INSIGHT INTO SIGHT:

DECIPHERING THE HETEROGENEITY OF HEREDITARY RETINAL DEGENERATION IN THE CANINE POPULATION.

A dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Orly Goldstein

January 2014

INSIGHT INTO SIGHT:

DECIPHERING THE HETEROGENEITY OF HEREDITARY RETINAL DEGENERATION IN THE CANINE POPULATION.

Orly Goldstein, Ph.D.

Cornell University 2014

As of October 2013, around 285 million people are visually impaired worldwide. For an important subset of these, this visual impairment is genetic. That is, they have inherited mutant genes that prevent sight by affecting an element critical for vision. Understanding the genetic basis of these diseases offers insight into the mechanisms involved in photoreceptor development, function, and maintenance as well as significant potential for therapies addressed at cure or prevention. In recent decades, understanding of the cellular and molecular mechanisms of vision has expanded dramatically. In particular, many genes have been discovered that are vital for normal function of the retina, the delicate, multilayered, light-sensitive layer at the back of the eye that connects by the optic nerve to the brain.

Retinitis pigmentosa (RP) is a subgroup of inherited eye diseases causing retinal degeneration. In the U.S.A alone, an estimated 100,000 people have inherited RP, either as an autosomal dominant, autosomal recessive, or X-linked disease. Among these, about 50% of cases of autosomal recessive RP cannot be explained by so far identified genes, and for many of the known genes, their role in vision is unclear.

The dog also suffers from inherited eye diseases. The canine disease homolog of RP is referred to as Progressive Retinal Atrophy (PRA). Because of the unique population structure and genetic

differences within and among the various breeds of dogs, this animal model offers a remarkable tool for discovering genetic mechanisms in vision and serves as a therapeutic model for potential gene therapy.

In the present work, we have investigated nine different canine diseases (OSD, *erd*, *prcd*, *cd*, *crd1*, *crd2*, *crd3*, Basenji PRA, and Italian Greyhound PRA), characterized them, discovered the genes responsible for their phenotype, and determined the broad spectrum of breeds affected by them. These studies utilized classical genetic methods such as linkage and candidate gene approaches, and were expanded to employ Linkage Disequilibrium and Association Studies. We discovered two novel genes, PRCD and STK38L, both of which cause PRA in dogs, but had not previously been recognized as involved in vision or visual disorders, and thereby identified novel pathways critical for vision, as well as genes potentially responsible for human RP. We also identified novel mutations in six known genes (COL9A2, COL9A3, ADAM9, PDE6B, IQCB1, and SAG) that cause five different diseases, and thereby established new animal models for potential gene therapy in their human counterparts. We identified the exact deletion points of the *cd* disease in a broad spectrum of breeds affected by this disease, showing that they are all inherited Identical By Descent (IBD). We discovered the potential involvement of a microRNA in retinal degeneration in the Italian greyhound PRA, the first such evidence in a large animal model, and the first suggestion that retinal degeneration can be caused by alteration in gene expression regulated by microRNA. We developed screening tests for all the above diseases so these diseases can be eliminated from affected breeding lines. We showed that, in the dog, pooled samples might be used for association studies, when research budgets are limited.

BIOGRAPHICAL SKETCH

Orly Goldstein was born in Jerusalem, Israel in 1964. After attending "Leyada" High School, and serving the Israeli defense army for two years, Orly attended the Hebrew University in order to pursue a Bachelor degree in Biological Sciences and Statistics, receiving the degree in 1989, with a major in Immunology. However, her real passion was Genetics and, as an identical twin, was always curious about how genes make people who they are. She continued her higher education in the genetic field, receiving her Master of Science degree, *Summa cum Laude*, from the Human Genetics Department of Tel-Aviv University, in 1993. After graduating, she continued to work at Tel-Hashomer Institute in Ramat-Gan, first as a research assistant, and then as the Acting Head of the Fragile-X Diagnostic Laboratory. Life carried her to the other side of the world, where she joined the team of Professor Tarantal, at the University of California, Davis, exposing her research experience to the animal model. In Dr. Tarantal's laboratory at the Primate Center, she learned about fetal/neonatal growth regulation and restriction, and *in utero* and postnatal therapies. When she returned to Israel in 2000, she ventured outside the academia and joined the team of Compugen, where she worked with a young and energetic team of researchers from different disciplines.

In 2002 she returned to the USA. Wanting to stay in the field of Medical Genetics, she joined the group of Drs. Gustavo Aguirre and Gregory Acland at the Veterinary School, Cornell University, and in 2007 she enrolled in the Graduate School in Comparative Biomedical Sciences in order to achieve a Ph.D. degree through the Employee Degree Program. In this laboratory she found a way to help dogs and humans in the field of Ophthalmology, researching blindness and its genetic components.

I would like to dedicate this work to my wonderful children, Nitsan, Roey, and Daniel, who supported me through this journey, and encouraged me when it wasn't easy.

All my love -- to you.

ACKNOWLEDGEMENTS

I always knew that my goal was to get a Ph.D. in the genetics and genomics field and to become an independent researcher. Life had its own plans, however, and the opportunity to achieve this goal came at an age when most people think about retirement. Even though my motivation was high, I would have not been able to achieve this goal if it wasn't for all the wonderful people who supported me throughout this journey.

First and foremost, I thank my advisor and committee chair, Dr. Gregory Acland, for believing in me, and teaching me so much. The most important lessons I have learned from him are to identify the questions I want to ask, choose the correct ones (those which can be answered), decide on the best approach to get the answers, and never forget what the questions were. These skills are essential for any researcher, and I am glad he was able to teach them to me. I thank him for sharing his experience with me, providing his guidance when I needed it, and engaging in many philosophical conversations about vision, science, and life in general.

During my time in the laboratory I had the honor to work with an outstanding group of people. I would like to thank Dr. Gustavo Aguirre, who was always willing to help, comment, and share his opinions about our research; to Sue Pearce- Kelling, an amazing colleague and a wonderful friend, who always helped with anything and everything that I needed, and for her endless support; to Anna Kukekova, for being a great mentor and role model; to Julie Ann Jordan for her dedicated work in histology, DNA extraction and managing all the samples- she provided invaluable technical support; to Jennifer Johnson for her great tips and graphical

support, as well as sharing her expertise in Next-Generation-Sequencing analysis; to Liz Corry and the DNA Bank at Cornell University; and to Dr. Barbara Zangerl, Dr. Duska Sidjanin, and Dr James W. Kijas, for being great team workers.

Thanks to Amanda Nickle, Karla Carlisle and the staff of the retinal Disease Studies Facility, for their commitment and professionalism that was essential to my research.

Thanks to my committee members, Dr. Vicki Meyers-Wallen, Dr. John Schimenti, and Dr. Scott Coonrod, for providing useful guidance and comments, quick responses and support, and to Dr. David Lin for always willing to meet with me when I needed help.

My work was also dependent on the contribution of several collaborators. In particular I am grateful to Dr. Andy Clark and Dr. Xu Wang from Cornell University, for sharing their expertise in Allele-specific-expression measurements, and Next-Generation-Sequencing; to Optigen and its team for sharing research samples with me; to Dr. Jason Mezey, Dr. Chuan Gao and Dr. Adam Boyko from Cornell University for statistical analysis and genotype calling, to Dr. Richard Guyon and Dr. Simone Iwabe from University of Pennsylvania, and to Dr. Connie Yeh, and Dr. Andras Komaromy from University of Pennsylvania and Michigan University.

During my research I received endless support from the Biotechnology Resource Center, which managed to sequence my samples in a very efficient and timely manner. I especially want to thank Dr. Peter Schweitzer for his generous help and advice in the SNPChip and Next-Generation-Sequencing fields.

I would like to thank the National Institute of Health; the Foundation Fighting Blindness; the Morris Animal Foundation; Hope for Vision; the Glenn of Imaal Terrier Club of America and the Van Sloun Fund for Canine Genetic Research for providing funding for this research.

Finally, this work would have not been possible without the moral support of all my Ithacanian friends, and my family, especially my children, Nitsan, Roey and Daniel, and my sister Nurit. I can't thank you all enough.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 Retinal diseases.

1.1.1 The retina, phototransduction and visual cycle

Vision is one of our five senses and is defined by the ability to interpret the surrounding environment by processing information encoded in visible light. There are various physiological components involved in vision, which are the focus of much research in psychology, cognitive science, neuroscience, and molecular biology. Light incident on the cornea, at the front of the eye, is focused by the lens onto the retina at the back of the eye. The retina mediates our visual perception. This approximately 0.2 mm thick central nervous tissue, is the first station in many, in a cascade of events that ends in the brain and the ability to see (Figure 1.1). It has a complexed- layered- structure consisting of neuronal and supporting cells: Retinal pigment Ephithelium cells in the RPE, photoreceptor cells in the photoreceptor layer, bipolar cells, Amacrine cells and Horizontal cells in the inner nuclear layer, outer and inner plexform layers, Ganglion cells layer, and the nerve fiber layer.

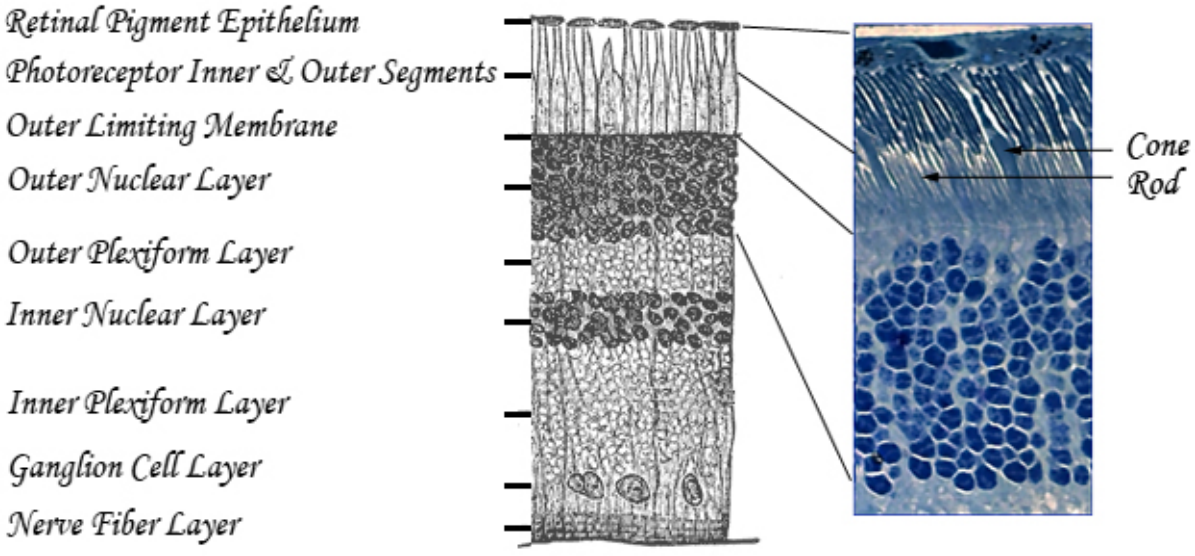


Figure 1.1. The retina. Several layers of cells are observed. In the photoreceptor layer different segments are identified: the outer nuclear layer, the inner segment, and the outer segment. Cones and rods are denoted by black arrows. Image is modified from Wikipedia, public domain.

The photoreceptor cells are categorized into two groups: the rods and the cones. These lie in the outer part of the retina, the region farthest from incoming light. Light travels through the transparent inner retinal layer (all cells mentioned above) before it can be captured by the photoreceptor cells. The retinal pigment epithelial (RPE) cells then absorbs scattered light, or light not absorbed by the photoreceptor cells. The photoreceptor cells have a unique and immaculate organization, which allow the cells to execute their function. They are polarized cells that consist of a synaptic region, a cell body, an inner segment (IS) and an outer segment (OS). The nuclei are adjacent to the bipolar cells, and are present in a ratio of about 1:20 cones to rods. In humans, the macula lays in the center of the retina, a region where 100% of the cells are cones. The "molecular machinery" involved in biosynthesis, energy metabolism, and membrane trafficking is taking place in the IS, which is connected to the OS by a connecting cilium. The OS is comprised of membranous discs surrounded by a plasma membrane. The discs are orderly arrayed perpendicular to the axis of the OS. In humans, about 1,000 discs are stacked in one rod cell. Over 90% of the total protein in the OS is rhodopsin, a G-protein coupled receptor, densely packed within the disc lamellae (~25,000 molecules/μm^2).

Vision in vertebrates begins when a light is absorbed by the rhodopsin, the rod photoreceptor visual pigment, which consists of an apoprotein, opsin, and a chromophore, 11-*cis* retinal. This is the beginning of the phototransduction. The light initiates the isomerization of 11-*cis* retinal and as a result activates rhodopsin. Photoactivated rhodopsin activates the heterotrimeric G protein transducin by catalyzing the exchange of GDP to GTP. The dissociated alpha unit of transducin then activates cGMP-phosphodiesterase, which rapidly hydrolyzes cytoplasmic cGMA. The decrease in cGMP concentration in the cell causes the cGMP-gates cation channels to close, and as a result, the cell becomes hyperpolarized and releases less glutamate transmitters to its

connected bipolar cells. This continues to cause chain reactions from the bipolar cells to the ganglion cells, and the message is then transferred to the brain through the optic nerve.

When light hits a rhodopsin molecule it changes 11-*cis* retinal conformation by isomerization to *all-trans* retinal. *All-trans* retinal is then liberated from opsin, reduced to *all-trans* retinol, leaves the rods and "diffuses" to the RPE, where it undergoes a series of enzymatic reactions to be converted back to 11-*cis*-retinal. 11-*cis* retinal returns back to the OS where it regenerates rhodopsin and completes the visual cycle. For further review on the structure and biology of the retinal cells- refer to reference 1.

Rods mediate vision in dim light and if they are dysfunctional, can cause night blindness, and cones mediate vision in day light and the perception of colors, and if dysfunctional can result in day blindness. A percentage of blindness occurs due to genetic mutations in genes essential for normal function of these photoreceptor cells and the RPE cells. These are usually inherited as autosomal recessive (ar), autosomal dominant (ad) or X-linked traits. The diseases are generally classified based on whether the disease primarily affects the rods or the cones, though this task is problematic since the diseases show substantial clinical and genetic overlap.

1.1.2 Retinitis pigmentosa

Retinitis pigmentosa (RP) represents the most frequent cause of inherited visual impairment, with a worldwide prevalence of about 1:4,000[2]. In RP, the rods are the cells that are initially affected, resulting in night blindness and tunnel vision (peripheral vision is lost, center vision is preserved). Later, with the progression of the disease, the cones are affected as well, which can result in a complete blindness. It encompasses a clinically and genetically heterogeneous group, with a large variability in the age of onset, rate of progression, retinal appearance, and final outcome. The disease can be inherited by any of the three classical modes of inheritance, ar (20-

25%), ad (15-20%), and X-linked (10-15%), as well as mitochondrial inheritance[3] (rare). Furthermore, almost half of all RP cases are simplex (only one person in the family is affected) in which the inheritance pattern cannot be determined, and no family history is reported[4]. RP patients can also be categorized into two main groups: "typical" or "non-syndromic" where the disease is confined to the eye, and "complicated" or "syndromic" where the patients suffer from non-ocular findings.

Currently, RP is known to be caused by mutations in over 50 genes (RetNet, https://sph.uth.edu/retnet/), which presents significant progress since the first mutation associated with RP in human was identified[5] (a proline to histidine change at amino acid position 23 in rhodopsin). Additional genes causing RP in large portions of the patients are RDS[6], ABCA4[7], RPE65[8], PDE6A[9], PDE6B[10], USH2A[11], RPGR[12], and RP2[13]. The more genes and mutations are discovered, and the more patients screened, the more it becomes apparent that phenotype and genotype relationship is inconsistent, and neither can be used to predict the other. Not only can different mutations in different genes cause the same phenotype, the same mutations can cause different phenotypes, and not only between families, but also within families. Interaction between genes may also affect the final outcome, and modifiers can influence the level of penetrance in dominant diseases with incomplete penetrance.

1.1.3 Other retinal degeneration diseases

Photoreceptor cell degeneration can be caused by other diseases, which have their own characterizations, but also share many symptoms with RP. Those include macular degeneration (MD), cone degeneration (CD), and cone-rod degeneration (CRD). These disorders can present at any stage of life but predominantly cause severe visual loss in early to middle age. The most

common photoreceptor degeneration though is age-related macular degeneration (AMD), which accounts for greater than one-half of blindness and visual impairment in industrialized countries[14,15]. As in RP, these diseases are mostly recognized as monogenic forms of progressive photoreceptor cell death, but most of the recognized gene mutations account for only a small fraction of cases. Almost all causal mutations are rare (minor allele frequency (MAF) <<0.01). The diseases start with cone cell malfunction, resulting in day blindness, which can then progress to night blindness (CRD) or not (CD). Also like RP, they can be expressed as part of a syndrome. ABCA4 is the most common mutated gene in autosomal recessive cone-rod dystrophy, CNGA3 is the most common in autosomal recessive cone dystrophy, and in both diseases there are still 60-90% cases with unknown genetic causality. Some diseases are congenital and as such it is unclear if the disease starts in the cone or the rod cells. An example of the latter is Leber congenital amaurosis (LCA), where the patients are severely visually impaired or blind from birth. They present with nystagmus and a retinal appearance that varies from normal to mild pigment mottling with mild vascular attenuation to severe pigmentation and vascular attenuation that resembles the fundus in RP-like dystrophies. Their ERG shows no function of either the cones or the rods in the first year of life. The distinction between some retinal dystrophies can be very subtle or even arbitrary, and mutation of a single gene can result in varied clinical diagnoses. For example, at end stage CD can hardly be distinguished from CRD and LCA clinically, and molecular-genetically overlaps with CD, CRD and RP.

1.2 The Dog as an animal model

In humans, gene discovery is still not near saturation yet for either monogenic disorders or complex traits. Although monogenic diseases present a simpler goal, complex traits harbor challenges that researchers struggle to overcome. Recent studies have used strategies of

hypothesis- free fine mapping of genes and loci to identify underlying factors in common complex diseases with major impact on public health (Wellcome trust case control consortium, 2007). The successful reports mainly involved the collection of very large study cohorts for any individual trait and international collaborations on an unprecedented scale. That said, detecting genes underlying diseases might not always need large global population samples. Samples of individuals from genetically isolated populations (population isolates) have proved immensely useful in the identification of rare recessive disease genes by way of homozygosity analysis. One of the most impressive examples is the data in genetic studies provided by the company Decode Genetics in Iceland (http://www.decode.com). They have identified by linkage, and more recently by genome- wide association studies (GWAS), an impressive number of variants contributing to the development of common/complex diseases and traits in the Iceland population.

Other isolated populations with proven value in gene mapping are found in Finland, where 35 monogenic diseases have been mapped and cloned, as well as complex traits such as schizophrenia[16]; the Micronesian Island of Kosrae[17], Northern Sweden[18], and the Chamorro people of Guam[19] where complex traits such as plasma plant sterol, Schizophrenia, and Guam neurodegenerative disease were mapped.

Systematic studies of common genetic variants are facilitated by the fact that Linkage Disequilibrium (LD) exists in the human population, because of shared common ancestry and historical and prehistorical population bottlenecks. Early information documented the existence of LD in the human genome[20,21] limited to a small number of regions in the genome and to a small number of individuals. With the development of advanced technology and statistical tools, and the explosion of data from the rest of the genome, it became clear that the human genome

displayed more LD than previously predicted, and the LD varied across regions, showing a segmented structure[22-24].

These observations suggest that researchers can benefit from LD- based methods and at the same time emphasize the need to study LD structure at high-density resolution and in more populations. The HapMap project phase I characterized 1.3 million human SNPs, and 3.1 million in phase II, in 270 individuals from 4 geographically diverse populations (The international HapMap Consortium 2005; The international HapMap Consortium 2007). The paradigm underlying association studies is that linkage disequilibrium can be used to capture association between markers and nearby un-typed SNPs. However, the Phase II HapMap data shows that there is much more complexity in the structure of the human genome: there are long-range similarity among haplotypes but more striking is the fact that about 0.5-1.0% of all high-frequency SNPs are untaggable (no other SNP within 100Kb has an r^2 of at least 0.2). As a result, one can imagine that in the research of common diseases, with no homogenous population in particular, GWAS will require a much more condensed SNP genotyping than is available, with high risk of missing the rare alleles. The causative allele might be located on several haplotypic backgrounds, thereby diluting its signal to an extent that precludes its identification by genetic means.

However, this is not observed in populations that are genetically more homogenous such as found in populations that are geographically isolated. In those populations the LDs are larger, and as a result, GWAS requires fewer markers (up to 30% less). In these populations, the use of a dense genotype platform would result in a good coverage of the genome and small number of gaps, and sometimes could be very beneficial in cases of rare alleles[25].

Dogs evolved through mutually beneficial relationships with humans, and their history can be traced back from 15,000 to 100,000 years ago[26]. The needs of the humans in different times and places for dogs with specific traits, ranging from hunting and herding to guarding and companionship, resulted in selection forces that have created about 400 modern dog breeds. As a consequences of these stringent breeding programmes and periodic population bottlenecks, the breeds mimic "human isolated populations" as far as the structure of their genetic pool, giving rise to a high prevalence of breed-specific diseases, including cancer, blindness, heart diseases, epilepsy, hip dysplasia and deafness. Most of these diseases are also commonly seen in human populations, with similar clinical symptoms.

The first draft of the canine Boxer genome was published in 2004 with a 7.6X coverage[27]. Since then two additional assemblies have been released (CanFam2- 2005 and CanFam3- 2011) improving the annotation and the SNP database. Subsequent contributions to the canine genome after the release of the first assembly have focused on better understanding canine chromosomal structure and LD. Gray and her group[28] showed that only 5% reduction in nucleotide diversity is a result of domestication whereas the loss of nucleotide diversity with breed formation averaged 35%. Other studies suggested that the Linkage disequilibrium (LD) within breeds extended over distances of several megabases, but across breeds only tens of kilobases[27,29]. Sutter and his group showed that LD depends on the breed under study, and can vary up to 10 fold in breeds that range from rare to popular, and whose population histories feature a range of popular sire and bottleneck effects. Together with the strong selection that breeders had imposed in order to achieve the specific traits desired in each breed, the breeds now have an excess of inherited diseases, expressed on a relatively homogenous genetic background.

As such, the canine can be used as a model for human diseases, with population structural advantages that can be exploited.

1.3 PRA and other retinal diseases in the dog population

Dogs suffer from various forms of naturally occurring retinal diseases causing blindness, with the most common one known as progressive retinal atrophy (PRA). Over 100 breeds are affected with PRA, and some minor and major differences in appearance can be recognized between them, from age of onset to rate of progression, to clinical abnormalities. The disease was recognized as the homolog to RP in humans, and the disease of both species shared genetic, clinical and pathological features[30].

Dogs are born with an undeveloped retina that reaches maturity around 5-8 weeks postnatally. As a mature retina, the structure is broadly similar to that of man, consisting of layers of specialized cells. The major difference between human and canine retina is in the localization of the cones: in humans an area named the fovea contains only cone cells, whereas such an area does not exist in the canine retina. In the canine retina, the closest parallel to the human fovea is a region identified with a higher concentration of cones than other more cone-dilute regions in the retina. Despite these differences in eye development and structure, the phenotypic similarities and heterogeneity observed in humans are very close to the ones observed in dogs. The retinal diseases in dogs, as in humans, can be classified as progressive, stationary, or developmental. All forms of inheritance are observed: autosomal dominant PRA (ADPRA), autosomal recessive PRA (ARPRA), and X-linked. The clinical observation of any of these types of diseases can be assessed in several ways: (i) Ophthalmologic examination by an Ophthalmologist, evaluating the health of the retina by its ability to reflect light back (increased reflectivity of the tapetal region

indicates a diseased retina), the number of visible branchpoints in blood vessels in the back of the eye (retinal vascular attenuation indicates disease) and appearance of the optic nerve (pallor, atrophy of the optic nerve head= disease); (ii) A physiological examination: electroretinogram (ERG) measures the ability of the rods and cones to generate an electrical signal when exposed to light; (iii) a histological morphology is possible only on research colony dogs or donated retinas from privately owned dogs, where the disease can be accurately assessed in comparison to the unique and well defined structure of a healthy retina. Retinal diseases can start in the cones and then affect the rods (cone-rod dystrophies, CRD), start in the rods and then affect the cones (rod-cone dystrophies, RCD) or can be confined to only one type of photoreceptor cells. Since the retina of the dog is not fully developed when it is born, it is easier to identify where and when the disease initiated, which many times is difficult to do in humans.

PRA is characterized by initial degeneration of the rod photoreceptor cells. Mutations causing PRA can affect the rods only or rods and subsequently cones. The dominance of rod involvement in PRA affected dogs manifests as night blindness, and can progress to complete blindness in cases where cones are secondarily affected. In contrast, cone-rod dystrophies in dogs affect predominantly the cones, which would result initially in day blindness, with rods being affected later or to a lesser extent.

There is a broad age of onset range in various breeds. All diseases can be categorized to two main groups: early onset, and late onset. In dogs, early onset PRA/CRD manifests during the postnatal retinal differentiation period (2-6 weeks) and results from abnormal to arrested retinal development or progressive degeneration immediately after retinogenesis. These diseases typically result in relatively rapid progression toward end-stage retinal degeneration, and can be clinically evident in young adult dogs (1-2 years of age). In contrast, late-onset PRA/CRD shows

pathological changes after the retina is fully developed and mature, and sometimes these diseases do not appear until well after reproductive maturity. Usually the progression is very slow, and some dogs do not show any behavioral symptoms until very late in life.

1.4 Statement of purpose

Hereditary retinal degenerations (HRD), caused by mutations in essential genes expressed in the retina, are characterized by dysfunction and often death of rod and cone photoreceptors. These HRDs are genetically heterogeneous, with more than 200 causative loci (including 202 identified genes) recognized to date in humans (RetNet: https://sph.uth.tmc.edu/retnet, 2013). The diseases can be subdivided by clinical phenotype, and include retinitis pigmentosa (RP), cone-rod dystrophy (CRD), retinal dysplasia (RD), and Leber congenital amaurosis (LCA), among many others.

My goal was to use the dog population, suffering from similar diseases, to accelerate the discovery of genes responsible for HRD in the dog, with the aim to successfully identify the locus and the mutation for each disease, and to evaluate the effect of the mutation on RNA stability and the morphology of the retina at different ages of retinal development.

This was undertaken in several breeds segregating different specific ocular diseases, using meiotic linkage analysis (colony-derived informative pedigreed samples); linkage disequilibrium analysis (population-based samples); and genome-wide association analysis on both individual samples (colony-derived, and population-based), and pooled samples (colony-derived, and population-based) (Figure 1.2).

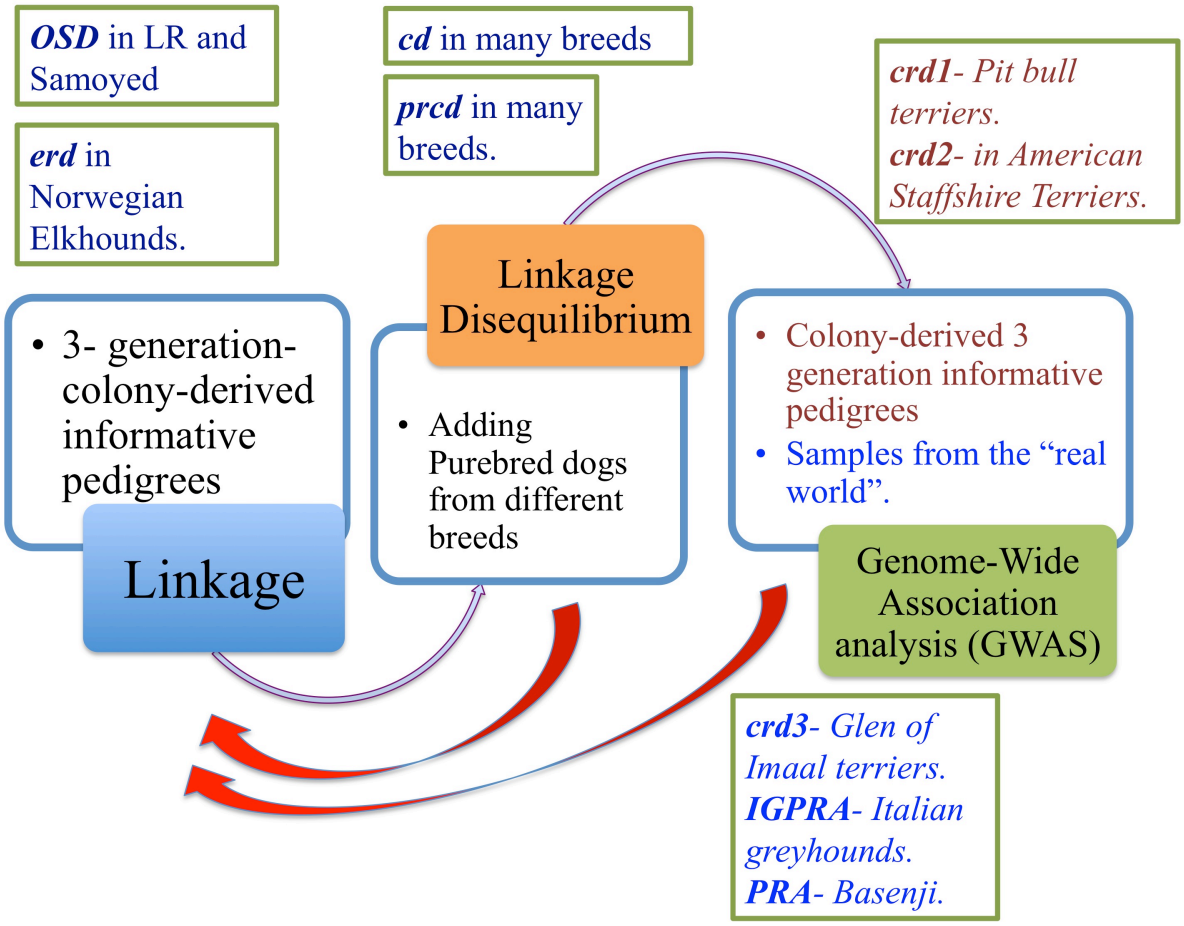


Figure 1.2. Flow chart of all the studies, the breeds affected, and the methods in which the loci were mapped. OSD and *erd* were mapped by linkage analysis on colony-pedigrees; *prcd* and *cd* were mapped by linkage analysis on colony-pedigrees with a combination of linkage disequilibrium using affected dogs from other breeds; *crd1* and *crd2* were mapped by GWAS using colony-derived dogs; and *crd3*, IGPRA and PRA in the Basenji breed were mapped by GWAS on purebred dogs.

In chapter 2 I describe how useful and efficient these approaches were to map the genes and discover the mutations for eight different diseases: in subchapters 2.2 and 2.3 I describe how using meiotic linkage analysis in colony-based pedigrees enabled us to identify three mutations in three different genes (Figure 1.3): a deletion mutation in Col9A2 causing Oculoskeletal dysplasia (OSD) in the Samoyed breed and an insertional mutation in Col9A3 causing OSD in

the Labrador retriever breed (**Subchapter 2.2**); a novel gene was identified, STK38L, causing early retinal degeneration (*erd*) disease in the Norwegian Elkhound dogs (**Subchapter 2.3**). A SINE insertion in exon 4 of the gene, results in exon skipping and the removal of S100B binding site from the mRNA, as well as an important phosphorylation residue. This is the first time that this gene has been associated with retinal degeneration, or any vision problem, and suggests a role for new pathways and genes in the vision cascade.

The emerging data from studies on the genome structure of purebred dogs[27-29] suggested that we could use Linkage Disequilibrium (LD) to reduce regions responsible for diseases that are shared among several breeds. In subchapters 2.4 and 2.5, using Linkage Disequilibrium (LD) across pure breeds, I was able to identify disease- associated- haplotypes inherited identical by descent (IBD) in multiple breeds sharing the same disease, and consequently reduce the LD intervals (Figure 1.3). An affected haplotype composed of 98 polymorphisms was shared between 14 dog breeds suffering from progressive rod- cone degeneration (*prcd*), and the LD was reduced from 1.5 Mb interval to 106 Kb (**Subchapter 2.4**). This work excluded all known genes within this interval, suggesting that the causative mutation for *prcd* is a novel gene, and focused our attention on the conserved sequences. We identified a novel gene, PRCD, that was mutated in these dogs[31], characterized the molecular structure of the gene, and the effect of the mutation on the retina, and established its potential to cause RP in humans. Following our discovery, three unrelated families were identified with mutations in the PRCD gene[31-33]. In **Subchapter 2.5**, I was able to identify the specific deletion point of the cone-degeneration (*cd*) disease affecting Malamutes, previously mapped to CFA29[34] and identified a 0.5 Mb shared- haplotype among the affected Malamutes, an affected Miniature Australian Shepherd (MAS), a Siberian Husky and Alaskan sled- dogs. This work shows that *cd*, which was previously thought to be breed specific,

is caused by an identical mutation in other breeds (IBD). This suggests that other canine diseases, thought be to limited to one breed, may be genetically homogeneous in closely, or even distantly, related breeds.

