 95% confidence interval, U95 - upper 95% confidence interval, P - p-value of association for that SNP, P_{strict} - p-value of association in stricter-control sample set.

Support for association of a region on CFA10 persisted in the second sample set of BOC. The strongest finding was at chr10.31603131, with a p-value of 9.09×10^{-5} and a strong "protective" OR of 0.008 (95% confidence interval 0.0006 – 0.09). The next strongest signal was detected on CFA10 at chr10.57383571, with a p-value of 0.001, OR 0.46. Additional local support was also found for both of these CFA10 regions (Table 3.6). The next most supported region was at 22Mb

on CFA8; the strongest association for CFA8 was ranked 16th at chr8.22135168, with a p-value of 0.007, OR 6.29.

Rank	CFA	SNP	OR	P	N_{set}	Perm p_{set}	Set
1st	10	chr10.31603131	0.008 (0.0006 - 0.09)	9.09E-05	427	0.002	chr10.31603131, chr10.36054599, chr10.34179508, chr10.32495782, **chr10.35153426**, chr10.35778878
2nd	10	chr10.57383571	0.46 (0.007 - 0.30)	0.001	749	0.174	chr10.57383571, chr10.48582562, chr10.49715500, chr10.55339145, chr10.49593364, chr10.51831584, chr10.54559962, chr10.49745955, chr10.56254773, **chr10.57685226**, chr10.51540995, chr10.50733918, chr10.49267100, chr10.48824301, chr10.55800066
15th	8	chr8.22135168	6.29 (1.65 - 24.05)	0.007	619	0.4458	chr8.22135168, chr8.21861892, chr8.21875109, chr8.25072329, chr8.22122802, chr8.27445059, chr8.23383109, chr8.23279180, chr8.28626478, chr8.23536911, chr8.22944048, chr8.22391661, chr8.28821931
18th	5	chr5.54582216	4.6 (1.42 - 14.92)	0.010	1465	0.702	chr5.54582216, chr5.52683997, **chr5.51683930**, chr5.40293087, chr5.46408387, chr5.38703372, chr5.53701768, chr5.53687444, chr5.52671293

Table 3.7: Top associations for each replication/fine-mapping chromosomal region.

Top associated hit for each chromosomal region, plus permuted p-value for the fine-mapping SNP set. There are several associations in the two CFA10 regions that demonstrate association in the replication sample. The 35Mb sets' permuted p-value reaches a level of significance for 4 sets tested ($p < 0.0125$) and includes a SNP (bold) implicated in epistatic interactions with the CFA10 at 58Mb region and with the CFA5 51.6Mb region (SNPs also in bold). Fine-mapping rank - rank by p-value of association for 3,260 SNPs in fine-mapping/replication analysis (all chromosome sets combined). CFA - chromosome, SNP - marker name (chromosome and base position), OR - odds ratio with 95% confidence interval in parenthesis, P - p-value of point-wise association, N_{set} - number of SNPs in set, Perm p_{set} - permuted p-value for set (EMP1), set - list of markers making up set.

Since this was not a true replication, a traditional $p < 0.05$ threshold for significance may not be applicable for this analysis; however, a Bonferroni correction for 3,260 tests (significance threshold $p < 1.53 \times 10^{-5}$) may be too strict considering these regions were chosen based on previous suggested association in the primary analysis. We thus performed set-based permutation analysis (four total sets – CFA5, CFA8, CFA10-35Mb, CFA10-55Mb) as another means of evaluating statistical significance. This analysis calculates the number of times a permuted set-

statistic exceeds the statistic for an original set composed of independent ($r^2 < 0.5$) SNPs that all reach a $p < 0.05$ threshold. The empirical p-value of the highest performing set was 0.002, which meets the significance threshold set at 0.0125 for four total sets tested. This set was composed of 6 SNPs spanning the region ~31.6Mb – 36Mb on CFA 10 (Table 3.7). The difference in coverage across platforms precluded a joint analysis with overlapping SNPs.

CFA	BP	first SNP	last SNP	P
10	57353121 - 57383571	chr10.57353121	chr10.57383571	0.0014
10	57362025 - 57384581	chr10.57362025	chr10.57384581	0.0014
10	58667130 - 58680393	chr10.58667130	chr10.58680393	0.0043
10	58673713 - 58688936	chr10.58673713	chr10.58688936	0.0043
10	58635986 - 58650185	chr10.58635986	BICF2G630492166	0.0057
10	58641713 - 58667130	chr10.58641713	chr10.58667130	0.0057
10	59368167 - 59391859	chr10.59368167	chr10.59391859	0.0130
10	58018214 - 58041409	chr10.58018214	chr10.58041409	0.0151
10	58029905 - 58071141	chr10.58029905	chr10.58071141	0.0151
10	58041409 - 58075689	chr10.58041409	chr10.58075689	0.0165
10	58071141 - 58088621	chr10.58071141	chr10.58088621	0.0165
10	58650185 - 58673713	BICF2G630492166	chr10.58673713	0.0168
10	48496503 - 48530062	chr10.48496503	chr10.48530062	0.0185
8	21835629 - 21861892	chr8.21835629	chr8.21861892	0.0192
8	21842264 - 21875109	chr8.21842264	chr8.21875109	0.0192
10	58088621 - 58110655	chr10.58088621	chr10.58110655	0.0194
10	56874515 - 56891092	chr10.56874515	chr10.56891092	0.0223
5	51624631 - 51670566	chr5.51624631	chr5.51670566	0.0242
5	51639046 - 51683930	chr5.51639046	chr5.51683930	0.0242
8	22135168 - 22155172	chr8.22135168	chr8.22155172	0.0267
8	22314192 - 22337065	chr8.22314192	chr8.22337065	0.0271
10	54668066 - 54696119	chr10.54668066	chr10.54696119	0.0282
8	22288469 - 22326392	chr8.22288469	chr8.22326392	0.0290
10	58484252 - 58501054	chr10.58484252	chr10.58501054	0.0292
10	58500421 - 58520322	chr10.58500421	chr10.58520322	0.0292

Table 3.8: Top 25 hits for 3-SNP haplotypic logistic regression with 1 MDS covariate.

The majority of top hits are in the CFA10-58Mb or CFA8-21Mb regions. CHR - chromosome, bp - base position range represented by haplotype, first SNP - first SNP of 3-SNP haplotype, last SNP - last SNP of 3-SNP haplotype, p - p-value for haplotypic association.

We investigated the two CFA10 candidate regions further to determine whether associations resulted from extended linkage disequilibrium (LD) in the region. Our findings suggested that the

signals were in fact in modest but extended LD, with $r^2 = 0.07$ and D' = 0.43 for chr10.31603131 and chr10.57383571, which are the top SNPs for each region and the top two associations in the fine-mapping analysis. A 3-SNP haplotypic logistic regression demonstrated further support for the ~58Mb region on CFA10, with the strongest signal ~57.3Mb, p = 0.001 (Table 3.8 and Figure 3.4, top panel). There was also modest haplotypic support on CFA8 around the 22Mb region (Table 3.8 and Figure 3.4, bottom panel) and on CFA5 at ~51.6Mb, with p-values around 0.02 (Table 3.8).

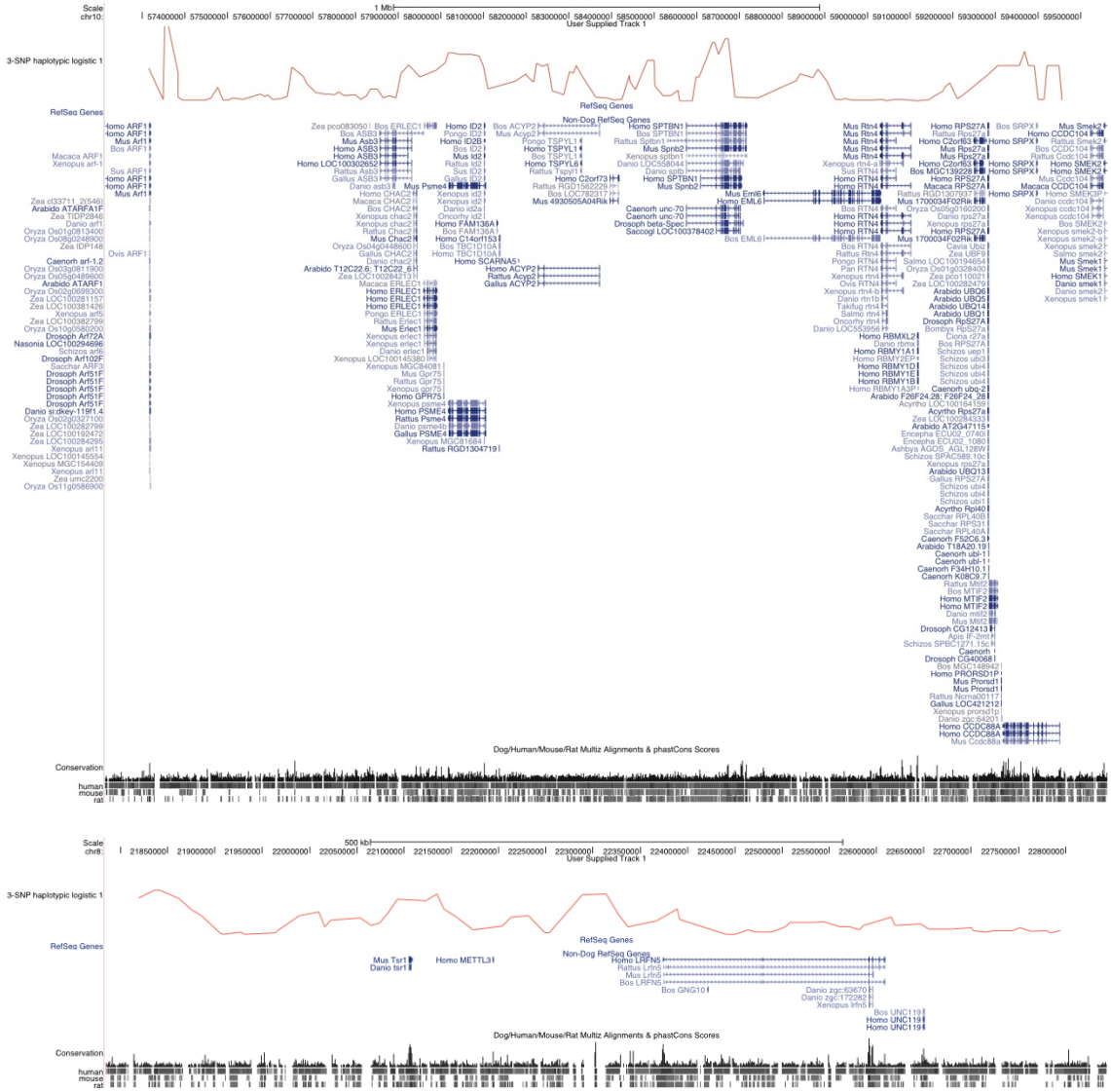


Figure 3.4: Manhattan plot of the 58Mb region on CFA10 and the 22Mb region on CFA8 for BOC noise phobia replication/fine-mapping.

For the 3-SNP haplotypic logistic regression with 1 MDS covariate in an independent sample of 50 BOC, the -$\log_{10}$(p-value) is plotted in red on the corresponding UCSC genome browser's putative genomic position on canFam2. Genes annotated in other species are shown below the plotted line. For CFA10, the strongest signal (2.85) is ~65kb upstream from the gene *ARF1* (as annotated in other species). The second strongest haplotype signal (2.37) lies over the putative *SPTBN1* gene region, a gene that is highly expressed in the brain. There is also haplotypic support on CFA8 (lower plot, signal = 1.72), ~200kb upstream from *LRFN5* (as annotated in other species), which is only expressed in the brain.

We assessed the relationship between these four candidate regions for epistatic interactions. We first analyzed only the SNPs present in the four regional sets (n = 43, Table 3.7). The top epistatic interaction was between chr5.51683930 and chr10.35153426, with a p-value of 0.01 and $OR_{interaction}$ = 0.13. Interestingly, the next strongest interaction was between the same CFA10 ~35Mb marker, chr10.35153426 and a SNP in the other CFA10 region, chr10.57685226 at a similar p-value = 0.01 and strong $OR_{interaction}$ = 23.52. Of note, these two markers appear to be in modest, but distant LD (r^2 = 0.02, D' = 0.35), which is consistent with the LD findings for the top two fine-mapping association hits. We also detected an interaction at p ~ 0.05 between chr5.53701768 and chr8.22135168, with $OR_{interaction}$ = 0.14. As with the primary analysis, we explored a stricter definition of controls and found the results did not differ overall, although significance levels were lower in the strict-controls analysis, likely due to the reduction in sample size (Table 3.9).

rank	CFA	SNP	BP	OR	P_{strict}	N_{set}	perm $p_{set\text{-}strict}$	set
1st	10	chr10.31603131	31603131	0.01 (0.0009 - 0.12)	2.67E-04	427	0.024	chr10.31603131, chr10.36054599, chr10.33534843, chr10.32495782, chr10.32130810, chr10.33913149, chr10.33188495
2nd	10	chr10.57383571	57383571	0.05 (0.006 - 0.36)	0.003	749	0.186	chr10.57383571, chr10.49745955, chr10.48582562, chr10.49715500, chr10.51831584, chr10.49593364, chr10.57685226, chr10.56254773, chr10.51540995, chr10.49819838, chr10.55892034, chr10.52667964, chr10.51715584, chr10.52729290, chr10.49315986
5th	8	chr8.21861892	21861892	0.17 (0.05 - 0.57)	0.004	619	0.1734	chr5.54783511, chr5.46408387, chr5.54231348
46th	5	chr5.54783511	54783511	6.10 (1.27 - 29.28)	0.024	1465	0.761	chr8.21861892, chr8.21842264, chr8.22135168, chr8.22122802, chr8.23069662, chr8.25139379, chr8.22391661, chr8.22016264, chr8.25072329, chr8.27445059, chr8.27921468, chr8.25667093, chr8.22944048, chr8.30342375, chr8.23536911, chr8.28100809, chr8.21472660, chr8.26910112, chr8.28821931, chr8.27895878

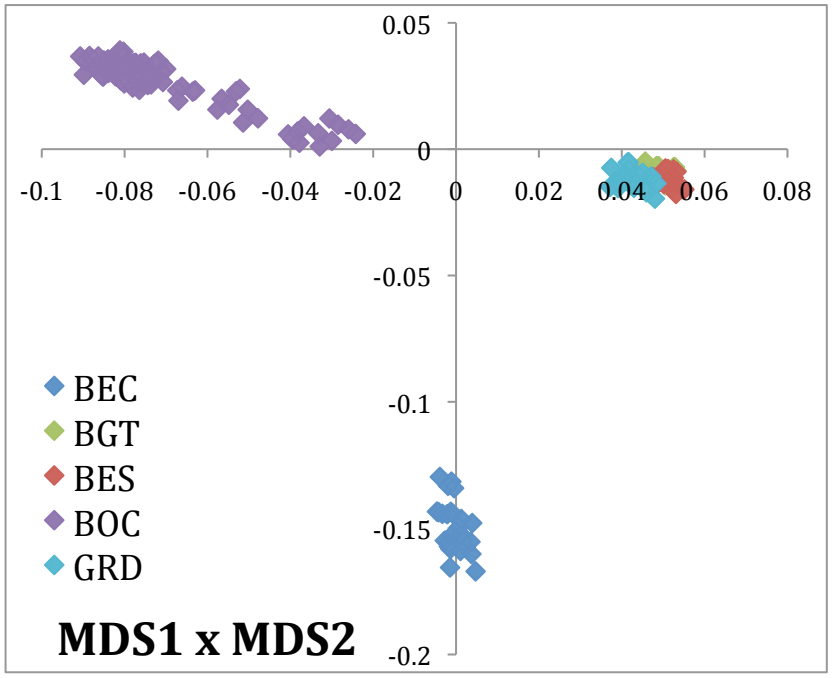
CFA	BP	first SNP	last SNP	P_{strict}
10	57353121 - 57383571	chr10.57353121	chr10.57383571	0.0032
10	57362025 - 57384581	chr10.57362025	chr10.57384581	0.0032
8	22288469 - 22326392	chr8.22288469	chr8.22326392	0.0068
10	48496503 - 48530062	chr10.48496503	chr10.48530062	0.0092
10	58635986 - 58650185	chr10.58635986	BICF2G630492166	0.0120
10	58641713 - 58667130	chr10.58641713	chr10.58667130	0.0120
10	58667130 - 58680393	chr10.58667130	chr10.58680393	0.0121
10	58673713 - 58688936	chr10.58673713	chr10.58688936	0.0121
8	21835629 - 21861892	chr8.21835629	chr8.21861892	0.0124
8	21842264 - 21875109	chr8.21842264	chr8.21875109	0.0124

Table 3.9: Findings for replication/fine-mapping with stricter control criterion.