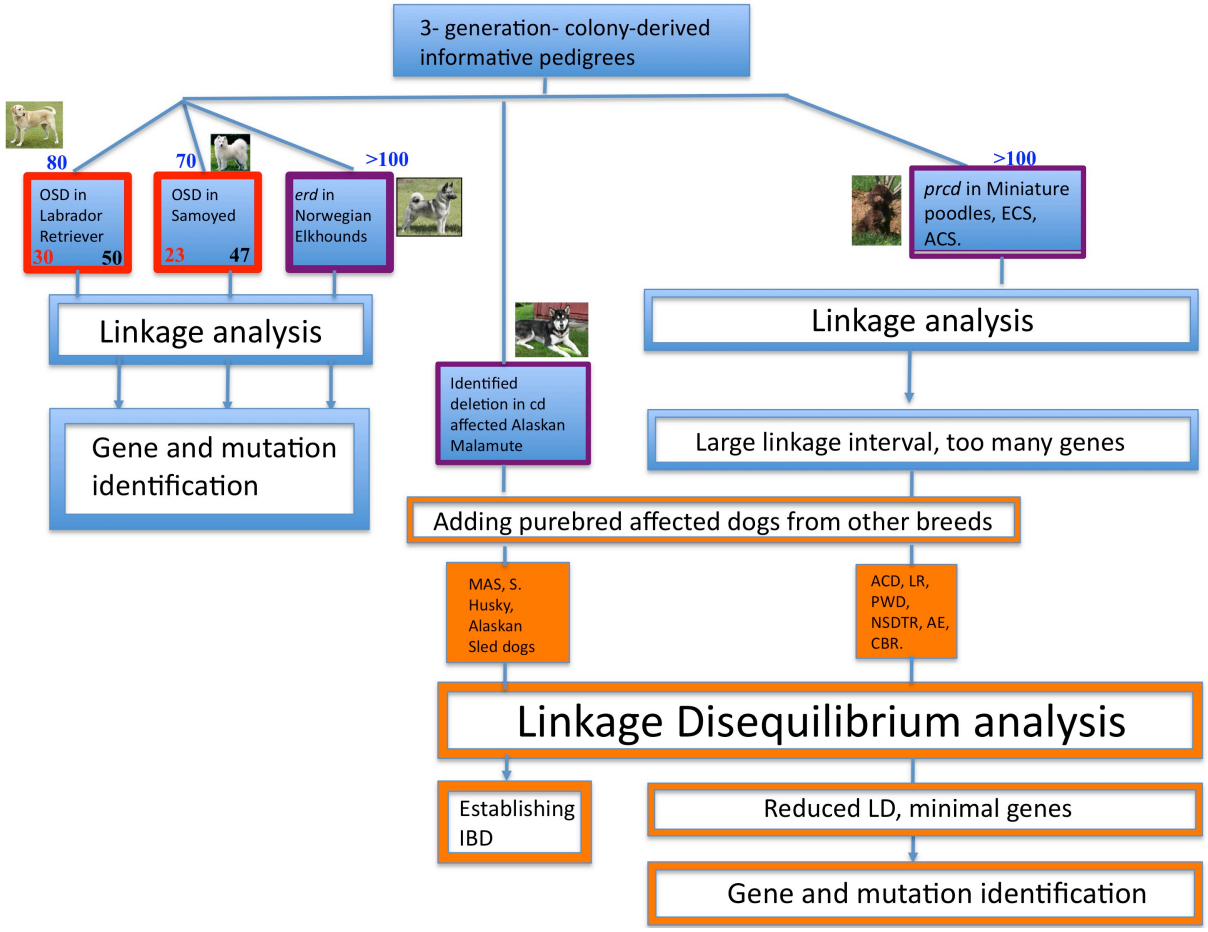


Figure 1.3. Schematic presentation of work- flow on four different diseases, number of dogs used in each study, and final outcome. OSD in Labrador retriever and Samoyed and erd in Norwegian Elkhounds were mapped by linkage analysis while cd and prcd were mapped by using linkage and linkage disequilibrium.

The fast growing database of single nucleotide polymorphisms (SNPs) and the development of high throughput Chips in the canine model, made it possible for the research community to explore the possibility of mapping disease-genes in dog populations by means of association, and move away from families and linkage. In **Subchapters 2.6, 2.7, 2.8, and Chapter 3**, I explored that idea and was able to identify the loci and causative mutations for four different retinal diseases (*crd3*, *crd1* and *crd2*, and Italian greyhound PRA) using population-based samples (*crd3* and IGPRA) or colony-pedigreed dogs (*crd1* and *crd2*) (Figure 1.4).

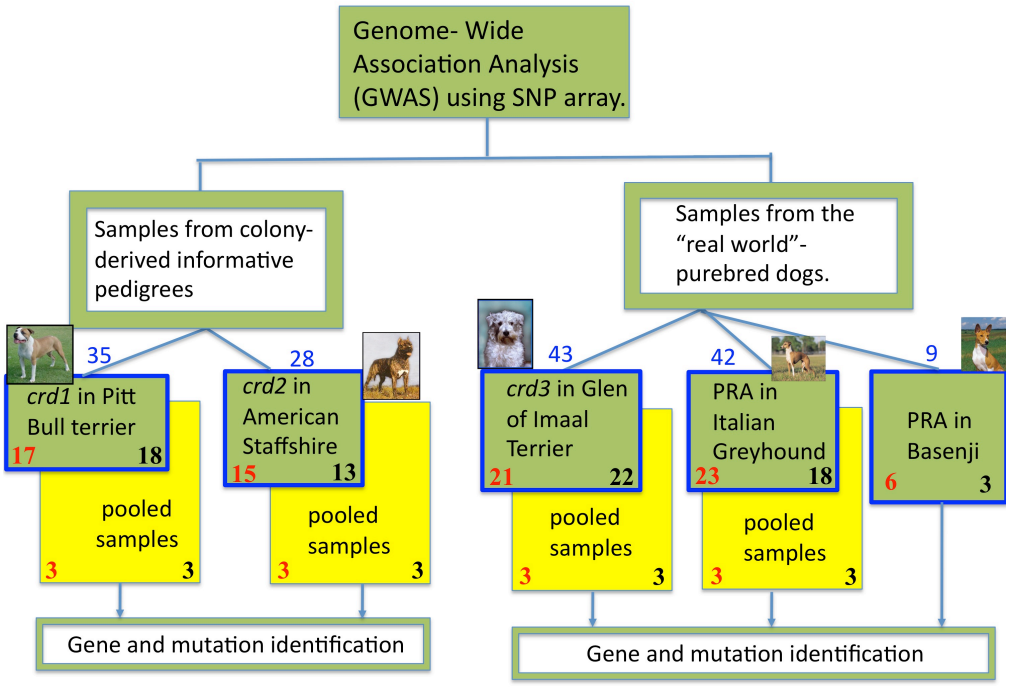


Figure 1.4. Schematic presentation of work-flow on five different diseases in five different breeds mapped by Genome-Wide- Association. The total number of dogs (Blue), number of affected dogs (red), and number of control dogs (black), as well as number of pooled samples are noted. *crd1* and *crd2* are rare diseases and colony-derived dogs were used, while *crd3*, IGPRA and PRA in the Basenji were mapped using DNA from purebred dogs.

We discovered a novel mutation in ADAM9 that is responsible for a late-onset cone-rod dystrophy in the Glen of Imaal terrier breed and is the homolog of human cone-rod dystrophy 9

(**Subchapter 2.6**). ADAM9 belongs to a family of genes that was only recently recognized as a major group of genes causing retinal degeneration. We characterized its gene structure, the mutation effect on the cone and rod cells, the progress of the disease, and suggested this animal model as a gene therapy model for the human counterpart. In **Subchapter 2.7** we show and disscuss how two different genes (PDE6B, and IQCB1), working in separate pathways, caused similar diseases in two closely related breeds. This emphasizes the heterogeneity of retinal degeneration and the need for genetic screening to better understand the cause of each disease, even when the symptoms are similar. These phenotypic similarities were seen also in the late onset PRA in the Basenji breed and the Italian Greyhound, as they exhibit the same phenotype as dogs affected with *prcd*. After proving these two diseases are non allelic to *prcd*, we mapped the loci by GWAS (**Subchapters 2.8 and Chapter 3**).

The unique genetic structure of the Basenji breed[35] inspired me to try to map the gene with a very small number of dogs (Figure 1.4). We mapped the disease using only 6 affected dogs and three controls, all related to each other but not too closely, which allowed a higher level of heterozygosity. We identified a non-stop mutation in the SAG gene and developed a screening test (**Subchapter 2.8**). We mapped the IGPRA to a region on chromosome 11, and identified a new pathway to RP: down regulation of Col27A1 and mir455 were retina- specific and associated with this late onset PRA (**Chapter 3**).

I was also curious to see if the unique genetic structure of dog breeds would permit researchers to run association studies using pooled samples when studying simple mendelian traits, to benefit laboratories with small budgets. I compared GWAS results from individual samples and pooled samples in four diseases: *crd1*, *crd2*, *crd3* and IGPRA (Figure 1.4), and the results (**Chapter 4**) clearly show that genotyping pooled samples (5 dogs in each pooled sample) and using them in

GWAS analysis was successful in mapping the genes. The correct locus was identified, but the trade-off was a larger LD, as rare recombinant animals were missed.

For all of these 10 genes, I developed genetic tests, which can be used to screen dogs, identifying carriers, affected and genetically normal dogs, thus helping breeders to create a healthy breeding program that would prevent the birth of blind dogs.

More importantly, we discovered new genes and pathways involved in vision and suggested potential therapeutic approaches for blindness in humans.

1.5 References

1. Sung CH, and Chuang JZ. The cell biology of vision. JCB 2010, 190 (6): 953-963.

2. Hamel C (2006) Retinitis pigmentosa. Orphanet J Rare Dis 1:40.

3. Mansergh FC, Millington-Ward S, Kennan A, Kiang AS, Humphries M, Farrar GJ, Humphries P, Kenna PF (1999) Retinitis pigmentosa and progressive sensorineural hearing loss caused by a C12258A mutation in the mitochondrial MTTS2 gene. Am J Hum Genet 64(4):971–985.

4. den Hollander AI, Black A, Bennett J, Cremers FP. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. J Clin Invest. 2010 Sep;120(9):3042-53.

5. Dryja TP, McGee TL, Reichel E, Hahn LB, Cowley GS, Yandell DW, Sandberg MA, Berson EL. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature. 1990 Jan 25;343(6256):364-6.

6. Keen T J, Ingleheran C F. Mutations and polymorphisms in the human peripherin-RDS gene and their involvement in inherited retinal degeneration. Hum. Mutat. 1996;8:297–303.

7. Klevering B J, Yzer S, Rohrschneider K, Zonneveld M, Allikmets R, van den Born L I, Maugeri A, Hoyng C B, Cremers F P. Microarray-based mutation analysis of the ABCA4 (ABCR) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. Eur. J. Hum. Genet. 2004;12:1024–1032.

8. Wang Q, Chen Q, Zhao K, Wang L, Wang L, Traboulsi EI. Update on the molecular genetics of retinitis pigmentosa. Ophthalmic Genet. 2001 Sep;22(3):133-54. Review.

9. Dryja TP, Finn JT, Peng YW, McGee TL, Berson EL, Yau KW. Mutations in the gene encoding the alpha subunit of the rod cGMP-gated channel in autosomal recessive retinitis pigmentosa. Proc Natl Acad Sci U S A. 1995 Oct 24;92(22):10177-81.

10. McLaughlin ME, Sandberg MA, Berson EL, Dryja TP. Recessive mutations in the gene encoding the beta-subunit of rod phosphodiesterase in patients with retinitis pigmentosa. Nat Genet. 1993 Jun;4(2):130-4.

11. Weston MD, Eudy JD, Fujita S, Yao S, Usami S, et al. 2000. Genomic structure and identification of novel mutations in usherin, the gene responsible for Usher syndrome type IIa. Am. J. Hum. Genet. 66:1199–210.

12. Fujita R, Swaroop A. RPGR: part one of the X-linked retinitis pigmentosa story. Mol Vis. 1996 Jun 4;2:4.

13. Branham K, Othman M, Brumm M, Karoukis AJ, Atmaca-Sonmez P, Yashar BM, Schwartz SB, Stover NB, Trzupek K, Wheaton D, Jennings B, Ciccarelli ML, Jayasundera KT, Lewis RA, Birch D, Bennett J, Sieving PA, Andreasson S, Duncan JL, Fishman GA, Iannaccone A, Weleber RG, Jacobson SG, Heckenlively JR, Swaroop A. Mutations in RPGR and RP2 account for 15% of males with simplex retinal degenerative disease. Invest Ophthalmol Vis Sci. 2012 Dec 13;53(13):8232-7.

14. Swaroop, A., Chew, E. Y., Rickman, C. B. & Abecasis, G. R. Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration. Annu. Rev. Genomics Hum. Genet. 10, 19–43 (2009).

15. Ting, A. Y., Lee, T. K. & MacDonald, I. M. Genetics of age-related macular degeneration. Curr. Opin. Ophthalmol. 20, 369–376 (2009).

16. St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, Gosden C, Evans HJ. illness. Lancet. 1990 Jul 7;336(8706):13-6.

17. Kenny EE, Gusev A, Riegel K, Lütjohann D, Lowe JK, Salit J, Maller JB, Stoffel M, Daly MJ, Altshuler DM, Friedman JM, Breslow JL, Pe'er I, Sehayek E. Systematic haplotype analysis resolves a complex plasma plant sterol locus on the Micronesian Island of Kosrae. Proc Natl Acad Sci U S A. 2009 Aug 18;106(33):13886-91.

18. Maaike Alaerts; Shana Ceulemans; Diego Forero; Lotte N. Moens; Sonia De Zutter, BE; Lien Heyrman; An-Sofie Lenaerts; Karl-Fredrik Norrback; Peter De Rijk;Lars-Göran Nilsson; Dirk Goossens; Rolf Adolfsson; Jurgen Del-Favero. Support for NRG1 as a Susceptibility Factor for Schizophrenia in a Northern Swedish Isolated Population. Arch Gen Psychiatry/vol 66: 8, Aug 2009.

19. Sieh W, Choi Y, Chapman NH, Craig UK, Steinbart EJ, Rothstein JH, Oyanagi K, Garruto RM, Bird TD, Galasko DR, Schellenberg GD, Wijsman EM. Identification of novel susceptibility loci for Guam neurodegenerative disease: challenges of genome scans in genetic isolates. Hum Mol Genet. 2009 Oct 1;18(19):3725-38.

20. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am J Hum Genet. 2001 Jul;69(1):1-14.

21. Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res. 2000 Oct;10(10):1435-44.

22. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA.

23. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet. 2001 Oct;29(2):229-32.

24. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science. 2002 Jun 21;296(5576):2225-9.

25. Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin MR, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet. 2006 May;38(5):556-60.

26. Wayne RK, Geffen E, Girman DJ, Koepfli KP, Lau LM, Marshall CR. Molecular systematics of the Canidae. Syst Biol. 1997 Dec;46(4):622-53.

27. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshatsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolen O, Elong R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M, Foley K, Foley C, Franke A, Friedrich D, Gage D, Garber M, Gearin G, Giannoukos G, Goode T, Goyette A, Graham J, Grandbois E, Gyaltsen K, Hafez N, Hagopian D, Hagos B, Hall J, Healy C, Hegarty R, Honan T, Horn A, Houde N, Hughes L, Hunnicutt L, Husby M, Jester B, Jones C, Kamat A, Kanga B, Kells C, Khazanovich D, Kieu AC, Kisner P, Kumar M, Lance K, Landers T, Lara M, Lee W, Leger JP, Lennon N, Leuper L, LeVine S, Liu J, Liu X, Lokyitsang Y, Lokyitsang T, Lui A, Macdonald J, Major J, Marabella R, Maru K, Matthews C, McDonough S, Mehta T, Meldrim J, Melnikov A, Meneus L, Mihalev A, Mihova T, Miller K, Mittelman R, Mlenga V, Mulrain L, Munson G, Navidi A, Naylor J, Nguyen T, Nguyen N, Nguyen C, Nguyen T, Nicol R, Norbu N, Norbu C, Novod N, Nyima T, Olandt P, O'Neill B, O'Neill K, Osman S, Oyono L, Patti C, Perrin D, Phunkhang P, Pierre F, Priest M, Rachupka A, Raghuraman S, Rameau R, Ray V, Raymond C, Rege F, Rise C, Rogers J, Rogov P, Sahalie J, Settipalli S, Sharpe T, Shea T, Sheehan M, Sherpa N, Shi J, Shih D, Sloan J, Smith C, Sparrow T, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Stone S, Sykes S, Tchuinga P, Tenzing P, Tesfaye S, Thoulutsang D, Thoulutsang Y, Topham K, Topping I, Tsamla T, Vassiliev H, Venkataraman V, Vo A, Wangchuk T, Wangdi T, Weiand M, Wilkinson J, Wilson A, Yadav S, Yang S, Yang X, Young G, Yu Q, Zainoun

J, Zembek L, Zimmer A, Lander ES. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005 Dec 8;438(7069):803-19.

28. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK. Linkage disequilibrium and demographic history of wild and domestic canids. Genetics. 2009 Apr;181(4):1493-505.

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

30. Aguirre GD and Acland GM. Models, Mutants, and Man: Searching for Unique Phenotypes and Genes in the Dog Model of Inherited Retinal Degeneration. In: Ostrander EA, Giger U, Lindblad-Toh K (eds). The dog and its genome. Cold Spring Harbor Laboratory Press, Cold Spring Harbor: 291-325.

31. 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. Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. Genomics. 2006 Nov;88(5):551-63.

32. Nevet MJ, Shalev SA, Zlotogora J, Mazzawi N, Ben-Yosef T. Identification of a prevalent founder mutation in an Israeli Muslim Arab village confirms the role of PRCD in the aetiology of retinitis pigmentosa in humans. J Med Genet. 2010 Aug;47(8):533-7.

33. Pach J, Kohl S, Gekeler F, Zobor D. Identification of a novel mutation in the PRCD gene causing autosomal recessive retinitis pigmentosa in a Turkish family. Mol Vis. 2013 Jun 13;19:1350-5.

34. Sidjanin DJ, Lowe JK, McElwee JL, Milne BS, Phippen TM, Sargan DR, Aguirre GD, Acland GM, Ostrander EA: Canine CNGB3 mutations establish cone degeneration as orthologous to the human achromatopsia locus ACHM3. Hum Mol Genet 2002, 11:1823–1833.

35. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, Reynolds A, Bryc K, Brisbin A, Knowles JC, Mosher DS, Spady TC, Elkahloun A, Geffen E, Pilot M, Jedrzejewski W, Greco C, Randi E, Bannasch D, Wilton A, Shearman J, Musiani M, Cargill M, Jones PG, Qian Z, Huang W, Ding ZL, Zhang YP, Bustamante CD, Ostrander EA, Novembre J, Wayne RK. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature. 2010 Apr 8;464(7290):898-902.

CHAPTER TWO

IDENTIFICATION OF NINE GENES INVLOVED IN EIGHT RETINAL DISEASES IN THE DOG POPULATION USING LINKAGE, LINKAGE DISEQULIBIRIUM AND GENOME-WIDE- ASSOCIATION ANALYSES.

2.1 Summary

The eye is an extremely complex, yet well defined organ that has an incredible responsibility: to capture light throughout the day and night, translate and transfer it to the brain so we can see. The retina, which is responsible mainly for the translation of the photon to an electrical signal and transfer it to the brain, is shown to express thousands of genes, however their role in vision is known and understood for only a small fraction of them. In this chapter I hope to show that we can use blind dogs to help us better understand how the retina functions, as well as helping the dog population by reducing the number of blind dogs. We had identified nine genes and mutations that are causing blindness in dogs, of which two are completely novel in sight. These newly discovered genes and pathways are candidates for RP and other ocular diseases in humans and can help explain a portion of the patients with no genetic resolution.

2.2 *Col9a2 and Col9a3 mutations in canine autosomal recessive oculoskeletal dysplasia.*

Oculoskeletal dysplasia segregates as an autosomal recessive trait in the Labrador retriever and Samoyed canine breeds, in which the causative loci have been termed *drd1* and *drd2*, respectively. Affected dogs exhibit short-limbed dwarfism and severe ocular defects. The disease phenotype resembles human hereditary arthro- ophthalmopathies such as Stickler and Marshall syndromes, although these disorders are usually dominant. Linkage studies mapped *drd1* to canine chromosome 24 and *drd2* to canine chromosome 15. Positional candidate gene analysis then led to the identification of a 1-base insertional mutation in exon 1 of COL9A3 that cosegregates with *drd1* and a 1,267-bp deletion mutation in the 5' end of COL9A2 that cosegregates with *drd2*. Both mutations affect the COL3 domain of the respective gene. Northern analysis showed that RNA expression of the respective genes was reduced in affected retinas. These models offer potential for studies such as protein-protein interactions between different members of the collagen gene family, regulation and expression of these genes in retina and cartilage, and even opportunities for gene therapy.

Following our discovery, a human family segregating Stickler syndrome disease was identified carrying a mutation in Col9A2[1], and a further family segregating Stickler syndrome disease was identified carrying a mutation in Col9A3[2].

For the complete research results please refer to:

Goldstein O, Guyon R, Kukekova A, Kuznetsova TN, Pearce-Kelling SE, Johnson J, Aguirre GD, Acland GM. COL9A2 and COL9A3 mutations in canine autosomal recessive oculoskeletal dysplasia. Mamm Genome. 2010 Aug;21(7-8):398-408.

2.3 Exonic SINE insertion in STK38L causes canine early retinal degeneration (*erd*).

Fine mapping followed by candidate gene analysis of *erd* — a canine hereditary retinal degeneration characterized by aberrant photoreceptor development — established that the disease cosegregates with a SINE insertion in exon 4 of the canine STK38L/NDR2 gene. The mutation removes exon 4 from STK38L transcripts and is predicted to remove much of the N terminus from the translated protein, including binding sites for S100B and Mob proteins, part of the protein kinase domain, and a Thr-75 residue critical for autophosphorylation. Although known to have roles in neuronal cell function, the STK38L pathway has not previously been implicated in normal or abnormal photoreceptor development. Loss of STK38L function in *erd* provides novel potential insights into the role of the STK38L pathway in neuronal and photoreceptor cell function, and suggests that genes in this pathway need to be considered as candidate genes for hereditary retinal degenerations. This is the first time that the STK38L/NDR2 pathway has been implicated in photoreceptor development or disease. With hindsight, however, it is reasonable to expect such a role, given the importance of the polarized, highly regulated, circadian growth and renewal cycle in photoreceptors, and the proposed roles for STK38L in other, especially neuronal, tissues. STK38L and its interacting partners, S100B and the Mob proteins, thus provide novel candidate genes for hereditary retinal disorders of humans and other species.

For the complete research results please refer to:

Goldstein O, Kukekova AV, Aguirre GD, Acland GM. Exonic SINE insertion in STK38L causes canine early retinal degeneration (erd). Genomics. 2010 Dec;96(6):362-8.

2.4 Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome.

Canine progressive rod–cone degeneration (*prcd*) is a retinal disease previously mapped to a broad, gene-rich centromeric region of canine chromosome 9. As allelic disorders are present in multiple breeds, we used linkage disequilibrium (LD) to narrow the ~6.4-Mb interval candidate region. Multiple dog breeds, each representing genetically isolated populations, were typed for SNPs and other polymorphisms identified from BACs. The candidate region was initially localized to a 1.5-Mb zero recombination interval between growth factor receptor-bound protein 2 (GRB2) and SEC14-like 1 (SEC14L). A fine-scale haplotype of the region was developed, which reduced the LD interval to 106 kb and identified a conserved haplotype of 98 polymorphisms present in all *prcd*-affected chromosomes from 14 different dog breeds. The findings strongly suggest that a common ancestor transmitted the *prcd* disease allele to many of the modern dog breeds and demonstrate the power of the LD approach in the canine model.

For the complete research results please refer to:

Goldstein O, Zangerl B, Pearce-Kelling S, Sidjanin DJ, Kijas JW, Felix J, Acland GM, Aguirre GD. Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. Genomics. 2006 Nov;88(5):541-50.

2.5 Genomic deletion of CNGB3 is identical by descent in multiple canine breeds and causes achromatopsia.

Achromatopsia is an autosomal recessive disease characterized by the loss of cone photoreceptor function that results in day-blindness, total colorblindness, and decreased central visual acuity. The most common causes for the disease are mutations in the CNGB3 gene, coding for the beta subunit of the cyclic nucleotide-gated channels in cones. CNGB3-achromatopsia, or cone degeneration (*cd*), is also known to occur in two canine breeds, the Alaskan malamute (AM) and the German shorthaired pointer. We characterized the achromatopsia phenotype in a new canine breed, the miniature Australian shepherd (MAS). Genotyping revealed that the dog was homozygous for a complete genomic deletion of the CNGB3 gene, as has been previously observed in the AM. Identical breakpoints on chromosome 29 were identified in both the affected AM and MAS with a resulting deletion of 404,820 bp. Pooled DNA samples of unrelated purebred Australian shepherd, MAS, Siberian husky, Samoyed and Alaskan sled dogs were screened for the presence of the affected allele; one Siberian husky and three Alaskan sled dogs were identified as carriers. An identical shared affected haplotype, 0.5 Mb long, was observed in all three breeds and defined the minimal linkage disequilibrium (LD) across breeds. Since the MAS is not known to be genetically related to the AM, other breeds may potentially carry the same *cd*-allele and be affected by achromatopsia.

For the complete research results please refer to:

Yeh CY*, Goldstein O*, Kukekova AV, Holley D, Knollinger AM, Huson HJ, Pearce-Kelling SE, Acland GM, Komáromy AM. Genomic deletion of CNGB3 is identical by descent in multiple canine breeds and causes achromatopsia. BMC Genet. 2013 Apr 20;14(1):27.

* First and second authors Contributed equally.

2.6 An ADAM9 mutation in canine cone-rod dystrophy 3 establishes homology with human cone-rod dystrophy 9.

Cone-rod dystrophies are severe hereditary retinal diseases characterized by primary dysfunction and loss of cone photoreceptors accompanying or preceding that of rods. A clinically similar disorder, termed canine cone-rod dystrophy 3 (*crd3*), segregates in the Irish Glen of Imaal Terrier (GIT) breed of dog as an adult onset trait of previously undetermined mode of inheritance. We report results of a GWAS that found significant association to *crd3* on canine chromosome 16 (CFA16), and led to identification of a deletion mutation in the canine ADAM9 gene that cosegregates with the disease. The mutation removes approximately 23 kb of genomic sequence, including exons 15 and 16, and results in a premature stop codon in exon 17. The mutant protein translated from this transcript is predicted to be truncated, lacking the last 287 amino acids of the C-terminus, part of the cysteine-rich domain, the complete epidermal growth factor (EGF)-like domain, the transmembrane domain, and the cytoplasmic tail. The association of this deletion mutation in canine ADAM9 with *crd3* establishes this canine disease as orthologous to CORD9 in humans, and offers opportunities for further characterization of the disease process, and the potential for genetic therapeutic intervention.

For the complete research results please refer to:

Goldstein O, Mezey JG, Boyko AR, Gao C, Wang W, Bustamante CD, Anguish LJ, Jordan JA, Pearce-Kelling SE, Aguirre GD, Acland GM. An ADAM9 mutation in canine cone-rod dystrophy 3 establishes homology with human cone-rod dystrophy 9. Mamm Genome. 2010 Aug;21(7-8):398-408.

2.7 IQCB1 and PDE6B mutations cause similar early onset retinal degenerations in two closely related terrier dog breeds.

We previously identified two early onset autosomal recessive retinal degenerations in American Staffordshire terrier dogs, and American Pit Bull Terrier dogs. In both diseases very young dogs (less than one year old) were affected by severe photopic and scotopic visual impairment, which progressed to more severe blindness in early adulthood. Because of the similarity of these two diseases, they were termed *crd1* and *crd2* (for cone-rod dystrophy 1, and 2, respectively). For the same reason, and because these two breeds of dog are physically similar and share common ancestry, a cross breeding complementation test was undertaken to prove that the two diseases were nonallelic. Structural photoreceptor abnormalities were observed in *crd1*-affected dogs as young as 11 weeks old. Rod and cone inner and outer segments were abnormal in size, shape and number. In *crd2*-affected dogs, rod and cone IS and OS were abnormal as early as 3 weeks of age, progressing with age to severe loss of the OS, and thinning of the ONL by 12 weeks of age. GWAS identified association at the telomeric end of CFA3 in *crd1*-affected dogs and on CFA33 in *crd2*-affected dogs. Candidate gene evaluation identified a three bases deletion in exon 21 of PDE6B in *crd1*-affected dogs, and a cytosine insertion in exon 10 of IQCB1 in *crd2*-affected dogs. These findings provide new large animal models for comparative disease studies and evaluation of potential therapeutic approaches for the homologous human diseases.

For the complete research results please refer to:

Goldstein O, Mezey JG, Schweitzer PA, Boyko AR, Gao C, Bustamante CD, Jordan JA, Aguirre GD, Acland GM. IQCB1 and PDE6B mutations cause similar early onset retinal degenerations in two closely related terrier dog breeds. Invest Ophthalmol Vis Sci. 2013 Oct 25;54(10):7005-19.

2.8 A non-stop S-antigen gene mutation is associated with late onset hereditary retinal degeneration in dogs.

In the Basenji breed, the adult onset of PRA was observed, with initial visual loss in dim light (night blindness), which gradually progressed to total blindness. Initial visual loss affects the peripheral visual field, but unless the dog is used for high visual performance tasks such as agility work, the reduction in the visual field (tunnel vision) may not be apparent.

Despite tunnel vision and night blindness, many Basenjis affected with PRA retain adequate forward daylight vision for many years, sometimes for their entire natural life. This phenotype highly resembles the progressive rod-cone degeneration (*prcd*) disease. However, a complementary breeding test for dogs affected with *prcd* excluded allelism between these two similar diseases, and this has been confirmed with PRCD genotyping. Blood samples from six affected cases and three nonaffected controls were collected, and DNA extraction was used for a genome-wide association study using the canine HD Illumina single nucleotide polymorphism (SNP) array. Homozygosity and linkage disequilibrium analyses favored one chromosome, CFA25, and screening of the S-antigen (SAG) gene identified a non-stop mutation (c.1216T>C), which would result in the addition of 25 amino acids (p.*405Rext*25).

Identification of this non-stop SAG mutation in dogs affected with retinal degeneration establishes this canine disease as orthologous to Oguchi disease and SAG-associated retinitis pigmentosa in humans, and offers opportunities for genetic therapeutic intervention.

For the complete research results please refer to:

Goldstein O, Jordan JA, Aguirre GD, Acland GM. A non-stop S-antigen gene mutation is associated with late onset hereditary retinal degeneration in dogs. Mol Vis. 2013 Aug 27;19:1871-84.

2.9 References

1. Baker S, Booth C, Fillman C, Shapiro M, Blair MP, Hyland JC, Ala-Kokko L. A loss of function mutation in the COL9A2 gene causes autosomal recessive Stickler syndrome. Am J Med Genet A. 2011 Jul;155A(7):1668-72.

2. Faletra F, D'Adamo AP, Bruno I, Athanasakis E, Biskup S, Esposito L, Gasparini P. Autosomal recessive stickler syndrome due to a loss of function mutation in the COL9A3 gene. Am J Med Genet A. 2013 Nov 22.

CHAPTER THREE: PROGRESSIVE RETINAL ATROPHY IN THE ITALIAN GREYHOUND IS ASSOCIATED WITH DOWN-REGULATION OF COL27A1 AND MIR455.

Manuscript from: Orly Goldstein, Julie Ann Jordan, Sue Pearce-Kelling, Xu Wang, Andrew Clark, Anna Kukekova, Jennifer Johnson, Simone Iwabe, Gustavo Aguirre, and Gregory M. Acland**.** Progressive retinal atrophy in the Italian greyhound is associated with down-regulation of Col27A1 and Mir455.

In preparation.

3.1 Summary

The Italian Greyhound breed suffers from a late onset progressive retinal atrophy (IG-PRA). Previously we excluded this disease as allelic to PRCD. Here we used population-based genome-wide-association- analysis of affected and control dogs, and mapped the disease to a locus on CFA11. Candidate gene analysis by retinal RT-PCR and exon scanning failed to find the causative genetic variation. A combination of DNA-next generation-sequencing, RNA and smallRNA-Seq, and Allele-specific pyrosequencing revealed down-regulation of Col27A1 and mir455 in dogs heterozygous for the risk allele, suggesting that the IGPRA is caused by low levels of mir455.

3.2 Introduction

The clinical term Progressive Retinal Atrophy (PRA) refers to a broad category of hereditary retinal degenerations that segregate in various breeds of dog, and represent the canine equivalent of Retinitis Pigmentosa in humans. Like RP, PRA in different families has differing and often family-specific distinguishing features such as mode of inheritance, age of onset, rate of progression and other aspects of clinical presentation. The archetypal presentation of PRA, and of RP, is as an autosomal recessive trait first manifesting in early adulthood with rod visual deficits (night-blindness) that progresses to deficits in cone-mediated vision, accompanied by ophthalmoscopic evidence of degenerative retinal thinning, and eventual total blindness. Such hereditary retinal degenerations account for about 50% of all forms of PRA, and of RP.

PRA in the Italian Greyhound (IG) is clinically a paradigmatic autosomal hereditary retinal degeneration. Affected dogs are visually, behaviorally, and ophthalmoscopically indistinguishable from nonaffected dogs when young (i.e from weaning at ~ 6-8 weeks of age,

through at least 2 years old). Ophthalmoscopic changes typical of PRA start to become apparent at about 3 to 5 years of age, and are indistinguishable from other canine adult onset forms of PRA, such as *prcd*[1] and *crd3*[2].

The Italian Greyhound population in the USA is a very small and inbred breed, and as in any such group, especially when the disease phenotype is noticeable only at an older age, the mode of inheritance is difficult to decode. We had previously suggested that the disease is autosomal recessive and showed that the Italian greyhound PRA (IGPRA) is non-allelic to *prcd*[3].