(Upper panel) Top associated hit for each chromosomal region, plus permuted p-value for the fine-mapping SNP set. No sets reach a level of significance for 4 sets tested ($p < 0.0125$). Fine-mapping rank - rank by p-value of association for 3,260 SNPs in replication/fine-mapping analysis (all chromosome sets combined). CFA - chromosome, SNP - marker name, BP - base position, OR - odds ratio with 95% confidence interval in parenthesis, P_{strict} - p-value of association in stricter-control sample set, N_{set} - number of SNPs in set, perm $p_{set\text{-}strict}$ - permuted p-value for set (EMP1) in stricter-control sample set, set - list of markers making up set. (Lower panel) Top 10 hits for 3-SNP haplotypic logistic regression with 1 MDS covariate for the finemapping regions in stricter-control sample set. In the analysis utilizing stricter control criterion there is more support for the CFA8 region. BP - base position range represented by haplotype, first SNP - first SNP of 3-SNP haplotype, last SNP - last SNP of 3-SNP haplotype, P_{strict} - p-value of association in stricter-control sample set.

3.3.2. Secondary analysis in multiple breeds. Within-breed analyses were performed in an additional six populations. Pseudoheritability values were used to assess breeds for which noise

phobia might be heritable (Table 3.3). Two populations were dropped (AUS and GSD) from analysis because of low estimates of pseudoheritability. BEC, BES, BGT and GRD were examined with the primary BOC sample in a meta-analysis to increase power and to evaluate trends for shared associated regions across breeds. We chose a meta-analysis over other methods because the distinct clustering of each breed (Figure 3.5) would have required numerous covariates to correct for inter-breed stratification.



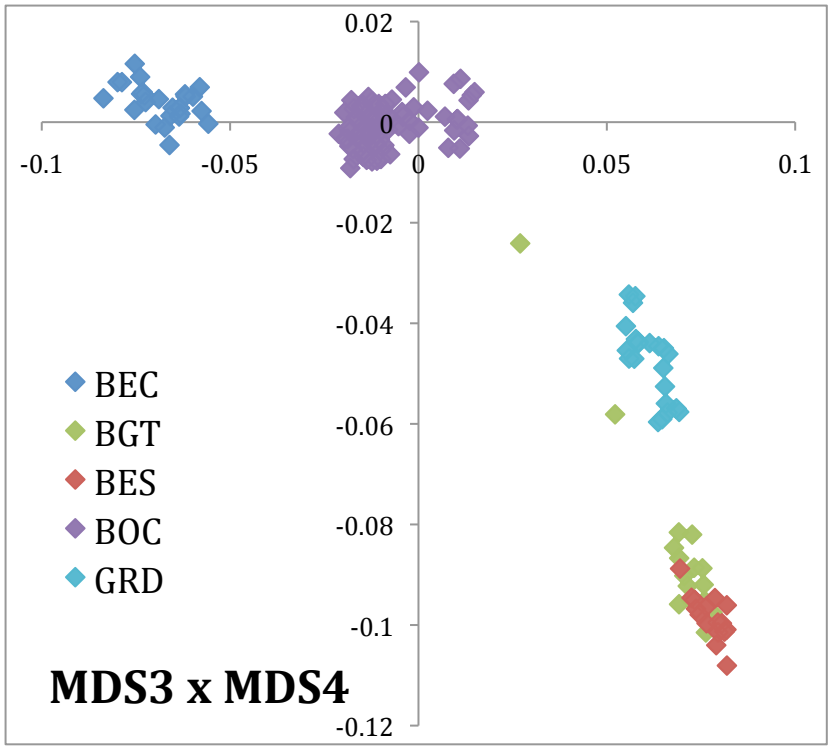


Figure 3.5: MDS plots for samples from all breeds analyzed for noise phobia.

Fifteen MDS covariates were calculated utilizing the 3-platform merged dataset of roughly 9,000 SNPs for ALL dogs genotyped, and the first four MDS covariates for only dogs included in the noise phobia analysis are shown here. Each breed, indicated by a different color, clusters separately, with BOC demonstrating the most within-breed variation. BEC - bearded collie, BES - Belgian shepherd, BGT - Belgian tervuren, BOC - Border collie, GRD - great Dane.

Meta-analysis with the five breeds resulted in a null effect in all regions where the BOC-only analysis demonstrated modest association signals, with no findings reaching genome-wide significance (Table 3.10). Each single-breed analysis demonstrated top findings around the same order of magnitude as the primary BOC analysis, and pseudoheritability calculations suggested that noise phobia was heritable within each sample (Table 3.3).

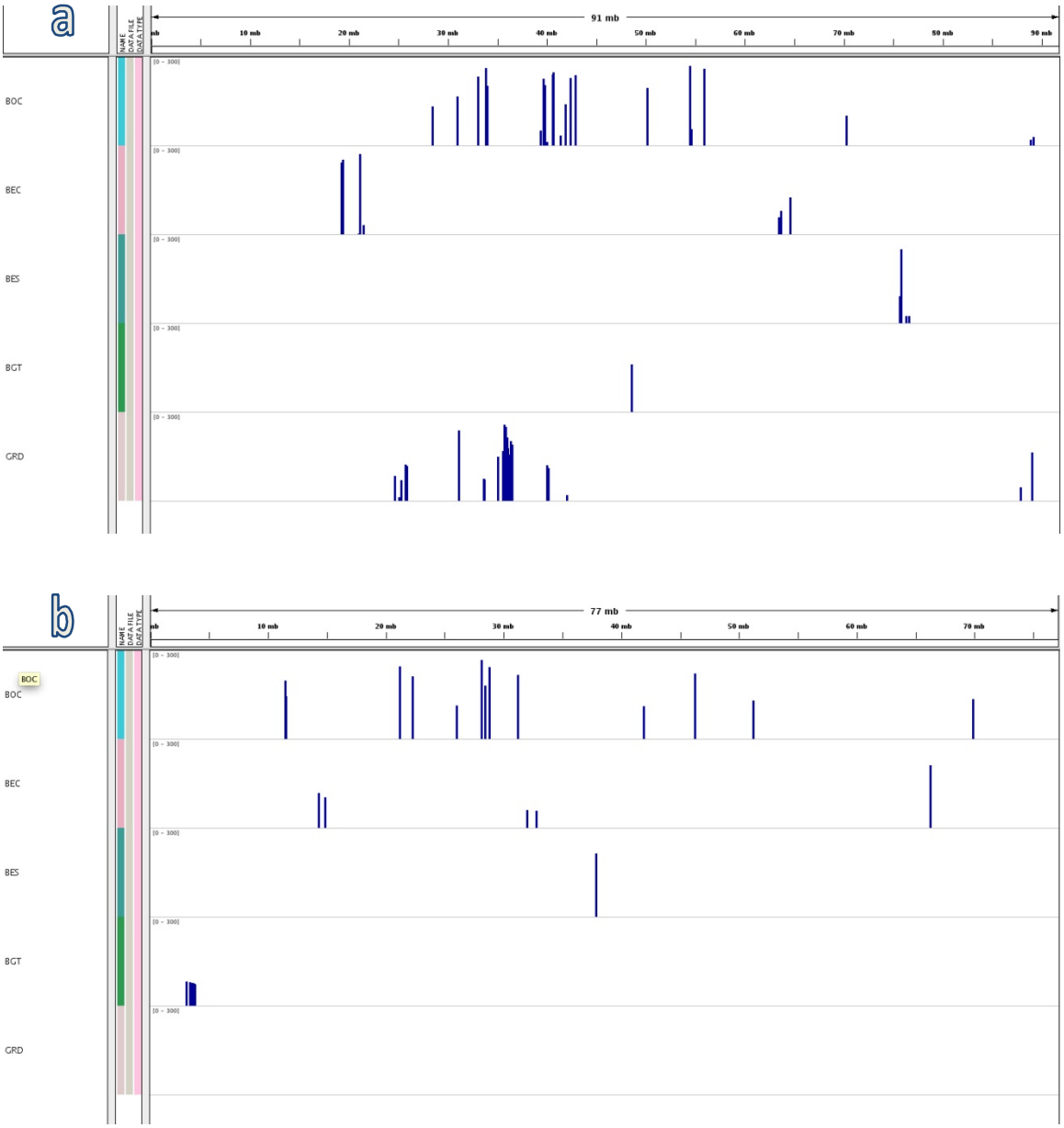
CFA	BP	SNP	N	P*	OR*	Q	BOC	BEC	GRD	BGT	BES
3	90817113	chr3.90817113	3	0.2776	0.5343	0.97	NA	NA	0.4482	0.6356	0.5354
34	37492667	chr34.37492667	3	0.2989	1.8216	0.9688	NA	NA	1.6214	2.2357	1.6675
36	26082312	chr36.26082312	3	0.3044	0.5527	0.965	NA	NA	0.4788	0.6845	0.5151
5	35897353	chr5.35897353	3	0.3084	1.8004	0.9806	NA	NA	1.711	2.1095	1.6171
5	35915206	chr5.35915206	3	0.3084	1.8004	0.9806	NA	NA	1.711	2.1095	1.6171
5	35920124	chr5.35920124	3	0.3084	1.8004	0.9806	NA	NA	1.711	2.1095	1.6171
4	52782032	chr4.52782032	2	0.3096	0.4875	0.9999	NA	NA	0.4876	NA	0.4875
29	16961809	chr29.16961809	3	0.312	0.5578	0.9567	NA	NA	0.4474	0.6801	0.5704
13	50291953	chr13.50291953	3	0.3124	1.7919	0.9484	NA	NA	1.5494	2.3367	1.5891
13	50293358	chr13.50293358	3	0.3124	1.7919	0.9484	NA	NA	1.5494	2.3367	1.5891
8	38066966	chr8.38066966	3	0.3127	0.5583	0.9619	NA	0.6529	0.4474	NA	0.5956
1	29638584	chr1.29638584	3	0.3137	1.789	0.9781	NA	NA	1.7791	2.0819	1.5458
8	45460871	chr8.45460871	3	0.3155	0.5602	0.9543	NA	NA	0.4474	0.6889	0.5704
36	26076543	chr36.26076543	3	0.3165	0.5608	0.9553	NA	NA	0.4788	0.7153	0.5151
28	19196259	chr28.19196259	3	0.3184	1.7789	0.963	NA	NA	1.5494	2.2214	1.6357
23	12966846	chr23.12966846	3	0.3227	0.565	0.908	NA	NA	0.4474	0.8039	0.5015
36	26027420	chr36.26027420	3	0.3233	0.5654	0.9589	NA	NA	0.4921	0.7153	0.5134
4	72670494	chr4.72670494	2	0.326	2.0028	0.9415	NA	NA	1.9015	2.1095	NA
21	45312922	chr21.45312922	2	0.326	2.0028	0.9415	NA	NA	1.9015	2.1095	NA
25	38139032	chr25.38139032	2	0.326	2.0028	0.9415	NA	NA	1.9015	2.1095	NA
20	28257988	chr20.28257988	3	0.3275	1.76	0.9637	NA	NA	1.5731	2.1974	1.5771
6	19178071	chr6.19178071	2	0.3278	1.9976	0.9386	NA	1.8916	NA	2.1095	NA
18	41416245	chr18.41416245	3	0.3282	1.7587	0.9742	NA	NA	1.5494	2.1095	1.6643
28	19101982	chr28.19101982	3	0.3288	1.7573	0.9704	NA	NA	1.5494	2.1413	1.6357
28	19226244	chr28.19226244	3	0.3288	1.7573	0.9704	NA	NA	1.5494	2.1413	1.6357

Table 3.10: Top 25 results from meta-analysis.

Results are reported for SNPs seen in ≥3 breed analyses. CHR - chromosome, BP - base pair, SNP - snp name, N - number of breeds in analysis, P - p-value, OR - odds ratio, Q - p-value for Cochrane's Q statistic; BOC - OR in Border Collie GWAS, BEC - OR in Bearded Collie GWAS, GRD - OR in Great Dane GWAS, BGT - OR in Belgian Tervuren GWAS, BES - OR in Belgian Shepherd GWAS. *p-values and OR for fixed effect vs. random effect model were the same and thus are listed only once.

We thus explored an alternate hypothesis of genetic heterogeneity for noise phobia across breeds. We compared the top findings from each breed-specific analysis to detect patterns in association between breeds. Because sample size and SNP coverage differed from breed to breed, we visualized findings based on ordinal rank rather than strength of p-value. When comparing EMMAX results from the five different breed analyses – BEC, BES, BGT, BOC, GRD – we found suggestions of both locus and allelic heterogeneity (Figure 3.6). Some chromosomes demonstrated top findings in every breed analysis, but at different SNPs within a ~10-200kb region in each breed, suggesting allelic heterogeneity (Figure 3.6a). However, in other regions an

association was only suggested by one or a subset of breeds while other breeds demonstrated no support in their top 300 findings (Figure 3.6b). Upon closer examination of the top findings for each breed, we found that only five (0.3%) of the top ranked findings were seen across two different analyses (and none were seen in more than two analyses).



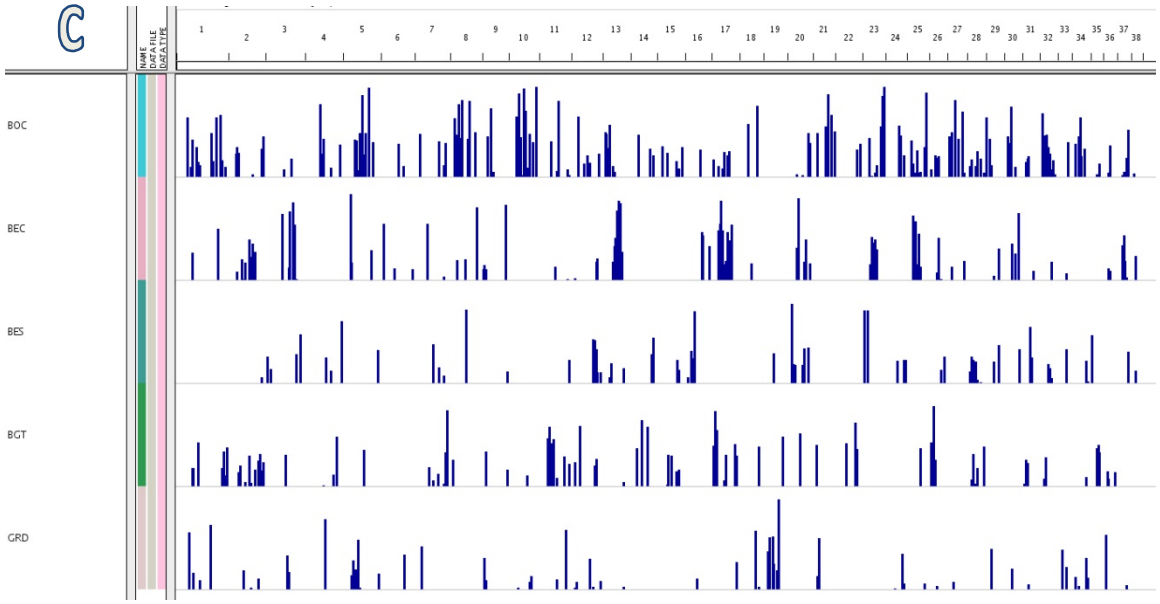


Figure 3.6: Visualization of top results from five different breed GWAS's for noise phobia.

Height of bar represents rank (1st is tallest, 300th is shortest) in top findings of breed-specific GWAS for noise phobia. (a) Support for association with CFA5 is seen across all five breeds but at varying locations. (b) Support on CFA8 is seen in BOC and only three of four other breeds at modest levels. (c) Whole-genome results demonstrate some regional support across multiple breeds but some breed-specific signals as well. For example, CFA10 demonstrates multiple top association signals in BOC but support is not demonstrated in the other four breeds. BOC - Border collie, BEC - bearded collie, BES - Belgian shepherd, BGT - Belgian tervuren, GRD - great Dane.

3.4. Discussion

Our study of noise phobia in small samples of herding- and utility-breed dogs suggests that this behavioral disorder is moderately heritable. Although none of our results reached genome-wide significance, these findings nevertheless suggest multiple genomic regions of potential interest, though much larger sample sizes of single breeds would be required given the heterogeneity observed across our multi-breed sample.

We identified two regions on CFA10 that demonstrate associations with noise phobia in BOC, plus continued support for associations on CFA5 and CFA8. While the findings for these regions in the original sample demonstrated only modest associations, their continued support in an independent sample of BOC suggests that genes in these regions may play a role in increased susceptibility to noise phobia in this breed. The first region on CFA10 is from ~31.6 – 36Mb, and contains one SNP, chr10.35153426 that appears to have epistatic interactions with two different SNPs: one on CFA5, chr5.51683930, and one in the second CFA10 candidate region around 58Mb, chr10.57685226. Chr10.35153426 is located in the predicted intronic region of *RFX4*, which encodes a transcription factor important in early brain development [19, 20]. This SNP is part of the only marker set that reached statistical significance in the set-based permutation analysis (Table 3.7). Chr5.51683930 is located in the predicted intronic region of *NFIA* (as annotated in other species), a DNA-binding protein and a member of a family of key regulators of central nervous system development and brain function [21]. *Nfia* knockout mice display a striking brain phenotype that includes agenesis of the corpus callosum and malformations of midline glial populations, which are required to guide axons of the corpus callosum across the midline of the developing brain [21]. Lower signal intensity in the corpus callosum has been reported in adolescents and adults with bipolar disorder [22]. Chr10.57685226 is located between two putative genes, *ARF1* and *ASB3* (as annotated in other species). *ARF1* is ~340kb away and is an ADP-ribosylation factor that is highly expressed in the brain and plays a critical role in eukaryotic cells regulating Golgi-membrane trafficking in the secretory pathway. *ASB3* is ~170kb away and thought to be a negative regulator of TNF-R2 –mediated cellular responses to TNF-α [23].

In addition to suggestive epistatic findings, there is also regional haplotypic support around this genomic region (Figure 3.4, Table 3.8). This includes a haplotypic signal above the putative gene *SPTBN1*, a gene that is highly expressed in brain and leads to altered TGF-β signaling when

disrupted [24]. Finally, the 22Mb region on CFA8 also shows continued support in the fine-mapping analyses, with haplotypic support (Figure 3.4) and a suggested interaction at chr8.22135168 and a SNP on CFA5, chr5.53701768. Chr8.22135168 is located 60kb upstream of the putative protein *METTL3* and 200kb upstream of putative *LRFN5* (both annotated in other species). *METTL3* is a critical subunit in a multicomplex enzyme that catalyzes methylation of internal adenosine residues in eukaryotic mRNA. *LRFN5* is expressed exclusively in the brain, and long-distance epigenetic silencing of *LRFN5* in a patient with a rare translocation is hypothesized to be the cause of the patient's autism [25].

Using Gene Ontology annotations, we found that both *NFIA* and *RFX4* are DNA-binding dependent transcription regulators. These roles, in addition to *METTL3*'s role in RNA methylation and in light of the multiple epistatic interactions detected between all three chromosomes, generate hypotheses regarding a larger network of gene/expression regulation. In particular, *SPTBN1* and *LRFN5,* which are expressed predominantly in the brain, make attractive candidates for altered expression, as does *ARF1* with its role in vesicle-mediated transport and receptor signaling. In fact, further analysis of all finemapping SNPs for epistasis suggests an even more complex picture, with multiple regions on each of the three candidate chromosomes (CFA5, 8 and 10) demonstrating predicted interactions with each other. A cartoon depicting the epistatic interactions detected between these three chromosomes is shown in Figure 3.7.

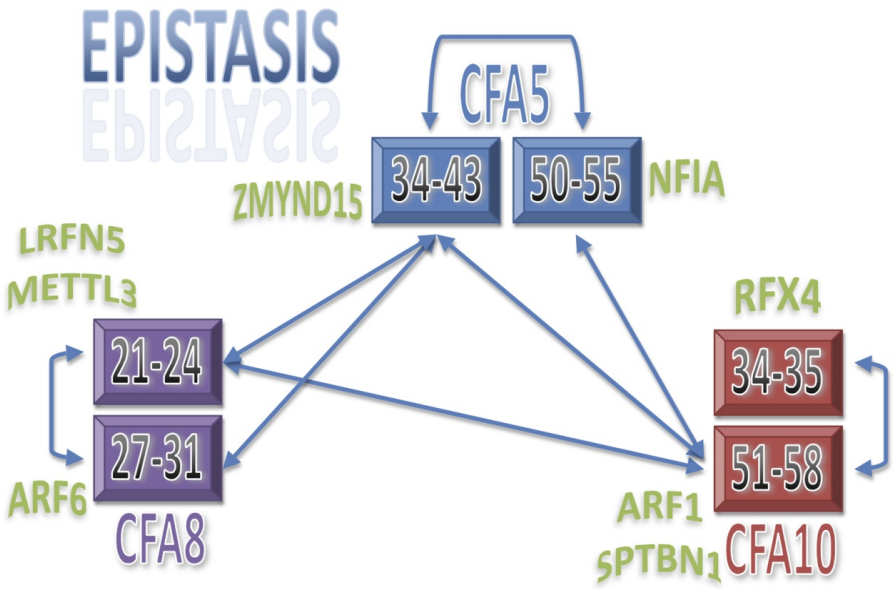


Figure 3.7: Epistatic interactions between three canine chromosomes.

Epistasis analysis in the replication/fine-mapping sample suggests possible interactions (indicated by blue arrows) between three different canine chromosomes (each represented by a different color, with position in black, in mega-bases) in noise phobia in BOC. Candidate genes (in green) within the implicated genomic regions include a number of regulatory genes that are expressed in the brain. Notably, there are also two members of the same gene family – *ARF1* and *ARF6* – implicated on CFA10 and CFA8, respectively. Both family members play different roles in membrane trafficking.

There are many limitations to the present study. First, our control samples were allowed to have subclinical noise response symptoms, which raises the possibility of controls actually containing some cases, reducing our power. Analyses utilizing a stricter definition for controls led to similar results, with slight improvements in statistical significance. Second, the number of samples is small, with 84 samples for our most numerous breed, and for many breeds the number of cases versus controls is not balanced. If the genetic architecture of the phenotype is polygenic, power will be limited even when there is prominent heritability for the trait. Our sample may simply not

be large enough to reliably detect association. In a similar study of a canine systemic lupus erythematosus-related disease complex, GWAS of 81 cases and 57 controls led to findings that achieved genome-wide significance only with the addition of new cases and controls [26]. Future efforts at sample collection may thus improve our power to detect association. Our small sample size may also artificially inflate the estimated pseudoheritability for noise phobia. Third, samples were genotyped on three different SNP platforms. While reproducibility was demonstrated for a common set of SNPs on all platforms, the use of three different sets of SNPs may have resulted in the loss of genotypic information when merging datasets to obtain the maximum number of samples to analyze. For example, in BOC genotyped on the HD platform, 10,085 SNPs overlapped with the 12K data set (v2.0/SNP20), and those SNPs tagged an additional 49,562 SNPs ($r^2 \geq 0.5$ for pairwise LD within 5Mb), suggesting that those SNPs cover approximately 40.8% of the 146K SNPs genotyped (data not shown). It may be possible to increase our genomic coverage by genotyping all samples with higher density arrays, or less optimally with genotype imputation. Additional heterogeneity in breed-associated behavior may be due to cultural differences between dog owners across continents, as well as to differences in the work/purpose for which the subject dogs were intended (or not). The challenge in phenotyping—namely determining whether a particular dog has a 'genetic' noise phobia or whether the cause is traumatic experience from loud noises—may also obscure the results of genetic investigations.

Despite promising candidate genes implicated in the fine-mapping/replication analyses in BOC, we did not see support for these loci in other breeds (Table 3.10). Our results also provide evidence that the genetic loci underlying noise phobia in BOC are distinct from those underlying noise phobia in other breeds. The history of selective breeding in BOC may have resulted in a distinct subset of noise phobia risk loci that increased in frequency only within this breed. The presence of a finding shared between BOC populations of different geographical origins suggests a common vulnerability. This is a strong possibility given that the majority of our BOC samples

are from the herding community, where selection regimes for specific behavioral suites may enrich the population for traits such as reactivity that may ultimately predispose dogs to noise sensitivity [9]. The observed *locus* heterogeneity would support the expectation that while genetic determinants of selectively bred traits may be shared across breeds, this would not necessarily be the case for maladaptive traits or disorders. A second possibility is that there is *allelic* heterogeneity across the different breeds; in other words, although the same genomic regions may be implicated in noise phobia across all breeds of dog, slight differences within each breed may result in association signals at different alleles distinct to each breed. This latter possibility, which would still demonstrate specific genomic regions as common risk loci for noise phobia, would be undetectable in meta-analysis, consistent with the lack of association observed in our own meta-analysis. This heterogeneity may be common. For example, Wood and colleagues [27] observed similar results in their examination of candidate gene markers associated with canine atopic dermatitis. In their study, meta-analysis across eight breeds rendered only one significant association that generalized across all eight breeds, despite multiple findings common to several (but not all) breeds within the study. Also, it has been suggested that dogs of different breeds may display variability in response to pharmacologic substrates, suggesting fundamental differences in physiology [28]. Such locus and allelic heterogeneity may be critical when designing studies of complex disease and behavior, and will be very important to consider should genetic tests ever be designed for detection of risk loci in breeding dogs.

We have provided an estimate for the heritability of noise phobia in canines, which is remarkably similar to previous estimates for gun shyness [29, 30], and are comparable to heritability estimates made for another behavioral disorder in dogs, aggression, utilizing an owner questionnaire [31]. Although pseudoheritability is not a true heritability measure in the traditional sense, it provides an estimate of the genetic contribution to the trait under study. In fact, pseudoheritability is highly correlated with heritability for many disease and quantitative

traits [16]. For traits or conditions such as noise phobia for which a heritability value is unclear, or may differ between breeds, estimates of pseudoheritability provide a basis upon which further genetic studies may be designed. In fact, we used pseudoheritability as a means of determining, prior to meta-analysis, available breed samples for which this trait may be heritable. The advantage of this strategy is the potential for discarding samples in which the presence of genetic factors is less likely, although small samples may show reduced pseudoheritability. The clear disadvantage of this strategy is, of course, that it assumes that pseudoheritability is a valid measure of heritability. If this assumption is wrong, we may discard samples that are in fact informative. However, because the calculation of pseudoheritability uses empirical genomic data drawn from the sample cohort itself, we believe this approach to be a reasonable method for determining in advance which samples to include or exclude from genetic studies to increase the likelihood of detecting association signals.

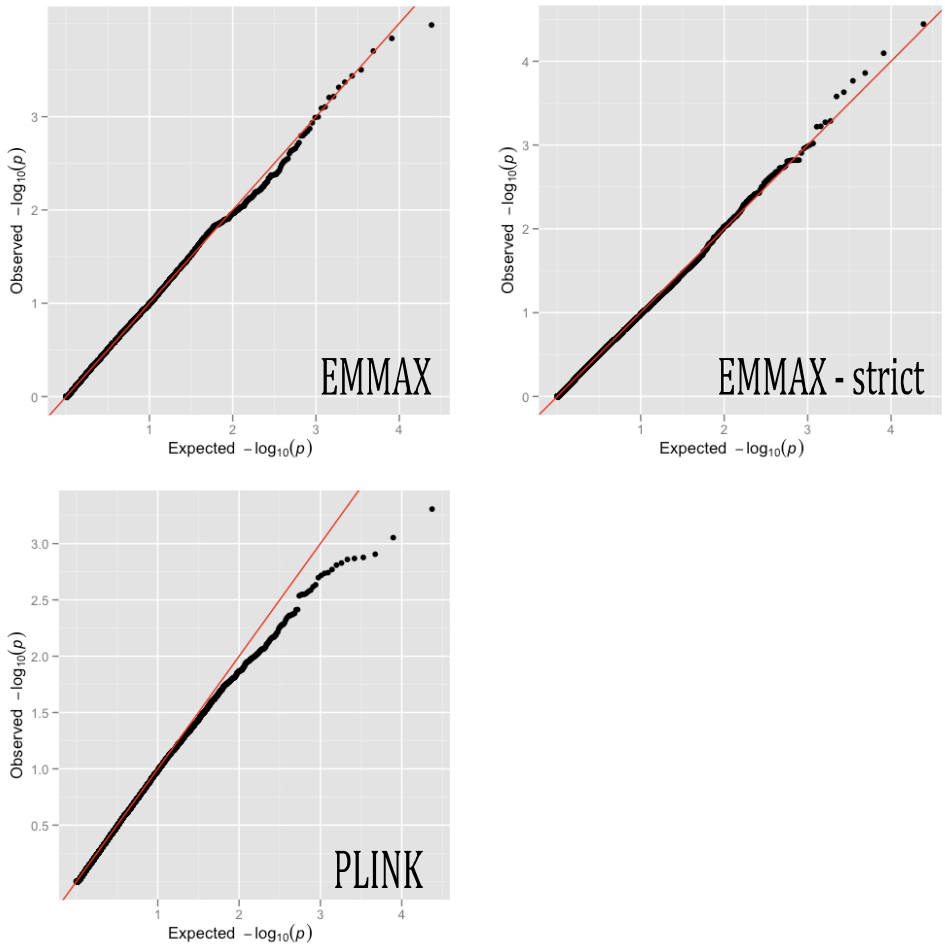


Figure 3.8: Q-Q plots for BOC analyses.