Here we established the phenotypic characterization of the disease and mapped the locus to a region on chromosome 11, by genotyping an affected and a control group using the Illumina High-definition canine SNP array. Association analysis identified the minimum linkage disequilibrium. Candidate gene analysis and RNA evaluation suggest a novel mechanism for retinal degeneration: downregulation of microRNA expression. This opens the door not only to a large group of candidate genes for RP, but also suggests a potential gene therapy, and offers a large animal model in which to test therapy.

3.3 Material and Methods

3.3.1 Animals

The IGPRA strain of dogs, maintained at the Retinal Disease Studies Facility (RDSF) in Kennett Square, PA, derives from one affected Italian greyhound dog, which was outbred to a beagle. Their progeny was used to generate pedigrees segregating IGPRA. Selected dogs from this strain were evaluated by DNA and RNA analysis (Figure 3.1). Blood samples were also collected from 95 purebred Italian greyhound privately owned dogs, as well as other breeds, and DNA extracted. All procedures involving animals were undertaken according to IACUC approved

protocols at Cornell University and the University of Pennsylvania, and in adherence to the ARVO Resolution for the Use of Animals in Ophthalmic and Vision Research.

3.3.2 Disease phenotype evaluation

Purebred, privately owned dogs were evaluated by certified Ophthalomologist, mostly by Drs Aguirre and Acland. A routine Ophtahalmoscopic evaluation was carried out and results reported.

***3.3.2.1 ERG analysis*-** in selected cases, electroretinography was undertaken either to confirm diagnosis, or to establish diagnosis before ophthalmoscopic evidence of disease, using methods described previously[4,5], or by using a modified Ganzfeld dome fitted with the LED light stimuli transferred from a ColorDome stimulator (Diagnosys LLC, Lowell, MA, USA), as described by Komáromy[6].

***3.3.2.2 Morphological analysis*-** from selected colony dogs, eyes were enucleated post mortem and processed for morphologic evaluation using a triple-fixative protocol, essentially as described previously[2]. For selected retinas, in vivo retinal imaging was performed under general anesthesia using a combined cSLO/sdOCT (Spectralis HRA/ Optical Coherence Tomography-OCT, Heidelberg, Germany) instrument. Near-infrared *en face* imaging was done using near infra-red and short wave autofluorescent mode with the 55^{o} lens, and overlapping images were acquired. Vertical and horizontal sdOCT parallel raster scans were obtained using a 30^{o} lens. All four central quadrants (superior, inferior, nasal and temporal) were scanned with the following settings: 30^{o} x 20^{o} scanned area with 49 sequential B-scans separated by a 120 μm distance and an average of 9 ART per scan. As well, single scans averaged (n=100) sequential B-scans were taken in a 30 Degree area (line) on the four cardinal points and centered in the optic disc.

3.3.3 SNP array and Genome-wide association analysis

We genotyped 23 affected and 18 unaffected dogs on the Illumina canine HD SNP array following the manufacturer's standard protocol (Illumina Inc, San Diego, California). The dogs were mostly related to each other, with a few dogs that had no parents or grandparents in common. For each dog in the affected group we matched a relative in the control group. Genotypes were called using the GenomeStudio algorithm (Illumina Inc, San Diego, California). We included another dog that was a duplicate of one of the affected dogs, and also a duplicate of one of the non-affecteds, to check for accuracy and consistency. The association analysis included only one copy of data of the duplicated animal. Out of the 23 affected dogs, 19 were purebred Italian greyhound dogs, diagnosed with PRA by certified Ophthalmologists, and of which we were therefore confident of their disease status; two were from our colony-derived pedigree. An additional two dogs were included, each having a phenotype that was not typical of the other affected dogs. These dogs had a tentative diagnosis of PRA, as we could not be confident that the diagnosis was accurate, or that the disease was the same as in the other Italian Greyhounds. We ran an association analysis using only the 21 dogs of which we were convinced of their status, excluding the two dogs with the tentative diagnosis. Genotype calls were converted into a Plink-format file and associations were tested using the PLINK[7] association command without pedigree or sex information (http://pngu.mgh.harvard.edu/purcell/plink/).

3.3.4 Candidate gene evaluation

Retinal RNA was extracted from a 9 weeks old heterozygous dog (IG72, Figure 3.1), and several control normal dogs, as previously described[8]. Comparison between IG72 cDNA PCR products and the sequences from the control dogs was performed with Sequencher® 4.2.2 Software (Gene

Codes Corporation, Ann Arbor, MI). The genes that were screened were: zinc finger protein 618 (ZNF618), alpha-1-microglobulin/bikunin precursor (AMBP), kinesin family member 12 (KIF12), collagen, type XXVII, alpha 1 (Col27A1), microRNA455 (mir455), orosomucoid 1 (ORM1), AT-hook transcription factor (AKNA), and Deafness autosomal recessive 31 (DFNB31). Each gene was identified on the CanFam 2.0 seqeunce and primers were designed to amplify over lapping fragments of the predicted cDNA, and in some cases, to amplify exons from genomic DNA (Supplement Table 3.1). Two polymorphisms in AKNA gene were screened on a large number of dogs: a SNP in exon 3, at position 71,624,795 (CanFam 2.0), and a 21 bp deletion in intron 4 at position 71,618,401-421 (CanFam 2.0). For exon 3 SNP, a restriction test was developed: PCR product of primers flanking the SNP (Supplement Table 3.2, primer pair 1) was digested with MaeIII restriction enzyme for 2 hours at 37^0C, and load on an 8% acrylamide gel for size evaluation. The wild-type allele (Guanine) creates a restiction digestion site, resulting in two visible bands of 83 and 26 bp, while the risk allele (adenine) eliminates the restriction site and the amplicon is visible as a 109 bp band. Primers flanking the 21 bp deletion in intron 4 of the AKNA gene (Supplement Table 3.2, primer pair 2) were used to amplify a 239 bp product from a normal chromosome, and a 218 bp band from the affected chromosome.

Col27A1 exon 1 was missing from the CanFam 2.0 and 3.0. Our genome-comparison between the dog, human and mouse suggested that exon 1 is within a 1.45 Kb gap, between 71,444,080 and 71,445,532 (CanFam 2.0). Primers were designed flanking the gap as well as primers located in the predicted exon 1 based on the most conserved bases between human, mouse, and cat homologs (Supplement Table 3.3A).

After retrieving part of exon 1 from RNA-seq data, primers were designed to amplify the gap at the 5' end of the gene based on the newly retrieved sequences (Supplement Table 3.3B), as well

as a combination of known sequence and new sequence (Supplement Table 3.3C). PCR was run using KAPA2G Robust HotStart ready mix kit (KAPA Biosystems), 95^0C for 3 minutes, following by 40 cycles of 95^0C- 15 sec, 58^0C- 15 sec, and 72^0C- 15 sec, ending with 10 minutes at 72^0C. Products were run on agarose gels, cleaned, sequenced by the Sanger sequencing, and aligned. A contig was created with Sequencher® 4.2.2 Software.

Primers were also designed to cover conserved regions upstream to Col27A1 (Supplement Table 3.4).

3.3.5 Next genenration sequencing

We used Illumina next-generation-Sequencing (NGS) technology to sequence DNA, retinal RNA, and retinal small RNA from selected dogs. RNA and small RNA libraries were done using the same RNA extractions. For assessing the run quality of each lane and estimating the base-calling error rate, a small % of the Illumina control sample, phiX, was added to each lane of the flowcells. The error rate was estimated by aligning the reads from each lane to the phiX reference genome.

3.3.5.1 Next-generation-sequencing- genomic DNA

Creating libraries

We included DNA from one affected and one normal purebred Italian Greyhounds. These dogs were chosen from the 41 dogs genotyped by the SNP array. The affected dog was the proband of the colony pedigree (Blue square, Figure 3.1). We also included one colony heterozygous dog (IG73, Figure 3.1). Each genomic DNA was converted to an Illumina sequencing library using Illumina's TruSeq DNA LT Sample Prep kit v2 according to the manufacturer's protocol. Briefly, genomic DNA was sheared using a Covaris sonicator and indexed Illumina adaptors

were ligated onto the ends of the DNA fragments, followed by enrichment by PCR. These genomic DNA libraries were pooled and run in two lanes of a HiSeq2500 rapid run flowcell, on a paired-end, 2 x 100 bp run. Reads were loaded onto GenomeBrowse and compared between affected and reference genomes, affected and normal and affected and a heterozygous dog.

DNA-seq analysis

Reads were mapped to CanFam3 using Bowtie2[9]. Reads with mapping quality <20 were removed with Samtools[10]. Duplicated reads were then removed using the MarkDuplicates tool from picard [http://picard.sourceforge.net.]. Finally, we performed local-realignment around indels using the RealignTargetCreator and IndelRealigner tools from GATK[11,12]. SNPs were then called in the three samples, between the samples, and vs CanFam3, using the UnifiedGenotype tool of GATK[11,12].

3.3.5.2 Next-generation-sequencing- RNA and small RNA

Creating libraries

We extracted RNA from six female dogs: three IG-colony-derived heterozygous dogs (IG72, IG73, IG74, Figure 3.1) and three control colony-derived dogs, each control having a different genetic background. RNA quality and quantity was measured on a spectrophotometer, and 1 ul was run on a formaldehyde agarose gel to assess RNA integrity. Each RNA sample was converted to an Illumina sequencing library using Illumina's TruSeq RNA Sample Prep kit v2 according to the manufacturer's protocol. Briefly, polyA+ RNA was isolated, treated with divalent cations to fragment, and converted to double-stranded cDNA with random priming for both first and second strand synthesis. Indexed Illumina adaptors were ligated onto the ends of the cDNA fragments, followed by enrichment by PCR. These RNA-seq libraries were pooled and run in a single lane of a HiSeq2000 high output flowcell on a single-end, 100 bp run. For the

small RNA analysis, the RNA was converted to an Illumina sequencing library using Illumina's TruSeq Small RNA Sample Prep kit according to the manufacturer's protocol. Briefly, Illumina adaptors were ligated onto the ends of the RNA fragments and reverse transcribed and amplified with indexed PCR primers. After size selection for small RNA, these libraries were pooled and run in a single lane of a HiSeq2000 high output flowcell on a single-end, 50 bp run.

SmallRNA-seq analysis

Reads were trimmed to 22 bases from the start of the reads with fastx_trimmer [FASTX-Toolkit, http://hannonlab.cshl.edu/fastx_toolkit/index.html]. They were then aligned to CanFam3 with Bowtie2[9]. The alignment files were compared using Cufflinks[13] using the three normal animals as replicates of a 'sample' and the three heterozygotes as replicates of a second 'sample'. To check for significant differences in level of expression between the three heterozygous IG and the three normal dogs we ran Cuffdiff.

RNA-seq alignments and differential allelic expression quantification

TruSeq adapter sequences were trimmed and quality filtered using the trimmomatic software[14]. RNA-seq reads were aligned to the dog reference genome (canFam3, http://genome.ucsc.edu/) using TopHat v2[15] with three mismatches. Total gene expression level, quantified by FPKM (Fragments Per Kilobase-pair of exon Model), was calculated for all samples using Cufflinks software[13] based on all mapped reads from the TopHat alignments. The multiple mapped reads were weighted using the "-u" parameter in Cufflinks. The expression level was normalized across samples using quartile normalization. Differentially expressed genes between the disease and control samples were called using Cuffdiff program. De novo SNP calling was performed on combined RNA-seq data only from reads that mapped uniquely to the dog reference sequence with ≥30 read depth, using SAMtools software[16]. Problematic SNPs, such as those with a third

allele, near an indel position, or at the exon-intron junctions, were excluded from the analysis. Allele-specific expression ratio was quantified in each sample, as the number of reference allele-containing reads divided by the total coverage at each identified SNP position[17].

RNA-seq analysis to retrieve exon 1 of Col27A1

The sequences of human, mouse and cat Col27a1 Exon1 were obtained from the UCSC genome browser [http://genome.ucsc.edu/]. Human and mouse Col27a1 was available as RefSeq, cat was determined by alignment to human and mouse RefSeq. The Illumina sequences were blasted against the human and cat exon1 and sequences matching exon1 were aligned in Sequencher v4.2.2 software (Gene Codes Corporation, Ann Arbor, MI). We used the UCSC browser blat function to confirm alignment of the canine contig to human Col27A1 exon 1.

3.3.6 Allele-specific pyrosequencing

Retinal RNA was extracted from 11 dogs: four littermates produced by crossing an affected dog to a normal (IG70, IG72, IG73, IG74 in Figure 3.1), and 7 controls. For two animals, IG70 and one of the control dogs, RNA was extracted also from kidney, cerebellum, ear cartilage, and Knee cartilage. Three of the 7 controls were the 3 control dogs chosen for RNA and small RNA seq experiment. Two ug of each RNA sample was treated with DNase to minimze DNA contamination. cDNA was then prepared with Oligo-dT as previously described[8].

We designed allele-specific pyrosequencing PCR and sequencing primers to target four informative exonic SNP positions in *COL27A1* using PyroMark Assay Design Software Version 2.0.1.15 (Qiagen, CA): one on exon 29, one on exon 32, and two on the last exon, exon 61 (Supplement Table 3.5). The primers were checked to make sure they do not contain any SNPs to eliminate potential biased amplification. We assayed the allelic percentages in both genomic DNA and total RNA samples. Pyrosequencing PCR amplification was carried out in 50 µl

system using Ampli-Taq Gold polymerase (Life Technologies) under the following cycling conditions: 1 cycle of 95^{o} C for 5 min, 45 cycles of 95^{o} C-45 sec, 60^{o} -68^{o} C-30 sec, 72^{o} C-20 sec, followed by 1 cycle of 72^{o} C for 10 min. Six ul of PCR products were checked on an agarose gel to confirm positive products in the RT samples and no product in the RT(-) negative control samples and then prepared according to the manufacturer's protocol and then loaded on the PSQ 96MA Pyrosequencer (Qiagen, CA) with the PyroMark Gold Reagents (Qiagen, CA) using the Allele Quantification method (AQ). Two technical replicates were done for each gene in each sample.

3.4 Results

3.4.1 Disease phenotype

The Italian greyhound strain was derived from an affected IG dog (Blue square, Figure 3.1) that was bred to a normal beagle. Their heretozygous F1 progeny were then intercrossed and backcrossed to generate a pedigree segregating the PRA phenotype.

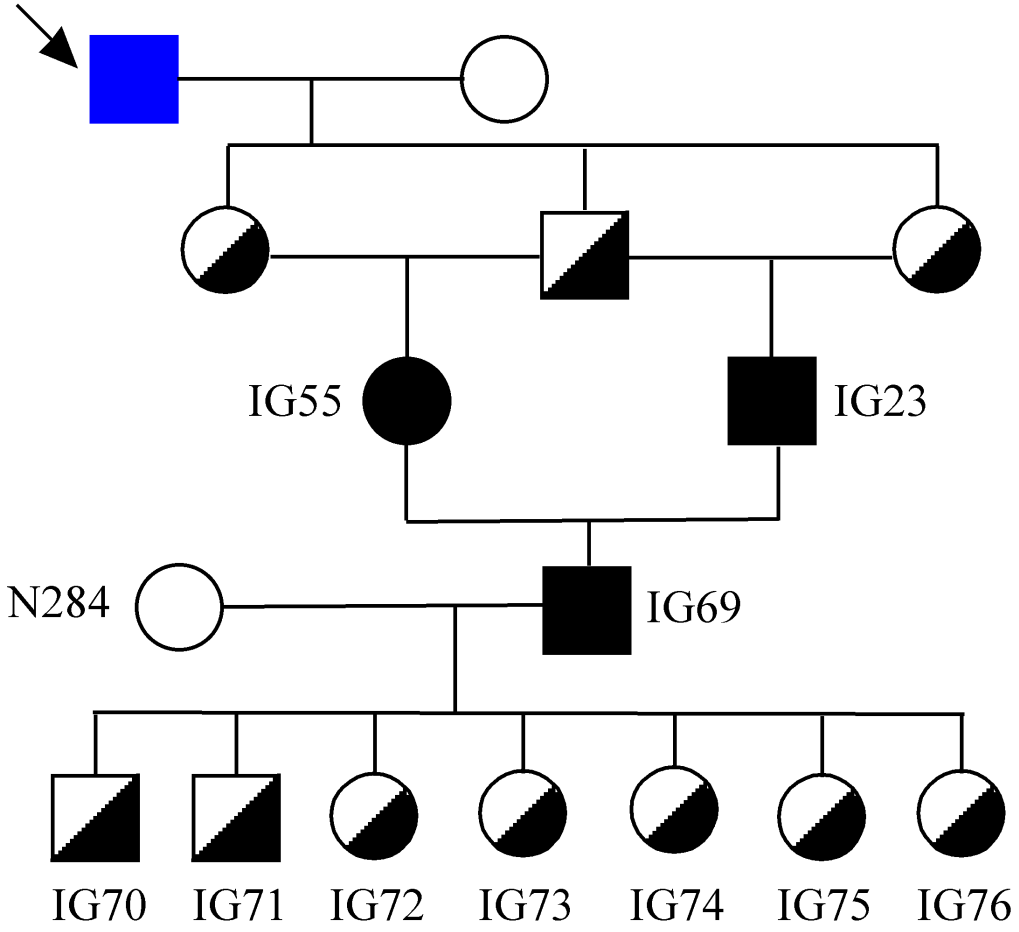


Figure 3.1. Colony-derived-three-generation-pedigree segregating IGPRA. A purebred affected Italian greyhound dog (Blue square and arrowed) was bred to a normal beagle. F1 heterozygous dogs were intercrossed. Their affected offspring were then outcrossed to normal beagles to generate more heterozygous dogs. Circle= females, Squares= males. Filled symbols are affected dogs. Half filled symbols are heterozygotes.

3.4.1.1 Electroretinogram analysis

Selected dogs from the colony were checked for their retina function using electroretinography as previously described[4,5] . At two years of age, the ERG showed normal functions of cones and rods (Figure 3.2, A). At approximately 5-6 years of age, ERGs started to show abnormal responses (data not shown) and by 9 years of age, no responses were recorded (Figure 3.2, B).

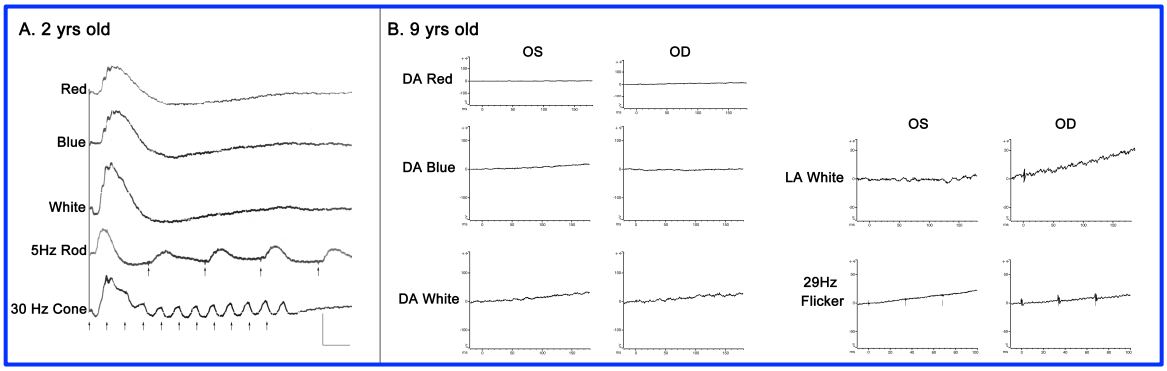


Figure 3.2. Electroretinograms from a PRA-affected Italian Greyhound.
A. ERG recorded at 2 years of age, using methods described previously[4,5]. Traces present, in descending order, ERG responses to single flashes of red, blue and white light; 5 Hz low intensity white flashes (Rod); and 30 Hz high intensity white light flicker (Cone). Red and White traces represent mixed rod-cone responses; Blue and Rod traces are rod-specific; and Cone traces are cone-specific. All responses are broadly within normal limits, although, normal response parameters specific for the Italian Greyhound breed have not been defined. **B**. ERG recorded at 9 years of age, using a modified Ganzfeld dome fitted with the LED light stimuli transferred from a ColorDome stimulator (Diagnosys LLC, Lowell, MA, USA), as described previously[6]. All responses are essentially nonrecordable, including rod specific (DA Blue), cone specific (LA White, 29 Hz Flicker), and mixed rod-cone traces (FA Red, DA White).

3.4.1.2 Morphological analysis

Histological analysis of IGPRA retinas showed a substantial loss of Outer Nuclear Layer (ONL) already at 19 months of age (Figure 3.3. B). Photoreceptor outer segments are markedly reduced in both number and size. The end stage of disease is observed in a 9 year old IGPRA affected retina with severe retinal degeneration where no photoreceptors are observed (Figure 3.3. C).

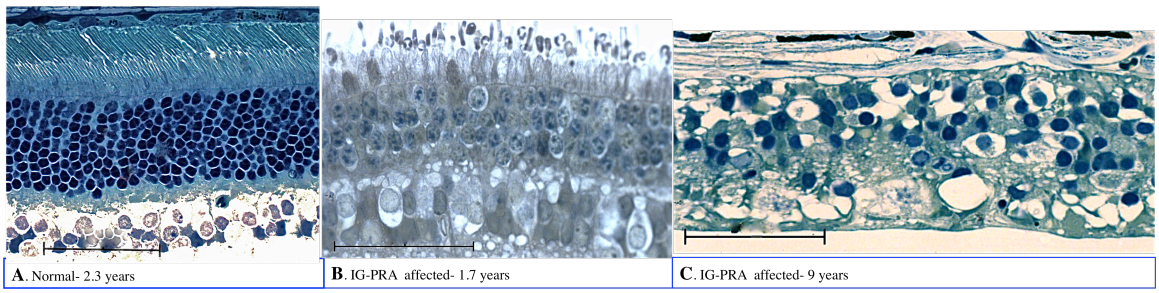


Figure 3.3. Retinal Morphology of Normal and PRA affected dogs. 1 micron plastic sections of canine retinas, fixed in glutaraldehyde / formaldehyde /cacodylate buffer, post fixed in osmium tetroxide; stained with methylene blue/Azure II and counterstained with paraphenylenediamine. **A**. A normal retina of a 2.3 years old dog. The outer nuclear layer (ONL) is thick, inner and outer segments (IS, OS) are well organized, and aligned. **B**. The retina of a 19- months- IGPRA-affected dog is artefactually detached. There is substantial loss of Outer Nuclear Layer (ONL) nuclei. Photoreceptor outer segments are markedly reduced in both number and size. **C**.The end stage of disease is observed in a 9 year old IGPRA affected retina with severe retinal degeneration. No photoreceptors are observed. Bar = 50 microns.

The retinas were also evaluated by fundus images and OCT in selected cases. The degeneration of the retina can be seen using OCT imaging. At 9 years of age, an affected IGPRA retina is markedly thin, much thinner than near the optic nerve (Figure 3.4 A and B). At an early age, ophthalmoscopic changes can become apparent at about 3-5 years of age, evidenced by irregular reflectivity progressing to hyperreflectivity of the tapetal fundus (the first clinical indications of significant retinal thinning). These early changes are followed at later ages by progressive retinal vascular attenuation and pallor of the optic nerve head (Figure 3.4 C).

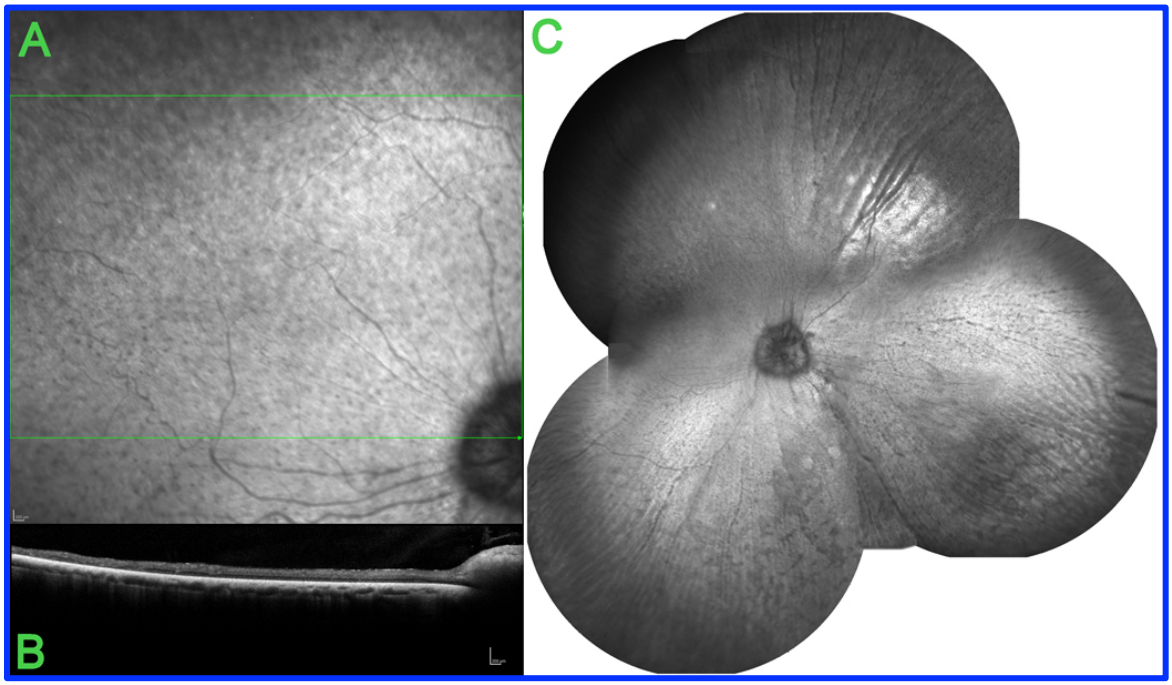


Figure 3.4. Retina of the right eye of a 9- year- old- PRA-affected Italian Greyhound.
A, B. OCT image from temporal fundus. The green arrow in panel A indicates the region scanned to produce the high- resolution cross-sectional retinal image in panel B. The retina is markedly thinned - clearly much thinner than the edge of the optic nerve head, and hardly thicker than the retinal vessels traversed by the OCT scan. This is typical of endstage PRA, and representative of scans recorded across the entire retina of both eyes of this dog. **C.** Montage of cSLO images from the same eye, demonstrating generalized retinal thinning, and advanced retinal vascular attenuation, typical of endstage PRA. The fundus of the left eye was essentially identical.

3.4.2 SNP array and Genome-wide association analysis

All 42 samples in the SNP array results passed quality control. Duplications had reproducibility frequency of 1.0 (all genotype calls were the same between duplicates). Among the 42 samples, 9 had one parent genotyped, and 6 had both parents genotyped, and heritability analysis showed an average of 0.999876 Heritability Frequency, suggesting no mixup of samples, and high accuracy in genotype calls. Genome-wide association analysis between 21 affected dogs and 18 controls showed a hit on chromsome 11, with 37 SNPs carrying a $-\log_{10}(P)$ higher than 4.0 (Table 3.1, Figure 3.5).

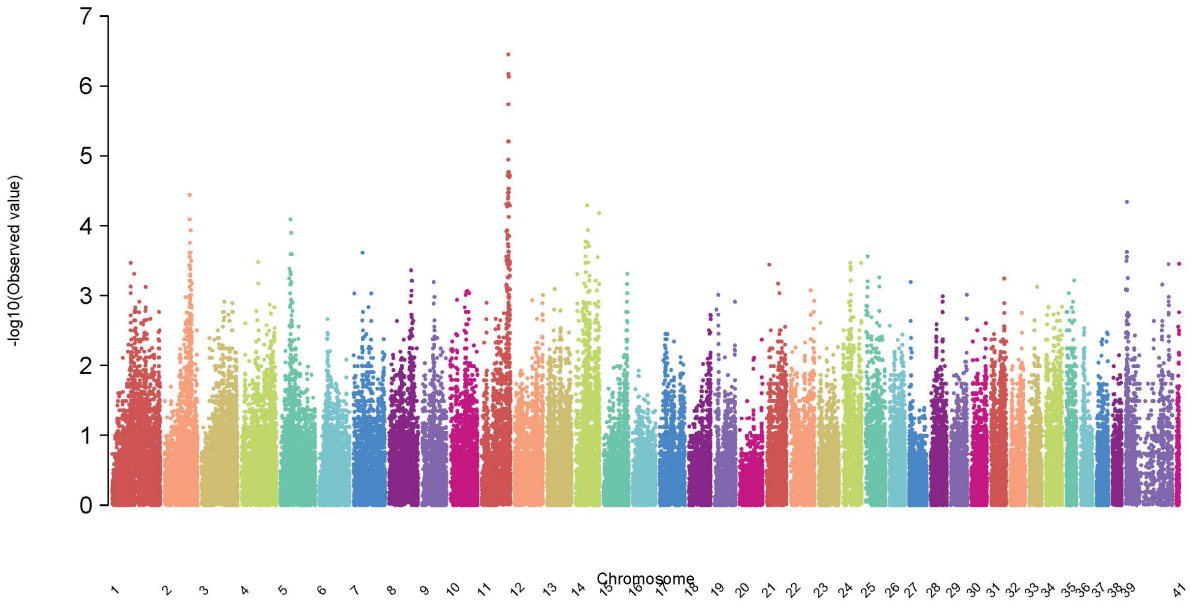


Figure 3.5. Manhattan plots summarizing results of GWAS in the IGPRA. The X-axis shows canine chromosomes 1-38, plus the X-chromosome represented as chromosomes 39 and 41. The Y-axis is the probability statistic ($-\log_{10}$[observed P]). The highest peak is observed on chromosome 11.

Table 3.1. Top hits ($-\text{Log}_{10}(P)>5$) of GWAS in the IGPRA Study.

Number	SNP	Chromosome	Location (CanFam 2)	$-\text{Log}_{10}(p)$
1	BICF2S23353375	11	71,517,224	6.450
2	BICF2P480550	11	71,676,384	6.173
3	TIGRP2P153210_rs8848930	11	72,141,083	6.158
4	BICF2S2304129	11	72,471,087	6.132
5	BICF2P871149	11	72,605,864	6.132
6	BICF2S23549389	11	71,166,278	5.737
7	BICF2P1149744	11	71,319,056	5.737
8	BICF2P482199	11	71,524,550	5.737
9	TIGRP2P152528_rs8622499	11	70,872,927	5.207
10	BICF2P85632	11	70,907,305	5.207
11	BICF2P970535	11	72,514,090	5.207

The genotype calls from chromosome 11 were retrieved and haplotypes were assembled to identify the minimum LD interval. Since the age of onset is late, and the breed is very small, two modes of inheritance models were equally considered: autosomal recessive (AR) and autosomal

dominant with incomplete penetrance (ADIP). We then analyzed the genotype results in two ways, each time assuming a different mode of inheritance.

Under the AR model, we looked for the homozygous block shared by all affected dogs. There was no block of homozygosity shared by all affected dogs. We identified a block that was shared by 17 dogs, while 5 dogs were heterozygous, and one dog was homozygous to a different haplotype. This block was identified between 71,590,374 and 71,732,636, 142.3Kb long (Supplement Table 3.6). Fine mapping and discovery of additional polymorphisms reduced the interval to 135.1 Kb, between 71,590,418 and 71,725,500 (data not shown). This interval included two genes, DFNB31 and AKNA.

Under the ADIP model, we compared the genotypes across all dogs to identify the minimal LD block, that is, where we can identify dogs that carry the risk haplotype but have one SNP in the proximal or distal end that does not share alleles with the rest of the risk haplotype. The LD was established between 71,319,056 and 71,676,384 (357.3Kb, Supplement Table 3.7) and after additional genotyping it was reduced to 285.6 Kb, between 71,347,359 and 71,632,998 (Supplement Table 3.7). This interval excluded DFNB31 coding sequence, and included a potential 3' end of DFNB31 gene, AKNA gene in partial, ORM1, Col27A1, mir455, KIF12, AMBP and ZNF618 in partial.

3.4.3 Candidate gene analysis

We did not have access to a retina from a young affected dog. For all the RNA-based experiments we used F1 retinas (IG72, Figure 3.1). A total of eight genes were screened by retinal cDNA amplification of IG72 and several normal dogs. We also screened some of the genes by exon scanning, genotyping affected and normal dogs. AMBP, KIF12, and ORM1 did

not show PCR products in any of the dogs, suggesting these genes are not expressed in the retina, at least not at the age that retinas were harvested (9-12 weeks old). Many exonic SNPs were found in the rest of the genes. In Table 3.2 we listed only those where the affected allele is different from that in the deposited Boxer genome.

Table 3.2. Coding polymorphisms found in the five genes screened in an affected dog that were different from the database.

#	Alleles	Location	Exon	Allele in database	IGPRA affected allele	Syn/Nonsyn	Comments
A. DFNB31 gene.							
1	G/C	71,651,318	10	C	G	Syn.	
B. AKNA gene.							
1	A/G	71,628,236	2	G	A	Syn	rs22170430
2	**A/G**	**71,624,795**	**3**	**G**	**A**	**Nonsyn.**	**Valine to Methionine**
3	A/G	71,608,978	10	A	G	Syn	
4	C/T	71,608,592	11	T	C	Syn	rs22145577
5	**A/G**	**71,605,654**	**14**	**A**	**G**	**Nonsyn**	**Threonine to Alanine**
6	C/T	71,605,625	14	C	T	Syn	
7	C/T	71,600,884	16	T	C	Syn	
8	**T/A**	**71,597,104**	**20**	**T**	**A**	**Nonsyn**	**Cysteine to Serine**
9	A/G	71,596,770	21	A	G	Syn	
10	**T/C**	**71,596,732**	**21**	**T**	**C**	**Nonsyn**	**rs22163584, Valine to Alanine**
11	**A/G**	**71,596,706**	**21**	**G**	**A**	**Nonsyn**	**Valine to Isoleucine**
C. Col27A1 gene.							
1	**GCG repeat**	**In Gap**	**1**	**4+5**	**4+8**	**Nonsyn**	**The normal allele is 9-interrupted Alanine, and the risk allele is 12.**
2	C/T	71,454,607	3	C	T	Syn	rs22165112
3	C/T	71,455,246	3	T	C	Syn	
4	**C/T**	**71,455,763**	**3**	**T**	**C**	**Nonsyn**	**rs22124807, Serine to Proline**
5	**A/G**	**71,455,803**	**3**	**A**	**G**	**Nonsyn**	**rs22124806, Histidine to Argenine**
6	C/G	71,516,225	19	G	C	Syn	
7	**A/G**	**71,516,229**	**19**	**A**	**G**	**Nonsyn**	**Isoleucine to Valine**
8	**C/T**	**71,540,880**	**32**	**C**	**T**	**Nonsyn**	**Proline to Leucine**
9	C/T	71,545,574	36	C	T	Syn	
10	**C/G**	**71,567,725**	**55**	**G**	**C**	**Nonsyn**	**rs22127846, Glycine to Argenine**
11	A/C	71,567,736	55	C	A	Syn	
12	C/T	71,576,710	61	T	C	Syn	rs22124152
D. ZNF618 (exon 3 to last exon)							
1	A/G	71,302,910	3	A	G	Syn	rs22165000
2	C/T	71,316,759	5	T	C	Syn	rs22117863
3	A/G	71,330,124	10	A	G	Syn	

Exonic non-synonymous SNPs were further investigated and the set of 41 dogs were genotyped for the SNP A/G in exon 3 of the AKNA gene, at position 71,624,795. We also identified a 21 bp deletion in intron 4 of that gene (position 71,618,401-421) and included it in our screening. The two SNPs were in a complete linkage in this set of 41 dogs, and segregated with the disease. Expanding the screening to a large set of Italian greyhound dogs, as well as a control group of 269 dogs from over 36 different breeds (Supplement Table 3.8) showed that one copy of the 21 bp deletion polymorphism is present in 14 dogs not affected with PRA from nine breeds which do not segregate this disease, suggesting that this deletion is not a causative allele. The risk allele "A" on exon 3 was not found in any of the control dogs, and the genotype AA was never observed in IG dogs with a normal phenotype at the risk age or older. The risk allele changes a Valine to Methionine, although the human and mouse counterparts carry the A allele and a methionine at the protein level.

Col27A1 5' end

The Col27A1 5' end is missing from CanFam3 and all previous assemblies. This gap includes the 5'UTR as well as exon 1 and part of intron 1. Amplifying across the gap with primers designed based on the sequence available failed to give any products. After retrieving some of the missing sequences by DNA-Seq analysis and RNA-seq analysis (see results in these sections to follow) we were able to amplify across the genomic gap and identified exon 1 of Col27A1 in affected, carrier and normal dogs. The affected allele has a 9 nucleotides insertion compared to the database (GCGGCGGCG, Figure 3.6). The insertion, in frame with the ORF, adds 3 alanine molecules to an alanine repeat, expanding the repeat from 9 to 12 amino acids. Although this expands the alanine repeat from 9 to 12 and compared to other species (6, 2, and 9 alanine amino acids in human, mouse and cat, respectively) seems high, it does not segregate with the disease,

and other dogs from other breeds also showed 12 repeats of alanine in this position (Beagle and Border Collie, data not shown).

PCR of conserved regions upstream to Col27A1 found no SNPs or potential mutations.

Figure 3.6. Canine Col27A1 exon 1 coding sequence in a normal Boxer dog and in an affected IG, its translation to protein, and the comparison to human, cat, mouse and rat. N=number of interrupted amino acid alanine. In red is the addition 9 bases (GCGGCGGCG) resulting in an addition of three alanines.

Boxer- exon 1 (CD) (**n=9**):

```
ATG GGA GCG GGA GCG GCG CGG GGG GCC CGA GGC ACG GCG GCG GCG GCG GAG
  M   G   A   G   A   A   R   G   A   R   G   T   A   A   A   A   E
GCG GCG GCG GCG GCG CGC GGG GG
  A   A   A   A   A   R   G
```

An affected IG- exon 1 (CD) (**n=12**):

```
ATG GGA GCG GGA GCG GCG CGG GGG GCC CGA GGC ACG GCG GCG GCG GCG GAG
  M   G   A   G   A   A   R   G   A   R   G   T   A   A   A   A   E
GCG GCG GCG GCG GCG GCG GCG GCG CGC GGG GG
  A   A   A   A   A   A   A   A   R   G
```

Human first exon (**n=6**):

```
ATG GGA GCG GGA TCG GCG CGG GGG GCC CGA GGC ACA GCG GCG GCG GCG GCG
  M   G   A   G   S   A   R   G   A   R   G   T   A   A   A   A   A
GCG CGC GGG GG
   A   R   G
```

Cat (**n=9**):

```
ATG GGA GCG GGA TCG GCG CGG GGG GCC CGA GGC ACA GCG GCG GCG GCG GCG
  M   G   A   G   S   A   R   G   A   R   G   T   A   A   A   A   A
GCG GCG GCG GCG CGC GGG GG
  A   A   A   A   R   G
```

Mouse (**n=<2?**):

```
ATG GGC ACG GGA TTC GCG CGG GGG GCC CGA GGC ACA GCG GCG TCA GGA CCC
  M   G   T   G   F   A   R   G   A   R   G   T   A   A   S   G   P
GGG GG
  G
```

Rat (**n=<2?**):

```
ATG GGC CTG GCG CGG GCG ACC GCG GGG CTG GGG CCG TGC TGT CCG CCT GCT
  M   G   L   A   R   A   T   A   G   L   G   P   C   C   P   P   A
CCG GCG CTC CTG GGC GCA GGG CTG CGC TGG GG
  P   A   L   L   G   A   G   L   R   W
```

Col27A1 allelic expression bias

In sequencing retinal cDNA from an IGPRA-heterozygous dog (IG72, 9 weeks old) three SNPs in the Col27A1 coding sequence were found. Inspection of chromatogram peaks showed that the risk allele of each SNP is underexpressed (exon 29, risk allele G- position 71,539,783; exon 32, risk allele T- position 71,540,880; exon 61, risk allele C- position 71,576,758, CanFam2). The same phenomenon was also observed for IGPRA-heterozygous dogs IG73 and IG74 (Figure 3.7-A). Sequence of genomic DNA from the latter two dogs confirmed that they are heterozygous for these three SNP alleles (Figure 3.7-A). The allelic difference in expression of Col27A1- retinal-mRNA as observed by the difference in the height of the peaks in the chromatograms was not observed for any of the neighboring genes in these dogs (data not shown).

To confirm the bias in the risk allele expression on exon 61, we ran RT-PCR on retinal cDNA from IG70, IG72, IG73, and IG74, using different primer pair combinations (Supplement Table 3.1,D-I, primer pairs 8, 9, 16,17). The same difference in the peak of the risk allele, where the C (risk) peak was always much lower than the T (wildtype) peak, was identified in all PCR products amplified from retinal RNA of obligate heterozygote dogs. This bias in allele expression was not observed in PCR products from a retinal cDNA extracted from an age-matched normal dog, which tended to exclude imprinting of the gene (Figure 3.7-B). Moreover, we extracted RNA from other tissues known to express Col27A1 (kidney, cerebellum, ear cartilage and knee cartilage) from two animals: IG70 and a normal dog. The bias in the risk allele expression was retina-specific (Figure 3.7-C): in the kidney, cerebellum, ear cartilage and knee cartilage of IG70 dog both alleles were expressed roughly in equal amount as observed from the height of the peaks, with both forward and reverse sequencing reactions.

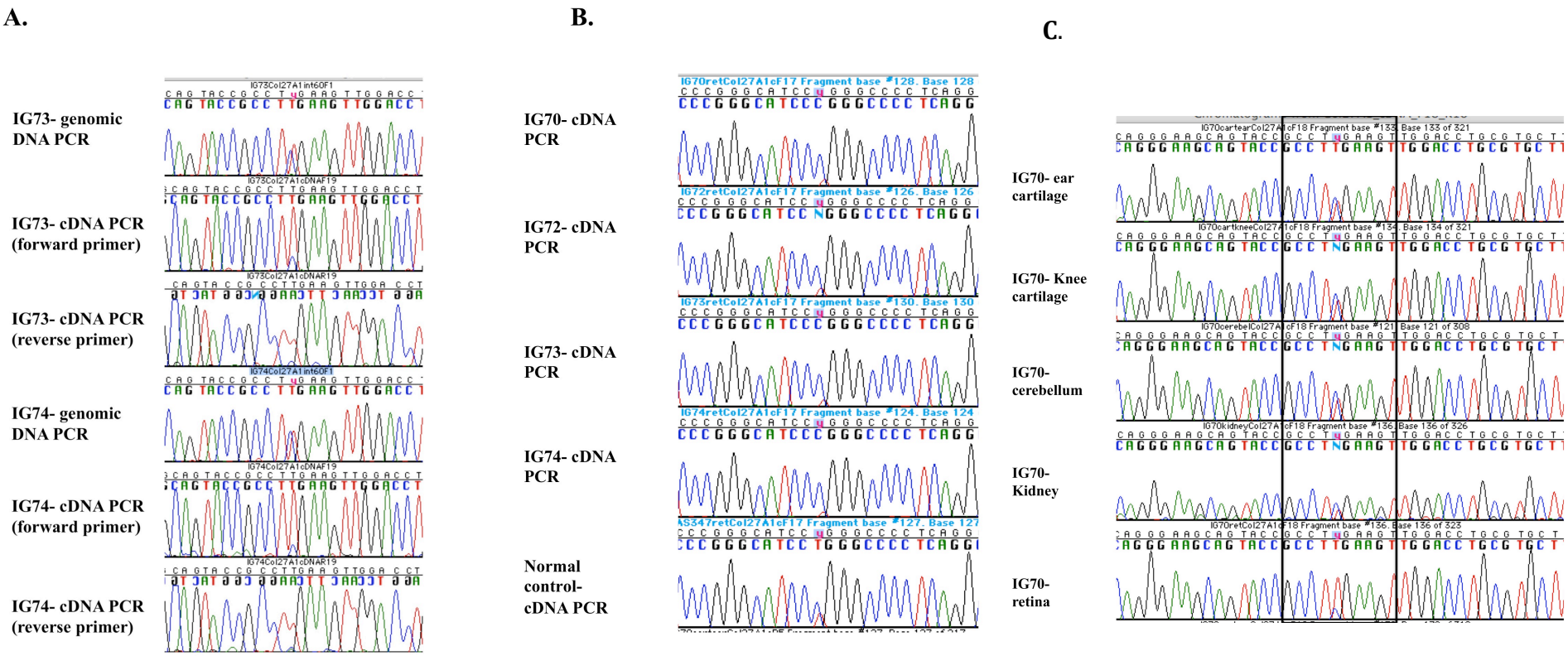


Figure 3.7. Chromatograms of genomic and cDNA PCR products of Col27A1 amplicons. Level of expression is roughly measured by the height of the peaks in genomic and cDNA PCR products. **A**. IG73 and IG74 genomic and retinal cDNA PCR chromatogram of a SNP in exon 61. The risk allele is cytosine (C) **B**. Retinal cDNA PCR chromatograms of IG70, 72, 73 and 74- all heterozygous to the risk allele tyrosine (T) in exon 32, and a normal retina. **C**. IG70 cDNA PCR chromatogram of brain, kidney, cartilage from ear and knee, and retina of exon 61 SNP. The risk allele is cytosine in exon 61 (C). The bias in expression is retina specific: only in the retina is the C allele under represented, with a much lower peak.

To quantify and confirm these observations, we researched this on two levels: (i) we looked at the reads of all the genes within the LD interval from RNA-Seq NGS experiments and compared the number of reads of each allele in the heterozygous IGPRA dogs, to those of the control group; and (ii) we ran allele-specific pyrosyquencing measurements for Col27A1 using the same dogs and an additional three control dogs (see results in the following sections).

9.4.4 Next Generation Sequencings (NGS)

Three genomic DNA and six retinal RNA samples were converted to an Illumina sequencing library and sequenced. The overall estimated error rate was less than 1% in the DNA and mRNA experiments, but higher than 1% in the smallRNA (Table 3.3).

Table 3.3 Statistics on Illumina multiplex NGS experiments. **A**. Genomic DNA samples, three dogs multiplexed in one lane, run in duplicate, 100 bp paired-end reads. **B**. Retinal mRNA samples- six dogs multiplexed in one lane, 100 bp- single-end reads. **C**. Retinal SmallRNA samples- six dogs multiplexed in one lane, 50 bp- single-end reads.

Lane	Dog ID	Bar Code	Number of reads	% PF *	% Perfect Index Reads	% of >=Q30 Bases (PF)	Mean Quality Score (PF)
A. Genomic DNA samples.							
1	IGPRA affected	AGTCAA	168,012,398	90.25	79.3	88.32	34.87
1	IG normal	AGTTCC	117,448,060	90.9	79.27	89.1	35.13
1	IG heterozygous (IG73)	ATGTCA	114,738,530	90.89	75.76	88.16	34.83
Estimated error rates: 0.21% and 0.55%.							
2	IGPRA affected	AGTCAA	174,219,396	89.55	95.39	88.21	34.84
2	IG normal	AGTTCC	121,639,254	90.24	95.34	89.04	35.1
2	IGPRA heterozygous (IG73)	ATGTCA	119,253,536	90.21	94.67	87.97	34.77
Estimated error rates: 0.19% and 0.52%.							
B. Retinal mRNA samples.							
8	IGPRA Het 1 (IG72)	GTGGCC	43,163,139	88.97	97.67	90.24	35.32
8	IGPRA Het 2 (IG73)	GTTTCG	35,122,816	87.91	97.75	90.52	35.45
8	IGPRA Het 3 (IG74)	CGTACG	27,771,266	89.15	96.52	90.78	35.53
8	Control 1	GAGTGG	38,265,540	89.76	96.28	91.17	35.66
8	Control 2	ACTGAT	32,887,071	89.87	97.97	90.83	35.52
8	Control 3	ATTCCT	39,633,028	89.16	97.77	90.92	35.56
Estimated error rates: 0.54%							
C. Retinal Small RNA samples.							
2	IGPRA Het 1 (IG72)	ATCACG	22,893,484	77.97	94.97	88.94	35.86
2	IGPRA Het 2 (IG73)	CGATGT	15,523,954	78.25	95.73	88.99	35.86
2	IGPRA Het 3 (IG74)	TTAGGC	18,786,152	78.32	95.57	89.3	36.03
2	Control 1	ACAGTG	20,392,073	78.14	95.41	89.16	35.93
2	Control 2	GCCAAT	18,976,310	77.7	94.16	88.75	35.78
2	Control 3	CAGATC	23,139,376	77.86	95.35	89.05	35.89
Estimated error rates: 1.2%							

* PF= Pass filter.

9.4.4.1 DNA whole-genome-sequencing

DNA from one affected IG (Blue square, Figure 3.1), one normal IG dog, and one F1 colony dog (IG73) were barcoded, pooled together, and run twice (see methods). The affected dog had a combined number of 342,231,794 reads; the normal dog 239,087,314 reads; and the heterozygous colony dog 233,992,066 reads (Table 3.3). The average estimated error rate for the first read was 0.2% and 0.535% for the second read. The percentage of reads that passed filter was between 89.55 and 90.9. We identified 1,427 SNPs within the LD interval (68,316,896-68,603,033 on CanFam 3.0, data not shown). 356 SNPs were identified where the affected dog is homozygous for the alternate allele, the normal dog is homozygous for the reference allele or the SNP reads do not contradict this criterion (Supplement Table 3.9A). Those SNPs have the potential to be a causative allele for IGPRA. If the risk allele was observed in the normal dog but only in one read, we included it as well. We also identified 28 indels and repeats that followed the above criterion: 3 microsatellites, 5 insertions or deletions of Cytosine in a stretch of Cs, 14 indels, and 6 insertions or deletions of 2 or more bases (Supplement Table 3.9B).

Col27A1 5' end gap

Search for DNA- Seq reads that align to the sequence proximal and distal to the 5' end gap of Col27A1 extended the 5' end of the genomic DNA gap by 57 bp, and the 3' end by 171 bp (Supplement Figure 3.1). PCR amplification with GC-rich enzyme and primers designed on the newly retrieved sequences of a boxer and a colony-derived affected dog was successful and alignment of the two sequences to each other revealed several polymorphisms (Supplement Figure 3.2).

3.4.4.2 RNA Next-generation-sequencing

The numbers of reads were between 27,771,266 reads and 43,163,139 (Table 3.3), and the estimated error rate was 0.54%. About 90% of the reads mapped uniquely to the genome. Within the LD interval, no reads were found for AMBP and KIF12, a result that agrees with the RT-PCR experiment. Although we did not detect any PCR products for ORM1 on an agarose gel, a small number of reads were present that mapped to this gene, which might suggest that the gene is expressed in the retina but in an amount not detectable by gel electrophoresis.

Relative allelic expression analysis:

We used the mRNA data to estimate the level of expression of each allele for any of the genes within the LD interval in animals that were informative, i.e. had both alleles in a specific exonic SNP. The overall expression of ZNF618 and ORM1 was very low compared to Col27A1, AKNA and DFNB31 genes: the average number of reads per SNP was 51.57, 41.71, for ZNF618 and ORM1 respectively, and 276.69, 1266.65 and 924 for Col27A1, AKNA, and DFNB31, respectively (Table 3.4). We asked two questions: (i) within each of the heterozygous IG dogs, was there an allelic expression bias of Col27A1, and if so, was it restricted to Col27A1, or was it shared by all genes in the LD interval. (ii) For each of the genes in the LD interval, were there significant differences in allelic expression bias between the two groups: the heterozygous IG group and the control group. To answer these questions we calculated the risk allele expression ratio for all the informative SNPs in the three IG dogs, and the minor allele ratio in the control dogs (Table 3.4). For the absolute numbers - see Supplement Table 3.10). We looked at the average of these ratios to determine bi-allelic or mono-allelic expression. We then calculated the deviation from 0.5, the ratio that we expected if there was no allelic expression bias, and ran a t-test to check for significant differences between the two groups in question.

In all genes except for Col27A1 the allelic expression bias was not significantly different in the IG heterozygous dogs from that in control dogs (Table 3.5), and all showed bi-allelic expression (average risk allele expression ratio between 0.45 and 0.64 in IG group and average minor allele expression ratio in the control group between 0.38 and 0.5 (Table 3.4). Col27A1 allelic expression bias was significantly different in the IG heterozygous dogs compared to the control, with a P- value of 0.0006 (Table 3.5). The risk allele was always under-represented with an average risk allele expression ratio of 0.21, 0.27 and 0.18 for IG72, IG73 and IG74 respectively (Table 3.4).

Table 3.4. Relative allele expression in genes within the LD interval in informative dogs as derived from RNA-seq reads data.

SNP_ID	CanFam3 position	Reference allele	Alter allele	Total counts			CanFam2 position	Risk allele expression ratio (#of risk allele/total # of reads)			Minor allele expression ratio (# of minor allele/# of total reads)			Risk allele
				Coverage	Reference count	Alternate count		IG72	IG73	IG74	Control 1	Control 2	Control 3	
ZNF618														
rs22165000	68,272,447	A	G	48	4	44	71302910	*	*	*	0.33	*	*	G
rs22164999	68,272,466	A	G	39	32	7	71302929	*	*	0.75	*	*	*	A
chr11_68299661	68,299,661	A	G	41	11	30	71330124	*	*	0.5	*	*	*	G
chr11_68332484	68,332,484	G	A	77	66	10	71362947	*	*	0.67	*	0.41	*	G
chr11_68332860	68,332,860	C	T	47	38	9	71363323	*	*	*	*	*	*	C
rs8968699	68,333,123	C	T	47	34	13	71363586	*	*	*	0.43	*	*	C
rs8968700	68,333,381	T	C	62	4	58	71363844	*	*	*	0.43	*	*	C
			Average:	**51.57**			**Average:**			**0.64**	**0.40**	**0.41**		
ORM1														
rs9188097	68,557,019	T	C	37	9	28	71587323	*	*	*	0.25	0.5	*	C
chr11_68557358	68,557,358	G	A	40	27	13	71587662	0.73	0.5	0.33	0.5	*	*	A
chr11_68557816	68,557,816	G	A	54	40	14	71588120	*	*	*	*	*	*	G
chr11_68558082	68,558,082	C	T	31	24	7	71588386	0.25	0.4	0.75	*	*	*	C
chr11_68558749	68,558,749	A	G	35	13	22	71589053	*	*	*	*	*	*	G
chr11_68559346	68,559,346	C	T	46	19	27	71589650	*	*	*	*	*	*	T
chr11_68559366	68,559,366	G	A	49	19	30	71589670	*	*	*	*	*	*	A
			Average:	**41.71**			**Average:**	**0.49**	**0.45**	**0.54**	**0.38**	**0.50**		
Col27A1														
chr11_68425869	68,425,869	A	G	94	33	61	71456098	*	0.27	*	*	*	*	A
chr11_68509555	68,509,555	G	A	97	64	33	71539783	0.18	*	0.14	*	0.43	*	G
chr11_68510591	68,510,591	C	T	76	53	23	71540880	0.08	0.1	0.09	*	*	*	T
chr11_68525287	68,525,287	C	T	-	-	-	71555591	*	*	*	*	0.34	*	C
chr11_68544801	68,544,801	A	C	-	-	-	71575105	*	*	*	0.41	*	*	A
rs22124152	68,546,406	T	C	454	161	293	71576710	*	*	*	*	*	*	C
chr11_68546454	68,546,454	C	T	459	269	190	71576758	0.22	0.26	0.15	*	0.29	*	C
chr11_68546511	68,546,511	C	G	290	97	193	71576815	*	*	*	*	*	*	G
chr11_68546704	68,546,704	A	C	146	45	99	71577008	*	*	*	*	*	*	C
chr11_68546777	68,546,777	A	G	439	140	299	71577081	*	*	*	*	*	*	G
chr11_68547035	68,547,035	G	A	195	94	101	71577339	0.3	0.33	0.33	0.35	*	*	G
rs22139258	68,547,670	G	A	340	207	133	71577974	0.29	0.34	0.25	*	0.35	*	G
chr11_68547749	68,547,749	T	C	182	145	37	71578053	0.18	0.3	0.12	0.36	*	*	C
			Average:	**252.00**			**Average:**	**0.21**	**0.27**	**0.18**	**0.37**	**0.35**		
AKNA														
chr11_68561787	68,561,787	C	T	1626	1416	208	71592091	*	*	*	*	0.47	*	C
rs22069543	68,562,441	G	A	1265	254	1008	71592745	*	*	*	0.46	0.45	*	A
chr11_68569222	68,569,222	C	T	1755	1603	150	71599526	*	*	*	*	0.5	*	C
chr11_68570580	68,570,580	A	G	821	128	692	71600884	*	*	*	*	*	*	G
chr11_68575321	68,575,321	G	A	1628	310	1318	71605625	0.55	0.54	*	*	*	*	A

chr11_68577401	68,577,401	**G**	**A**	1301	979	321	71607705	0.45	0.47	*	*	*	*	**G**
chr11_68578296	68,578,296	**C**	**T**	1238	932	300	71608600	0.46	0.5	*	*	*	*	**C**
chr11_68578674	68,578,674	**T**	**C**	930	330	600	71608978	0.53	0.49	*	*	*	*	**C**
chr11_68580620	68,580,620	**G**	**A**	931	810	120	71610924	*	*	*	*	*	*	**G**
chr11_68591406	68,591,406	**A**	**G**	1297	590	706	71621710	0.45	0.43	*	*	*	*	**A**
rs22122505	68,593,585	**G**	**C**	1089	931	158	71623889	*	*	*	*	*	*	**G**
rs22122506	68,593,599	**C**	**T**	948	815	131	71623903	*	*	*	*	*	*	**C**
chr11_68593630	68,593,630	**G**	**A**	1116	781	334	71623934	0.49	0.5	*	*	0.48	*	**G**
rs22122509	68,593,713	**C**	**A**	874	744	128	71624017	*	*	*	*	*	*	**C**
chr11_68594017	68,594,017	**C**	**T**	1327	952	374	71624321	0.53	0.53	*	*	0.5	*	**C**
rs22122510	68,594,098	**C**	**G**	1398	1170	225	71624402	*	*	*	*	*	*	**C**
chr11_68594225	68,594,225	**A**	**G**	1546	980	566	71624529	0.46	0.53	*	*	*	*	**A**
rs22122511	68,594,462	**G**	**A**	1514	1077	431	71624766	0.52	0.46	*	*	0.47	*	**G**
chr11_68594491	68,594,491	**C**	**T**	1578	1151	426	71624795	0.48	0.47	0.47	*	*	*	**T**
			Average:	**1272.7**			**Average:**	**0.49**	**0.49**	**0.47**	**0.46**	**0.48**		
DFNB31														
chr11_68617832	68,617,832	**T**	**C**	1016	605	410	71647797	0.57	0.47	0.5	*	*	*	**C**
chr11_68617936	68,617,936	**T**	**C**	1585	899	684	71647901	0.53	0.57	0.51	*	*	*	**C**
chr11_68618606	68,618,606	**A**	**G**	1032	896	136	71648571	*	*	*	*	*	*	**A**
chr11_68619188	68,619,188	**A**	**T**	1063	926	130	71649153	*	*	*	*	*	*	**A**
chr11_68621353	68,621,353	**G**	**C**	481	283	198	71651318	0.52	0.55	0.5	*	0.48	*	**C**
chr11_68635719	68,635,719	**T**	**C**	553	403	148	71665684	0.68	0.56	0.63	*	0.38	*	**T**
chr11_68671292	68,671,292	**C**	**T**	738	505	233	71701257	0.6	0.52	*	*	*	*	**C**
			Average:	**924**			**Average:**	**0.58**	**0.53**	**0.54**		**0.43**		

* Dog is not informative for that SNP

Table 3.5. T-test results for significant differences between allelic expression bias between the heterozygous IG dogs and the control dogs for each gene.

Gene	**ZNF618**		**ORM1**		**Col27A1**		**AKNA**		**DFNB31**	
Statistics/groups	IG	Control	IG	Control	IG	Control	IG	Control	IG	Control
Mean	0.14	0.1	0.1667	0.0833	0.2817	0.1386	0.031	0.0243	0.055	0.07
SD[a]	0.1277	0.0476	0.1001	0.1443	0.0897	0.0463	0.0173	0.019	0.0523	0.0707
SEM[b]	0.0737	0.0238	0.0409	0.0833	0.0211	0.0175	0.0038	0.0072	0.014	0.05
N[c]	3	4	6	3	18	7	21	7	14	2
P-value[d]	0.5809 not significant		0.3377 not significant		**0.0006 significant**		0.3962 not significant		0.7179 not significant	

a- Standard Deviation
b- Standard Error of the Mean
c- Number of data points
d- Two tailed P-value unpaired t-test.

Relative abundance of transcripts:

The RNA-Seq fragment counts can be used as a measure of relative abundance of transcript. Expression analysis of all RNA-seq data using Cufflinks generated normalized Fragments Per Kilobase of exon per Million fragments mapped (FPKM). COL27A1 expression in control individuals was 1.5-fold higher than the F1 heterozygous dogs on average, suggesting potential reduction of total expression in affected samples as well. This ratio of 1.5 is very close to the value expected if the risk allele expression is reduced approximately 10-20%. The average control/IG-het FPKM ratio for the other genes in the LD interval was 0.74, 0.88 and 0.71 for ZNF618, AKNA, and DFNB31, respectively, which suggests no difference in total expression of these genes between the two groups. Cuffdiff analysis identified 106 genes with significant differences in level of expression between the two groups (Supplement Table 3.11). The top score of p=5.00E-05 was shared by 59 genes (Table 3.6). Outside the LD region, the top five genes with total expression reduction in IG-het samples were: ANKS4B, DPY19L2, RASL11A, CRYM, and GADD45G with average control/IG-het FPKM ratio of 7.89, 7.07, 6.32, 5.23 and 4.3, respectively (log2(fold_change)= -2.98, -2.82, -2.66, -2.38, -2.1) . The top five genes with higher expression in IG-het samples compared to controls were NOTCH4, TNNI2, MYLPF, CAPN6 and MKRN2-AS1 with average control/IG-het FPKM ratio of 5.67, 5.62, 4.76, 4.59, and 4.22, respectively (log2(fold_change)= 2.5, 2.49, 2.25, 2.2, 2.08), (Table 3.6)

Table 3.6. Cuffdiff analysis of RNA-seq reads between control dogs and IG heterozygous dogs. The list includes the genes with the most significant difference in expression between the two groups (p-value-uncorrected for multiple testing=5.00E-05, q-value corrected=0.0110218). In bold is the higher number of the two.