EMMAX plot is on the top left and logistic regression with 1 MDS covariate is on the bottom. Top right plot represents EMMAX analysis utilizing stricter control criterion. EMMAX appears to effectively correct for stratification within the BOC sample with less diminishment of statistical power, and stricter criterion for controls further appears to enhance power to detect association signals.

EMMAX was able to account for within-breed stratification in our sample of BOC (Figure 3.8). This was critical given the sample included dogs from multiple geographic locales as well as from a diverse range of purpose-bred lines, all of which have the potential to confound GWAS results [32, 33]. In addition, we found EMMAX to provide a better correction with less loss of power than the more traditional logistic regression with MDS covariate method (Figure 3.8).

Finally, we found that utilizing stricter control requirements also modestly increased our power to detect associations (Figure 3.8), although the findings remained overall largely unchanged (Table 3.5). Taken together, these findings suggest that EMMAX is a highly effective tool for performing GWAS in dogs in which cryptic relatedness and intra-breed stratification may be of particular concern, although we still assume that susceptibility variants contribute small effects.

In summary, we have conducted the first genetic study of the canine anxiety disorder noise phobia. We have identified multiple candidate regions in Border collies that may confer risk for noise phobia, including loci on CFA5, CFA8 and two loci on CFA10 that all appear to demonstrate epistatic interactions. Although none of these signals reached genome-wide significance, we feel they are worth exploring further in larger samples of herding dog populations, particularly given the continued support of these findings in an independent BOC replication sample. The modest genomic coverage of our combined SNP dataset greatly reduced our ability to identify regional support for suggestive findings in our primary analysis through joint analysis, and highlights opportunities for future investigations that may utilize new, high-density SNP arrays. Additionally, the genetic heterogeneity demonstrated in our exploratory multi-breed analyses emphasizes the need for sizeable primary samples from one breed in the investigation of complex behavioral traits. Denser coverage of a larger number of samples we are actively pursuing in multiple breeds will likely uncover several new genetic loci, including those regions for which this pilot study has hinted, providing important insight into a troubling disorder for both dogs and humans.

3.5 References

1. Spady TC, Ostrander EA (2008): Canine behavioral genetics: pointing out the phenotypes and herding up the genes. Am J Hum Genet 82: 10-18.

2. Dodman NH, Karlsson EK, Moon-Fanelli A, Galdzicka M, Perloski M, Shuster L et al (2010): A canine chromosome 7 locus confers compulsive disorder susceptibility. Mol Psychiatry 15: 8-10.
3. Bozdagi, O., Wang, X. b., Nikitczuk, J. S., Anderson, T. R., Bloss, E. B., Radice, G. L., Zhou, Q., Benson, D. L., & Huntley, G. W. (2010) Persistence of Coordinated Long-Term Potentiation and Dendritic Spine Enlargement at Mature Hippocampal CA1 Synapses Requires N-Cadherin. *Journal of Neuroscience* **30**, 9984-9989.
4. Overall KL, Hamilton SP, Chang ML (2006): Understanding the genetic basis of canine anxiety: phenotyping dogs for behavioral, neurochemical, and genetic assessment. J Vet Behav: Clin Appl Res 1: 124-141.
5. Sherman, B. L. & Mills, D. S. (2008) Canine Anxieties and Phobias: An Update on Separation Anxiety and Noise Aversions. *Veterinary Clinics of North America: Small Animal Practice* **38**, 1081-1106.
6. American Psychiatric Association, American Psychiatric Association, & Task Force on DSM-IV. (2000) *Diagnostic and statistical manual of mental disorders DSM-IV-TR*. American Psychiatric Association, Washington, DC.
7. Smoller JW, Gardner-Schuster E, Covino J (2008): The genetic basis of panic and phobic anxiety disorders. Am J Med Genet C Semin Med Genet 148: 118-126.
8. Tuber DS, Hothersall D, Peters MF (1982): Treatment of fears and phobias in dogs. Vet Clin North Am Small Anim Pract 12: 607-623.
9. McCobb EC, Brown EA, Damiani K, Dodman NH (2001): Thunderstorm phobia in dogs: an Internet survey of 69 cases. J Am Anim Hosp Assoc 37: 319-324.
10. Overall KL, Dunham AE, Frank D (2001): Frequency of nonspecific clinical signs in dogs with separation anxiety, thunderstorm phobia, and noise phobia, alone or in combination. J Am Vet Med Assoc 219: 467-473.
11. Seksel K, Lindeman MJ (2001): Use of clomipramine in treatment of obsessive-compulsive disorder, separation anxiety and noise phobia in dogs: a preliminary, clinical study. Aust Vet J 79: 252-256.

12. Branson NJ, Rogers LJ (2006): Relationship between paw preference strength and noise phobia in Canis familiaris. J Comp Psychol 120: 176-183.

13. Mathersul D, Williams LM, Hopkinson PJ, Kemp AH. (2008) Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety. Emotion. 8(4):560-72.

14. Oertel V, Knöchel C, Rotarska-Jagiela A, Schönmeyer R, Lindner M, van de Ven V, Haenschel C, Uhlhaas P, Maurer K, Linden DE. (2010) Reduced laterality as a trait marker of schizophrenia--evidence from structural and functional neuroimaging. J Neurosci. 30(6):2289-99.

15. Cottam, N. & Dodman, N. H. (2009) Comparison of the effectiveness of a purported anti-static cape (the Storm Defender«) vs. a placebo cape in the treatment of canine thunderstorm phobia as assessed by owners' reports. *Applied Animal Behaviour Science* **119**, 78-84.

16. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong Sy, Freimer NB et al (2010): Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348-354.

17. Foulley JL, Gianola D, Im S (1987): Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. Theoretical and Applied Genetics 73: 870-877.

18. Olesen I, Perez-Enciso M, Gianola D, Thomas DL (1994): A comparison of normal and nonnormal mixed models for number of lambs born in Norwegian sheep. J Anim Sci 72: 1166-1173.

19. Zhang D, Stumpo DJ, Graves JP, DeGraff LM, Grissom SF, Collins JB, Li L, Zeldin DC, Blackshear PJ. (2006) Identification of potential target genes for RFX4_v3, a transcription factor critical for brain development. J Neurochem. 98(3):860-75.

20. Zhang D, Zeldin DC, Blackshear PJ. Regulatory factor X4 variant 3: a transcription factor involved in brain development and disease. (2007) J Neurosci Res. 85(16):3515-22.

21. Mason, S., Piper, M., Gronostajski, R., & Richards, L. (2009) Nuclear Factor One Transcription Factors in CNS Development. *Molecular Neurobiology* **39**, 10-23.

22. Caetano SC, Silveira CM, Kaur S, Nicoletti M, Hatch JP, Brambilla P, Sassi R, Axelson D, Keshavan MS, Ryan ND, Birmaher B, Soares JC. (2008) Abnormal corpus callosum myelination in pediatric bipolar patients. *J Affect Disord.* 108(3):297-301.

23. Chung, A. S., Guan, Y. J., Yuan, Z. L., Albina, J. E., & Chin, Y. E. (2005) Ankyrin repeat and SOCS box 3 (ASB3) mediates ubiquitination and degradation of tumor necrosis factor receptor II. *Mol Cell Biol* **25**, 4716-4726.

24. Tang, Y., Katuri, V., Dillner, A., Mishra, B., Deng, C. X., & Mishra, L. (2003) Disruption of Transforming Growth Factor-beta Signaling in ELF beta -Spectrin-Deficient Mice. *Science* **299**, 574-577.

25. de Bruijn DR, van Dijk AH, Pfundt R, Hoischen A, Merkx GF, Gradek GA, Lybæk H, Stray-Pedersen A, Brunner HG, Houge G. (2010) Severe Progressive Autism Associated with Two de novo Changes: A 2.6-Mb 2q31.1 Deletion and a Balanced t(14;21)(q21.1;p11.2) Translocation with Long-Range Epigenetic Silencing of LRFN5 Expression. *Mol Syndromol.* 1(1):46-57.

26. Wilbe M, Jokinen P, Truve K, Seppala EH, Karlsson EK, Biagi T et al (2010): Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet 42: 250-254.

27. Wood SH, Ollier WE, Nuttall T, McEwan NA, Carter SD (2010): Despite identifying some shared gene associations with human atopic dermatitis the use of multiple dog breeds from various locations limits detection of gene associations in canine atopic dermatitis. Veterinary Immunology and Immunopathology In Press, Corrected Proof.

28. Fleischer S, Sharkey M, Mealey K, Ostrander E, Martinez M (2008): Pharmacogenetic and Metabolic Differences Between Dog Breeds: Their Impact on Canine Medicine and the Use of the Dog as a Preclinical Animal Model. The AAPS Journal 10: 110-119.

29. Goddard, M. E. & Beilharz, R. G. (1985) A multivariate analysis of the genetics of fearfulness in potential guide dogs. *Behav Genet* **15**, 69-89.

30. van der Waaij, E. H., Wilsson, E., & Strandberg, E. (2008) Genetic analysis of results of a Swedish behavior test on German Shepherd Dogs and Labrador Retrievers. *J Anim Sci* **86**, 2853-2861.

31. Liinamo AE, van den Berg L, Leegwater PAJ, Schilder MBH, van Arendonk JAM, van Oost BA (2007): Genetic variation in aggression-related traits in Golden Retriever dogs. Applied Animal Behaviour Science 104: 95-106.

32. Quignon P, Herbin L, Cadieu E, Kirkness EF, Hedan B, Mosher DS et al (2007): Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. PLoS ONE 2: e1324.
33. Chang ML, Yokoyama JS, Branson N, Dyer DJ, Hitte C, Overall KL et al (2009): Intrabreed Stratification Related to Divergent Selection Regimes in Purebred Dogs May Affect the Interpretation of Genetic Association Studies. J Hered 100: S28-S36.

Chapter 4: Genome-wide Association Study of Adult-onset Deafness in Herding Border Collies Implicates Human Deafness Locus[****]

In Chapter 4, I present our work mapping risk loci for adult-onset deafness in herding Border collies. The motivation for this work is two-fold: first, fueled by an increasing aging human population; second, rooted in the sheepdog herding community. Both groups seek common answers – to identify risk loci underlying age-related hearing impairment. In this final data chapter, I will outline the genome-wide study that identified a strong candidate region on *Canis familiaris* chromosome (CFA) 6. I will also describe the fine-mapping we performed using target enrichment and next-generation sequencing technology. Finally, I will highlight the implications of our findings for both the human and canine communities.

4.1. Introduction

"Deafness" is a lack or deficiency in the sense of hearing [1]. Acquired hearing loss is typically caused by disease, toxicity or noise trauma. There are also two forms of deafness that are affected by genetics: congenital and late onset. Congenital deafness occurs when one is born with a hearing deficit, while late onset, as the name implies, occurs later in life. Age-related hearing loss (ARHL, also known as presbycusis) is a continuing problem in aging human populations, and with a prevalence of around 40% in those older than 65 years of age. ARHL can lead to problems with communication, isolation, depression and possibly even dementia in the severely affected

[****]Portions of this section appear in the manuscript: Yokoyama JS, Lam ET, Erdman CA, Corneveaux J, Ruhe AL, Robertson KR, Chang ML, Overall KL, Huentelman MJ, Lohi H, Hamilton SP and Neff MW. "Genome-wide association study of adult-onset deafness in herding Border collies." *(In preparation)* **JSY** - sample ascertainment, genotype generation and QC, study design, statistical analysis for primary GWAS, manuscript preparation, next-gen sequencing (NGS) sample prep; ETL – NGS sample prep and data analysis, manuscript input for all NGS portions; CAE – candidate sequencing, technical support; JA and MJH – whole-genome NGS (not presented here); all other co-authors – sample ascertainment and/or funding.

[2]. Although known to have a hereditary component, much research is still required to uncover the breadth of genetic variation that may confer risk to this common disorder.

Although a natural deterioration of the hearing system is expected with the aging process, it is believed that exposure to ototoxins and environmental noise likely contribute the most to hearing loss [2]. In addition, there is an extensive genetic contribution to hearing variation [3], estimated at 35-55% [4]. There have been many studies to identify genetic contributors of risk for ARHL in humans, which have uncovered various mutations in mitochondrial [5, 6] and autosomal DNA [reviewed in 4]. Most recently, a genome-wide association study (GWAS) was performed in an isolated Finnish population to discover novel risk loci for presbycusis [7]. The study identified the candidate gene *IQGAP2*, and also found modest support for another, previously indentified candidate from GWAS, *GRM7* [7, 8]. Despite these advances in knowledge for ARHL, the breadth of congenital deafness loci that have been mapped in the human genome suggests that there are many other possible mechanisms for conferred risk to ARHL [4], which presumably may affect pathways related to the different types of presbycusis that may arise independently: sensory, neural, strial or metabolic, and cochlear conductive [9]. In addition to each of the four independent types, combinations (termed "mixed" presbycusis by Schuknecht [9]) may also arise, highlighting the heterogeneity seen in this disorder in human populations.

The domestic dog offers a unique opportunity to explore the genetics of disorders analogous to human disease. Many disorders with low overall prevalence in humans are often enriched in a subset of dog breeds as a by-product of the selective breeding regimes set up approximately 200 years ago [10]. Increased disease prevalence within certain breeds may suggest a genetic component conferring a large effect. Increased prevalence also allows for the collection of larger numbers of affected dogs for studies that are hoped to ultimately inform researchers and breeders alike in the genetic components underlying troubling disorders. Genetic studies of this nature can

be particularly informative when the disorder of interest is more simplified in the context of the dog. Information gleaned from such studies would thus inform on both genetic risk factors as well as underlying biological pathways that may be involved in overall disease presentation or pathology.

Genetic investigations of deafness in the dog have been predominantly related to *congenital* sensorineural hearing loss, culminating most recently in strong association of markers in the *MITF* gene with congenital deafness in dalmatians [11]. As mentioned in Chapter 1, in 2007, Karlsson, *et al.* [12] mapped the white color/spotting locus to the same gene, providing continuity for a long line of research relating coat and eye pigmentation to congenital hearing loss in different breeds [reviewed in 11].

As in humans, all dogs are expected to have some hearing deterioration related to the aging process, with an onset around 8-10 years [14] and corresponding with physiological changes in critical systems in the ear [15]. In the dogs examined by Shimada, *et al.,* all morphological changes in dogs with hearing loss demonstrated the four different types of lesions found in humans as described by Schuknecht [9]: sensory, neural, strial and cochlear conductive. These findings suggested that the "mixed" presbycusis described by Schuknecht [9] is also seen in the dog [15]. Importantly, physiological measurements of hearing ability (as measured using brainsteam auditory evoked response or BAER) in the dog do demonstrate similar patterns as those seen in humans, where high and mid-range frequencies are most severely affected with age-related hearing deficits [14, 16]. This latter point suggests that a similar phenotype is also observed in both humans and dogs, although no consistent patterns of lesions were found to correlate directly with auditory dysfunction in a sample of 23 household dogs [15]. In humans there are marked differences in the prevalence and severity of presbycusis [17-19]; in dogs it is as

yet unknown whether similar sex differences exist [16, 20], though the limited data that does exist for dogs does not demonstrate a readily apparent difference [15].