Number	Gene_id	Gene	Locus	Value_1[a]	Value_2[b]	Log2(fold_change)	Test_stat[c]
1	ENSCAFG00000017810	ANKS4B	chr6:23657103-24217826	**14.7401**	1.86649	-2.98135	-2.85372
2	ENSCAFG00000003176	DPY19L2	chr14:46787464-46881015	**2.65667**	0.375765	-2.82172	-4.90261
3	ENSCAFG00000006798	RASL11A	chr25:12315962-12318638	**22.3406**	3.53283	-2.66077	-6.23349
4	ENSCAFG00000017807	CRYM	chr6:23657103-24217826	**24.8682**	4.75495	-2.3868	-4.47202
5	ENSCAFG00000002174	GADD45G	chr1:96744757-96746752	**11.1038**	2.58121	-2.10494	-5.14077
6	ENSCAFG00000031664	KCNT2	chr38:3305568-3382572	**3.37359**	0.85451	-1.98111	-4.06635
7	ENSCAFG00000019134	PPL	chr6:36497461-36541298	**2.28724**	0.581794	-1.97503	-4.14373
8	ENSCAFG00000006714	NLRP14	chr21:30722678-30756070	**1.83171**	0.501719	-1.86824	-3.20277
9	ENSCAFG00000023684	-	chrX:121060226-121074003	**1.9279**	0.547474	-1.81617	-3.19666
10	ENSCAFG00000000492	-	chr12:1013177-1021690	**3.55046**	1.02932	-1.78632	-2.98578
11	ENSCAFG00000013763	Q2LC20_CANFA	chr20:42231834-42263250	**3.97622**	1.17822	-1.75479	-3.77628
12	ENSCAFG00000025465	TRPV6	chr16:6691520-6705704	**2.92284**	0.893114	-1.71045	-3.26907
13	ENSCAFG00000000538	VIP_CANFA	chr1:42941848-42950730	**4.86565**	1.5423	-1.65754	-3.45848
14	ENSCAFG00000029483	C17orf67	chr9:31635471-31654119	**62.1904**	21.1012	-1.55937	-6.38461
15	ENSCAFG00000007199	-	chr27:2221688-2229112	**11.5803**	3.96361	-1.54679	-4.01687
16	ENSCAFG00000004404	-	chr19:22979380-23044911	**1.58121**	0.551966	-1.51838	-3.09268
17	ENSCAFG00000007375	PARP4	chr25:18447412-18538508	**5.03721**	1.7896	-1.49299	-3.78555
18	ENSCAFG00000028905	CDKN2C	chr15:10218823-10225053	**55.7227**	24.1923	-1.20372	-3.82904
19	ENSCAFG00000004130	NEBL	chr2:12126777-12229218	**17.9102**	8.15899	-1.13432	-3.38298
20	ENSCAFG00000007539	ADML_CANFA	chr21:33472109-33474520	**12.7947**	6.43024	-0.992601	-2.78398
21	ENSCAFG00000032483	LY96	chr29:22493775-22515117	**34.2359**	17.6329	-0.95724	-3.38232
22	ENSCAFG00000017664	SLC14A1	chr7:45378306-45408131	**15.0399**	7.84565	-0.938832	-2.81636
23	ENSCAFG00000012904	CDK1	chr4:12885999-12898957	**70.5831**	38.1764	-0.886641	-3.30066
24	ENSCAFG00000020110	Q4W6L5_CANFA	chr6:54709521-54720797	**44.7585**	25.2714	-0.824653	-2.76379
25	ENSCAFG00000015239	C1orf114	chr7:29141745-29173572	**83.2129**	47.3228	-0.814271	-2.85471
26	ENSCAFG00000023759	MT2_CANFA	chr2:59607925-59608825	**179.274**	102.14	-0.811619	-4.39619
27	ENSCAFG00000008873	-	chr26:9970455-9975820	**255.426**	149.951	-0.768418	-3.006
28	ENSCAFG00000006193	-	chr16:27598816-27614612	**39.8399**	23.6591	-0.751817	-2.52272
29	ENSCAFG00000018849	SNX29	chr6:30317097-30822479	14.9604	**24.6087**	0.718015	2.32246
30	ENSCAFG00000025128	TRAC	chr8:2943108-2952167	130.913	**222.552**	0.765537	3.01681

31	ENSCAFG00000009901	-	chr38:1474699-1542627	14.7048	**25.1243**	0.772789	2.40359
32	ENSCAFG00000011477	BSN	chr20:39612348-39628313	4.16641	**7.27205**	0.803557	2.3408
33	ENSCAFG00000031504	UBAP1L	chr30:29475497-29489123	81.1432	**142.778**	0.815232	3.12021
34	ENSCAFG00000016397	VPS13D	chr2:83901305-84132453	24.8109	**45.0315**	0.85996	2.75266
35	ENSCAFG00000004038	HIPK2	chr16:9090726-9244272	28.3184	**51.801**	0.871238	2.79822
36	ENSCAFG00000004375	NUP210	chr20:3636670-3722019	3.68209	**6.74686**	0.873691	2.48661
37	ENSCAFG00000020254	NFAT5	chr5:80003597-80115099	4.54118	**8.3556**	0.879677	2.42235
38	ENSCAFG00000006985	Q2EFX7_CANFA	chr26:5315659-5476358	5.99268	**11.0752**	0.886055	2.48337
39	ENSCAFG00000018418	RHBDL3	chr9:40633359-40660779	25.6011	**49.5661**	0.953149	3.2839
40	ENSCAFG00000005194	PFKFB3	chr2:29732444-29748084	58.6269	**114.273**	0.962848	3.47457
41	ENSCAFG00000013739	TET1	chr4:19768258-19889904	2.27243	**4.57145**	1.00842	2.57845
42	ENSCAFG00000016302	VWCE	chr18:55104014-55125461	3.81192	**7.89682**	1.05076	2.79413
43	ENSCAFG00000015078	FAM184B	chr3:62753439-62887350	2.38213	**5.0254**	1.07698	2.91737
44	ENSCAFG00000014201	DNAAF2	chr8:26251159-26295095	4.63608	**9.78829**	1.07815	2.76583
45	ENSCAFG00000016786	DNAH2	chr5:32600483-32659183	0.740779	**1.63042**	1.13813	2.6472
46	ENSCAFG00000016065	-	chrX:44508996-44509863	49.836	**121.907**	1.29052	4.55046
47	ENSCAFG00000029478	-	chr9:10287951-10304725	5.76627	**15.1094**	1.38973	4.20572
48	ENSCAFG00000008253	ACTA1	chr4:9812781-9815574	2.35588	**6.48882**	1.46169	3.23504
49	ENSCAFG00000000296	GPR126	chr1:33969265-34107014	0.757815	**2.38019**	1.65116	3.86754
50	ENSCAFG00000011417	RPL4	chr27:21855204-21856473	4.82751	**17.8928**	1.89003	4.31819
51	ENSCAFG00000031904	MKRN2-AS1	chr20:6027419-6034350	0.618453	**2.60856**	2.07652	3.50834
52	ENSCAFG00000018175	CAPN6	chrX:84351484-84377324	0.251976	**1.15798**	2.20025	3.34512
53	ENSCAFG00000016543	MYLPF	chr6:17719682-17721750	3.0976	**14.7401**	2.25052	5.48009
54	ENSCAFG00000010068	TNNI2	chr18:46077957-46079538	0.688972	**3.87324**	2.49103	4.04836
55	ENSCAFG00000000791	NOTCH4	chr12:1586323-1608575	1.2213	**6.92773**	2.50397	5.79464
56	ENSCAFG00000023827	CRBB2_CANFA	chr26:19298946-19308447	**1.20963**	0	-	nan
57	ENSCAFG00000001705	MB	chr10:28574067-28583590	0	**1.04489**	inf	nan
58	ENSCAFG00000008062	-	chr8:8320212-8323009	0	**17.9946**	inf	nan
59	ENSCAFG00000028985	-	chr1:111083447-111083951	0	**5.22049**	inf	nan

a- FPKM of the gene in the control dog group
b- FPKM of the gene in the IG heterozygous dog group
c- The value of the test statistic used to compute significance of the observed change in FPKM

Col27A1 5' end gap

When screening Col27A1, we identified all predicted exons except exon 1. After comparing the dog genome to human and mice, we concluded that canine exon 1 was missing from the assembly, and mapped to an approximately 1.5 Kb gap. PCR amplification with primers flanking the gap failed. We then used the mRNA-seq data to try to obtain exon 1 sequence by looking for reads overlapping exon 2, and homologous to human, mouse, and cat exon 1. Fishing exon 1 with exon 2 yielded no reads. Homology analysis found reads homologous to human, mouse and cat exon 1, and alignment of those reads retrieved 127 bp from Col27A1: 87 bp from the 5' UTR and the first 40 bp of the coding sequence (Supplement Figure 3.1)

9.4.4.3 SmallRNA Next-generation-sequencing

To evaluate the total level of microRNA expression in IGPRA affected dogs and normal control dogs, we ran SmallRNA next generation sequencing experiment in three IGPRA-heterozygous dogs and three control dogs. There were between 15,523,954 and 23,139,376 reads (Table 3.3). Although the percentages that passed filter were lower than in the DNA and retinal mRNA experiments, for those reads that passed filter, the percentage of bases receiving or exceeding a minimal quality score of 30 was high. Cuffdiff analysis showed 12 SmallRNA genes that significantly differed in level of expression between the IG-heterozygous dogs and control dogs (Table 3.7A). Of those, mir3609 expression was the most significant with a 6.66 IG-het/control FPKM ratio. SNORD12 and SNORD15b were also significantly different with ratios of 4.208, and 6.462, respectively. Mir455 showed 2.16 ratio of expression between controls and IG-heterozygous dogs, which correlates with almost mono-allelic expression of the mir gene in the IG-het dogs. When comparing this to other mir genes (mir96, mir1, mir133, mir142, mir365-a,

and mir184) located on different chromosome, ratios between the higher and lower expressions were between 1.1 and 1.8 suggesting no difference in level of expression in these mir genes (Table 3.7B).

Table 3.7. SmallRNA genes with significant fold change between IG heterozygous dogs and control dogs,and several mir genes. **A**. SmallRNA-seq reads with significan expression change. **B**. Results for six mir genes. In bold is the higher number of the two.

#	Gene_id	Gene	Locus	Location (CanFam3.0)	value_1[a]	value_2[b]	log2 (fold_change)	test_stat[c]	p_value[d]	q_value[e]	Signi fican t	Comment
A. Significant SmallRNA												
1	XLOC_002016	mir3609	chr6	10632457-10632481	706450	**4.70E+06**	2.7346	-4.90613	9.29E-07	0.00251641	yes	
2	XLOC_002521	no good match in human-maybe intron in BC048201	chrUn_AAEX03025000	1106-1136	**388142**	14184.5	-4.77419	4.66165	3.14E-06	0.00330688	yes	
3	XLOC_002533	3' UTR of LSAMP	chrUn_AAEX03025644	1881-1909	**736876**	38808.8	-4.24697	4.62969	3.66E-06	0.00330688	yes	
4	XLOC_000383	no good match in human-maybe intron in chrM JA760600/602	chr13	24498136-24498180	**87752.6**	10127.9	-3.11511	4.20764	2.58E-05	0.0149019	yes	
5	XLOC_001124	SNORD12	chr24	35959255-35959338	1732.88	**7292.91**	2.07332	-4.17577	2.97E-05	0.0149019	yes	
6	XLOC_001533	intron in SLC39A8	chr32	23782524-23782547	787580	**4.28E+06**	2.44362	-4.15167	3.30E-05	0.0149019	yes	
7	XLOC_000859	-	chr2	78054717-78054744	**1.19E+06**	271100	-2.13643	4.01856	5.86E-05	0.0226604	yes	Same sequence exists also on CFA35, CFA13, and CFA14.
8	XLOC_001949	intron in PDE4B	chr5	44325505-44325527	**1.02E+07**	2.08E+06	-2.2934	3.97385	7.07E-05	0.0239473	yes	
9	XLOC_000970	SNORD15b	chr21	23181792-23181839	33475.1	**216303**	2.6919	-3.91969	8.87E-05	0.0247581	yes	
10	XLOC_001987	not conserved	chr5	78836580-78836605	**999469**	228319	-2.13011	3.91238	9.14E-05	0.0247581	yes	
11	XLOC_002517	RNA45S5-45S pre-ribisomal5	chrUn_AAEX03024183	8316-9719	**347.533**	34.2786	-3.34177	3.84871	0.000118741	0.0287679	yes	
12	XLOC_001550	not conserved	chr33	17014722-17014747	**1.67E+06**	412238	-2.01788	3.83137	0.000127433	0.0287679	yes	Same seqeunce exist on CFA21, and partial seq also on

												other chromosomes.
B. mir455, mirs involved in RP, and randomly chosen mir.												
1	XLOC_000295	**mir455**	chr11	68461904-68461971	**109364**	50585	**-1.11236**	1.72886	0.0838349	0.996532	no	
2	XLOC_000438	mir96	chr14	7068725-7068831	636555	510625	-0.318021	0.598127	0.549755	0.996532	no	
3	XLOC_002189	mir-1	chr7	66214417-66214439	3.39E+06	3.07E+06	-0.139106	0.244712	0.80668	0.996532	no	
4	XLOC_001132	mir-1	chr24	46481115-46481137	3.08E+06	3.85E+06	0.322895	-0.593623	0.552765	0.996532	no	
5	XLOC_002190	mir133	chr7	66217499-66217560	27281.7	34806	0.351404	-0.671885	0.501657	0.996532	no	
6	XLOC_001133	mir133	chr24	46490513-46490574	26609.6	32937.6	0.30779	-0.596623	0.550759	0.996532	no	
7	XLOC_002439	mir142	chr9	32977392-32977455	10500.3	12967.2	0.304441	-0.665962	0.505436	0.996532	no	
8	XLOC_002038	mir365-a	chr6	28873144-28873206	77.1565	138.998	0.849202	-0.335066	0.737575	0.996532	no	
9	XLOC_001400	mir184	chr3	57914448-57914509	847.245	801.287	-0.080461	0.0889373	0.929132	0.996532	no	

a- FPKM of the gene in the control dog group
b- FPKM of the gene in the IG heterozygous dog group
c- The value of the test statistic used to compute significance of the observed change in FPKM
d- The uncorrected p-value of the test statistic
e- The False-discovery-rate-adjusted p-value of the test statistic

3.4.5 Allele-specific pyrosequencing.

To further evaluate the accuracy of the RT-PCR sequencing and RNA-seq results, suggesting almost mono-allelic expression of Col27A1, we executed allele-specific pyrosequencing experiments of four informative SNPs in the gene. We compared the levels of allelic expression between retinas of heterozygous dogs (n=4) and normal dogs (n=7), and between retina and other tissues (kidney, cerebellum, ear cartilage and knee cartilage of one IG heterozygous and one control dog). Primers targeting the SNP on exon 29 did not work and produced low to no PCR products, and PCR on ear cartilage for Col27A1 showed no product. One control dog (control dog number 4) was not informative to any of the SNPs. Control dog number 7 was only informative for the third SNP but its results were not reliable due to positive products in the negative control of the RT. The rest of the results of the allele-specific pyrosequencing experiment support the results we observed in RNA-seq. In the IG heterozygous dogs we observed lower-allelic expression of the Col27A1 risk allele (12-30% expression), while bi-allelic expression was observed in the control dogs (0.3-0.65%), (Table 3.8). Moreover, the bias in expression was retina specific. In heterozygous IG we observed bi-allelic expression in the kidney, cerebellum and ear cartilage, where Col27A1 is expressed (0.5-0.7, Table 3.8). These differences were statistically significant with P-value of 0.0003 between retinal and non-retinal tissues in IG70, and 0.0116 between retinas of IG heterozygous dogs, and control retinas (Table 3.9).

Table 3.8. Average risk allele expression ratio in IG heterozygous and control dogs in retinal and non-retinal tissues and genomic DNA. Numbers were corrected according to background reads in homozygous dogs, or genomic DNA.

Sample number	Tissue	Exon 32- SNP 71,540,880		Exon 61- SNP 71,576,758		Exon 61- SNP 71,577,339	
		DNA genotype	Average of T allele	DNA genotype	Average of C allele	DNA genotype	Average of C allele
1	IG72 retina	C/T	0.12	C/T	0.28	C/T	0.26
2	IG73 retina	C/T	0.12	C/T	0.26	C/T	0.25
3	IG74 retina	C/T	0.12	C/T	0.25	C/T	0.22
4	IG70 retina	C/T	0.30	C/T	0.28	C/T	0.24
5	IG70 kidney	C/T	0.70	C/T	-	C/T	-
6	IG70 cerebellum	C/T	0.62	C/T	0.51	C/T	0.53
7	IG70 knee cartilage	C/T	0.64	C/T	0.50	C/T	0.50
8	Control-1 retina	C/T	0.63	C/T	0.65	C/T	0.53
9	Control-1 kidney	C/T	0.60	C/T	0.69	C/T	0.57
10	Control-1 cerebellum	C/T	0.49	C/T	0.61	C/T	-
11	Control-1 knee cartilage	C/T	NA	C/T	0.61	C/T	0.53
12	Control-2 retina	C/C	0.00	C/C	0.97	C/T	0.39
13	Control-3 retina	C/C	0.00	C/T	0.31	T/T	0.08
14	Control-4 retina	T/T	1.00	C/C	1.04	C/C	0.88
15	Control-5 retina	C/T	0.60	C/T	0.47	C/T	0.33
16	Control-6 retina	C/T	1.12	C/C	1.00	C/T	0.63
17	Control-7 retina	C/C	0.00	C/C	0.95	C/T	-
18	IG73 gDNA (20 ng)	C/T	NA	C/T	0.50	C/T	0.53
19	IG74 gDNA (20 ng)	C/T	NA	C/T	0.50	C/T	0.51
20	IG70 gDNA (20 ng)	C/T	NA	C/T	0.48	C/T	0.47
21	Control-1 gDNA (20 ng)	C/T	NA	C/T	0.54	C/T	0.50
22	Control-2 gDNA (20 ng)	C/C	NA	C/C	1.00	C/T	0.47

(-) = results were not available due to DNA contamination in the RT-PCR (RT negative control was not zero).
NA= results were not available due to failed PCR.
Red allele is the risk allele in IG dogs.
Blue allele is the paternal allele in control dogs.

Table 3.9. Statistical analysis of the differences of pyrosequencing results between groups. **A.** T-test of the differences between IG70 retina results and other tissues of this dog (kidney, cerebellum, and knee cartilage). **B.** T-test of the differences between IG heterozygous dogs and control dogs of their deviation from 0.5 allele ratio.

A. T-test between average risk allele expression ratio in IG70 retina compare to non-retina tissues.		
	Retina	Non-retina tissues
Mean	0.2733	0.5714
SD	0.0306	0.0809
SEM	0.0176	0.0306
N	3	7
Unpaired two-tailed t-test:	**0.0003- significant**	
B. T-test between average allele expression deviation from 0.5 in IG heterozygous dogs and control dogs in retina.		
	IGPRA-heterozygous	Controls
Mean	0.275	0.154
SD	0.0665	0.1325
SEM	0.0192	0.0419
N	12	10
Unpaired two-tailed t-test:	**0.0116- significant**	

Out of the 42 dogs that were run on the Illumina SNP array, five dogs carried one copy of the risk haplotype in the locus of Col27A1 and mir455 and were affected with IGPRA, while 13 dogs carried one copy of the risk haplotype but showed no symptoms of disease, and were considered phenotypically normal. We ran an association between these two groups. The highest P-value was shared among 10 SNPs with P-value of 0.0001021 (-log10(P)=3.99), six on CFA3 and four on chromosome X (Table 3.10A). Interestingly, in the middle of the 811.7 Kb LD interval on CFA3 (CFA3: 60,345,668-60,345,668) mir184 is located (Table 3.10B) and ADAMTS7 is located in the distal end of the LD. We sequenced mir184 on a subset of dogs (n=12, Supplement Table 3.11A). Although we did not find polymorphisms in the pre-miRNA conserved sequence (83 bp at position 60,776,998-60,777,080 CanFam 2), we found 3 SNPs in the pri-miRNA sequence (226bp, 204 bp, and 34 bp upstream to the conserved pre-miRNA;

Supplement Table 3.11B). Expanding this genotyping to the complete IG population is needed to be able to evaluate if any of the SNPs can explain some of the imprinting.

Table 3.10. Top hits of GWA analysis between IG dogs carrying one copy of the risk haplotype in Col27A1 locus that are affected with PRA (n=5) and that are not affected with PRA (n=13). **A.** The most significant SNPs ($-\log_{10}(P)=3.99$) and their location. **B.** Genes within the significant interval on CFA3.

A. Most significant SNPs and location.		
CFA	**SNP**	**Position**
3	BICF2G630342052	60,345,668
3	BICF2G630342252	60,485,857
3	BICF2G630342336	60,604,161
3	BICF2P635067	61,006,592
3	BICF2G630342761	61,145,506
3	BICF2P314519	61,157,377
X	BICF2G630533117	26,859,917
X	BICF2G630533113	26,864,307
X	BICF2G630532418	28,194,061
X	BICF2G630532410	28,202,238
B. Genes within the significant locus on CFA3.		
Gene	**Name**	**Position**
MTHFS	5,10-methenyltetrahydrofolate synthetase (5-formyltetrahydrofolate cyclo-ligase) (MTHFS), transcript variant 1, mRNA	60,344,001-60,376,340
KIAA1024	KIAA1024	60,598,504-60,610,818
TMED3	transmembrane emp24 protein transport domain containing 3	60,695,345-60,701,640
ANKRD34C	ankyrin repeat domain 34C	60,708,285-60,723,156
Mir184	Mir 184	60,776,998-60,777,070
RASGRF1	Ras protein-specific guanine nucleotide-releasing factor 1	60,858,018-60,951,418
CTSH	cathepsin H	60,968,460-60,978,236
MORF4L1	mortality factor 4 like 1	60,991,592-61,013,174
ADAMTS7	ADAM metallopeptidase with thrombospondin type 1 motif, 7	61,053,256-61,087,513
TBC1D2B	TBC1 domain family, member 2B	61,089,422-61,135,752

3.5 Discussion

MicroRNAs are small single-stranded RNA molecules used by the cell to regulate gene expression through base-pairing to partially complementary mRNA sequence. This serves to enhance RNA decay or to inhibit translation[18-20]. They are transcribed in the nucleus, either independently or as part of the introns of protein-coding genes[21,22]. The single strand RNA goes through several processes to release pre-miRNA (usually about 70-100 bp long) into the cytoplasm out of which Dicer cuts it to 20-24 bp long. The short molecule binds to target molecules with complementary sequence and represses translation, or enhances rate of decay. We mapped the Italian greyhound PRA disease to a distinct interval on CFA11. Within the LD interval a significant RNA expression level difference of Col27A1 gene and mir455 were observed between F1 IGPRA dogs and normal dogs. Mir455 is a microRNA gene located in intron 10 of the Col27A1 gene. The Col27A1 risk allele was expressed at levels of about 20% of the total transcript. The total level of expression of mir455 in controls relative to IG heterozygous was 2.16, almost mono-allelic expression in the heterozygous dogs, suggesting a very low level of mir455 in affected dogs. We did not have retinal samples from homozygous affected dogs to check their level of expression of Col27A1 and mir455 and we will explore that once these animals become available. Our observations from both RNA-seq and micro-RNA seq that both Col27A1 and mir455 are down-regulated in the affected chromosome almost to the same digree, suggest that mir455 is transcribed from the Col27A1 intron, and most likely does not have an independent promoter. This agrees with the observations made by Monk and his group[23], where levels of Col27A1 and mir455 were upregulated in macrophages after *candida* exposure at the same level.

In recent years, more than 250 retinal microRNA genes have been identified[24], with studies in small animal models showing that miRNA genes have a major role in eye development and retina[25]. More specifically, six different retinal miRNA genes showed changes in expression in four different mouse models for retinitis pigmentosa[26,27]. To the best of our knowledge, this is the first time that altered mir gene expression has been reported to be associated with retinal degeneration in a large animal model, and the first time where tissue specific allelically biased expression is associated with a disease.

Interestingly, in our IGPRA model, we not only observed a low level of mir455 expression, but we also observed a significantly high level of mir3609 expression. Lower levels of mir455 would result in higher expression of its target coding genes compared to controls, and a high level of mir3609 would result in low expression of its target coding genes compared to controls. Many of mir3609's target genes are known to cause retinal degeneration when mutated: these include Rho, ABCA4, PDE6A, and LCA5 (Supplement Table 3.13). We did not find these target genes for mir455 of mir3609 to be significantly differentially expressed in IG heterozygous dogs compared to the controls from RNA-seq analysis. Further investigation of altered expression of the mir genes and their target genes are necessary, testing a larger number of dogs, and specifically dogs that are affected and are homozygous for the risk haplotype in Col27A1.

The three IG heterozygous dogs used for the RNA-seq, SmallRNA-seq and allele-specific pyrosequencing experiments were all littermates of an affected sire bred to a normal beagle dam (as well as the fourth sibling in the pyrosequencing experiment) (Figure 3.1). In all four dogs the allele that was down-regulated was the risk allele transmitted from the affected father. This suggests that the mutation yet to be found is affecting the level of expression of Col27A1 and as

a result also mir455. We do not believe this is random parental imprinting, especially since we see both paternal and maternal alleles expressed in the controls (Supplement Table 3.12). One control dog though showed a complete monoallelic expression in exon 32 (control dog number 6, Supplement Table 3.12) expressing only the paternal allele, but that bias was not observed in an exon 61 informative SNP. We do not currently have an explanation for this monoallelic expression in exon 32, except that a primer bias might be present in that specific dog. A repetition of the experiment with different primers is in progress. We do not have evidence for imprinting from the control dogs but to completely rule out an imprinting scenario, we would design the reciprocal breeding experiment of breeding, mating an affected female to a normal male, and measure the level of risk allele expression in the heterozygous offsprings, to determine whether the same bias in allelic expression is observed.

To understand the factors underlying the incomplete penetrance of the disease we compared the genome wide SNP alleles in the five affected dogs heterozygous for the Col27A1 risk allele to the 13 dogs heterozygous to the risk allele and phenotypically normal. Interestingly the autosomal locus that was the most significant was CFA3, in a region that includes mir184. Several targets for miR-184 have been described, including that of mediators of neurological development, and apoptosis, and a mutation in the seeded sequence (the first 8 bases of the mature microRNA, important for recognition of the target mRNA) of mir184 that causes eye disease and cataract[28-30]. It was also observed that mir184 has an imprinted paternal expression in some cases and it is known that mir genes compete on shared targets[31,32]. This information makes the modifier hit on mir184 a very interesting candidate for such, and encourages further investigation.

We did not find an obvious mutation that can explain retina specific down regulation of Col27A1 and mir455. A list of polymorphisms in the affected chromosome was extracted from the DNA-seq experiment. A few regions were also not covered by the deep sequencing, and need to be retrieved by individual PCR experiments. An interesting finding though was the poly-alanine expansion in exon 1 of Col27A1. It was reported that in the dog polyglutamine and polyalanine are variable in numbers and that transcriptional activities of both of these types of amino acid repeat domains are present, with polyglutamines reported to drive transcription and polyalanines repressing it[33,34]. Although it seems that the high number of alanine repeats in the exon (n=12) is fixed in the IG, and it is not unique to this breed, it might be necessary but not sufficient to cause down-regulation of Col27A1.

The microRNA world is far from being understood at the molecular level and the network of all their targets, as well as the redundancy of their function. Knockout models and rapidly accumulating human miRNA-disease association data are suggesting that miRNA genes can function as modifiers[35]. We believe that the IGPRA dog model offers a unique opportunity to explore the effect of mir455 on retinal degeneration, and also explore other mir genes as modifiers for the disease. This might shed light on the mir cooperation network in the retina, and the pathways in which they react. Moreover, miRNA genes are now considered therapeutic agents in inherited retinal diseases, especially after the substantial progress in developing efficient delivery systems for small RNA molecules[36,37]. These dogs can offer a large animal model for microRNA gene therapy for late onset PRA.

3.6 References

1. 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. Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. Genomics. 2006 Nov;88(5):551-63.

2. Goldstein O, Mezey JG, Boyko AR, Gao C, Wang W, Bustamante CD, Anguish LJ, Jordan JA, Pearce-Kelling SE, Aguirre GD, Acland GM. An ADAM9 mutation in canine cone-rod dystrophy 3 establishes homology with human cone-rod dystrophy 9. Mol Vis. 2010 Aug 11;16:1549-69.

3. Goldstein O, Zangerl B, Pearce-Kelling S, Sidjanin DJ, Kijas JW, Felix J, Acland GM, Aguirre GD. Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. Genomics. 2006 Nov;88(5):541-50.

4. Acland G.M., Aguirre G.D. Retinal degenerations in the dog: IV. Early retinal degeneration (erd) in Norwegian elkhounds. Exp. Eye Res. 1987;44:491-521.

5. Acland G.M.,. Ray K.,. Mellersh C.S.,. Langston A.A.,. Rine J., Ostrander E.A.,. Aguirre G.D. A novel retinal degeneration locus identified by linkage and comparative mapping of canine early retinal degeneration. Genomics 1999;59:134-142.

6. Komáromy AM, Alexander JJ, Rowlan JS, Garcia MM, Chiodo VA, Kaya A, Tanaka JC, Acland GM, Hauswirth WW, Aguirre GD. Gene therapy rescues cone function in congenital achromatopsia. Hum Mol Genet. 2010 Jul 1;19(13):2581-93.

7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559-75.

8. Goldstein O, Guyon R, Kukekova A, Kuznetsova TN, Pearce-Kelling SE, Johnson J, Aguirre GD, Acland GM. COL9A2 and COL9A3 mutations in canine autosomal recessive oculoskeletal dysplasia. Mamm Genome. 2010 Aug;21(7-8):398-408.

9. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.

10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.

11. 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. 20:1297-303.

12. DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D and Daly, M (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 43:491-498.

13. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511-515.

14. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W622-7.

15. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105-1111.

16. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

17. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

18. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet. 2011 Feb;12(2):99-110.

19. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem. 2010;79:351-79.

20. Fabian MR, Sundermeier TR, Sonenberg N. Understanding how miRNAs post-transcriptionally regulate gene expression. Prog Mol Subcell Biol. 2010;50:1-20.

21. Newman MA, Hammond SM. Emerging paradigms of regulated microRNA processing. Genes Dev. 2010 Jun 1;24(11):1086-92.

22. Siomi H, Siomi MC. Posttranscriptional regulation of microRNA biogenesis in animals. Mol Cell. 2010 May 14;38(3):323-32.

23. Monk CE, Hutvagner G, Arthur JS. Regulation of miRNA transcription in macrophages in response to Candida albicans. PLoS One. 2010 Oct 27;5(10):e13669.

24. Krol J, Busskamp V, Markiewicz I, Stadler MB, Ribi S, Richter J, Duebel J, Bicker S, Fehling HJ, Schübeler D, Oertner TG, Schratt G, Bibel M, Roska B, Filipowicz W. Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell. 2010 May 14;141(4):618-31.

25. Hackler L Jr, Wan J, Swaroop A, Qian J, Zack DJ. MicroRNA profile of the developing mouse retina. Invest Ophthalmol Vis Sci. 2010 Apr;51(4):1823-31.

26. Loscher CJ, Hokamp K, Kenna PF, Ivens AC, Humphries P, Palfi A, Farrar GJ. Altered retinal microRNA expression profile in a mouse model of retinitis pigmentosa. Genome Biol. 2007;8(11):R248.

27. Loscher CJ, Hokamp K, Wilson JH, Li T, Humphries P, Farrar GJ, Palfi A. A common microRNA signature in mouse models of retinal degeneration. Exp Eye Res. 2008 Dec;87(6):529-34.

28. Foley NH, Bray IM, Tivnan A, Bryan K, Murphy DM, Buckley PG, Ryan J, O'Meara A, O'Sullivan M, Stallings RL (2010). "MicroRNA-184 inhibits neuroblastoma cell survival through targeting the serine/threonine kinase AKT2.". Mol Cancer 9: 83.

29. Iliff BW, Riazuddin SA, Gottsch JD. A single-base substitution in the seed region of miR-184 causes EDICT syndrome. Invest Ophthalmol Vis Sci. 2012 Jan 25;53(1):348-53.

30. Bykhovskaya Y, Caiado Canedo AL, Wright KW, Rabinowitz YS. C.57 C > T Mutation in MIR 184 is Responsible for Congenital Cataracts and Corneal Abnormalities in a Five-generation Family from Galicia, Spain. Ophthalmic Genet. 2013 Oct 18.

31. Liu C, Teng ZQ, Santistevan NJ, Szulwach KE, Guo W, Jin P, Zhao X. Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. Cell Stem Cell. 2010 May 7;6(5):433-44.

32. Hughes AE, Bradley DT, Campbell M, Lechner J, Dash DP, Simpson DA, Willoughby CE. Mutation altering the miR-184 seed region causes familial keratoconus with cataract. Am J Hum Genet. 2011 Nov 11;89(5):628-33.

33. H. P. Gerber, K. Seipel, O. Georgiev, M. Höfferer, M. Hug, S. Rusconi, and W. Schaffner. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science, 263(5148):808–811, 1994.

34. Albrecht AN, Kornak U, Böddrich A, Süring K, Robinson PN, Stiege AC, Lurz R, Stricker S, Wanker EE, Mundlos S. A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. Hum Mol Genet. 2004 Oct 15;13(20):2351-9.

35. Bandiera S, Hatem E, Lyonnet S, Henrion-Caude A. microRNAs in diseases: from candidate to modifier genes. Clin Genet. 2010 Apr;77(4):306-13.

36. Kaiser PK, Symons RC, Shah SM, Quinlan EJ, Tabandeh H, Do DV, Reisen G, Lockridge JA, Short B, Guerciolini R, Nguyen QD; RNAi-based treatment for neovascular age-related macular degeneration by Sirna-027. Am J Ophthalmol. 2010 Jul;150(1):33-39.

37. Miller AD. Delivery of RNAi therapeutics: work in progress. Expert Rev Med Devices. 2013 Nov;10(6):781-811.

Supplemental material:

Supplement Tables 3.1-3.12

Supplement Figures 3.1-3.2

Supplement Table 3.1. Primers used to screen candidate genes for IGPRA**.**

Primer pair	Forward Primer name	Forward primer sequence	Forward primer location	Reverse Primer name	Reverse primer sequence	Reverse primer location	Size
A. DFNB31 gene.							
A-i. Primers used to amplify retinal cDNA.							
1	DFNB31_F1	GTTTCCCTCGGACCCCAGGAC	5' UTR	DFNB31_R1	GACTTGTCGTTGACGCGCAGAAT	Exon1	779
2	DFNB31_F2	GGCATCTACGTGTCGCTGGTG	Exon 1	DFNB31_R2	GAACTGGCGATCCACTTGGTCT	Exon 4	641
3	DFNB31_F3	CTCAAGTCATCCCAGCACCTCAT	Exon 4	DFNB31_R3	ACAGCCTCACCAGTTTCCTCCAG	Exon 8	626
4	DFNB31_F4	CACCTACTCCGTGGTCTCCTACAG	Exon 7	DFNB31_R4	CTCCTTGCTACACCTGCTCTTGGT	Exon 10	838
5	DFNB31_F5	CCTCCACACTGTCCCAGCTCTC	Exon 10	DFNB31_R5	ATGTCTTCCTGCTACCCTGGCTAC	3'UTR	604
6	DFNB31_F6	ATGAAGGCGCAGCTGGACGGCCT	Exon1	DFNB31_R6	TGTAGATGTGGTTGGTGATGTAGCC	Exon2	757
7	DFNB31_F7	CCAGCTGCTCTTCGACCAGTACAC	Exon 1	DFNB31_R7	CACCAGGTTCACCTTCTTCTCGTC	Exon2+3	568
8	DFNB31_F8	CACTGCCTCAACGCCTACCAC	Exon 1	DFNB31_R8	CACGGACAACAGCAGCTTCTTAGAG	Exon2	501
9	DFNcDNAF1	CGGGGACGAGAAGAAGGTGAAC	Exon2+3	DFNcDNAR1	GGAACTGGCGATCCACTTGGT	Exon 4	313
10	DFNcDNAF2	GGGCTCAGAAGCAGAGAGCAGT	Exon3	DFNcDNAR2	GAGAACTTGGCGTTGGTGTTGAA	Exon6+7	489
11	DFNcDNAF3	GAAGGAACAAGCAAGCTGGGATT	Exon5+6	DFNcDNAR3	GCTGTAGGAGACCACGGAGTAGGT	Exon 7	381
12	DFNcDNAF8	CCTGGTGCTGAGACGGGAGAT	Exon 7	DFNcDNAR8	GGGGACACAGAGGGCATAGACG	Exon 10	528
13	DFNcDNAF4	CCACTCGGGCATCGTCTTCTC	Exon 10	DFNcDNAR4	GAGCTGGGACAGTGTGGAGGCT	Exon 10+11	415
14	DFNcDNAF4	CCACTCGGGCATCGTCTTCTC	Exon 10	DFNcDNAR5	AGCTGGGACAGTGTGGAGGCG	Exon 10+11	412
15	DFNcDNAF6	GGGCACGAACCAGCACTTTGT	Exon 10	DFNcDNAR6	GCACTTTTCTTCACACGGACCAC	Exon 12	324
16	DFNcDNAF7	AGACAGATCGCCTCCACCAAGA	Exon 11	DFNcDNAR7	GCCCTTGAGGAGCTTATAGTGCTG	Exon 13-3' UTR	473
A-ii. Primers used to amplify genomic DNA							
1	DFNB31_intron1F	CTGGGGAAGAAAAACAAAGGTTGG	Intron 1	DFNB31_intron2R	CTGGGTCAGGCAAGGAGAACAG	Intron 2	489
2	DFNB31_intron2F	GTTCACGTCACCCACCCTCCA	Intron 2	DFNB31_intron3R	CAGCTCCTGCCCTCTCTCTCC	Intron 3	254
3	DFNB31_intron3F	GAGCCAGTGTCCTCCAAGCAG	Intron 3	DFNB31_intron4R	TCTAACTCCCAGACCCCTTTCCTC	Intron 4	573

4	DFNB31_intron4F	CTATGGGCTGCTGTCTCCTGTCCT	Intron 4	DFNB31_intron5R	GCGGGTGTGATTTTCTGTGTCTA	Intron 5	585
5	DFNB31_intron5F	GTCTGAGATGGGGCCTGTAACG	Intron 5	DFNB31_intron6R	CACCCGATTCTTTGGGTTCTCA	Intron 6	461
6	DFNB31_intron6F	CAAGATGCCAGCGAGGACGAG	Intron 6	DFNB31_intron7R	CAAGGGACAGAGCCAGGAAGG	Intron 7	533
7	DFNB31_intron7F	GGGGAGATAGGGGATTTTAGAGAA	Intron 7	DFNB31_intron8R	AGACATAAGGGGGTGGCAGGAAG	Intron 8	286
8	DFNB31_intron8F	GGAGAGCCCCAAACAAAAAGC	Intron 8	DFNB31_intron9R	GGCAAACAAACAGCCAGAGTCC	Intron 9	493
9	DFNB31_intron9F	CCCCTGTTGTTGTTGTTTTTCCAT	Intron 9	DFNB31_intron10R	ACACTGGCTCCCACTTTGTGC	Intron 10	970
10	DFNB31_intron10F	ACAGCATCCCAGCTCTATCACCTC	Intron 10	DFNB31_intron11R	CTCACTTCTTCTCCGTGGTGGTC	Intron 11	380
11	DFNB31_intron11F	CTGAGCGGGTGTCCTGGCTAA	Intron 11	DFNB31_exon13R	CGATGTAGTCTCTCTCCTTGGTCTTG	Exon 13	822
12	DFNB31_intron12F	CACTGAGCCTGTTTTCCTAATCA	Intron 12	DFNB31_3UTRR	GCTTTCTTCTCCTGGTTCTGG	3' UTR	499
B. AKNA gene							
B-i. Primers used to amplify retinal cDNA.							
1	AKNA_cDNA_F1	GTGGCAGCTCCGTCTCCTGAAG	Exon 2-5' UTR	AKNA_cDNA_R1	GACCAGTCTCATGCTCCAGTCTTG	Exon3	484
2	AKNA_cDNA_F2	AGAAAGTTCCATCGTGCCTCTCAG	Exon 3	AKNA_cDNA_R2	GGGGTGGTAGCTCTCATCTTTGG	Exon 3	817
3	AKNA_cDNA_F3	CTGCCGATGCCTCCAAGTACG	Exon 3	AKNA_cDNA_R3	GGCCCAAGAAGCCAATTCCTG	Exon 6	671
4	AKNA_cDNA_F4	CAGTCGGTGGTATGGTGGTCAG	Exon 5	AKNA_cDNA_R4	CTCACTGCTGGTGGCGCTAAG	Exon 10	666
5	AKNA_cDNA_F5	CAGAAGCAGCCAGAGCCTGAG	Exon 9	AKNA_cDNA_R5	ACTTCCCTCCAAGCTGGTCAGA	Exon 12	687
6	AKNA_cDNA_F6	CCCTGGCTGAAAGAAGTCACAGG	Exon 11	AKNA_cDNA_R6	AGACGCTTGGAAATCCGCTTC	Exon 15	670
7	AKNA_cDNA_F7	GaGGTCCCCGGCTCAGAGTTC	Exon 14+15	AKNA_cDNA_R7	ACTCACGGACAGTTGGGGCATC	Exon 17+18	459
8	AKNA_cDNA_F8	GGCGAGCCAGGTCTTCCTCAG	Exon 17	AKNA_cDNA_R8	CTCTGCGTGCGTTTCTCCTTG	Exon 21	453
9	AKNA_cDNA_F9	CTCCCACTGACACTGTGCGATGT	Exon 20	AKNA_cDNA_R9	GCCACCCTGTGTTGCTGAGTCT	Exon 22-3'UTR	573

B-ii. Primers used to amplify genomic DNA							
1	AKNA_F1	AAACCCCCTCCTTCTGACAAACAC	5' UTR	AKNA_R1	GGACTAACTCTCCAAGTGCCGATG	Intron 1	563
2	AKNA_F2	CCGTCATCCTCCTGGTTTCTCTAA	Intron 1	AKNA_R2	AACACCCTCCATTCCCCACCT	Intron 2	629
3	AKNA_F3	GGTGGTGTTTCCATTTGCTGAGAC	Intron 2	AKNA_R3	GTGTTTCCATTGCCCCCTGAG	Exon 3	431
4	AKNA_F4	CTCAACTGGCCGTGACTGAAGAG	Exon 3	AKNA_R4	GTTCTGGGGTTTGCGTGTAGC	Exon 3	403
5	AKNA_F5	TACCTGGGAAGGGGAAACTGATG	Exon 3	AKNA_R5	TGCCTAGAGACGTGAGTCCAGCTT	Intron3	700
6	AKNA_F6	GAGGTTCCCCAAAGATGAGAGC	Exon 3	AKNA_R6	GCAGAGAGAGAGGCAGAGACACAG	Intron 3	557
7	AKNA_F7	GGGGTAGAGGCGGGGAAGAGT	Intron 3	AKNA_R7	GGGACCTTGTGGTAGATGGAGGAC	Intron 4	468
8	AKNA_F8	GTGATGGTGTGAATGAAGGTGAGG	Intron 4	AKNA_R8	GCTGCTCTGGCTCTGGGTAAGAT	Intron 5	433
9	AKNA_F9	AGCTGTGTGATCTGGGGCAGAC	Intron 5	AKNA_R9	CTTTCAAGCACAACTGGAGGGAGT	Intron 6	586
10	AKNA_F10	GGATTCACGGAGTGTGTTTCCTTC	Intron 6	AKNA_R10	GACAGTGTGTCTGGGATTCTCTGG	Intron 7	223
11	AKNA_F11	GCCCAGGAGATGGGTCTCTGC	Intron 7	AKNA_R11	GACTTGGGTAATCCCTCCCCACT	Intron 8	496
12	AKNA_F12	TTCCCTTTTCACCTTGTTGTTTCG	Intron 8	AKNA_R12	AGCCCCCTCGCTTTCCATACT	Intron 9	551
13	AKNA_F13	CGTGAAAGGGTGAGAGGGTTG	Intron 9	AKNA_R13	GATACCTACAGAGCACAGAGAGGAACA	Intron 10 +Exon11	403
14	AKNA_F14	CTGGAGCGGTGAGTACCAGGAC	Exon 10 +Intron 10	AKNA_R14	TAATACCATCCAACCCTGCCTCTG	Intron 11	485
15	AKNA_F15	TACCCAAGTTCCTGAAGCCTGAGA	Intron 11	AKNA_R15	GAGGCAGCACTGGGCATAAGTC	Intron 12	672
16	AKNA_F16	ATCCTTCCCCTCTTCCTCAGC	Intron 12	AKNA_R16	GCACCTTGATTCCCAGGCTACC	Intron 13	442
17	AKNA_F17	GGGTGAAGGATGGAGTGACAGAC	Intron 13	AKNA_R17	TTTGTTGTTGTTGTTGTTGGTTTGA	Intron 14	683
18	AKNA_F18	GTGCTGAAGTGGTGGACAGATGG	Intron 14	AKNA_R18	GAGGAGAAATACAGGGAGCGAGGT	Intron 15	670
19	AKNA_F19	GCAGGAGAGGGACAGGGTTGG	Intron 15	AKNA_R19	TGAGAGTAGGAGAGGGGTGTGAGTT	Intron 16	383
20	AKNA_F20b	GCGTAAAATCTCACCCACTGACATC	Intron 16	AKNA_R20b	CGTCCATCCACGTCCTTGTCT	Intron 17	465
21	AKNA_F21	GTGTGCGTCGGCTGCTGTTACTA	Intron 17	AKNA_R21	CGCTCTGAAGTCCTACTGTGTGTCA	Intron 18	689
22	AKNA_F22b	GCCCTGTTAGCCGTTGCTTAG	Intron 18	AKNA_R22b	CGATGCCTGAGTGAATGAATGAAC	Intron 19	583
23	AKNA_F20	CCCAGTGACCCAGACATGACATC	Intron 19	AKNA_R20	AGGAAGGTGGGAGAGGAAGGTTCT	Intron 21	621
24	AKNA_F22	CCATTTTCGAGGTGAAGACACTGA	Intron 21	AKNA_R22	GAGACTCAGGAGGGGTCTGTGG	Exon 22-3' UTR	498
C. ORM1 gene- Primers used to amplify retinal cDNA.							
1	ORM1_cDNA_F1	TGCCCACACACAGCCCGTTAT	5' UTR	ORM1_cDNA_R1	GTTCCTTCAGCCTCTCCTCCTTGT	Exon 6	673
2	ORM1_cDNA_F1	TGCCCACACACAGCCCGTTAT	5' UTR	ORM1_cDNA_R2	GCAACTGAGACCAAAACCTGAGAT	3'UTR	776
3	ORM1_cDNA_F1	TGCCCACACACAGCCCGTTAT	5' UTR	ORM1_cDNA_R3	TGAGCGACTGAATGAGCCAAC	3'UTR	835

D. Col27A1 gene							
D-i. Primers used to amplify retinal cDNA.							
1	Col27A1_cDNA_F2	CAATCGGGCTTCATCTTTACGC	Exon 3	Col27A1_cDNA_R2	GTGGGTGGGGCTGACTTTGTG	Exon 3	986
2	Col27A1_cDNA_F3	TCGGTCCCTTCTCCTTCTGTGATA	Exon 3	Col27A1_cDNA_R3	GGTCCCCTTTCTGTCCCTTGG	Exon 5+6	880
3	Col27A1_cDNA_F4	TCCTCCTGGGCCTTATGGAAAC	Exon 5	Col27A1_cDNA_R4	GCTTACCCTTCATGCCTTCTGG	Exons 20+21	824
4	Col27A1_cDNA_F5	GCCTCTGGGGAAAATTGGAGA	Exons 18+19	Col27A1_cDNA_R5	GTCACCCTTCATCCCGGCTTC	Exons 34+35	841
5	Col27A1_cDNA_F6	ACTTGGCCCTCCAGGAGAGATG	Exon 33	Col27A1_cDNA_R6	TGATGCCTTTGTGTCCATCGTAAC	Exons 40+41	533
6	Col27A1_cDNA_F7	GTCCTGAAGGAAAACCTGGGAAG	Exon 38	Col27A1_cDNA_R7	GTAACCCACTTTCGCCCTTGAATC	Exons 48+49	774
7	Col27A1_cDNA_F8	GGGAAGAGAGGAAATCCAGGTGTG	Exon 48	Col27A1_cDNA_R8	GCGTCTTAATGCTCTGGATGAGGT	Exon 58	630
8	Col27A1_cDNA_F9	ATGGACACAAATGGAGCACTGAAA	Exon 57	Col27A1_cDNA_R9	CAGGAGGTCAGAGGTCAGAGGAAG	Exon 61	689
9	Col27A1_cDNA_F10	CACCATCCACTGCCTTAACATGAC	Exon 60	Col27A1_cDNA_R10	CTTCTCTCACCTCTTTCCCAGGTG	3' UTR	844
10	Col27A1_cDNA_F12	GCGATTCCTGTCCTCCAGTGC	Exon 3	Col27A1_cDNA_R12	GATAGCCCTTGTGTCCCTTTCGT	Exons 7+8	356
11	Col27A1_cDNA_F13	GTCCTCCTGGGCCTTATGGAA	Exon 5	Col27A1_cDNA_R13	GACCTCGTTCCCCAAAGTCTCC	Exon 13	435
12	Col27A1_cDNA_F14	GTAGGAGACCCTGGCCCCAAAG	Exon 10	Col27A1_cDNA_R14	GACACACCTGGAACCCCTTGTTC	Exon 17	353
13	Col27A1_cDNA_F15	GTGTGTCAGGAGACCCTGGATTC	Exon 17	Col27A1_cDNA_R15	GTCACCCTTTTCTCCCATGATCC	Exons 25+26	455
14	Col27A1_cDNA_F16	AACCTGGGAGGAAGGGGTTTC	Exon 23	Col27A1_cDNA_R16	CATCTCTCCTGGAGGGCCAAGT	Exon 33	554
15	Col27A1_cDNA_F17	ATGAGGGGAGCAAAGGGACTCT	Exon 28	Col27A1_cDNA_R17	ATCTCACCCAGCTGTCCTTGGTAG	Exon 39	635
16	Col27A1_cDNA_F18	AGATCTTTGAAGCAGGGGGTCAGT	Exon 60	Col27A1_cDNA_R18	CCACTCCTACAGGGCAGACAACAG	3' UTR	362
17	Col27A1_cDNA_F19	TCAGATGAACTTTCTGCACCTGCT	Exon 60	Col27A1_cDNA_R19	GCCTCCTCCTCCTCCTCTTCC	3' UTR	402
18	Col27A1_cDNA_F14	GTAGGAGACCCTGGCCCCAAAG	Exon 10	Col27A1_cDNA_R13	GACCTCGTTCCCCAAAGTCTCC	Exon 13	160
19	Col27A1_cDNA_F17	ATGAGGGGAGCAAAGGGACTCT	Exon 28	Col27A1_cDNA_R5	GTCACCCTTCATCCCGGCTTC	Exons 34+35	336
20	Col27A1_cDNA_F20	GAAAGGGACACAAGGGCTATCCTG	Exons 7+8	Col27A1_cDNA_R20	CCCTCTTGCCAGGAACACCAG	Exon 12	231
21	Col27A1_cDNA_F21	TCCTGGCAAGAGGGGCAAGAT	Exon 12	Col27A1_cDNA_R21	ACCTGGAACCCCTTGTTCACC	Exon 17	250
22	Col27A1_cDNA_F22	ACTTGGCCCTCCAGGAGAGATG	Exon 33	Col27A1_cDNA_R22	AGCTGTCCTTGGTAGCCCTTGG	Exon 39	381

D-ii. Primers used to amplify genomic DNA.							
1	Col27A1_intron1_F	CTGCCCTGTGACCTTGGCTTATC	Intron 1	Col27A1_intron2_R	CCCTGCCCCTTACAAAACTTCAAA	Intron 2	570
2	Col27A1_intron2_F	TGTGCCTACAATGAGGATGTCCAG	Intron 2	Col27A1_exon3_R8	GCAAAAAGGAAGGCATGGTTCAC	Exon 3	400
3	Col27A1_intron8_F	TCACATTCCTGCCCTTCTCTTTGT	Intron 8	Col27A1_intron9_R	ACCAGAACCCCTTTCTTCTGCTCT	Intron 9	404
4	Col27A1_intron9_F	AGAAGAAAGGGGTTCTGGTGCTGT	Intron 9	Col27A1_intron10_R	GATCCCCCATCACCTGTCCAC	Intron 10	489
5	Col27A1_intron11_F	CAGGCCCTCTGTGACTTCCTCTC	Intron 11	Col27A1_intron12_R	CCCAGACACTACTGACCCAGACAA	Intron 12	477
6	Col27A1_intron12_F	CCAGTAGCCAGTCCCTGAGAAGTT	Intron 12	Col27A1_intron13_R	TGGGCCTCCCATTATTTACAAACTC	Intron 13	664
7	Col27A1_intron13_F	CTTCTCGGACACTGACAAGTGGAA	Intron 13	Col27A1_intron14_R	GAACTATGCAGGAAAGGGCTCAGA	Intron 14	514
8	Col27A1_intron14_F	ATTGCACTGGTCCCCCTTTTAGAT	Intron 14	Col27A1_intron15_R	CAGGAGACTCAAGCTCAGCACAAG	Intron 15	509
9	Col27A1_intron15_F	CACCTGGGGCATAGAAACTGTACC	Intron 15	Col27A1_intron16_R	CAAGGAAGAAGCTGGGGAATACAG	Intron 16	661
10	Col27A1_intron34_F	CTTTGGTCCTTGGGAACGAGGT	Intron 34	Col27A1_intron35_R	CTACGTTGTGGAGATGCTGGTTTG	Intron 35	340
11	Col27A1_intron35_F	GGTCTGCTCCTCGTTTCTGAAGAG	Intron 35	Col27A1_intron36_R	CCCTTACCCCAGAGAGGCTGTAGT	Intron 36	576
12	Col27A1_intron36_F	ATCCAAACGGAGCATTGAGGTTTA	Intron 36	Col27A1_intron37_R	AGCAGCAAGTTCTGGTGAGGAATC	Intron 37	614
E. mir455 gene.							
1	MIR455F1	AGCCTCGGTTTCCTTGTCTGTG		MIR455R1	TTGTCCCCACTCTCCCTTACATTG		731
2	MIR455F2	AGCATCTAACCCATCCGAGAGC		MIR455R2	GGGACCTGACTGCCTTGACTC		517
F. KIF12 gene- Primers used to amplify retinal cDNA.							
1	KIF12_cDNA_F1	CTTCTCCTGCACCGTCTTCACCTT	5' UTR	KIF12_cDNA_R1	AGGCTGCGGTTGATGCTGTTC	Exon 5	585
2	KIF12_cDNA_F2	GCTCTACATCAGCCGCCAAACT	Exon 4	KIF12_cDNA_R2	GGCACCATCACACAGGGACAC	Exon 11	759
3	KIF12_cDNA_F3	GTGCCTCCTCTGTGCCTGCTAC	Exon 11	KIF12_cDNA_R3	ATTTCATCACCTGTTCCTGGCTTC	3' UTR	780
F. ZNF618 gene.							
F-i. Primers used to amplify retinal cDNA.							
1	ZNF618_cDNA_F1	GCAGCCAGAAGAGCACCAAGG	Exon 3	ZNF618_cDNA_R1	TACTGATGGGGGCGTGTAAGG	Exon 10	720
2	ZNF618_cDNA_F2	ACTGACGAGGTGAAGGAGGAACC	Exon 8	ZNF618_cDNA_R2	CTCTTCTACGACAGCAGGTGGTGA	Exon 13	423
3	ZNF618_cDNA_F3	AAAACTCCAGCGAACCCTACACCT	exon 11+12	ZNF618_cDNA_R3	TACACCGTCCTGATTTCCGACA	Exon 14	688
4	ZNF618_cDNA_F4	CGAGGGCAACCACATCAAGAG	Exon 14	ZNF618_cDNA_R4	GCAGGTAATCATACACTTCGTTTTTCC	Exon 14	923
5	ZNF618_cDNA_F5	GCCCTACCAGCACGAGGAGAT	Exon 14	ZNF618_cDNA_R5	CAAGGACAGAGAGCGTCAAGTCAC	3'UTR	970
F-ii. Primers used to amplify genomic DNA.							
1	ZNF618F1	AGTGGGACTGGGGGTTGGTTT	Intron 5	ZNF618R1	GAGAGAGAGGACACCTGGATGGAG	Intron 6	447
2	ZNF618F2	AAGGATGGGGAGGATGACGAG	Intron 6	ZNF618R2	AGAACCCTGTTTGTGGGATGAGTG	Intron 8	519
3	ZNF618F3	GACAGATGATGTCAGGGAGAGAAGC	Intron 8	ZNF618R3	TGCTCAGGAAATGTGTGGTGATCT	Intron 9	455
4	ZNF618F4	GGAGTTGGAGGGAGGGAGCAT	Intron 9	ZNF618R4	GAGGGAACCGTCTTGTCTAGGTGA	Intron 10	498

5	ZNF618F5	GGGTGGCAGGTCCAAATGACTTAG	Intron 10	ZNF618R5	ATGGGAAGAGACCAAGAACAGCAG	Intron 11	664
6	ZNF618F6	AAGTGCAGAGCCCCAGTGTGT	Intron 11	ZNF618R6	CCACAGCCTTCCAGAGCCTTCT	Intron 12	617
7	ZNF618F7	CGCCATGCTGTAGTTCTTGTTTTG	Intron 12	ZNF618R7	TCACAATTCCCAAGGATGAGGTCT	Intron 13	369
8	ZNF618F8	CCAGTTTGTGTGCAAGATGAAACC	Intron 13	ZNF618R8	TTTAGCGCCAGTGTGTTGAAGTTG	Exon 14	537
9	ZNF618F9	CGTGGTGAGTGGGAAGGAGTTC	Exon 14	ZNF618R9	CAGCAGCAACGAGTCTGTGACC	Exon 14	655
10	ZNF618F10	CTGTGCCTTGAACTCGGTGGT	Exon 14	ZNF618R10	GCAGGTAATCATACACTTCGTTTTTCC	Exon 14	725
11	ZNF618F11	GCCCTACCAGCACGAGGAGAT	Exon 14	ZNF618R11	CACGCACACACTGAAAAAGCACT	3' UTR	603
12	ZNF618F12	CACTGTCGCAAACAAAGAAAAGGA	3' UTR	ZNF618R12	TCCTACCAATCTGCACACCTGAAA	3' UTR	404
13	ZNF618F13	CTGAGTGGGAAAGAGGGAAGACAC	3' UTR	ZNF618R13	TTGAGGAACAAAAAGGGCAACTGA	3' UTR	652
14	ZNF618F14	CTCTGCTAGGAGGCACTGAATGTC	3' UTR	ZNF618R14	ACTACTGTCGTGGAAGGAGGCTCA	3' UTR	592
15	ZNF618F15	GGGGTTCTGGTCGTTCTCTGG	3' UTR	ZNF618R15	GGGACACCGTTCTGGTCTACAGTC	3' UTR	605
16	ZNF618F16	TAAGTCCGCTCCATCCCTTCCT	3' UTR	ZNF618R16	CCCAGCAGCAAAGCAAAACAG	3' UTR	521
17	ZNF618F17	GAGCATCCGCATACCCTTTGAT	3' UTR	ZNF618R17	CTGCCACCTTGTCACCGTATCTG	3' UTR	511
18	ZNF618F18	GCAGGTAGAAGGTAGACGCCACTG	3' UTR	ZNF618R18	GAGCCAGCAGAAGACCAAAGAGC	3' UTR	503
19	ZNF618F19	GATGTGTCCTGCTCCCCAGTTCT	3' UTR	ZNF618R19	CTTCCTCGGACGGTTCTCTCCT	3' UTR	636
20	ZNF618F20	GTCTGTGAGCAAGGAGGCTTGG	3' UTR	ZNF618R20	GGTTCACCATAACTGGCCTAGCAA	3' UTR	695
21	ZNF618F21	GTGCTCGCAGAGACGGGAAGA	3' UTR	ZNF618R21	GGGCTGTGCCCCTTAAAGAACT	3' UTR	664
22	ZNF618F22	ACAGAGGAGGCCCTGTTAGGTCTG	3' UTR	ZNF618R22	TTGAGTTTCTTGCACTCGGAGAGA	3' UTR	690
23	ZNF618F23	CACCCCCGTCCTTTCTCAGTAGTT	3' UTR	ZNF618R23	CATCACGAAACTGATTTTGGCTGT	3' UTR	643
24	ZNF618F24	TGTTTGTCCTTTGGGTTTTTCGTT	3' UTR	ZNF618R24	ATGAGGGACTCAGGTCAAGCTCAC	3' UTR	667

Supplement Table 3.2. Primers designed for high-throughput screening of two polymorphisms in AKNA.

Primer pair number	Polymorphism	Location on chromosome 11 (CanFam 2)	Forward primer	Reverse primer	Restriction Enzyme used for digestion	Product size of a normal allele	Product size of an affected allele
1	SNP- A/G (A is the risk allele)	71,624,795 (AKNA, exon 3)	AGTTTGAAGCGGAGGATGTGGAC	GCCTCATCCGGTTCCTCTTCAG	MaeIII	83 bp +26 bp	109 bp
2	21 bp deletion	71,618,401-421 (AKNA intron 4)	GGCCAGAGACAGATCCCCAAAG	CTTTTCCATCAGCACCCCTTTCCT	None	239 bp	218 bp

Supplement Table 3.3. Primers used to amplify through the gap at the 5' end of Col27A1**.**

Primer pair	Forward Primer name	Forward primer sequence	Forward primer location	Reverse Primer name	Reverse primer sequence	Reverse primer location	Expected Size
A. Primers ordered before NGS new sequence data							
1	Col27A1_5_gap_F1	ggacaggctagtgtgcatttgtgt	Pre-gap	Col27A1_5_gap_R1	gcacagggtcttctggagaaagtg	Intron 1	2448
2	Col27A1_5_gap_F2	gcgtgtgactgagtttggggtaag	Pre-gap	Col27A1_5_gap_R2	gctctcgcgttttctctctcacac	Intron 1	1768
3	Col27A1_5_gap_F3	gcctgggtcaaggtgtgcagt	Pre-gap	Col27A1_5_gap_R3	tttctctctcacacgcgcaaaat	Intron 1	1600
4	Col27A1_5_gap_F4	gttgtggttgtgcgaacttgct	Pre-gap	Col27A1_5_gap_R4	actcgctcgccctttacaacc	Intron 1	1783
5	Col27A1_5_gap_F2	gcgtgtgactgagtttggggtaag	Pre-gap	Col27A1_h_exon1_R1	CCGCGCCGATCCCGCTCCCAT	Human exon 1	
6	Col27A1_h_exon1_F1	TCGGCGCGGGGGGCCCGAGGCACAG	Human exon 1	Col27A1_gap_R7	tttacaaccaggccaccagctc	Intron 1	
7	Col27A1_h_exon1_F2	**GCCATGGGAGCGGGATCGGCG**	Human exon 1	Col27A1_5_gap_R4	actcgctcgccctttacaacc	Intron 1	
8	Col27A1_h_exon1_F3	**GCCATGGGAGCGGGATCG**	Human exon 1	Col27A1_5_gap_R2	gctctcgcgttttctctctcacac	Intron 1	
B. Primers ordered based on new sequence was retrieved from DNA-seq and RNA-seq NGS data.							
1	k9exon1F1	GCTCGGGAGCATGAAGTAGGG	Exon 1	newgapR1	gctctcgcgttttctctctcacac	Intron 1	
2	k9exon1F1	GCTCGGGAGCATGAAGTAGGG	Exon 1	newgapR2	GGCTCCCCGGAGCAGGGACGT	Intron 1	
3	newgapF1	GTTGAGCTCCCGACTCTTACC	5' end	K9exon1R1	CCGCTCCCGCTCCCATGG	Exon 1	
4	newgapF1	GTTGAGCTCCCGACTCTTACC	5' end	K9exon1R2	GGGCCCCTACTTCATGCT	Exon 1	
C. Primer combinations of A and B.							
1	Col27A1_5_gap_F3	gcctgggtcaaggtgtgcagt	Pre-gap	K9exon1R2	GGGCCCCTACTTCATGCT	Exon 1	
2	Col27A1_5_gap_F4	gttgtggttgtgcgaacttgct	Pre-gap	K9exon1R1	CCGCTCCCGCTCCCATGG	Exon 1	
3	Col27A1_5_gap_F4	gttgtggttgtgcgaacttgct	Pre-gap	K9exon1R2	GGGCCCCTACTTCATGCT	Exon 1	
4	k9exon1F1	GCTCGGGAGCATGAAGTAGGG	Exon 1	Col27A1_5_gap_R3	tttctctctcacacgcgcaaaat	Intron 1	
5	k9exon1F1	GCTCGGGAGCATGAAGTAGGG	Exon 1	Col27A1_gap_R7	tttacaaccaggccaccagctc	Intron 1	
6	k9exon1F1	GCTCGGGAGCATGAAGTAGGG	Exon 1	Col27A1_5_gap_R4	actcgctcgccctttacaacc	Intron 1	

Supplement Table 3.4. Primers used to amplify conserved regions upstream to Col27A1**.**

Primer pair	Forward primer name	Forward primer sequence	Reverse primer name	Reverse primer sequence	Amplicon location (CanFam2)	Size (bp)
1	**Col27A1_cons_F1**	cttcggcttgggacttttcctaaa	**Col27A1_cons_R1**	acacaaatgcacactagcctgtcc	71443034-71443704	671
2	**Col27A1_cons_F2**	agaagaatgctctgggcagcttt	**Col27A1_cons_R2**	caaggtgtctggtttcagcaaact	71442562-71443080	519
3	**Col27A1_cons_F3**	gccttcctgtgccttgaccac	**Col27A1_cons_R3**	cctgccttgaaacctgagcagt	71429485-71430315	831
4	**Col27A1_cons_F4**	ccccttaaatgccccaggaat	**Col27A1_cons_R4**	gtcctgctccctgtggacctg	71423177-71423549	373
5	**Col27A1_cons_F5**	caggtggacatcgagggagataat	**Col27A1_cons_R5**	ttccattagcagaaggaccagagg	71414326-71415140	815
6	**Col27A1_cons_F6**	ctggccggctctcaagtgatg	**Col27A1_cons_R6**	aattcccagaggggttggggtact	71410672+71411535	864

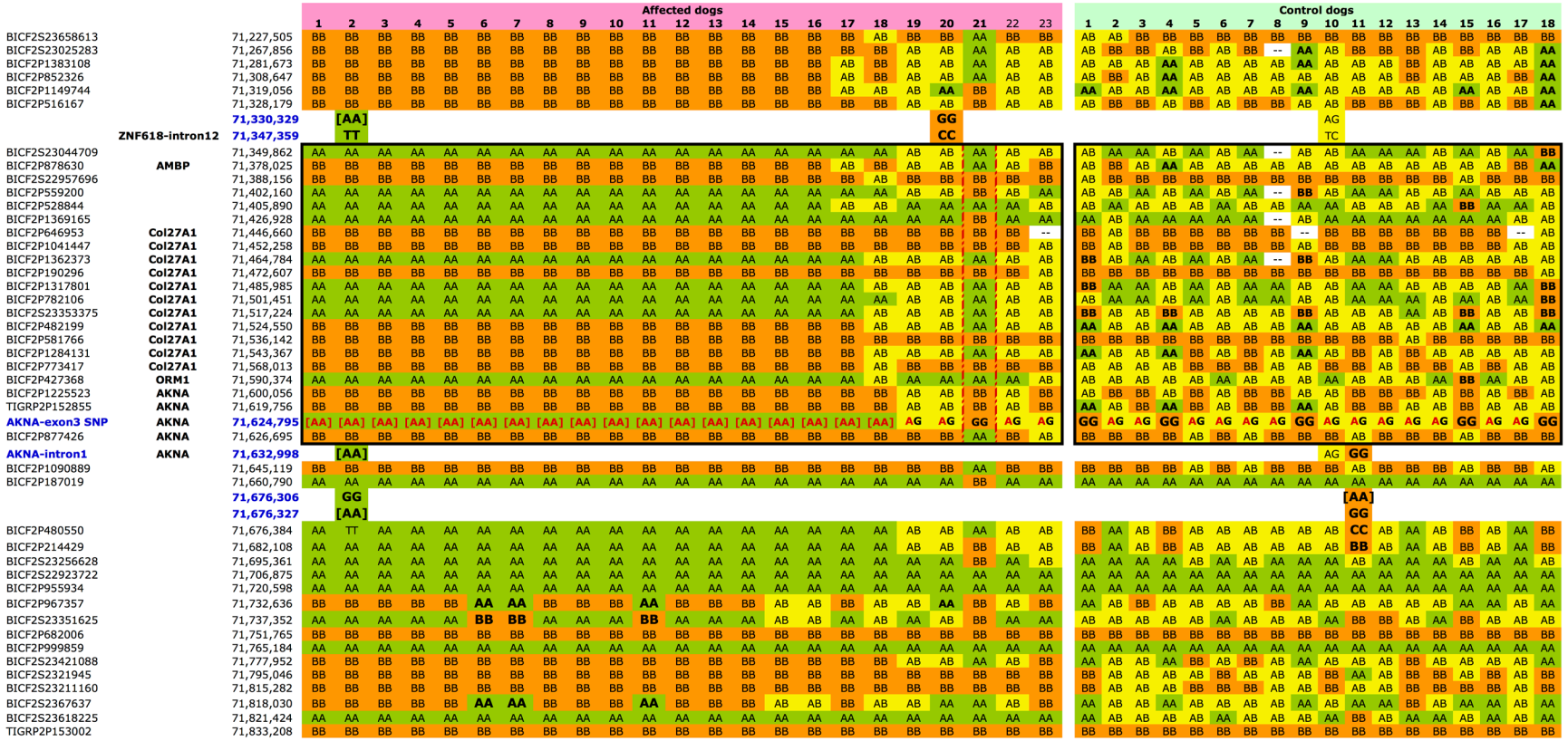
Supplement Table 3.5. Primers used in allele-specific pyrosequencing assay.

Primer ID	Sequence	Location	Length	Amplicon size (bp)	Type	Use_with	Pyro_amplicon_name	Targeted_SNP	CanFam2
2164	/5Biosg/GAGGGGAGCAAAGGGACTCT	Exon 28	20	263	PCR_F_biotin	2165	COL27A1_exon29_SNP68509555	chr11_68509555	71,539,783
2165	TCTCTCCTGGAGGGCCAAGT	Exon 33	20		PCR_R	2164			
2166	CCCCGGCTGGTCCGT	Exon 29	15		Pyro_primer (reverse)	2164-2165			
2167	ACTCGGGCACGCGAGGCT	Exon 31	18	119	PCR_F	2168	COL27A1_exon32_SNP68510591	chr11_68510591	71,540,880
2168	/5Biosg/CTCTCCTGGAGGGCCAAGTG	Exon 33	20		PCR_R_biotin	2167			
2169	GCTTCCCGGGCATCC	Exon 31+32	15		Pyro_primer (forward)	2167-2168			
2170	CTTTTCACCTTCCGGACCCA	Exon 61	20	116	PCR_F	2171	COL27A1_exon61_SNP68546454	chr11_68546454	71,576,758
2171	/5Biosg/AGGAAGCACGCAGGTCCAAC	Exon 61-3UTR	20		PCR_R_biotin	2170			
2172	GAAGCAGTACCGCCT	Exon 61	15		Pyro_primer (forward)	2170-2171			
2173	/5Biosg/CCTGGGAAAGAGGTGAGAGAAGG	Exon 61-3UTR	23	117	PCR_F_biotin	2174	COL27A1_exon61_SNP68547035	chr11_68547035	71,577,339
2174	TTCACAGTGGGAGGTAGCTAGCA	Exon 61-3UTR	23		PCR_R	2173			
2175	ATCCCACCCTCTGGC	Exon 61-3UTR	15		Pyro_primer (reverse)	2173-2174			

Supplement Table 3.6. Homozygosity block in the affected dogs. Genotype calls are denoted as AA, AB or BB. The homozygous block observed in 18 affected dogs is boxed. Dogs 22 and 23 from the affected group were not used in the GWAS due to unclear phenotype.