Although uncommon, adult-onset deafness has become an increasing problem within the herding Border collie (BOC) community, where dogs' ability to hear keenly directly affects their ability to work effectively [21]. In particular, the later onset of this disorder makes it difficult for breeders to know in advance which dogs may be at greater risk, and thus dogs are often bred prior to knowledge of potential hearing impairments, further propagating the problem. Adult-onset deafness in herding BOC has an earlier onset than that resulting from physiological aging of hearing organs, at an average of 3-5 years [21; present work]. It has been suggested that adult-onset deafness is genetic in nature, passing through families by an autosomal dominant mode of inheritance [21]. Anecdotally, this form of deafness is seen in family lineages, thus further suggesting a familial component. Unlike in humans, no mitochondrial deafness has yet been identified in dogs [20]. Given the problematic nature of adult-onset deafness in the herding BOC community, as well as the potential to inform on the analogous human disorder, presbycusis, we undertook a genomic investigation for risk loci for this troubling form of hearing loss in the dog.

4.2. Materials and Methods

4.2.1. Samples

Whole blood samples (3-8 ml) were collected from a total of 48 purebred BOC for the primary GWAS. Twenty affected BOC were collected specifically for this genetic survey of risk loci for adult-onset deafness from sheepdog herding trials. Twenty-eight unrelated (at the grandparental level as per pedigree analysis) controls were collected at sheepdog trialing events or sent directly to the laboratory from owners and breeders in the context of ongoing genetic studies of canine behavior and complex disease. The 20 adult-onset deafness cases consisted of 9 males and 11 females, and the 28 controls were composed of 15 males and 13 females (mean control age = 6.6

years). One of the cases and two controls were also sequenced using next-generation sequencing (NGS) technology. All DNA was extracted in-house using standard protocols. As outlined in Chapter 2, SNP genotyping was performed on the Affymetrix Custom Canine Array v2.0, a perfect-match only array targeting 127K SNPs chosen from the SNP list generated as part of the dog genome project (Affymetrix, Santa Clara CA).

4.2.2. Phenotypes

'Adult-onset deafness' phenotypes were assigned based on owner response to verbal questioning as to whether the sampled dog was noted to have had hearing loss that onset during adulthood (i.e., was not present at birth). Hearing loss was demonstrated indirectly by owners' observations that their working dog was previously responsive to verbal and whistle commands given both in home and working conditions, but had demonstrated significant decrease in response, or inability to hear commands in adult life. Often, this loss of hearing ability was observed to take place over the course of several months or years. Some owners noted that they did not notice any significant changes in their dogs' hearing ability until much later in the dog's life, but suspected that the dog was "compensating" in the work environment by observing the handlers' physical cues, or by moving into closer proximity of the handler when commands were being given, so as to avoid belying any deficits in their working ability. When possible, indirect 'proof' of the dogs' ability to hear earlier in adulthood was demonstrated through sheep trialing awards obtained through competition. Controls for the adult-onset deafness study were herding BOC selected on two criteria: (1) genetic clustering in the same region as affected dogs (i.e., "genetic matching"), and (2) no hearing loss indicated in the health section of behavioral questionnaires completed by owners at the time of sample collection.

4.2.3. Next-generation Sequencing

4.2.3.a. Sample preparation. Genomic library sample preparation was performed using Illumina's Single-end library sample preparation kit (Illumina Inc., San Diego, CA). Sample preparation was carried out as per manufacturer's instructions, except the solution-capture samples' preparation was modified as follows: 3 μg of genomic DNA was subjected to shearing via sonication for a total input of ~58,000 joules using the following protocol: 6 cycles of 3 minutes of sonication with 10 second On/Off pulses at 40% power, and a 2-minute pause between each cycle (S-4000 with 2.5" diameter cup horn, Misonix, Inc., Farmingdale, NY); all purification steps were performed using Agencourt AMPure XP magnetic beads (Beckman Coulter, Inc., Brea, CA); seven cycles of ligation-mediated PCR were used for library amplification. Sample libraries were run on a Bioanalyzer 2100 with manual peak integration for quantitation and confirmation of fragment size distribution (High Sensitivity DNA Kit, Agilent Technologies, Waldbronn, Germany).

4.2.3.b. Target-capture. For targeted sequencing of the ~6.1Mb candidate region, we performed solution-based capture (SureSelect Target Enrichment System Kit, Agilent). Briefly, a custom panel of 120bp cRNA oligos was designed to target 1000bp upstream and downstream of 75 predicted genes based on mammalian alignments (or in one case, frog) to CanFam2 (version 2 of the canine genome) in the candidate region on CFA6. Regions containing known repetitive sequences and segmental duplications were excluded in the design. Prepared genomic libraries were hybridized to the panel of biotin-labeled 'bait' oligos for 24 hours then targets were pulled down via streptavidin magnetic beads. Targets were purified then enriched through 13 cycles of amplification.

4.2.4. Analysis

4.2.4.a. GWAS. Primary GWAS analysis was performed utilizing the beta version of Efficient Mixed-Model Association eXpedited [22]. We chose to utilize a mixed model based analysis to account for any population stratification or cryptic relatedness that may have been present in the sample (as pedigrees were not available for the case samples). In addition, allelic association with one million permutations were performed in PLINK (v1.07, [23]) to further assess association strength and rule out false-positives. For this analysis, we did not correct for within-breed stratification by incorporation of principal components or multi-dimensional scaling vectors since the analyzed dogs were selected to be genetically matched (Figures 4.1-4.2). Finally, multi-marker analysis was performed in PLINK (v1.07) to identify a common risk haplotype which was visualized using the Genome Variation Server (GVS) application available for use online (http://gvs.gs.washington.edu/GVS/index.jsp).

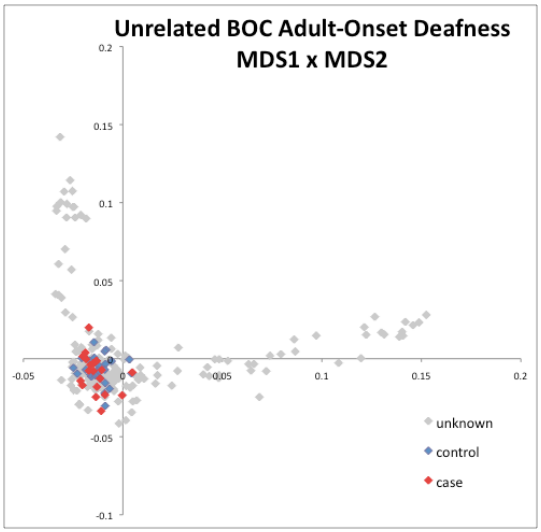


Figure 4.1: Multi-dimensional scaling (MDS) vector plots of BOC used for deafness analysis.

MDS1 x MDS2 based on data from ALL unrelated BOC genotyped for common dataset (see Chapter 2, section 2.3). Matched controls (blue) were selected based on genetic similarity (matching) to cases (red). Samples in gray were not utilized for this study.

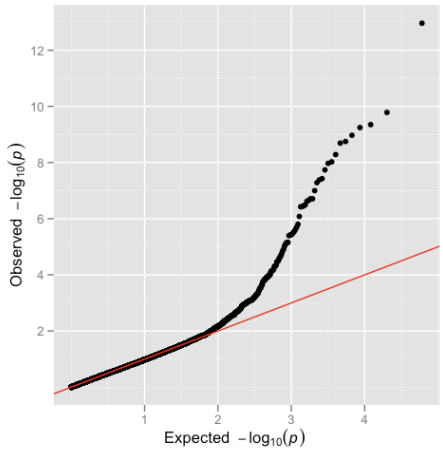


Figure 4.2: Q-Q plot of GWAS analysis for adult-onset deafness in BOC.

Expected versus observed $-\log_{10}$(p-value) for the primary GWAS are plotted for each marker, where the red line indicates the null distribution. The Q-Q plot for this analysis suggests that there is minimal population stratification in this sample, as the majority of points lie on the null distribution. Only around 2,2 does the high tail rise off the null towards higher p-values, which reflects the strong associations seen on CFA6 (these SNPs comprise only a small subset of the total number of markers analyzed).

4.2.4.b. NGS variant calling and assessment. For NGS data processing, Bowtie [24] was used for read alignment against CanFam2. The "-n" mode was used with a seed length of 60 bases. SAMtools [25], Picard (not published), BEDTools [26], and the Genome Analysis Toolkit (GATK, [27]) were all used for post-alignment processing. Multi-sample realignment around potential insertion/deletions (indels) and base quality score recalibration were both performed prior to variant calling by GATK's Unified Genotyper. Indel calling was performed using Dindel [28]. ANNOVAR [29] was used to annotate and prioritize variants found. Phastcons4way scores, which assess conservation based on phylogenetic analysis of multiple species, were obtained from the UCSC Genome Browser [30].

4.3. Results

4.3.1. Samples

Exact age of onset for hearing deterioration is difficult to determine in a community environment since subtle changes in hearing ability are likely to go unnoticed at first, and dogs may also compensate for hearing losses [20]. Nonetheless, when samples were collected, owners were asked to give their best estimate of when their dogs' hearing began to deteriorate. Interestingly, we found that the average estimated age of onset for adult-onset deafness in our cohort was 4.3 years of age, with a range of 1-9 years. Most owners estimated onset around 3-5 years; this finding is consistent with previous work by Chu and Schmutz [21]. The average age of onset we observed is intriguing given that presbycusis in dogs is estimated to begin around 8-10 years, when marked increases in hearing thresholds are observed at all frequencies [14]. The earlier onset of adult-onset deafness observed in our cohort suggests that this form of deafness may be more severe than what is typically observed in geriatric dogs. In addition, many of the affected dogs in our study have one or more first-degree family members that also demonstrated similar-onset deafness phenotypes, further suggesting we may be investigating a more severe form of ARHL demonstrating a simple inheritance pattern as previously noted by others [21].

4.3.2. Genome-wide Association Study

After quality control (QC, see Chapter 2, section 2.3.2), 30,231 SNPs remained for analysis. Pseudoheritability, an estimate of the genetic contribution to phenotypic variation based on the matrix of identity-by-state of analyzed samples, was estimated by EMMAX [22] to be 0.99 for this trait. Association analysis in EMMAX rendered two regions on CFA6 demonstrating strong regional support (Figure 4.3). In total, there were 25 markers that reached significance at a strict Bonferroni threshold of 1.65×10^{-6} for 30,231 tests. The strongest hit was at chr6.25819273 with a p-value of 1.09×10^{-13} (Table 4.1). The closest predicted gene to this SNP is *HS3ST2* (as

annotated in other species), about 24kb downstream. *HS3ST2* is a member of the heparan sulfate biosynthetic enzyme family, and is expressed predominantly in the brain [31].

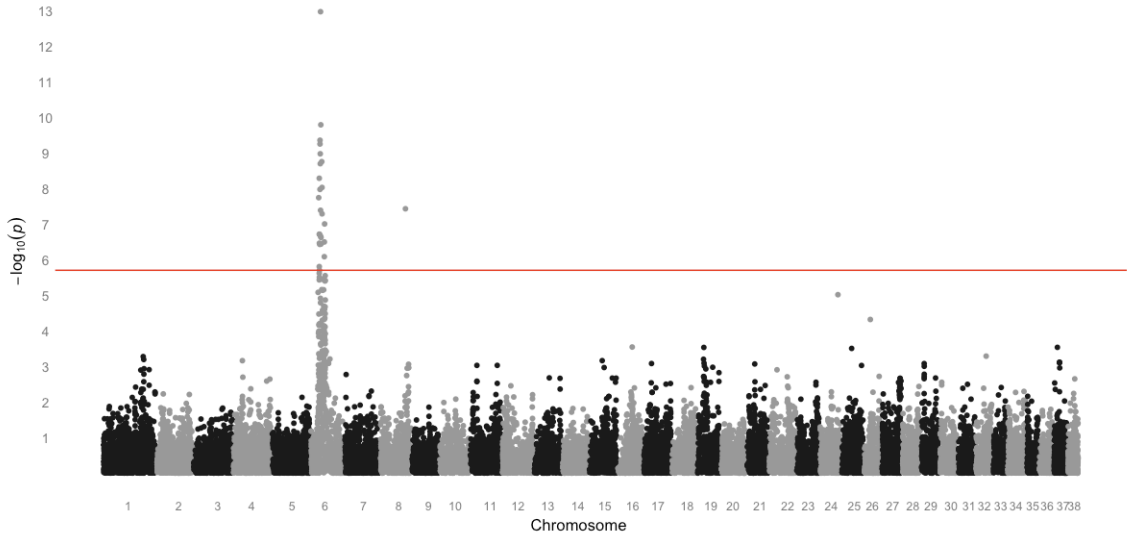


Figure 4.3: Manhattan plot of GWAS for adult-onset deafness.

Markers are plotted on the X-axis in order and shaded by chromosome. The -$\log_{10}$(p-value) is plotted on the y-axis. The red line indicates significance at the Bonferroni-corrected level for ~30,000 SNPs. There is extensive regional support for an association on CFA6.

In addition to extensive regional support around the ~25Mb region on CFA6, there was also regional support for an independent association signal around the 35Mb region (Table 4.1). The top hit for this region was the marker chr6.35491820, which was ranked 15th overall at a p-value of 1 x 10^{-7}. This SNP is putatively located intronic to *GRIN2A* (as annotated in other species), an ionotropic glutamate receptor subunit. These receptors are known to play an important role in synaptic transmission [32]. Polymorphisms in *GRIN2A* have also been associated with psychiatric disease [33, 34], while rare mutations have been found in individuals with epilepsy and mental retardation [35, 36].

CFA	SNP	BP	P
6	chr6.25819273	25819273	1.09E-13
6	chr6.26517587	26517587	1.64E-10
6	chr6.24591869	24591869	4.46E-10
6	chr6.24577002	24577002	5.68E-10
6	chr6.25174415	25174415	1.07E-09
6	chr6.28753894	28753894	1.78E-09
6	chr6.25181733	25181733	2.03E-09
6	chr6.22844453	22844453	5.21E-09
6	chr6.29363433	29363433	9.45E-09
6	chr6.24570819	24570819	1.07E-08
6	chr6.21475826	21475826	1.83E-08
8	chr8.62484232	62484232	3.75E-08
6	chr6.25913101	25913101	4.14E-08
6	chr6.29470484	29470484	5.20E-08
6	chr6.35491820	35491820	1.00E-07
6	chr6.23160353	23160353	1.95E-07
6	chr6.23166082	23166082	1.95E-07
6	chr6.25900591	25900591	2.21E-07
6	chr6.26959216	26959216	2.40E-07
6	chr6.34915222	34915222	3.21E-07
6	chr6.23177930	23177930	3.48E-07
6	chr6.26917473	26917473	3.67E-07
6	chr6.24104844	24104844	3.76E-07
6	chr6.34819558	34819558	8.32E-07
6	chr6.22861769	22861769	1.58E-06

Table 4.1: Top 25 ranked findings from analysis for adult-onset deafness in BOC.

Analysis was performed in EMMAX. All top 25 hits reach statistical significance at the Bonferroni-corrected level, and all but one are on CFA6. CFA - chromosome, SNP - single nucleotide polymorphism, BP - base position, P - p-value of association.

To further evaluate the significance of these findings, we performed permutation testing in PLINK. One million permutations rendered genome-wide permutated p-values (EMP2) that still reached significance, suggesting that both regions' findings were not due to chance (Figure 4.4). In addition, inspection of allele frequency differences between cases and controls demonstrated the large differences one would expect to achieve p-values at the level observed given the small sample size resulting in odds ratios (OR) of large effect (Table 4.2).