		Affected dogs																					Control dogs																		
SNP name	Location (CFA11, CanFam2)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
BICF2P1383108	71,281,673	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	AB	AB	AA	AB	AB	AB	AB	AB	AA	AB	AB	AB	AB	AA	AB	AB	AB	BB	AB	AB	AB	AB	AA
BICF2P852326	71,308,647	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	AB	AB	AA	AB	AB	AB	BB	AB	AA	AB	AB	AB	AB	AB	AB	AB	AB	AB	BB	AB	AB	AB	BB	AA
BICF2P1149744	71,319,056	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	AB	AA	BB	AB	AB	AA	AB	AB	AA	AB	AB	AB	AB	AA	AB	AB	AB	AB	AB	AA	AB	AB	AA
BICF2P516167	71,328,179	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	BB	AB	AB	AB	BB	BB	AB	BB	AB	BB	BB	AB	AB	BB	BB	BB	AB	BB	AB	BB	AA
BICF2P1013392	71,339,307	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2S23044709	71,349,862	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AA	AB	AB	AB	AA	AA	AB	AA	AB	AA	--	AB	AB	AA	AA	AA	AB	AA	AB	AA	BB
BICF2P759856	71,360,535	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2P878630	71,378,025	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	AB	AB	AA	AB	BB	AB	BB	AB	AA	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	BB	AA
BICF2S22957696	71,388,156	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	BB	BB	BB	BB	AB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	BB	BB
BICF2P559200	71,402,160	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	BB	AB	AA	AB	AA	AB	AA	AB	AA	AB	AA	--	BB	AB	AA	AA	AB	AB	AA	AB	AB	AB
BICF2P528844	71,405,890	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AA	AA	AA	AA	AB	AB	AA	AB	AB	AB	AA	AB	AB	AA	AA	AB	AB	AB	AA	BB	AA	AA	AB
BICF2P1369165	71,426,928	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AB	AA	AA	AA	AA	AA	AA	--	AB	AA	AA	AA	AA	AA	AA	AB	AB
BICF2P646953	71,446,660	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	AB	BB	BB	BB	BB	BB	BB	--	BB	BB	BB	BB	BB	BB	BB	BB	--	AB
BICF2P646950	71,447,358	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2P1041447	71,452,258	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	AB	BB	BB	BB	BB	BB	BB	AB	BB	BB	BB	BB	BB	BB	BB	AB	AB
BICF2P1362373	71,464,784	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AA	AB	AB	BB	AB	AA	AB	AA	AB	AA	--	BB	AB	AA	AA	AB	AB	AB	AB	AB	AB
BICF2P190296	71,472,607	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB
BICF2P1317801	71,485,985	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AB	AA	AB	AB	BB	AA	AA	AB	AA	AB	AA	AA	AB	AB	AA	AA	AB	AB	AB	AB	AA	BB
BICF2P782106	71,501,451	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AA	AB	AB	AB	AA	AB	AB	AB	AA	AA	AB	AB	AA	AB	AB	AA	AB	AA	AB	AA	BB
BICF2S23353375	71,517,224	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AB	AA	AB	BB	AB	AB	BB	AB	AB	AB	AB	BB	AB	AB	AB	AA	AB	BB	AB	AB	BB
BICF2P482199	71,524,550	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	AB	AB	AA	AB	AB	AA	AB	AB	AA	AB	AB	AB	AB	AA	AB	AB	AB	AB	AB	AA	AB	AB	AA
BICF2P581766	71,536,142	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	BB	BB
BICF2P1284131	71,543,367	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	AA	AB	AB	AA	AB	AB	AA	AB	BB	AB	BB	AB	AA	AB	BB	AB	BB	AB	AB	AB	AB	AB
BICF2P625812	71,555,966	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2P773417	71,568,013	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	BB	BB	BB	BB	BB	AB	AB	AB	AB	BB	BB	BB	AB	AB	BB	BB	AB	BB	BB	AB	BB	AB	AB
BICF2P592354	71,577,551	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2P427368	71,590,374	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AA	AA	AA	AA	AB	AB	AB	AB	AB	AB	AA	AB	AB	AB	AA	AB	AB	AB	AA	BB	AA	AB	AB
BICF2P1225523	71,600,056	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	BB	AB	BB	AB	BB	BB	AB	BB	AB	BB	BB	AB	AB	BB	BB	BB	AB	BB	AB	BB	AB
TIGRP2P152855	71,619,756	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	BB	AB	AB	AA	AB	BB	AA	BB	AB	BB	BB	AA	AB	BB	BB	AA	AB	BB	AB	AB	AB
BICF2P877426	71,626,695	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AA	BB	AB	BB	BB	BB	AB	BB	AB	BB	BB	BB	BB	AB	BB	BB	BB	AB	BB	BB	BB
BICF2P1090889	71,645,119	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AA	BB	BB	BB	BB	BB	AB	BB	AB	BB	BB	BB	AB	BB	BB	BB	AB	BB	BB	BB	AB
BICF2P187019	71,660,790	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
BICF2P480550	71,676,384	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AA	AB	AB	BB	AA	AB	BB	AB	AB	AB	AB	AB	AB	AB	BB	AB	AA	AB	BB	AB	AA	BB
BICF2P214429	71,682,108	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	BB	AB	AB	BB	AA	AB	BB	AB	AB	AB	AB	AB	AB	AB	BB	AB	AA	AB	BB	AB	AA	BB
BICF2S23256628	71,695,361	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	BB	AA	AB	AA	AA	AA	AB	AA	AB	AA	AA	AA	AA	AB	AA	AA	AB	AA	AA	AB	AA	AB
BICF2S22923722	71,706,875	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
BICF2P955934	71,720,598	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
BICF2P967357	71,732,636	BB	BB	BB	BB	BB	AA	AA	BB	BB	BB	AA	BB	BB	BB	AB	AB	BB	AB	AB	AA	BB	AB	BB	AA	AB	BB	AB	AB	AB	AB	BB	AA	AB	AB	AB	AB	AB	AB	AA	AA	AB	AA
BICF2S23351625	71,737,352	AA	AA	AA	AA	AA	BB	BB	AA	AA	AA	BB	AA	AA	AA	AB	AB	AA	AB	AA	AB	BB	AA	AA	AB	AB	AB	AB	AB	AA	AB	AB	AB	AA	BB	BB	AB	AA	BB	AB	AB	AB
BICF2P682006	71,751,765	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2P999859	71,765,184	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
BICF2S23421088	71,777,952	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	AA	AB	BB	AA	AB	AB	AA	AB	BB	AB	BB	AB	AA	AB	AB	AB	AB	AB	AB	AB	AA
BICF2S2321945	71,795,046	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	AB	AB	BB	AB	AB	AB	BB	AA	AB	BB	BB	AB	BB	AB	BB	BB
BICF2S23211160	71,815,282	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	AB	BB	BB	BB	AB	AB	BB	AB	AB	BB	BB	BB	BB	AB	BB
BICF2S2367637	71,818,030	BB	BB	BB	BB	BB	AA	AA	BB	BB	BB	AA	BB	BB	BB	AB	AB	BB	AB	AB	AA	AA	AB	AB	AA	AB	AB	AB	AB	AA	AB	AB	AA	AA	AB	AB	AA	AA	AB	AA	AA	AA	AA
BICF2S23618225	71,821,424	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AB	AB	AB	AB	AA	AB	AB	AB	AA	BB	AB	AA	AA	AB	AA	AB	AA	AA
TIGRP2P153002	71,833,208	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
BICF2P1249136	71,843,346	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	AA	AB	BB	AB	BB	--	AA	AB	AB	AB	AB	--	AA	AB	BB	AB	AB	AB	BB	AB	AB	AB
BICF2P968910	71,855,058	BB	BB	BB	BB	BB	AA	AA	BB	BB	BB	AA	BB	BB	BB	AB	AB	BB	AB	BB	AB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	AB	AB	BB	BB

Supplement Table 3.7. LD block under autosomal dominant with incomplete penetrance mode of inheritance model. Boxed is the LD interval. Dog 21 has a different haplotype not shared with any of the affected. In the control group, dogs 2,3,5-8,10-14, 16 and 17 are heterozygous. [AA] is homozygouse adenine. AA is homozygous to one allele denoted on the SNP array as A (compare to B- the alternate allele). In blue are new SNPs that were not on the SNP array and define the LD block.



Supplement Table 3.8. Number of dogs and breeds screened for AKNA- exon 3 SNP at position 71,624,795. Part of this group was also genotyped for the deletion in intron 4 at position 71,618,401-421.

Number	Breed	Number of dogs
1	American Cocker Spaniel	5
2	American Eskimo	5
3	Australian Cattle dog	13
4	Australian Shepherd	5
5	Basenji	5
6	Berger Picard	5
7	Border Collie	6
8	Borzoi	39
9	Chesapeake Bay Retriever	4
10	Chinese Crested	11
11	Corgi, Welsh Pembroke	4
12	Elkhound	1
13	English Cocker Spaniel	13
14	English Mastiff	1
15	English Springer Spaniel	5
16	Entlebucher Mountain Dog	5
17	Glen of Imaal Terrier	9
18	Golden Retriever	7
19	Gordon Setter	5
20	Greyhound	9
21	Ibizan hound	2
22	Irish Wolfhound	14
23	Kuvasz	2
24	Labrador Retriever	5
25	Miniature Pinscher	8
26	Nova Scotia Duck Tolling Retreiver	5
27	Old English Sheepdog	9
28	Papilion	6
29	Pomeranian	8
30	Portuguese Water Dog	6
31	Shetland Sheepdog	8
32	Siberian Husky	5
33	Soft Coated Wheaton Terrier	5
34	Tibetan Terrier	5
35	Toy Manchester Terrier	2
36	Whippet	22
	Total	**269**

Supplement Table 3.9. SNPs identified in the LD interval that have the potential to be causative of IGPRA.

Location	Reference allele	Alternate allele	IG73 (F1 heterozygous dog)		Normal dog		Affected dog	
			Genotype[a]	Alleles counts[b]	Genotype[a]	Alleles counts[b]	Genotype[a]	Alleles counts[b]
68,317,690	T	C	1/1	0,6	0/0	6,0	1/1	0,9
68,317,704	C	T	1/1	0,7	0/0	6,0	1/1	0,9
68,318,054	T	G	1/1	0,9	0/0	7,0	1/1	0,7
68,318,079	C	G	1/1	0,7	0/0	6,0	1/1	0,6
68,318,088	T	C	1/1	0,7	0/0	5,0	1/1	0,8
68,318,157	A	T	1/1	0,6	0/0	3,0	1/1	0,12
68,319,399	G	A	1/1	0,10	0/1	5,1	1/1	0,18
68,319,646	T	G	1/1	0,5	0/0	5,0	1/1	0,11
68,319,656	A	G	1/1	0,4	0/0	5,0	1/1	0,13
68,319,900	A	G	1/1	0,3	0/0	7,0	1/1	0,9
68,319,959	C	T	1/1	0,5	0/0	7,0	1/1	0,5
68,319,984	A	G	1/1	0,7	0/0	5,0	1/1	0,3
68,320,066	A	G	1/1	0,7	0/0	4,0	1/1	0,5
68,320,147	C	T	1/1	0,1	0/0	5,0	1/1	0,8
68,320,148	A	G	1/1	0,1	0/0	3,0	1/1	0,8
68,320,207	C	A	1/1	0,3	0/0	3,0	1/1	0,8
68,320,348	T	G	1/1	0,9	0/0	4,0	1/1	0,10
68,320,349	A	T	1/1	0,9	0/0	5,0	1/1	0,10
68,320,376	A	T	1/1	0,7	0/0	3,0	1/1	0,8
68,320,451	C	T	1/1	0,8	0/0	7,0	1/1	1,6
68,321,029	G	A	1/1	0,15	0/0	10,0	1/1	0,20
68,321,161	C	A	1/1	0,11	0/0	8,0	1/1	0,7
68,321,173	T	A	1/1	0,13	0/0	8,0	1/1	0,7
68,322,467	T	C	1/1	0,15	0/0	11,0	1/1	0,21
68,322,883	A	T	1/1	2,5	0/0	5,0	1/1	0,13
68,323,301	T	A	1/1	0,5	0/0	14,0	1/1	0,12
68,323,310	T	G	1/1	0,7	0/0	14,0	1/1	0,11
68,323,312	T	C	1/1	0,7	0/0	13,0	1/1	0,11
68,323,566	T	C	1/1	0,14	0/0	10,0	1/1	0,8
68,323,658	T	C	1/1	0,10	0/0	12,0	1/1	0,12
68,323,746	T	G	1/1	0,12	0/0	11,0	1/1	0,9
68,323,946	T	A	1/1	0,8	0/0	13,0	1/1	0,3
68,323,947	T	A	1/1	0,8	0/0	12,0	1/1	0,3
68,324,835	T	A	1/1	0,7	0/0	5,0	1/1	0,15
68,324,854	A	C	1/1	0,6	0/0	5,0	1/1	0,13
68,325,555	C	T	1/1	0,14	0/0	4,0	1/1	0,10
68,328,583	A	C	1/1	0,10	0/0	8,0	1/1	0,7
68,328,584	G	A	1/1	0,10	0/0	8,0	1/1	0,7

68,333,381	T	C	1/1	0,4	0/0	5,0	1/1	0,10
68,334,388	C	T	1/1	0,8	0/0	6,0	1/1	0,10
68,334,405	C	T	1/1	0,7	0/0	6,0	1/1	0,14
68,334,864	G	T	1/1	0,3	0/0	5,0	1/1	0,3
68,335,122	G	C	1/1	0,1	./.		1/1	0,1
68,335,501	A	G	1/1	0,1	./.		1/1	0,2
68,337,445	C	T	1/1	0,1	0/0	2,0	1/1	0,4
68,337,534	C	G	1/1	0,2	0/0	3,0	1/1	0,4
68,337,768	G	C	1/1	0,2	0/0	2,0	1/1	0,4
68,338,478	A	G	1/1	0,6	0/0	13,0	1/1	0,7
68,338,634	C	A	1/1	0,3	0/0	5,0	1/1	0,4
68,338,658	A	G	1/1	0,1	0/0	5,0	1/1	0,1
68,338,693	A	C	1/1	0,1	0/0	2,0	1/1	0,1
68,338,922	T	C	1/1	0,3	0/0	2,0	./.	
68,338,941	T	C	1/1	0,2	0/0	2,0	./.	
68,339,094	T	C	1/1	0,3	0/0	4,0	1/1	0,2
68,339,327	A	T	1/1	0,7	0/0	3,0	1/1	0,4
68,339,639	T	C	1/1	0,1	0/0	5,0	1/1	0,4
68,339,743	T	C	1/1	0,2	0/0	3,0	1/1	0,3
68,339,749	T	C	1/1	0,2	0/0	4,0	1/1	0,3
68,339,772	A	G	1/1	0,2	0/0	4,0	1/1	0,3
68,340,011	T	C	1/1	0,2	0/0	2,0	1/1	0,7
68,340,119	T	C	1/1	0,9	0/0	7,0	1/1	0,9
68,340,134	G	A	1/1	0,9	0/0	3,0	1/1	1,9
68,340,594	C	A	1/1	0,10	0/0	6,0	1/1	0,7
68,340,942	G	A	1/1	0,4	0/0	3,0	1/1	0,8
68,341,349	T	A	1/1	0,9	0/0	15,0	1/1	0,11
68,341,368	G	A	1/1	0,6	0/0	16,0	1/1	0,16
68,341,410	A	G	1/1	0,9	0/0	17,0	1/1	0,9
68,341,793	T	C	1/1	0,10	0/0	8,0	1/1	0,10
68,342,037	G	A	1/1	0,4	0/0	14,0	1/1	0,5
68,342,056	G	A	1/1	0,5	0/0	10,0	1/1	0,8
68,342,362	C	G	1/1	0,8	0/0	6,0	1/1	0,18
68,342,753	C	A	1/1	0,9	0/0	7,0	1/1	0,14
68,342,799	A	G	1/1	0,9	0/0	6,0	1/1	0,18
68,342,933	G	A	1/1	0,5	0/0	10,0	1/1	1,11
68,343,259	C	G	1/1	0,10	0/0	9,0	1/1	0,7
68,343,327	C	T	1/1	0,8	0/0	6,0	1/1	0,13
68,344,047	A	G	1/1	0,8	0/0	8,0	1/1	0,7
68,344,432	C	A	1/1	0,10	0/0	6,0	1/1	0,10
68,344,535	G	A	1/1	0,6	0/0	9,1	1/1	0,13
68,344,657	G	T	1/1	0,4	0/0	4,0	1/1	0,3
68,344,872	C	T	1/1	0,10	0/0	2,0	1/1	0,11
68,345,206	A	G	1/1	0,5	0/0	6,0	1/1	0,6
68,345,244	T	C	1/1	0,4	0/0	6,0	1/1	0,10

68,345,320	A	G	1/1	0,8	0/0	13,0	1/1	0,12
68,345,374	C	T	1/1	0,10	0/0	9,0	1/1	0,15
68,345,652	A	G	1/1	0,6	0/0	11,0	1/1	0,5
68,345,809	T	C	1/1	0,5	0/0	2,0	1/1	0,9
68,345,839	A	G	1/1	0,7	0/0	5,0	1/1	0,12
68,346,457	T	C	1/1	0,4	0/0	1,0	1/1	0,11
68,346,520	C	T	1/1	0,3	0/0	2,0	1/1	0,12
68,346,830	C	T	1/1	0,4	0/0	5,0	1/1	0,6
68,346,844	T	A	1/1	0,4	0/0	6,0	1/1	0,6
68,346,858	A	G	1/1	0,4	0/0	6,0	1/1	0,5
68,346,901	T	C	1/1	0,2	0/0	12,0	1/1	0,2
68,347,388	G	T	1/1	0,5	0/0	1,0	1/1	0,5
68,347,402	A	G	1/1	0,4	0/0	2,0	1/1	0,5
68,347,562	T	C	1/1	0,7	0/0	6,0	1/1	0,6
68,347,747	T	G	1/1	0,6	0/0	8,0	1/1	0,15
68,347,781	T	C	1/1	0,8	0/0	5,0	1/1	0,14
68,348,161	G	C	1/1	0,2	0/0	5,0	1/1	0,3
68,348,188	T	C	1/1	0,2	0/0	3,0	1/1	0,3
68,348,732	C	T	1/1	0,7	0/0	8,0	1/1	0,10
68,348,789	A	G	1/1	0,2	0/0	7,0	1/1	0,6
68,348,842	C	T	1/1	0,5	0/0	6,0	1/1	0,3
68,349,117	T	C	1/1	0,7	0/0	7,0	1/1	0,3
68,349,222	A	T	1/1	0,9	0/0	4,0	1/1	0,5
68,349,328	T	C	1/1	0,2	0/0	5,0	1/1	0,5
68,349,703	A	G	1/1	0,5	0/0	2,0	1/1	0,9
68,349,717	T	C	1/1	0,8	0/0	2,0	1/1	0,10
68,349,921	G	A	1/1	0,4	0/0	2,0	1/1	0,8
68,350,238	G	A	1/1	0,3	0/0	4,0	1/1	0,3
68,350,289	T	C	1/1	0,5	0/0	4,0	1/1	0,2
68,350,295	G	T	1/1	0,6	0/0	5,0	1/1	0,3
68,350,396	T	C	1/1	0,4	0/0	4,0	1/1	0,2
68,350,409	C	T	1/1	0,5	0/0	5,0	1/1	0,3
68,350,563	C	T	1/1	0,1	0/0	5,0	1/1	0,1
68,350,938	A	G	1/1	0,2	./.		1/1	0,4
68,350,955	G	A	1/1	0,1	./.		1/1	0,3
68,351,005	C	G	1/1	0,2	0/0	3,0	1/1	0,3
68,351,027	C	A	1/1	0,2	0/0	3,0	1/1	0,5
68,351,209	G	A	1/1	0,1	0/0	1,0	1/1	0,1
68,351,220	T	G	1/1	0,1	0/0	1,0	1/1	0,1
68,351,303	T	C	1/1	0,1	0/0	5,0	1/1	0,7
68,351,345	G	A	1/1	0,2	0/0	4,0	1/1	0,4
68,351,360	G	A	1/1	0,2	0/0	4,0	1/1	0,4
68,351,367	T	C	1/1	0,2	0/0	4,0	1/1	0,4
68,351,379	G	A	1/1	0,2	0/0	2,0	1/1	0,4
68,351,526	G	A	1/1	0,4	0/0	1,0	1/1	0,3

68,351,582	G	C	1/1	0,3	0/0	1,0	1/1	0,5
68,351,821	G	A	1/1	0,3	0/0	1,0	1/1	0,4
68,356,000	T	C	1/1	0,8	0/0	7,0	1/1	0,9
68,356,673	A	G	1/1	0,9	0/0	7,0	1/1	1,4
68,357,477	A	G	1/1	0,8	0/0	2,0	1/1	0,7
68,357,573	C	T	1/1	0,9	0/0	4,0	1/1	0,11
68,357,584	C	T	1/1	0,10	0/0	4,0	1/1	0,11
68,357,827	C	T	1/1	0,2	0/0	4,0	1/1	0,7
68,358,060	A	T	1/1	0,7	0/0	2,0	1/1	0,18
68,358,243	A	G	1/1	0,5	0/0	2,0	1/1	0,9
68,358,379	A	G	1/1	0,4	0/0	1,0	1/1	0,7
68,358,465	G	A	1/1	0,7	0/0	2,0	1/1	0,3
68,358,621	C	T	1/1	0,5	0/0	1,0	1/1	0,4
68,358,736	C	T	1/1	0,2	0/0	2,0	1/1	0,3
68,358,800	G	T	1/1	0,3	0/0	1,0	1/1	0,4
68,359,245	A	G	1/1	0,5	0/0	14,0	1/1	0,5
68,359,326	T	C	1/1	0,10	0/0	8,0	1/1	0,7
68,359,387	C	T	1/1	0,7	0/0	7,0	1/1	0,7
68,359,412	A	G	1/1	0,7	0/0	5,0	1/1	0,8
68,359,422	T	C	1/1	0,8	0/0	5,0	1/1	0,7
68,359,526	C	T	1/1	0,4	0/0	5,0	1/1	0,5
68,359,622	T	C	1/1	0,10	0/0	8,0	1/1	0,7
68,359,671	C	T	1/1	0,10	0/0	4,0	1/1	0,10
68,359,760	C	G	./.		0/0	2,0	1/1	0,6
68,359,773	C	T	./.		0/0	2,0	1/1	0,5
68,359,783	G	A	1/1	0,1	0/0	1,0	1/1	0,7
68,360,443	C	G	1/1	0,3	0/0	2,0	1/1	0,4
68,360,627	C	T	1/1	0,4	0/0	8,0	1/1	0,8
68,360,669	A	G	1/1	0,3	0/1	10,1	1/1	0,10
68,360,696	T	C	1/1	0,1	0/1	8,1	1/1	0,9
68,361,086	C	G	./.		./.		1/1	0,2
68,361,385	G	T	1/1	0,2	0/0	3,0	1/1	0,2
68,361,425	C	A	1/1	0,2	0/0	1,0	1/1	1,2
68,362,558	C	T	1/1	0,5	./.		1/1	0,4
68,362,585	G	T	1/1	0,3	./.		1/1	0,5
68,362,767	G	A	1/1	0,5	0/0	2,0	1/1	0,11
68,363,146	T	C	1/1	0,3	./.		1/1	0,1
68,364,390	C	T	1/1	0,3	0/0	4,0	1/1	0,6
68,364,646	C	T	1/1	0,5	0/0	4,0	1/1	0,8
68,364,814	C	T	1/1	0,4	0/0	5,0	1/1	0,5
68,364,912	G	A	1/1	0,5	0/0	4,0	1/1	0,7
68,365,034	G	A	1/1	0,3	0/0	5,0	1/1	0,4
68,365,055	G	A	1/1	0,6	0/0	5,0	1/1	0,6
68,367,586	A	G	1/1	0,2	0/0	3,0	1/1	0,3
68,370,040	G	A	1/1	0,9	0/0	6,0	1/1	0,15

68,370,526	A	G	1/1	0,6	0/0	4,0	1/1	0,11
68,370,984	C	T	1/1	0,2	0/0	4,0	1/1	0,8
68,371,086	C	T	1/1	0,11	0/0	6,0	1/1	0,13
68,371,164	A	T	1/1	0,7	0/0	7,0	1/1	0,7
68,371,564	T	G	1/1	0,5	0/0	4,0	1/1	0,10
68,375,183	G	A	./.		0/0	1,0	1/1	0,5
68,375,952	C	T	1/1	0,7	0/0	4,0	1/1	0,9
68,378,360	C	T	1/1	0,12	0/0	5,0	1/1	0,7
68,389,284	T	C	1/1	0,4	0/0	9,0	1/1	0,5
68,389,455	G	A	1/1	0,3	0/0	4,0	1/1	0,6
68,397,694	T	C	1/1	0,5	0/0	5,0	1/1	0,8
68,413,798	A	G	./.		./.		1/1	0,2
68,413,799	A	C	./.		./.		1/1	0,2
68,413,800	A	G	./.		./.		1/1	0,2
68,416,922	G	A	1/1	0,4	0/0	3,0	1/1	0,9
68,418,424	C	T	1/1	0,13	0/0	2,0	1/1	1,10
68,418,565	G	A	1/1	0,5	./.		1/1	0,10
68,419,840	T	C	1/1	0,2	./.		1/1	0,3
68,419,888	G	A	1/1	0,2	./.		1/1	0,3
68,427,362	C	G	1/1	0,9	0/0	6,0	1/1	0,5
68,427,367	A	C	1/1	0,9	0/0	4,0	1/1	0,4
68,428,111	A	G	1/1	0,7	0/0	10,0	1/1	0,4
68,428,119	A	T	1/1	0,6	0/0	9,0	1/1	0,4
68,429,781	G	A	./.		0/0	7,0	1/1	0,12
68,429,865	A	G	1/1	0,5	0/0	7,1	1/1	0,14
68,429,992	C	T	1/1	0,3	0/1	8,1	1/1	0,17
68,430,343	C	T	1/1	0,5	0/0	7,0	1/1	0,6
68,430,655	C	G	1/1	0,6	0/0	1,0	1/1	0,5
68,430,690	G	T	1/1	0,5	0/0	1,0	1/1	0,6
68,430,724	C	T	1/1	0,3	0/0	1,0	1/1	0,3
68,431,413	C	T	1/1	0,11	0/0	4,0	1/1	0,7
68,431,481	G	A	1/1	0,8	0/0	5,0	1/1	0,3
68,431,578	T	C	1/1	0,4	0/0	9,0	1/1	0,7
68,431,620	T	C	./.		0/0	8,0	1/1	0,5
68,432,110	T	A	1/1	0,7	./.		1/1	0,5
68,447,605	G	A	0/1	1,10	0/0	9,0	1/1	0,10
68,452,062	C	T	1/1	0,5	0/0	8,0	1/1	0,11
68,453,305	C	T	1/1	0,6	0/0	8,0	1/1	0,3
68,455,000	C	T	1/1	0,5	0/0	3,0	1/1	0,8
68,456,319	T	C	1/1	0,13	0/0	5,0	1/1	0,2
68,458,749	C	T	0/1	3,2	0/0	4,0	1/1	0,13
68,461,635	T	C	1/1	0,7	0/0	2,0	1/1	0,5
68,462,663	C	T	1/1	0,5	0/0	3,0	1/1	0,11
68,465,132	C	T	0/1	5,3	0/0	1,0	1/1	0,7
68,466,760	C	G	1/1	1,2	0/0	7,0	1/1	0,10

68,467,124	C	T	0/1	6,1	0/0	4,0	1/1	0,10
68,468,160	G	A	0/0	3,1	./.		1/1	0,6
68,469,058	C	A	0/1	7,1	0/0	2,1	1/1	0,3
68,469,105	C	G	1/1	0,6	0/0	4,0	1/1	0,3
68,469,267	C	T	0/1	4,3	0/0	5,0	1/1	0,1
68,470,478	C	T	1/1	0,8	0/0	5,0	1/1	0,6
68,472,696	A	T	0/1	4,3	0/0	9,0	1/1	0,10
68,472,715	A	C	1/1	1,2	0/0	6,0	1/1	0,8
68,472,795	T	C	0/1	2,1	0/0	3,0	1/1	0,11
68,472,804	T	C	0/1	2,1	0/0	4,0	1/1	0,11
68,473,119	G	A	0/1	3,3	0/0	6,0	1/1	0,8
68,473,268	T	C	0/1	2,1	0/0	5,0	1/1	0,13
68,475,417	G	A	1/1	0,11	0/0	4,0	1/1	0,9
68,475,589	G	A	1/1	0,9	0/0	11,0	1/1	0,13
68,476,026	C	A	1/1	0,6	0/0	6,0	1/1	0,5
68,478,047	T	C	1/1	0,3	0/0	4,0	1/1	0,4
68,478,536	C	T	1/1	0,5	0/0	3,0	1/1	0,5
68,478,735	C	T	1/1	0,2	0/0	4,0	1/1	0,7
68,479,093	C	T	1/1	0,2	0/0	9,0	1/1	0,11
68,479,546	C	G	1/1	0,6	0/0	3,0	1/1	0,3
68,481,860	C	T	1/1	0,2	0/0	6,0	1/1	0,5
68,483,358	A	G	1/1	0,6	0/0	1,0	1/1	0,5
68,484,044	C	T	1/1	0,3	0/0	5,0	1/1	1,8
68,485,091	C	T	1/1	0,12	0/0	6,0	1/1	0,9
68,485,254	C	T	1/1	0,5	0/0	11,0	1/1	0,5
68,485,540	C	T	1/1	0,7	0/0	4,0	1/1	0,5
68,485,600	C	T	1/1	0,8	0/0	5,0	1/1	0,7
68,485,747	T	C	1/1	0,11	0/0	7,1	1/1	0,8
68,485,757	C	T	1/1	1,12	0/0	7,0	1/1	0,8
68,485,797	A	T	1/1	0,12	0/0	5,0	1/1	0,6
68,485,945	G	T	1/1	1,4	./.		1/1	1,4
68,486,633	G	C	0/1	3,3	0/0	5,0	1/1	1,5
68,486,742	C	G	1/1	0,4	0/0	5,0	1/1	0,5
68,487,732	C	T	0/0	4,0	0/0	6,0	1/1	0,3
68,489,176	G	A	1/1	0,4	0/0	4,0	1/1	0,5
68,489,177	T	A	1/1	0,4	0/0	4,0	1/1	0,5
68,489,191	A	G	1/1	0,4	0/0	3,0	1/1	0,5
68,489,199	T	C	1/1	0,3	0/0	2,0	1/1	0,6
68,489,351	C	T	0/1	4,2	0/0	5,0	1/1	0,4
68,490,035	C	T	0/1	2,3	0/0	5,0	1/1	0,15
68,490,885	G	A	1/1	0,3	0/0	9,0	1/1	0,5
68,492,423	G	A	0/1	5,2	0/0	5,0	1/1	1,8
68,493,634	A	C	0/1	5,2	0/0	7,0	1/1	0,3
68,494,321	G	C	0/1	4,2	0/0	10,0	1/1	0,6
68,494,989	A	G	0/1	5,4	0/0	9,0	1/1	0,16