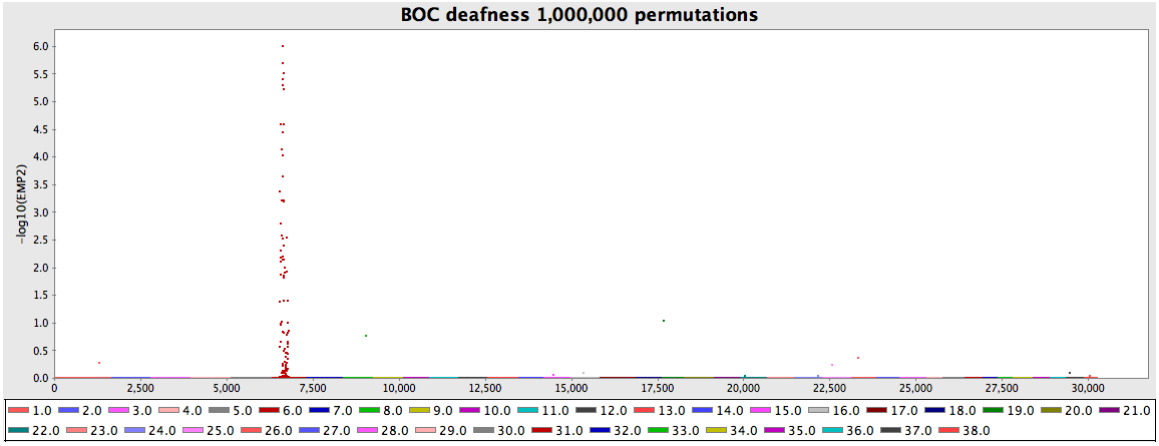


Figure 4.4: Manhattan Plot of P_{genome}.

One-million permutations in PLINK rendered permuted p-values at the genome-wide level (EMP2) that reach significance ($p < 0.05$), including numerous hits on CFA6, further supporting primary GWAS findings in EMMAX.

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR	SE	L95	U95	EMP2
6	chr6.25819273	25819273	C	0.00	0.78	A	56.24	6.42E-14	0.00	inf	0.00	nan	1.00E-06
6	chr6.26517587	26517587	A	0.83	0.14	G	44.37	2.71E-11	28.29	0.56	9.35	85.57	1.00E-06
6	chr6.24577002	24577002	C	0.89	0.20	T	44.24	2.91E-11	34.77	0.63	10.18	118.70	1.00E-06
6	chr6.25174415	25174415	C	0.10	0.78	G	42.24	8.06E-11	0.03	0.62	0.01	0.11	2.00E-06
6	chr6.25181733	25181733	G	0.89	0.21	A	41.95	9.38E-11	31.17	0.62	9.23	105.20	2.00E-06
6	chr6.24591869	24591869	T	0.10	0.77	C	41.65	1.09E-10	0.03	0.61	0.01	0.11	2.00E-06
6	chr6.28753894	28753894	A	0.78	0.13	G	41.22	1.36E-10	24.11	0.55	8.15	71.38	3.00E-06
6	chr6.24570819	24570819	T	0.10	0.75	G	39.5	3.28E-10	0.04	0.61	0.01	0.12	4.00E-06
6	chr6.25913101	25913101	C	0.00	0.61	T	37.6	8.67E-10	0.00	inf	0.00	nan	5.00E-06
6	chr6.29363433	29363433	T	0.78	0.15	G	37.19	1.07E-09	19.81	0.54	6.89	56.92	6.00E-06

Table 4.2: Top ten ranked findings for allelic association with 1,000,000 permutations in PLINK.

Odds ratios (OR) estimated in PLINK are very strong, which is what we would expect given the large allele frequency differences observed in cases versus controls. All of these top findings reach genome-wide significance after permutation ($p_{genome} < 0.05$). CHR - chromosome, SNP - marker name, BP - base position, A1 - reference allele, F_A - frequency of A1 in affecteds [cases], F_U - frequency of A1 in unaffecteds [controls], A2 - alternate allele, CHISQ - chi-square value, p - p-value, OR - odds ratio, SE - standard error of OR, L95 - lower 95% confidence interval of OR, U95 - upper 95% confidence interval of OR, EMP2 - genome-wide permuted p-value.

4.3.3. Fine-mapping

The large candidate region on CFA6 implicated by our top findings is syntenic to human 16p12.1-p12.3, which encompasses the autosomal recessive deafness locus DFNB22 (UCSC Genome Browser [30]). A candidate of immediate interest was the gene *OTOA*, found to be defective in prelingual sensorineural deafness in a consanguineous Palestinian family [37]. We performed PCR amplification of the 28 putative exons plus one highly conserved non-coding region [30] for direct Sanger sequencing to identify potential mutations of interest in deaf dogs. However, no polymorphisms or putative causative mutations discovered during sequencing segregated specifically in affected dogs (data not shown).

Given the large region of association and lack of segregating polymorphisms in the candidate gene *OTOA*, we attempted to narrow the critical region of interest in our BOC cohort for the ~25Mb region using haplotypic analysis. We found a 7-SNP haplotype that was homozygous in all cases and only one control sample (Figure 4.5). Additionally, a larger 11-SNP haplotype was also unique to 19 of 20 cases (and present in the same single control, Figure 4.5). We identified two adult-onset deafness cases that were homozygous for the extended 11-SNP risk haplotype at ~25Mb that was observed in a majority of the cases, and were also homozygous for an extended 15-SNP haplotype at the second CFA6 ~35Mb candidate region (Figure 4.6). We were able to obtain a new blood sample from one of these dogs and it was sequenced utilizing target capture as well as whole genome NGS (data for the whole genome NGS is not presented here). In addition, two control dogs that did not carry the risk haplotypes were identified, confirmed to be hearing intact at the time of fine-mapping via owner correspondence, and subjected to target capture and NGS.

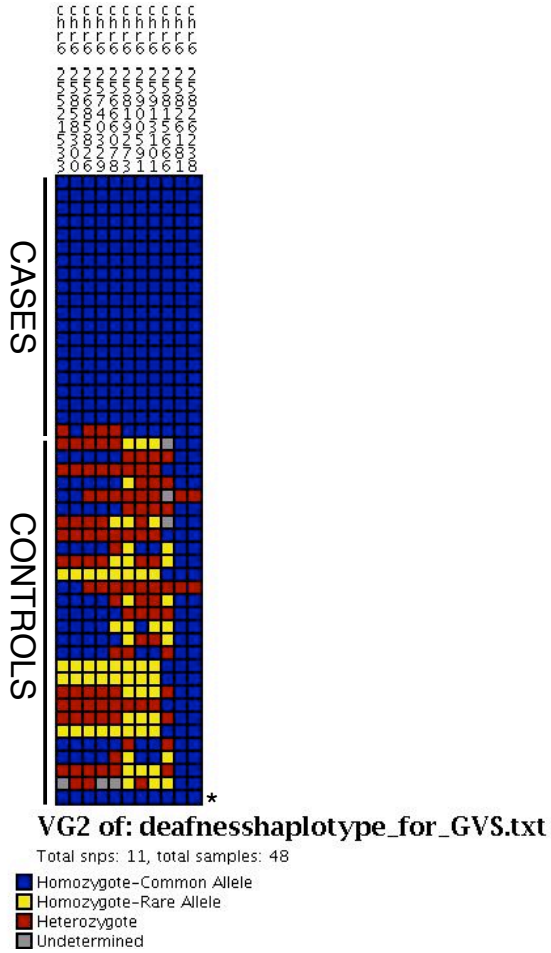


Figure 4.5: Haplotypes in CFA6 hit region ~25Mb.

Each color box represents a different genotype as indicated by the key, with dogs listed in rows and SNPs listed in columns. Case dogs are all homozygous for a single haplotype spanning 7 markers, and all but one case also share an 11-SNP haplotype (for which the single dog is heterozygous). One sample used as a control (marked with *), also carries the 11-SNP risk haplotype.

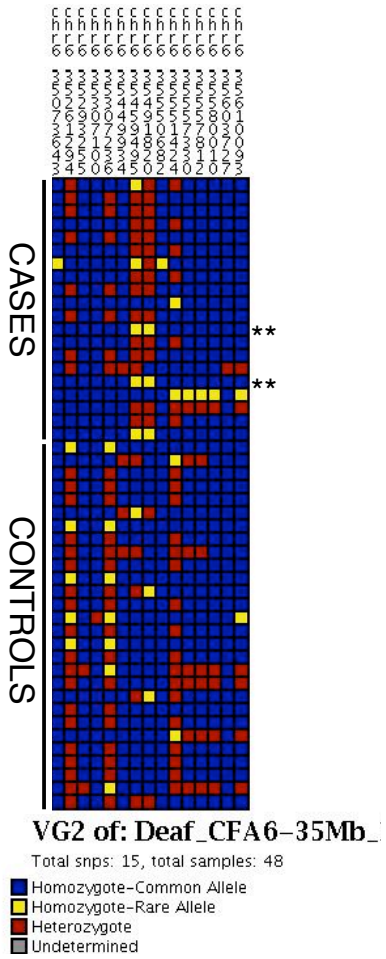


Figure 4.6: Haplotypes in CFA6 hit region ~35Mb.

Each color box represents a different genotype as indicated by the key, with dogs listed in rows and SNPs listed in columns. There were three case dogs that were all homozygous for a single haplotype spanning 15 markers, and two of those three (marked with **) also carry the 11-SNP risk haplotype at ~25Mb.

We used the extended risk haplotype spanning from ~25.52-25.91Mb, to guide further fine-mapping efforts to identify the causative mutation in this region hypothesized to play a critical role in adult-onset deafness. Utilizing the haplotypic data, we identified ~70 candidate genes and designed a solution-based target capture mixture to target the putative exons and upstream and downstream untranslated regions for these candidates. Of 3.67Mbp initially selected, the oligo capture mix was successfully designed to target 2.3Mbp. These included 74 predicted genes

around the ~25Mb region as well as 1 gene (*GRIN2A*) in the ~35Mb region. The targeted 2.3Mbp were covered utilizing approximately 43,000 probes, for a total of ~3.0X coverage with 75bp single-end sequencing.

For the solution capture, an average of 57 pmols of about 300bp-fragments of DNA were isolated for the targeted genomic regions and used as input for sequencing, with each sample run in one lane. Table 4.3 shows a summary of the target-capture NGS results. Overall, all three samples performed extremely well, with over 30 million reads per sample. Over 90% of each samples' reads aligned to the dog genome (CanFam2). Target enrichment was also very successful, resulting in >10X coverage for about 75% and >30X coverage in nearly 70% of the total targeted sequence (Table 4.3).

	Lane 1	Lane 2	Lane 3
Sample	CAN3148	CAN3149	1012544
Total Reads	36270529	30867026	32404825
Aligned Reads	33564330	27899663	30252775
%Aligned Reads	92.5%	90.4%	93.4%
Mean Bait Coverage	905.90	658.22	785.33
Mean Target Coverage	548.21	403.67	483.69
Fold Enrichment	868.34	759.03	835.17
%Target > 2X	77.6%	78.7%	81.6%
%Target >10X	75.7%	73.1%	77.1%
%Target > 20X	73.1%	68.6%	73.8%
%Target > 30X	71.1%	65.2%	71.1%
Bases on Bait	1841699115	1338166930	1596587940
Bases near Bait	158908041	135122788	168667197
Bases off Bait	550277059	647080612	533951957
Bases on Target	1997808859	1471059990	1762689753
Reads on Bait	24232883	17607460	21007736
Reads near Bait	2090895	1777931	2219305
Reads off Bait	7240488	8514219	7025684
Reads on Target	26286959	19356053	23193286
%Aligned Reads on Bait	72%	63%	69%
%Aligned Reads near Bait	6%	6%	7%
%Aligned Reads off Bait	22%	31%	23%
%Aligned Reads on Target	78%	69%	77%

Table 4.3: NGS statistics.

Each DNA sample was run in a single lane, with statistics for each provided in columns. Overall, all three samples performed very well in the sequencing reaction. Target sequence enrichment was also quite successful, with the majority of sequence mapping to the baited regions.

Variant calling resulted in a total of 183,289 SNPs and 1,473 indels across all three samples across the entire genome, with breakdowns by CFA6 (versus the rest of the genome) provided in Table 4.4. As expected, the average quality of SNP variant calls was very high since the coverage for these variants was around 150-250X; this is in contrast to SNP calls made on the rest of the genome (which was not baited), where average coverage around called variants was in the range

of 10-50X (which is still quite high and thus of sufficient to pass SNP-calling criterion). These findings reflect the utility of using even target-enriched NGS sequencing data for variant discovery across the genome (although not one of the main aims for the current study).

	CFA6				rest of genome			
	All samples	Control 1	Control 2	Case	All samples	Control 1	Control 2	Case
indels	127	73	93	84	1346	443	708	270
SNPs	14706	8704	11734	11147	168583	34632	97901	81296
SNP quality (avg)	3054.1	na	na	na	214.6	na	na	na

Table 4.4: Variant-calling statistics.

Statistics for indels and SNPs are provided for all samples, with breakdowns by sample. Indels were called using Dindel (results for indel analysis are not discussed in this dissertation; please refer to manuscript, in preparation). Overall, all three samples demonstrated a large number of variants, though interestingly Control 1 appears to have less variants called than the other two samples. This is unlikely related to sample performance since this sample showed the highest performance for overall sequence production and baiting (see Table 4.3).

Given the high number of SNPs called, we chose first to filter variants with regard to their genotype in cases and controls, filtering for variants called homozygous in the case sample and not homozygous for that variant in either controls (e.g., if the case was called homozygous alternate allele ("1/1"), controls had to be heterozygous reference-alternate ("0/1") or homozygous reference ("0/0") for that particular variant), which resulted in 2495 SNPs remaining. We next narrowed our search to only variants within the CFA6 targeted regions (n = 1369), then performed first-pass variant assessment in exonic SNPs only (n = 32). For a breakdown of SNPs as annotated in ANNOVAR, see Table 4.5.

	CFA6	rest of genome
total SNPs	1369	1126
SNP quality (avg)	7013.1	1467.3
downstream	24	18
exonic	32	45
ssSNP	*23*	*17*
nsSNP	*8*	*27*
stopgain	*1*	*1*
intergenic	451	759
intronic	829	265
upstream	19	15
UTR3	3	4
UTR5	11	11
ncRNA	0	10

Table 4.5: Breakdown of SNPs homozygous in cases and not in controls.

This table shows a breakdown of the SNP annotations using ANNOVAR for SNPs that were called homozygous in cases and did not have the same (case) genotype in controls. As expected, variants on CFA6 (which were targeted for enrichment) have higher average quality scores than those called throughout the rest of the genome, with higher scores reflecting higher confidence in the allele being different from the reference allele. Gray font in italics represents a breakdown of type for only exonic SNPs. ssSNP – synonymous SNP; nsSNP – non-synonymous SNP; UTR3 – untranslated 3'; UTR5 – untranslated 5'; ncRNA – non-coding RNA.

Of the 32 putative exonic SNPs, only 8 were annotated to be non-synonymous changes (nsSNPs, Table 4.6). Four nsSNPs were found in *ABCA14*, which was the gene with the most nsSNPs. *ABCA14* is a putative ATP binding cassette transporter gene that has only been annotated in the genomes of rodents [38]. Conservation scores for all four of these nsSNPs are quite low, suggesting that this gene may not be active in the canine genome and thus tolerates non-synonymous changes more readily than coding regions. Although it is possible that the phastCons4Way (phastCons; see Table 4.6 caption for more information) scores are lower simply because *ABCA14* is not observed in humans, we are not actively pursuing these variants at this time.

Table 4.6: Exonic variants for deafness on CFA6.

							Case		Control 1		Control 2		
Type	Gene	CFA	BP	Ref	Alt	SNP Score	genotype	coverage	genotype	coverage	genotype	coverage	phastCons Score
nsSNP	**RBBP6, exon18, p.T1397N**	**6**	**24500625**	**G**	**T**	**7471.4**	**1/1**	**226**	**0/0**	**249**	**0/0**	**72**	**0.001**
ssSNP	RBBP6, exon11, p.E445E	6	24508479	T	C	12802.2	0/0	226	1/1	248	1/1	123	0.925
ssSNP	ERN2, exon9, p.L260L	6	25178743	T	C	12191.9	1/1	240	0/0	236	0/1	241	0.871
ssSNP	ERN2, exon19, p.F578F	6	25185894	C	T	11553.3	0/0	214	1/1	233	0/1	223	0.925
ssSNP	ERN2, exon25, p.D838D	6	25188304	C	T	11937.6	0/0	244	1/1	208	0/1	245	0.949
ssSNP	PLK1, exon9, p.L511L	6	25189840	A	G	11207.4	0/0	244	1/1	233	0/1	236	0.792
ssSNP	PLK1, exon9, p.E488E	6	25189907	T	C	11908.6	0/0	249	1/1	242	0/1	247	0.831
ssSNP	PLK1, exon1, p.K97K	6	25199818	C	T	12886.7	0/0	241	1/1	238	0/1	233	0.971
nsSNP	**USP31, exon17, p.I847V**	**6**	**25714052**	**A**	**G**	**12119.9**	**1/1**	**245**	**0/0**	**245**	**0/1**	**246**	**0.950**
nsSNP	**EEF2K, exon1, p.N62K**	**6**	**26442657**	**G**	**T**	**11171.7**	**0/0**	**249**	**1/1**	**249**	**0/1**	**250**	**0.627**
nsSNP	**ABCA14 (Mus), exon26, p.M1292L**	**6**	**26909869**	**T**	**G**	**762.9**	**1/1**	**16**	**0/0**	**53**	**0/1**	**18**	**0.058**
nsSNP	**ABCA14 (Mus), exon23, p.L1134I**	**6**	**26924547**	**G**	**T**	**7577.2**	**1/1**	**189**	**0/0**	**230**	**0/1**	**63**	**0.013**
nsSNP	**ABCA14 (Mus), exon16, p.V699I**	**6**	**26951574**	**C**	**T**	**12400.1**	**1/1**	**247**	**0/0**	**243**	**0/1**	**247**	**0.045**
nsSNP	**ABCA14 (Mus), exon9, p.I472M**	**6**	**26972673**	**T**	**C**	**10207.9**	**1/1**	**242**	**0/0**	**220**	**0/1**	**114**	**0.145**
ssSNP	DNAH3, exon56, p.R2825R	6	27436889	G	A	13331.7	1/1	246	0/0	247	0/1	246	0.980
ssSNP	DNAH3, exon60, p.D3733D	6	27456701	C	T	11929.7	1/1	245	0/0	243	0/1	243	0.031
ssSNP	LOC57020, exon11, p.D462D	6	27898482	G	A	8681.1	1/1	242	0/0	244	0/0	246	0.352
ssSNP	C16orf62, exon26, p.T707T	6	28661889	C	T	9143.1	1/1	245	0/0	249	0/0	67	0.972
ssSNP	C16orf62, exon9, p.S232S	6	28709547	A	G	7111.7	1/1	246	0/0	243	0/0	226	0.972
ssSNP	CP110, exon3, p.K201K	6	28750373	C	T	8429.3	1/1	221	0/0	191	0/0	91	0.463
ssSNP	CP110, exon1, p.L142L	6	28753519	T	C	11474.0	0/0	230	1/1	248	0/1	148	0.918
ssSNP	CP110, exon1, p.T17T	6	28753894	G	A	8823.1	1/1	246	0/0	247	0/0	222	0.518
ssSNP	COQ7 (Sac), exon4, p.K131K	6	28948448	A	G	8298.8	1/1	249	0/0	240	0/0	249	0.923
ssSNP	ITPRIPL2, exon1, p.A365A	6	28978457	C	T	3129.3	1/1	88	0/0	55	0/0	149	0.880
ssSNP	ITPRIPL2, exon1, p.L414L	6	28978602	T	C	3019.7	1/1	102	0/0	80	0/0	165	0.141
ssSNP	ITPRIPL2, exon1, p.L415L	6	28978605	T	C	2859.0	1/1	104	0/0	77	0/0	156	0.063
ssSNP	TEKT5, exon6, p.E400E	6	35015026	G	A	1034.3	0/0	26	0/1	33	0/1	44	0.760
ssSNP	GRIN2A, exon1, p.T4T	6	35563900	A	G	6343.6	0/0	248	0/1	246	0/1	245	0.910
ssSNP	**GRIN2A, exon, p.A595A**	**6**	**35680749**	**G**	**A**	**17706.6**	**0/0**	**248**	**1/1**	**243**	**1/1**	**250**	**0.714**
ssSNP	GRIN2A, exon10, p.N665N	6	35703953	C	T	11329.3	0/0	246	1/1	237	0/1	244	0.988
stopgain*	**HNRNPA1 (Bos, Rat), exon2, p.R196X**	**6**	**39425712**	**C**	**T**	**9704.7**	**0/0**	**246**	**0/1**	**240**	**0/1**	**241**	**0.972**
nsSNP	**ZNF434, exon1, p.A4T**	**6**	**40872512**	**G**	**A**	**3524.7**	**0/0**	**37**	**0/1**	**59**	**1/1**	**46**	**0**

Table 4.6: Exonic variants for deafness on CFA6.

A list of the 32 exonic SNPs for CFA6 plus annotations appears on the preceding page. Gene annotations and predicted amino acid (AA) changes (single letter AA abbreviations flanking AA position) are given with reference to the gene in human unless the gene was not present in human, in which case it was given for another species as noted (Mus – mouse, Sac – yeast, Bos – cow, Rat – rat). The stopgain SNP (marked with an asterisk) is only exonic in the putative gene *HNRNPA1* annotated for cow and rat; in human, this SNP is intronic of the putative gene *ALG1L2*. Non-synonymous SNPs (nsSNP) plus the strongest candidate synonymous SNP (ssSNP) in *GRIN2A* are marked in bold. In addition to the called genotypes for each sample, the sequence coverage for that SNP is also provided. Finally, the phastCons4Way score provides a measure of conservation for the each sequence change, where values closer to 1 mean the base is more highly conserved across species. Conservation is based on alignment with human (hg17), mouse (mm6), and rat (rn3). Please note: the order of samples in this table is different from previous tables. CFA – canine chromosome; BP – base position; Ref – reference allele from genome; Alt – alternate allele observed in sample[s]; genotype – 0 = reference allele, 1 = alternate allele; phastCons score = phastCons4Way score from UCSC genome browser.

The strongest nsSNP candidate is located in putative exon 17 of *USP31*, a ubiquitin specific peptidase. Chr6.25714052 is an A > G SNP predicted to cause an I847V change in the resulting protein product. Chr6.25714052 is highly conserved, with a phastCons score of 0.95 (Table 4.6). This SNP changes the dog reference allele to the reference allele seen in other species, with genotype calls in our samples of G/G for the case and A/A and A/G for controls, respectively. This results in an AA change from isoleucine (the AA in dog, as per UCSC genome browser) to valine (which is the AA found in human, primate and mouse). We used the online program, SIFT, to evaluate whether this change may potentially affect protein function based on sequence homology and amino acid properties [39-42]; the corresponding change in humans is predicted to be tolerated (SIFT score = 0.66, where lower scores suggest less toleration). Also of note in *USP31* is an intronic G > T SNP at chr6.25681850 that shows very high conservation (phastCons

= 0.98) and is 5bp away from an intron-exon boundary. This SNP is called as T/T in the case and G/G in both controls (data not shown).

Another candidate nsSNP is in putative exon 18 in *RBBP6*, a retinoblastoma binding protein, and is predicted to code a threonine to asparagine change at amino acid (AA) 1397. Chr6.24500625 is a G > T SNP, and is called T/T in the case and G/G in both control dogs. Interestingly, this SNP changes the reference dog allele to the reference human allele at the analogous position on human chromosome 16 (GRCh37/hg19 assembly). SIFT predicts that the corresponding allele change in humans would be tolerated (SIFT score = 0.69). Although the conservation score for this SNP is quite low (phastCons = 0.001), *RBBP6* (also known as *PACT*) has been shown to play a critical role in ear development and hearing in the mouse [43].

There are three additional genes that contain nsSNPs, but because they are not readily linked to hearing function or expression we are not currently following them up at this time. These include: *EEF2K*, *ZNF434*, and the stopgain variant in *HNRNPH1* (annotated in rat and cow) which is expected to be intronic to *ALG1L2* (annotated in humans).

There were no predicted nsSNPs in the *GRIN2A* candidate region (Table 4.6). However, there were three ssSNPs, which we evaluated for predicted codon usage changes. Synonymous changes have been shown to affect protein function due to differences in tRNA codon usage [44]. Although all three ssSNPs in *GRIN2A* are moderately to very highly conserved (Table 4.6), predicted codon usage is only modestly different when comparing the reference allele to the alternate allele (Table 4.7).

SNP	**chr6.35563900**	**chr6.35680749**	**chr6.35703953**
predicted change	**T3T**	**A595A**	**N665N**
codon/alternate	ACA/G	GCC/U	AAC/U
case	0/0	0/0	0/0
control 1	0/1	1/1	1/1
control 2	0/1	1/1	0/1
ref (0) codon usage	0.28	0.40	0.53
alt (1) codon usage	0.11	0.27	0.47

Table 4.7: *GRIN2A* ssSNP codon usage.

The three ssSNPs in *GRIN2A* are listed with genotypes for case and control samples (where 0 is reference and 1 is alternate allele). For each SNP, the predicted AA change is provided, along with the codon (listed with the third position as "reference/alternate"), which is provided for the corresponding ***human*** reference sequence. Human codon use frequency is provided with regard to the reference (0) or alternate (1) allele. Please note, the human reference sequence is complementary (opposite strand) to the published dog genome sequence, so SNP alleles in Table 4.6 are ***not*** the same as those listed in this table.

Of note, NGS alignment for the candidate gene *OTOA* did not result in any variants that were homozygous in the case but not in controls, further suggesting that this gene does not play a role in adult-onset deafness in this sample of BOC.

4.4. Discussion

Our results represent the first GWAS of adult-onset deafness in the dog. The pseudoheritability estimated by EMMAX is consistent with the hypothesis that adult-onset deafness in Border collies is a genetically mediated disorder of simple inheritance [21], a point further supported by our strong association findings on CFA6. The allele frequency differences required to obtain such strong p-values and the subsequent large OR's associated with them highlight the magnitude of these genetic findings (Table 4.2). Furthermore, permutation in a complementary association framework rendered genome-wide permuted p-values at the maximum threshold for the number of permutations performed, suggesting that spurious association is highly unlikely (Table 4.2).

These findings are particularly relevant in the context of the relatively unsophisticated phenotyping scheme that was utilized for collecting affected dogs. It is worth noting that, in this case, accurate assessment of disease status was made despite heavy reliance on owner observations only. Also, in light of the strong genetic findings, it is now possible to go back to dogs and owners for further phenotypic characterization to identify more specific symptomatology, physiological, morphological and/or audiometric measures that may be particularly relevant in the clinical context. This type of phenotyping scheme has important implications for future genetic studies of disorders that may have similar "anecdotal" roots within breed communities.

The region implicated by our GWAS is syntenic to regions implicated in congenital sensorineural deafness in humans [28]. Because of the size of the regions demonstrating strong association with adult-onset deafness, and because of the multitude of strong gene candidates present in these regions, we elected to perform NGS on the top gene candidates within the ~3.5Mb region implicated by our top findings around 25Mb on CFA6. We additionally sequenced the strongest candidate gene in a second region at ~35Mb on CFA6 which demonstrated independent association signals and regional support. High-throughput sequencing allowed us to screen a large number of candidates in a much more time-efficient manner as compared to standard Sanger sequencing methods. There is some precedence implicating genes that may play a role in both congenital deafness and ARHL [45], and it thus follows that any congenital hearing loss gene (or region) may also be a candidate for ARHL [4]. This is not surprising given the complex mechanical and neuronal orchestration required for proper sound conduction and detection, and implies that gene modifiers likely nuance the extent to which hearing ability may be maintained throughout the aging process. In our sample, the magnitude of our findings on CFA6 in the primary GWAS likely overshadowed any other loci's signals, even if they play fairly substantial risk-modifying roles in adult-onset deafness. For the scope of this chapter, I have highlighted only

a small selection of the most promising candidate variants discovered through our target-capture NGS experiment. However, the wealth of sequence data provided by these next-generation technologies provides ample opportunity for further mutation discovery in adult-onset deafness (as well as genomic variation in BOC) in the future.

Specifically, we have identified a number of strong candidate coding and non-coding variants for adult-onset deafness. Our top candidate is chr6.25714052, a SNP in *USP31* (as annotated in humans), a ubiquitin-related gene that has been linked to Parkinson's disease in humans [46]. The implication of a ubiquitin-related gene playing a role in adult-onset deafness is particularly exciting given the histological findings of Shimada and colleagues [15], which included ubiquitin-positive granules in the neuropil of the cochlear nuclei of aged dogs. The nsSNP in the putative 17th exon is predicted to code a I847V change in the protein product. This variant is highly conserved, lies in the middle (@25.7Mb) of the region implicated by our risk haplotype (~25.5-25.9Mb), and is approximately 100kb away from our top associated GWAS hit, chr6.25819273. The top GWAS SNP had a 0.78 difference in allele frequency between cases and controls, suggesting a very strong effect by this locus. In addition to the nsSNP, we also identified the non-coding SNP chr6.25681850, which is located 5bp away from an intron-exon boundary and may potentially play a role in splicing. We are currently sequencing both of these variants directly to look for segregation in the full sample of cases.

There is also a nsSNP candidate variant in putative gene *RBBP6*. Despite low conservation and predicted toleration of the putative T1397N change, we feel the biological implications of this gene in hearing [43] warrant follow-up and also plan to directly sequence this SNP for segregation analysis in cases. Finally, we have identified several ssSNPs in *GRIN2A* (Table 4.6). *GRIN2A* is a tempting gene candidate due to its critical role in signal transduction [32]. Although preliminary analysis suggests only modest (if any) functional impact by any of these variants on

codon usage in humans (Table 4.7), moderate-to-high levels of conservation suggest that these ssSNPs could still play a role modifying disease risk. Given the independent signal of the ~35Mb region from the extended ~25Mb region, these and other variants in *GRIN2A* may therefore be followed up in the future. Additionally, there is an extensive list of other potential candidate variants (Tables 4.4 – 4.5) that have yet to be investigated.

There are several caveats to the present study. A recent human GWAS for presbycusis adjusted their phenotypes for hearing thresholds for age and sex based on observed differences in hearing threshold variability in male samples compared to females [7]. However, we elected not to correct for sex in our canine study as such sexual dimorphism has not yet been determined in aging dogs [14-16, 20]. Further, we did not adjust for age due to the earlier owner-estimated age of onset demonstrated by our sample cohort, which is likely a specific trait of this form of hearing loss [21]. The mean age for our control group was 6.6 years, which is still within the range of deafness onset. Thus, it is possible that dogs categorized as "controls" may, at later stages in life, demonstrate hearing losses similar to those observed in our case samples. However, misclassification of cases as controls would only reduce our power to detect significant genetic associations. Given the strength of the associations we identified on CFA6, this does not seem to be of concern. Another caveat stems from the fact that we performed target enrichment for only the "functional" regions (i.e., predominantly coding regions of the most biologically relevant candidate genes) of our extended association region. This means we are potentially missing other, non-coding and/or large structural variants within our associated region by assuming risk is more likely conferred by a traditional coding variant. Target capture also results in uneven coverage, so variants may still be missed since some bases are not covered as well. Finally, it is possible, as mentioned above, that there are other genetic regions of more modest strength that may have gone undetected in our primary GWAS given the small sample size. In the future, further loci in other genomic regions that may be contributing modest risk could potentially be identified by

conducting a GWAS after controlling for the signal on CFA6 (e.g., by removing the SNPs on CFA6 demonstrating strong association) and/or utilizing control samples that are older to ensure there is no confounding due to hearing loss. Moreover, whole-genome NGS (which our group also performed on the same case sample, but for which results are not presented here) may provide further insight into novel causative variants in other regions outside of those initially targeted by this experiment.

One sample classified as a control for the GWAS was later shown to carry the 11-SNP risk haplotype identified in deafness cases. Further inspection of the medical history reported by the owner of this dog at the time of sample collection revealed that, though the dog demonstrated normal hearing, it had several deaf siblings. Additional information provided by the dog's owner revealed that both congenital and adult-onset deafness were known to affect other members of this dog's extended family pedigree. Interestingly, the dog had also very recently been BAER tested and, though still demonstrating intact hearing, was reported to have an "abnormal" audiometric reading. This owner is still actively interested in participating in our research, and we hope to follow the progress of this dog's hearing through longitudinal study.

In sum, we have identified a very strong candidate region for adult-onset deafness in Border collies on CFA6. Causative variant fine-mapping is underway, and primary candidate SNPs in multiple exonic variants have already been identified through next-generation sequencing, many of which are in genes that have been implicated in cochlear development and aging, hearing, and signal transduction. For more details on the validation of these SNP findings, plus copy number analysis (copy number variant calling using genotype data; indel calls from NGS), please refer to the final manuscript, which is currently in preparation (December 2010).

4.5. References

1. http://www.merriam-webster.com/
2. Gates GA and Mills JH (2005). Presbycusis. Lancet. 366:1111-1120.
3. Karlsson KK, Harris JR, Svartengren M. (1997) Description and Primary Results from an Audiometric Study of Male Twins. Ear & Hearing. 18(2):114-20.
4. Liu XZ and Yan D. (2007) Ageing and hearing loss. J Pathol. 211:188-97.
5. Bai U, Seidman MD, Hinojosa R, Quirk WS. (1997) Mitochondrial DNA deletions associated with aging and possibly presbycusis: a human archival temporal bone study. Am J Otol. 18:449-53.
6. Fischel-Ghodsian N, Bykhovskaya Y, Taylor K, et al. (1997) Temporal bone analysis of patients with presbycusis reveals high frequency of mitochondrial mutations. Hear Res. 110:147-54.
7. Van Laer L, Huyghe JR, Hannula S et al. (2010) A genome-wide association study for age-related hearing impairment in the Saami. Eur J Hum Genet. 18:685-93.
8. Friedman RA, Van Laer L, Muentelman MJ et al. (2009) GRM7 variants confer susceptibility to age-related hearing impairment. Hum Mol Genet. 18:785-96.
9. Schuknecht HF and Gacek MR (1993) Cochlear pathology in presbycusis. Annals Otology, Rhinology and Laryngology. 102:1-16.
10. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M et al (2005): Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803-819.
11. Stritzel S, Wohlke A and Distl O. (2009) A role of the microphthalmia-associated transcription factor in congenital sensorineural deafness and eye pigmentation in Dalmatian dogs. J Anim Breed Genet. 126:59-62.
12. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet 39: 1321-1328.
13. De Risio L, Lewis T, Freeman J et al. (2010) Prevalence, heritability and genetic correlations of congenital sensorineural deafness and pigmentation phenotypes in Border Collie. The Veterinary Journal. doi:10.1016/i.tvil.2010.05.012.

14. ter Haar G, Venker-van Haagen AJ, van den Brom FJ et al. (2008) Effects of Aging on Brainstem Responses to Toneburst Auditory Stimuli: A Cross-Sectional and Longitudinal Study in Dogs. J Vet Intern Med. 22:937-45.
15. Shimada A, Ebisu M, Morita T, et al. (1998) Age-Related Changes in the Cochlea and Cochlear Nuclei of Dogs. J Vet Med Sci. 60(1): 41-8.
16. ter Haar G, de Groot JCMJ, Venker-van Haagen FJ et al. (2009) Effects of Aging on Inner Ear Morphology in Dogs in Relation to Brainstem Responses to Toneburst Auditory Stimuli. J Vet Intern Med. 23:536-43.
17. Jerger J, Chmiel R, Stach B, and Spretnjak M. (1993) Gender affects audiometric shape in presbycusis. J Amer Acad Audiology. 4:42-9.
18. Sajjadi H, Paparella MM, Canalis RF. Presbycusis. In: Canalis RF, Lambert PR (eds). The Ear: Comprehensive Otology. Philadelphia, PA: Lippincott Williams & Wilkins; 2000: 545-57.
19. Pearson JD, Morrell CH, Gordon-Salant S, et al. (1995) Gender differences in a longitudinal study of age-associated hearing loss. J Acoust Soc Am. 97:1196-1205.
20. Strain GM. (1996) Aetiology, Prevalence and Diagnosis of Deafness in Dogs and Cats. Br Vet J. 152, 17-36.
21. Chu T and Schmutz SM. Inherited Adult Onset Deafness in Border Collies. Poster session presented at: 4th International Conference: Advances in canine and feline genomics and inherited diseases; 2008 May 23; St. Malo, France.
22. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong Sy, Freimer NB et al (2010): Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348-354.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC, 2007. pLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81(3):559-75.
24. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. (2009) Genome Biology 10(3):R25.

25. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.
26. Quinlan AR, Hall IM. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841-2.
27. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297-303.
28. Albers CA, Lunter G, Macarthur DG, McVean G, Ouwehand WH, Durbin R. (2010) Dindel: Accurate indel calls from short-read data. Genome Res. Epub: 2010 Oct 27.
29. Wang K, Li M, Hakonarson H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38(16):e164. Epub 2010 Jul 3.
30. http://genome.ucsc.edu/cgi-bin/hgGateway
31. Shworak, N. W., Liu, J., Petros, L. M., Zhang, L., Kobayashi, M., Copeland, N. G., Jenkins, N. A., Rosenberg, R. D. (1999) Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase: isolation, characterization, and expression of human cDNAs and identification of distinct genomic loci. J. Biol. Chem. 274: 5170-5184.
32. Liu, L., Wong, T. P., Pozza, M. F., Lingenhoehl, K., Wang, Y., Sheng, M., Auberson, Y. P., Wang, Y. T. (2004) Role of NMDA receptor subtypes in governing the direction of hippocampal synaptic plasticity. Science 304: 1021-1024.
33. Tang J, Chen X, Xu X, Wu R, Zhao J, Hu Z, Xia K. (2006) Significant linkage and association between a functional (GT)n polymorphism in promoter of the N-methyl-D-aspartate receptor subunit gene (GRIN2A) and schizophrenia. Neurosci Lett. 409(1):80-2.
34. Cherlyn SY, Woon PS, Liu JJ, Ong WY, Tsai GC, Sim K. (2010) Genetic association studies of glutamate, GABA and related genes in schizophrenia and bipolar disorder: a decade of advance. Neurosci Biobehav Rev. 34(6):958-77.

35. Endele S, Rosenberger G, Geider K, et al. (2010) Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet. Epub ahead of print. PMID: 20890276.

36. Reutlinger C, Helbig I, Gawelczyk B, et al. (2010) Deletions in 16p13 including GRIN2A in patients with intellectual disability, various dysmorphic features, and seizure disorders of the rolandic region. Epilepsia. 51(9):1870-3.

37. Zwaenepoel, I., Mustapha, M., Leibovici, M., Verpy, E., Goodyear, R., Liu, X. Z., Nouaille, S., Nance, W. E., Kanaan, M., Avraham, K. B., Tekaia, F., Loiselet, J., Lathrop, M., Richardson, G., Petit, C. (2002) Otoancorin, an inner ear protein restricted to the interface between the apical surface of sensory epithelia and their overlying acellular gels, is defective in autosomal recessive deafness DFNB22. Proc. Nat. Acad. Sci. 99: 6240-6245.

38. Chen ZQ, Annilo T, Shulenin S, Dean M. (2004) Three ATP-binding cassette transporter genes, Abca14, Abca15, and Abca16, form a cluster on mouse Chromosome 7F3. Mamm Genome. 15(5):335-43.

39. Ng PC and Henikoff S. (2001) Predicting Deleterious Amino Acid Substitutions. Genome Res. 11(5): 863.874.

40. Ng PC and Henikoff S. (2002) Accounting for Human Polymorphisms Predicted to Affect Protein Function. Genome Res. 436-446.

41. Ng PC and Henikoff S. (2003) SIFT: predicting amino acid changes that affect protein function, Nucleic Acids Research. 31(13):3812-3814.

42. Kumar P, Henikoff S and Ng PC. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols. 4:1073-1081.

43. Rowe TM, Rizzi M, Hirose K, Peters GA and Sen GC. (2006) A role of the double-stranded RNA-binding protein PACT in mouse ear development and hearing. PNAS. 103(15):5823-5828.

44. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. (2007) A “silent” polymorphism in the MDR1 gene changes substrate specificity. Science. 315(5811):525-8. Erratum in: Science. 2007 Nov 30;318(5855):1382-3.

45. Van Laer L, Van Eyken E, Fransen E, *et al.*. (2008) The grainyhead like 2 gene (GRHL2), alias TFCP2L3, is associated with age-related hearing impairment. Hum Mol Genet. 17:159-169.

46. Lockhart PJ, Hulihan M, Lincoln S, Hussey J, Skipper L, Bisceglio G, Wilkes K, Farrer MJ. (2004) Identification of the human ubiquitin specific protease 31 (USP31) gene: structure, sequence and expression analysis. DNA Seq. 15(1):9-14.

Chapter 5: Concluding Remarks and Future Directions

As each chapter of this dissertation represents an independent research aim, the bulk of the research summation and future plans has already been highlighted in those sections. I will thus now only briefly reiterate the research results as they relate to the "bigger picture" of canine genetics research in our laboratory and beyond. I will also elaborate on the extended research goals for the two disease investigations outlined in Chapters 3 and 4, before finally concluding with the implications of this research for human psychiatric genetics and medicine.

5.1. Summary of Research Findings

In sum, this dissertation describes the first use of saliva-extracted dog DNA on high-throughput genotyping arrays, the first whole-genome survey of intra-breed stratification, the first genetic investigation of the canine anxiety disorder noise phobia, and the first whole-genome survey for loci underlying adult-onset deafness in the dog. Collectively, however, all of these studies form the research foundation for our laboratory's Canine Behavioral Genetics Project, whose ultimate goal is to advance knowledge of human neuropsychiatry.

In Chapter 2.1, we validated the use of saliva-derived DNA for whole-genome studies in the dog, which allows for: (a) easier sample ascertainment across international borders and from rural locales, (b) increased participation due to less invasive sample collection, and (c) less sample shipping and storage burdens. All of these allowances will result in larger sample sizes, which are of increasing importance due to the statistical burden of correcting for within-breed stratification to avoid Type I error as described in Chapter 2.2. Moreover, larger samples of single breeds may also be critical to the success of genomic studies of complex disease because of possible breed heterogeneity, like that observed in our survey for loci underlying noise phobia in Chapter 3. However it should be noted that required sample sizes for dog studies of complex traits are still

orders of magnitude smaller than those required for human studies, in the range of 100 affected dogs versus 1000 or even 10,000 affected people. Additionally, for traits with simplified genetic structure in the dog, even relatively small samples characterized by broad phenotypes may be used to detect associations with loci of strong effect, as was the case for our investigation of adult-onset deafness in herding Border collies (BOC), described in Chapter 4. This latter study highlights the distinct advantage of utilizing canine samples for genetic investigations of disease, particularly when research may provide critical insight into the even more complex genetic structure of the homologous disease in human populations.

5.2. Who Let the Dogs Out: The Future of Canine Genetics

Although there is much left to write in the storybook of canine noise phobia, the preliminary findings demonstrated in Chapter 3 suggest that genetic investigations of phobic anxiety disorders in the dog may in fact not only be doable, but also valuable. Despite small, heterogeneous samples and lackluster multi-breed results, the informal replication of association signals on canine chromosome (CFA) 10, plus suggestive replication signals on CFA5 and 8 all implicate a cornucopia of tantalizing candidate genes expressed in the brain and involved in membrane trafficking, neuronal development and gene regulation. Moreover, predicted epistatic interactions between all three chromosomes are exactly the type of picture one hopes to elucidate for complex multigenic disorders, particularly as they remain muddled in human psychiatric studies complicated by heterogeneous sample populations and inconsistent phenotyping methodologies. Denser genotyping of both existing and new collections of larger samples of single breeds, coupled with stricter control requirements, will hopefully allow for further depth and statistical power to identify risk loci and gene candidates for noise phobia in dogs and, ultimately, specific phobia and/or anxiety in humans.

For the deafness study described in Chapter 4, the obvious next step will be to identify the variant(s) that segregate in affected dogs, and then determine whether the same variants also confer risk in an independent sample. As risk variants are confirmed through direct sequencing in the rest of our cases as well as independent samples of BOC, it is the goal of our project to next embark on extensive longitudinal studies of adult-onset deafness. Working with interested breeders, we plan to systematically track hearing ability—both through observation-based owner questionnaires and physiologically using brainstem auditory evoked response—in dogs starting at puppyhood. This information would be used in conjunction with genetic "risk" profiles to determine if particular genetic variants may be useful predictors of deafness onset and/or severity. Only after proof of predictive validity will it then be reasonable to consider development of a genetic test for adult-onset deafness, though this is the ultimate goal of our group and the BOC herding community. However, it will be interesting to see whether the genetic results will be generalizable to other breeds as well, or if the genetic variants identified are risk loci within the (herding) BOC community only.

In addition to further research in the dog, it is also our hope to extend the knowledge we gain in our canine genetic studies to the human world at large. Specifically, genes implicated in noise phobia/anxiety and/or adult-onset deafness could be investigated in the respective corresponding human clinical populations to test the hypothesis that the biological mechanisms underlying these complex disorders are similar enough in dogs and humans alike that genetic findings in one population may be informative to both. Perhaps only then will man's best friend be truly that, pointing the way towards scientific innovation and medical breakthroughs that are as yet unachievable with humans alone.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement: *I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*

Author Signature

12/10/2010
Date