68,496,142	T	C	./.		0/0	1,0	1/1	0,8
68,496,601	C	T	0/1	4,3	0/0	5,0	1/1	0,6
68,497,722	T	C	0/1	5,3	0/0	6,0	1/1	0,6
68,498,417	C	G	0/1	1,4	0/0	5,0	1/1	0,3
68,509,606	C	G	0/1	2,1	0/0	4,0	1/1	0,3
68,509,608	C	A	0/1	3,1	0/0	4,0	1/1	0,3
68,509,853	G	A	./.		./.		1/1	0,3
68,509,889	C	T	./.		./.		1/1	0,5
68,509,962	C	T	0/0	2,0	0/0	3,0	1/1	1,7
68,510,591	C	T	./.		0/0	4,0	1/1	0,6
68,511,661	T	G	0/0	1,0	./.		1/1	0,3
68,513,145	G	A	0/1	4,1	0/0	3,0	1/1	0,7
68,514,803	T	C	0/1	3,5	0/0	4,0	1/1	0,13
68,527,629	C	T	0/1	4,8	0/0	4,0	1/1	0,1
68,527,639	C	T	0/1	4,8	0/0	4,0	1/1	0,6
68,530,931	G	A	1/1	1,5	0/0	4,0	1/1	0,8
68,531,510	T	C	0/1	4,2	1/1	1,1	1/1	0,8
68,546,406	T	C	1/1	0,2	./.		1/1	0,8
68,547,672	C	T	0/1	1,1	0/0	8,0	1/1	1,18
68,547,749	T	C	0/1	2,4	0/0	8,0	1/1	0,20
68,548,204	C	T	0/1	1,1	0/0	1,0	1/1	0,2
68,548,517	A	G	./.		./.		1/1	0,9
68,548,519	C	G	./.		./.		1/1	0,10
68,548,531	G	C	./.		./.		1/1	3,7
68,548,737	G	A	0/1	4,5	0/0	6,0	1/1	0,6
68,549,160	G	C	0/1	1,5	0/0	5,0	1/1	0,12
68,549,281	G	C	1/1	0,6	0/0	4,0	1/1	0,5
68,549,306	A	T	1/1	0,5	0/0	3,0	1/1	0,4
68,549,354	T	C	0/1	1,4	0/0	2,0	1/1	0,3
68,549,396	A	G	0/1	3,2	0/0	6,0	1/1	0,6
68,549,415	C	A	0/1	4,2	0/0	6,0	1/1	0,6
68,549,533	C	T	0/1	3,4	0/0	5,0	1/1	0,11
68,550,237	A	G	0/1	1,3	0/0	4,0	1/1	0,9
68,550,292	C	T	0/1	2,5	0/0	7,0	1/1	0,11
68,552,042	C	T	0/1	1,3	0/0	4,0	1/1	0,9
68,552,043	C	G	0/1	1,3	0/0	4,0	1/1	0,9
68,552,213	T	C	0/1	1,3	0/0	5,0	1/1	0,2
68,552,593	G	A	0/1	5,7	0/0	3,0	1/1	0,7
68,552,609	C	T	0/1	5,6	0/0	4,0	1/1	0,6
68,552,673	A	C	0/1	7,3	0/0	2,1	1/1	0,11
68,554,763	C	T	1/1	0,1	0/0	6,0	1/1	0,6
68,555,385	T	C	1/1	0,2	0/0	2,1	1/1	0,1
68,555,826	G	A	1/1	0,1	0/0	4,0	1/1	0,1
68,559,477	A	G	1/1	0,3	./.		1/1	0,2
68,559,881	T	C	0/1	1,1	./.		1/1	0,9

68,560,024	C	T	0/1	2,3	0/0	4,0	1/1	0,8
68,566,184	G	A	0/1	4,1	0/0	4,0	1/1	0,6
68,566,402	C	T	1/1	0,7	0/0	1,0	1/1	0,6
68,566,636	T	C	1/1	0,4	0/0	5,0	1/1	0,1
68,566,680	G	A	1/1	0,2	0/0	8,0	1/1	0,5
68,567,387	C	T	1/1	0,2	0/0	3,0	1/1	0,9
68,568,735	C	T	0/1	1,4	0/0	5,0	1/1	0,14
68,574,189	C	A	1/1	0,1	0/0	4,0	1/1	0,2
68,575,321	G	A	0/1	4,7	0/0	2,0	1/1	0,7
68,575,350	T	C	1/1	1,13	0/0	2,0	1/1	0,8
68,575,477	C	G	1/1	0,5	0/0	5,0	1/1	0,12
68,575,502	C	T	1/1	0,4	0/0	3,0	1/1	0,10
68,575,770	G	A	1/1	0,2	0/0	8,0	1/1	1,5
68,575,934	T	G	1/1	0,5	0/0	6,0	1/1	0,15
68,576,010	T	C	1/1	0,4	./.		1/1	0,6
68,576,518	G	A	0/1	6,4	0/0	12,0	1/1	0,7
68,576,764	C	T	0/1	9,4	0/0	9,0	1/1	0,10
68,577,616	C	T	1/1	1,1	0/0	1,0	1/1	0,2
68,577,629	G	A	1/1	0,2	0/0	1,0	1/1	0,2
68,577,649	G	A	1/1	0,4	0/0	1,0	1/1	0,3
68,577,690	C	T	1/1	0,5	0/0	8,0	1/1	0,5
68,578,674	T	C	0/1	1,3	0/0	2,0	1/1	0,8
68,578,990	C	T	0/1	2,3	0/0	3,0	1/1	0,8
68,579,013	G	A	0/1	2,2	0/0	3,0	1/1	0,10
68,579,702	C	G	0/1	2,1	0/0	1,0	1/1	0,6
68,580,200	C	T	0/1	3,2	0/0	3,0	1/1	0,8
68,580,248	G	A	0/1	2,1	0/0	1,0	1/1	0,6
68,580,425	A	T	1/1	2,1	0/0	1,0	1/1	0,2
68,580,940	C	G	0/1	4,4	0/0	4,0	1/1	0,5
68,584,468	T	C	1/1	0,2	./.		1/1	0,5
68,584,470	T	C	1/1	0,2	./.		1/1	0,5
68,584,492	C	A	1/1	0,3	./.		1/1	0,3
68,584,498	C	G	1/1	0,3	./.		1/1	0,3
68,584,505	C	G	1/1	0,2	./.		1/1	0,3
68,584,511	C	G	1/1	0,2	./.		1/1	0,2
68,589,035	T	C	1/1	0,5	./.		1/1	0,2
68,594,491	C	T	1/1	0,2	0/0	5,0	1/1	0,13
68,598,729	C	T	0/1	2,4	0/0	4,0	1/1	1,7
68,598,830	G	A	0/1	4,11	0/0	3,0	1/1	0,5
68,599,310	A	G	0/1	4,3	0/0	6,0	1/1	0,6
68,599,523	G	C	0/1	2,1	0/0	4,0	1/1	0,5
68,600,027	C	T	0/1	1,7	0/0	10,0	1/1	0,13
68,600,292	C	T	1/1	0,4	0/0	2,0	1/1	0,6
68,602,393	G	C	0/1	3,1	0/0	7,0	1/1	0,12
68,602,394	G	C	0/1	3,1	0/0	7,0	1/1	0,12

68,602,930	C	T	1/1	0,6	./.		1/1	0,6
68,603,010	A	C	1/1	0,10	./.		1/1	0,12
68,603,033	C	T	0/1	5,7	./.		1/1	0,10

[a] - 0 is the reference allele, 1 is the alternate allele. 0/0 is homozygous to the reference allele (only reference allele is observed in the reads), 0/1 is heterozygous (both alleles are observed in the reads) and 1/1 is homozygous to alternate allele (only alternate allele is observed in the reads). ./. = reads were not observed.
[b] - The first number is the number of reads with the reference allele, and the second number is the number of reads with the alternate allele.

Supplement Table 3.10. Relative allele expression for informative dogs within the LD interval from RNA-seq data.

SNP_ID	CanFam3 position	ref allele	alter allele	Total counts			canFam2 pos	SNP alleles						SNP counts in informative individuals +						Risk allele
				coverage	ref count	alter count		IG72	IG73	IG74	Control 1	Control 2	Control 3	IG72	IG73	IG74	Control 1	Control 2	Control 3	
ZNF618																				
rs22165000	68,272,447	A	G	48	4	44	71302910	G/G	G/G	G/G	A/G?	G/G	G/G	*	*	*	4/2	*	*	G
rs22164999	68,272,466	A	G	39	32	7	71302929	A/A	A/A	A/G?	A/A?	G/G?	G/G?	*	*	3/1	*	*	*	A
chr11_68299661	68,299,661	A	G	41	11	30	71330124	G/G	G/G	A/G	A/A	A/A	G/G	*	*	3/3	*	*	*	G
chr11_68332484	68,332,484	G	A	77	66	10	71362947	G/G	G/G	G/A	G/G	G/A	G/G	*	*	6/3	*	10/7	*	G
chr11_68332860	68,332,860	C	T	47	38	9	71363323	C/C	C/C	C/C	C/C	C/C	T/T	*	*	*	*	*	*	C
rs8968699	68,333,123	C	T	47	34	13	71363586	C/C	C/C	C/C	C/T	C/C	T/T	*	*	*	3/4	*	*	C
rs8968700	68,333,381	T	C	62	4	58	71363844	C/C	C/C	C/C	C/T	C/C	C/C	*	*	*	4/3	*	*	C
	Average			51.57			Informative?	NO	NO	YES	YES	YES	NO							
													Total			12/7	8/12	7/10		
													% of risk haplotype ^			63%	40%	41%		
ORM1																				
rs9188097	68,557,019	T	C	37	9	28	71587323	C/C	C/C	C/C	T/C?	T/C?	T/T	*	*	*	1/3	2/2	*	C
chr11_68557358	68,557,358	G	A	40	27	13	71587662	G/A?	G/A?	G/A?	G/A?	G/G	G/G	3/8	3/3	2/1	1/1	*	*	A
chr11_68557816	68,557,816	G	A	54	40	14	71588120	G/G	G/G	G/G	G/G	G/G	A/A	*	*	*	*	*	*	G
chr11_68558082	68,558,082	C	T	31	24	7	71588386	C/T?	C/T?	C/T?	NA	C/C	C/C	1/3	2/3	3/1	*	*	*	C
chr11_68558749	68,558,749	A	G	35	13	22	71589053	G/G	G/G	G/G	NA	A/A	A/A	*	*	*	*	*	*	G
chr11_68559346	68,559,346	C	T	46	19	27	71589650	T/T	T/T	T/T	NA	C/C	C/C	*	*	*	*	*	*	T
chr11_68559366	68,559,366	G	A	49	19	30	71589670	A/A	A/A	A/A	NA	G/G	G/G	*	*	*	*	*	*	A
	Average			41.71			Informative?	YES	YES	YES	YES	YES	NO							
													Total	9/6	5/6	4/3	2/4	2/2		
													% of risk haplotype ^	60%	45%	57%	33%	50%		
COL27A1																				
chr11_68425869	68,425,869	A	G	94	33	61	71456098	A/A	A/A	A/G	G/G	G/G	G/G	*	3/8	*	*	*	*	A
chr11_68509555	68,509,555	G	A	97	64	33	71539783	A/G	A/A?	A/G	G/G	A/G	G/G	2/9	*	2/12	*	8/6	*	G
chr11_68510591	68,510,591	C	T	76	53	23	71540880	C/T	C/T	C/T	C/C	C/C	T/T	11/1	9/1	10/1	*	*	*	T
chr11_68525287	68,525,287	C	T	-	-	-	71555591	C/C	C/C	C/C	C/C	C/T	C/C	*	*	*	*	25/13	*	C
chr11_68544801	68,544,801	A	C	-	-	-	71575105	A/A	A/A	A/A	A/C	A/A	A/A	*	*	*	23/16	*	*	A
rs22124152	68,546,406	T	C	454	161	293	71576710	C/C	C/C	C/C	C/C	C/C	T/T	*	*	*	*	*	*	C
chr11_68546454	68,546,454	C	T	459	269	190	71576758	C/T	C/T	C/T	C/C	C/T	C/C	10/35	12/34	10/55	*	27/66	*	C
chr11_68546511	68,546,511	C	G	290	97	193	71576815	G/G	G/G	G/G	G/G	G/G	C/C	*	*	*	*	*	*	G
chr11_68546704	68,546,704	A	C	146	45	99	71577008	C/C	C/C	C/C	C/C	C/C	A/A	*	*	*	*	*	*	C
chr11_68546777	68,546,777	A	G	439	140	299	71577081	G/G	G/G	G/G	G/G	G/G	A/A	*	*	*	*	*	*	G
chr11_68547035	68,547,035	G	A	195	94	101	71577339	G/A	G/A	G/A	G/A	A/A	G/G	3/7	6/12	7/14	15/28	*	*	G
rs22139258	68,547,670	G	A	340	207	133	71577974	G/A	G/A	G/A	G/G	G/A	G/G	14/34	14/27	10/30	*	23/42	*	G
chr11_68547749	68,547,749	T	C	182	145	37	71578053	T/C	T/C	T/C	T/C	T/T	T/T	53/12	41/18	51/7	47/27	*	*	C
	Average			252.00			Informative?	YES	YES	YES	YES	YES	NO							
													Total	42/149	54/131	37/172	58/98	69/141		
													% of risk haplotype ^	22%	29%	18%	37%	33%		
AKNA																				
chr11_68561787	68,561,787	C	T	1626	1416	208	71592091	C/C	C/C	C/C	C/C	C/T	C/C	*	*	*	*	185/204	*	C
rs22069543	68,562,441	G	A	1265	254	1008	71592745	A/A	A/A	A/A	A/G	A/G	A/A	*	*	*	94/80	159/129	*	A
chr11_68569222	68,569,222	C	T	1755	1603	150	71599526	C/C	C/C	C/C	C/C	C/T	C/C	*	*	*	*	148/146	*	C
chr11_68570580	68,570,580	A	G	821	128	692	71600884	G/G	G/G	G/G	A/A	G/G	G/G	*	*	*	*	*	*	G
chr11_68575321	68,575,321	G	A	1628	310	1318	71605625	G/A	G/A	A/A	A/A	A/A	A/A	196/242	105/125	*	*	*	*	A
chr11_68577401	68,577,401	G	A	1301	979	321	71607705	G/A	G/A	G/G	G/G	G/G	G/G	163/198	108/120	*	*	*	*	G
chr11_68578296	68,578,296	C	T	1238	932	300	71608600	C/T	C/T	C/C	C/C	C/C	C/C	176/205	94/95	*	*	*	*	C
chr11_68578674	68,578,674	T	C	930	330	600	71608978	T/C	T/C	C/C	C/C	T/T	C/C	118/132	85/83	*	*	*	*	C
chr11_68580620	68,580,620	G	A	931	810	120	71610924	G/G	G/G	G/G	A/A	G/G	G/G	*	*	*	*	*	*	G
chr11_68591406	68,591,406	A	G	1297	590	706	71621710	A/G	A/G	A/A	A/A	G/G	G/G	164/201	87/117	*	*	*	*	A
rs22122505	68,593,585	G	C	1089	931	158	71623889	G/G	G/G	G/G	G/G	G/G	C/C	*	*	*	*	*	*	G
rs22122506	68,593,599	C	T	948	815	131	71623903	C/C	C/C	C/C	C/C	C/C	A/A	*	*	*	*	*	*	C
chr11_68593630	68,593,630	G	A	1116	781	334	71623934	G/A	G/A	G/G	G/G	G/A	G/G	167/171	97/95	*	*	70/66	*	G
rs22122508	68,593,634	T	C	1151	647	504	71623938	T/C	T/C	T/T	T/T	T/C	C/C	165/172	99/99	*	*	75/75	*	T
rs22122509	68,593,713	C	A	874	744	128	71624017	C/C	C/C	C/C	C/C	C/C	A/A	*	*	*	*	*	*	C
chr11_68594017	68,594,017	C	T	1327	952	374	71624321	C/T	C/T	C/C	C/C	C/T	C/C	196/175	124/108	*	*	91/90	*	C
rs22122510	68,594,098	C	G	1398	1170	225	71624402	C/C	C/C	C/C	C/C	C/C	G/G	*	*	*	*	*	*	C
chr11_68594225	68,594,225	A	G	1546	980	566	71624529	A/G	A/G	A/A	A/A	G/G	A/A	211/251	139/124	*	*	*	*	A
rs22122511	68,594,462	G	A	1514	1077	431	71624766	G/A	G/A	G/G	G/G	G/A	G/G	219/201	112/133	*	*	86/95	*	G
chr11_68594491	68,594,491	C	T	1578	1151	426	71624795	C/T	C/T	C/T	C/C	C/C	C/C	226/209	136/123	105/92	*	*	*	T
	Average			1266.65			Informative?	YES	YES	YES	YES	YES	NO							
													Total	2044/2114	1191/1217	105/92	80/94	777/842		
													% of risk haplotype ^	49%	50%	53%	46%	48%		
DFNB31																				
chr11_68617832	68,617,832	T	C	1016	605	410	71647797	T/C	T/C	T/C	T/T	T/T	C/C	97/129	90/79	72/71	*	*	*	C
chr11_68617936	68,617,936	T	C	1585	899	684	71647901	T/C	T/C	T/C	T/T	T/T	C/C	160/179	134/175	118/122	*	*	*	C
chr11_68618606	68,618,606	A	G	1032	896	136	71648571	A/A	A/A	A/A	A/A	A/A	G/G	*	*	*	*	*	*	A
chr11_68619188	68,619,188	A	T	1063	926	130	71649153	A/A	A/A	A/A	A/A	A/A	T/T	*	*	*	*	*	*	A
chr11_68621353	68,621,353	G	C	481	283	198	71651318	G/C	G/C	G/C	G/G	G/C	G/G	66/72	45/55	36/36	*	32/35	*	C
chr11_68635719	68,635,719	T	C	553	403	148	71665684	T/C	T/C	T/C	T/T	C/T	T/T	98/47	56/43	48/28	*	49/30	*	T
chr11_68671292	68,671,292	C	T	738	505	233	71701257	C/T	C/T	C/C	C/C	C/C	T/T	132/87	65/60	*	*	*	*	C
	Average			924			Informative?	YES	YES	YES	NO	YES	NO							
													Total	610/457	430/372	277/254		62/84		
													% of risk haplotype ^	57%	54%	53%		43%		

\+ First number is the counts of the reads with the reference allele, second number is the counts of reads with the alternate allele.

***** Dog is not informative for that SNP

^ Risk haplotype was measured as sum of risk alleles devided by the total number of reads in IG dogs. In the control dogs the number was calculated as the sum of the minor alleles devided by the total number of reads.

In red are the number of reads of the risk allele.

? The real genotype of the dog based on the reads was not clear

Supplement Table 3.11. Cuffdiff analysis of RNA-seq reads between control dogs and IG heterozygous dogs. The list includes only the genes with statistically significant difference in expression between the two groups. In bold is the higher number of the two.

Number	gene_id	gene	locus	value_1[a]	value_2[b]	log2(fold_change)	test_stat[c]	p_value[d]	q_value[e]
1	ENSCAFG00000023827	CRBB2_CANFA	chr26:19298946-19308447	**1.20963**	0	-	nan	0.00005[f]	0.0110218
2	ENSCAFG00000001705	MB	chr10:28574067-28583590	0	**1.04489**	inf	nan	0.00005[f]	0.0110218
3	ENSCAFG00000008062	-	chr8:8320212-8323009	0	**17.9946**	inf	nan	0.00005[f]	0.0110218
4	ENSCAFG00000028985	-	chr1:111083447-111083951	0	**5.22049**	inf	nan	0.00005[f]	0.0110218
5	ENSCAFG00000000791	NOTCH4	chr12:1586323-1608575	1.2213	**6.92773**	2.50397	5.79464	5.00E-05	0.0110218
6	ENSCAFG00000010068	TNNI2	chr18:46077957-46079538	0.688972	**3.87324**	2.49103	4.04836	5.00E-05	0.0110218
7	ENSCAFG00000016543	MYLPF	chr6:17719682-17721750	3.0976	**14.7401**	2.25052	5.48009	5.00E-05	0.0110218
8	ENSCAFG00000018175	CAPN6	chrX:84351484-84377324	0.251976	**1.15798**	2.20025	3.34512	5.00E-05	0.0110218
9	ENSCAFG00000031904	MKRN2-AS1	chr20:6027419-6034350	0.618453	**2.60856**	2.07652	3.50834	5.00E-05	0.0110218
10	ENSCAFG00000011417	RPL4	chr27:21855204-21856473	4.82751	**17.8928**	1.89003	4.31819	5.00E-05	0.0110218
11	ENSCAFG00000000296	GPR126	chr1:33969265-34107014	0.757815	**2.38019**	1.65116	3.86754	5.00E-05	0.0110218
12	ENSCAFG00000008253	ACTA1	chr4:9812781-9815574	2.35588	**6.48882**	1.46169	3.23504	5.00E-05	0.0110218
13	ENSCAFG00000029478	-	chr9:10287951-10304725	5.76627	**15.1094**	1.38973	4.20572	5.00E-05	0.0110218
14	ENSCAFG00000016065	-	chrX:44508996-44509863	49.836	**121.907**	1.29052	4.55046	5.00E-05	0.0110218
15	ENSCAFG00000016786	DNAH2	chr5:32600483-32659183	0.740779	**1.63042**	1.13813	2.6472	5.00E-05	0.0110218
16	ENSCAFG00000014201	DNAAF2	chr8:26251159-26295095	4.63608	**9.78829**	1.07815	2.76583	5.00E-05	0.0110218
17	ENSCAFG00000015078	FAM184B	chr3:62753439-62887350	2.38213	**5.0254**	1.07698	2.91737	5.00E-05	0.0110218
18	ENSCAFG00000016302	VWCE	chr18:55104014-55125461	3.81192	**7.89682**	1.05076	2.79413	5.00E-05	0.0110218
19	ENSCAFG00000013739	TET1	chr4:19768258-19889904	2.27243	**4.57145**	1.00842	2.57845	5.00E-05	0.0110218
20	ENSCAFG00000005194	PFKFB3	chr2:29732444-29748084	58.6269	**114.273**	0.962848	3.47457	5.00E-05	0.0110218
21	ENSCAFG00000018418	RHBDL3	chr9:40633359-40660779	25.6011	**49.5661**	0.953149	3.2839	5.00E-05	0.0110218
22	ENSCAFG00000006985	Q2EFX7_CANFA	chr26:5315659-5476358	5.99268	**11.0752**	0.886055	2.48337	5.00E-05	0.0110218
23	ENSCAFG00000020254	NFAT5	chr5:80003597-80115099	4.54118	**8.3556**	0.879677	2.42235	5.00E-05	0.0110218
24	ENSCAFG00000004375	NUP210	chr20:3636670-3722019	3.68209	**6.74686**	0.873691	2.48661	5.00E-05	0.0110218
25	ENSCAFG00000004038	HIPK2	chr16:9090726-9244272	28.3184	**51.801**	0.871238	2.79822	5.00E-05	0.0110218
26	ENSCAFG00000016397	VPS13D	chr2:83901305-84132453	24.8109	**45.0315**	0.85996	2.75266	5.00E-05	0.0110218
27	ENSCAFG00000031504	UBAP1L	chr30:29475497-29489123	81.1432	**142.778**	0.815232	3.12021	5.00E-05	0.0110218
28	ENSCAFG00000011477	BSN	chr20:39612348-39628313	4.16641	**7.27205**	0.803557	2.3408	5.00E-05	0.0110218
29	ENSCAFG00000009901	-	chr38:1474699-1542627	14.7048	**25.1243**	0.772789	2.40359	5.00E-05	0.0110218
30	ENSCAFG00000025128	TRAC	chr8:2943108-2952167	130.913	**222.552**	0.765537	3.01681	5.00E-05	0.0110218
31	ENSCAFG00000018849	SNX29	chr6:30317097-30822479	14.9604	**24.6087**	0.718015	2.32246	5.00E-05	0.0110218
32	ENSCAFG00000006193	-	chr16:27598816-27614612	**39.8399**	23.6591	-0.751817	-2.52272	5.00E-05	0.0110218
33	ENSCAFG00000008873	-	chr26:9970455-9975820	**255.426**	149.951	-0.768418	-3.006	5.00E-05	0.0110218
34	ENSCAFG00000023759	MT2_CANFA	chr2:59607925-59608825	**179.274**	102.14	-0.811619	-4.39619	5.00E-05	0.0110218
35	ENSCAFG00000015239	C1orf114	chr7:29141745-29173572	**83.2129**	47.3228	-0.814271	-2.85471	5.00E-05	0.0110218
36	ENSCAFG00000020110	Q4W6L5_CANFA	chr6:54709521-54720797	**44.7585**	25.2714	-0.824653	-2.76379	5.00E-05	0.0110218
37	ENSCAFG00000012904	CDK1	chr4:12885999-12898957	**70.5831**	38.1764	-0.886641	-3.30066	5.00E-05	0.0110218
38	ENSCAFG00000017664	SLC14A1	chr7:45378306-45408131	**15.0399**	7.84565	-0.938832	-2.81636	5.00E-05	0.0110218
39	ENSCAFG00000032483	LY96	chr29:22493775-22515117	**34.2359**	17.6329	-0.95724	-3.38232	5.00E-05	0.0110218
40	ENSCAFG00000007539	ADML_CANFA	chr21:33472109-33474520	**12.7947**	6.43024	-0.992601	-2.78398	5.00E-05	0.0110218
41	ENSCAFG00000004130	NEBL	chr2:12126777-12229218	**17.9102**	8.15899	-1.13432	-3.38298	5.00E-05	0.0110218

42	ENSCAFG00000028905	CDKN2C	chr15:10218823-10225053	**55.7227**	24.1923	-1.20372	-3.82904	5.00E-05	0.0110218
43	ENSCAFG00000007375	PARP4	chr25:18447412-18538508	**5.03721**	1.7896	-1.49299	-3.78555	5.00E-05	0.0110218
44	ENSCAFG00000004404	-	chr19:22979380-23044911	**1.58121**	0.551966	-1.51838	-3.09268	5.00E-05	0.0110218
45	ENSCAFG00000007199	-	chr27:2221688-2229112	**11.5803**	3.96361	-1.54679	-4.01687	5.00E-05	0.0110218
46	ENSCAFG00000029483	C17orf67	chr9:31635471-31654119	**62.1904**	21.1012	-1.55937	-6.38461	5.00E-05	0.0110218
47	ENSCAFG00000000538	VIP_CANFA	chr1:42941848-42950730	**4.86565**	1.5423	-1.65754	-3.45848	5.00E-05	0.0110218
48	ENSCAFG00000025465	TRPV6	chr16:6691520-6705704	**2.92284**	0.893114	-1.71045	-3.26907	5.00E-05	0.0110218
49	ENSCAFG00000013763	Q2LC20_CANFA	chr20:42231834-42263250	**3.97622**	1.17822	-1.75479	-3.77628	5.00E-05	0.0110218
50	ENSCAFG00000000492	-	chr12:1013177-1021690	**3.55046**	1.02932	-1.78632	-2.98578	5.00E-05	0.0110218
51	ENSCAFG00000023684	-	chrX:121060226-121074003	**1.9279**	0.547474	-1.81617	-3.19666	5.00E-05	0.0110218
52	ENSCAFG00000006714	NLRP14	chr21:30722678-30756070	**1.83171**	0.501719	-1.86824	-3.20277	5.00E-05	0.0110218
53	ENSCAFG00000019134	PPL	chr6:36497461-36541298	**2.28724**	0.581794	-1.97503	-4.14373	5.00E-05	0.0110218
54	ENSCAFG00000031664	KCNT2	chr38:3305568-3382572	**3.37359**	0.85451	-1.98111	-4.06635	5.00E-05	0.0110218
55	ENSCAFG00000002174	GADD45G	chr1:96744757-96746752	**11.1038**	2.58121	-2.10494	-5.14077	5.00E-05	0.0110218
56	ENSCAFG00000017807	CRYM	chr6:23657103-24217826	**24.8682**	4.75495	-2.3868	-4.47202	5.00E-05	0.0110218
57	ENSCAFG00000006798	RASL11A	chr25:12315962-12318638	**22.3406**	3.53283	-2.66077	-6.23349	5.00E-05	0.0110218
58	ENSCAFG00000003176	DPY19L2	chr14:46787464-46881015	**2.65667**	0.375765	-2.82172	-4.90261	5.00E-05	0.0110218
59	ENSCAFG00000017810	ANKS4B	chr6:23657103-24217826	**14.7401**	1.86649	-2.98135	-2.85372	5.00E-05	0.0110218
60	ENSCAFG00000031467	-	chr14:22321435-22333226	0.166451	**1.10504**	2.73093	3.76508	1.00E-04	0.0177494
61	ENSCAFG00000011054	PLEKHG4B	chr34:11991028-12042856	0.372462	**1.09117**	1.55072	2.78831	0.0001	0.0177494
62	ENSCAFG00000002086	TPM2	chr11:52213493-52219776	4.1969	**10.0472**	1.2594	2.96098	0.0001	0.0177494
63	ENSCAFG00000029782	LRRC14B	chr34:11978120-11985313	0.633337	**1.37096**	1.11414	2.32659	0.0001	0.0177494
64	ENSCAFG00000001486	CPA5	chr14:6578484-6602924	6.41951	**12.5292**	0.964756	2.51603	0.0001	0.0177494
65	ENSCAFG00000010102	SEMA5A	chr34:4783837-5082084	2.49432	**4.84793**	0.958718	2.4014	0.0001	0.0177494
66	ENSCAFG00000008728	TET3	chr17:48982557-49083545	3.29968	**5.91888**	0.843	2.38692	0.0001	0.0177494
67	ENSCAFG00000017423	-	chr20:49954858-49964416	30.5757	**52.3138**	0.774807	2.60342	0.0001	0.0177494
68	ENSCAFG00000012713	KIF1A	chr25:50859012-50923122	54.7919	**88.7482**	0.695756	2.25231	0.0001	0.0177494
69	ENSCAFG00000004115	KIAA1549	chr16:9696591-9834339	10.8852	**17.6054**	0.693652	2.19115	0.0001	0.0177494
70	ENSCAFG00000005788	LAMTOR1	chr21:25981307-25986399	**223.401**	127.622	-0.807757	-4.01745	0.0001	0.0177494
71	ENSCAFG00000030087	KIAA0101	chr30:28835064-28847133	**16.0347**	9.13032	-0.812457	-2.42449	0.0001	0.0177494
72	ENSCAFG00000031808	-	chr12:69982482-69987647	**19.5553**	11.0961	-0.817508	-2.5532	0.0001	0.0177494
73	ENSCAFG00000019120	-	chr6:36388165-36402849	**33.3592**	18.2873	-0.86724	-2.67183	0.0001	0.0177494
74	ENSCAFG00000010559	SCGN	chr35:23602237-23638101	0.453832	**1.62812**	1.84298	2.90861	0.00015	0.0244054
75	ENSCAFG00000030662	-	chr8:72423344-72430357	0.479142	**1.26572**	1.40144	2.61226	0.00015	0.0244054
76	ENSCAFG00000010161	Q95KX1_CANFA	chr25:37079872-37103255	4.08055	**7.7625**	0.927758	2.43304	0.00015	0.0244054
77	ENSCAFG00000002838	CCDC104	chr10:56374144-56403051	**107.283**	66.5037	-0.689919	-2.61688	0.00015	0.0244054
78	ENSCAFG00000029474	GADD45A	chr6:77391624-77396500	**67.3317**	41.6368	-0.693426	-2.38804	0.00015	0.0244054
79	ENSCAFG00000009113	MT1_CANFA	chr2:59602960-59604676	**21.4034**	11.5276	-0.892752	-2.69557	0.00015	0.0244054
80	ENSCAFG00000000996	FABP7	chr1:62084556-62088231	**27.2359**	11.7661	-1.21088	-4.33737	0.00015	0.0244054
81	ENSCAFG00000009877	ITGB6	chr36:5835307-5909634	0.245485	**1.00717**	2.0366	3.09065	0.0002	0.0284729
82	ENSCAFG00000004229	-	chr19:21387334-21389006	1.2438	**4.20193**	1.75629	3.42786	2.00E-04	0.0284729
83	ENSCAFG00000031337	-	chr9:17335270-17341774	6.64902	**13.0362**	0.971307	2.63542	0.0002	0.0284729
84	ENSCAFG00000031544	-	chrUn_JH373476:43431-56878	6.94758	**12.5058**	0.848014	2.38296	0.0002	0.0284729
85	ENSCAFG00000032222	DLA-12	chr12:933311-936671	16.4384	**28.9237**	0.815181	2.57213	0.0002	0.0284729
86	ENSCAFG00000010692	KIAA2018	chr33:17949824-17994649	5.9698	**10.0535**	0.751941	2.26054	0.0002	0.0284729
87	ENSCAFG00000008199	FMN1	chr30:1781801-2184909	7.13286	**11.9821**	0.748325	2.23873	0.0002	0.0284729

88	ENSCAFG00000005775	-	chr17:25639733-25751938	9.18094	**15.3251**	0.739185	2.28845	0.0002	0.0284729
89	ENSCAFG00000011326	PRKCA	chr9:13644530-14030778	63.3657	**102.392**	0.692325	2.4178	0.0002	0.0284729
90	ENSCAFG00000009659	HR	chr25:35119991-35138180	12.5057	**19.5347**	0.643456	2.08639	0.0002	0.0284729
91	ENSCAFG00000004424	C3orf77	chr23:2009612-2086676	**3.49914**	1.88606	-0.891622	-2.207	0.0002	0.0284729
92	ENSCAFG00000024295	-	chr35:24092006-24092663	0.955554	**4.46507**	2.22427	4.08206	0.00025	0.0341675
93	ENSCAFG00000004058	CABP5	chr1:108147294-108155999	229.011	**374.813**	0.710752	3.51242	0.00025	0.0341675
94	ENSCAFG00000030812	MRPL51	chr27:38497870-38502110	**12.557**	7.62125	-0.720394	-2.18942	0.00025	0.0341675
95	ENSCAFG00000028512	-	chr31:29108349-29127830	**14.8485**	8.30609	-0.838075	-2.42566	0.00025	0.0341675
96	ENSCAFG00000011366	KCNB1	chr24:36025471-36119488	84.9956	**137.733**	0.696417	2.25255	0.0003	0.0390486
97	ENSCAFG00000018854	SEZ6	chr9:43121410-43166314	11.637	**18.6497**	0.680432	2.09438	0.0003	0.0390486
98	ENSCAFG00000011227	-	chr25:44160225-44165426	**75.7886**	46.0941	-0.717399	-2.34554	0.0003	0.0390486
99	ENSCAFG00000031492	MRPL21	chr18:49094651-49104533	**27.8634**	16.9082	-0.720648	-2.32742	0.0003	0.0390486
100	ENSCAFG00000007857	SCRG1	chr25:23987090-23989934	**45.7902**	18.8876	-1.2776	-6.94652	0.0003	0.0390486
101	ENSCAFG00000001488	MDGA1	chr12:6660180-6717647	3.19957	**5.64211**	0.818356	2.21118	0.00035	0.0438849
102	ENSCAFG00000014434	RC3H1	chr7:25265846-25305007	4.8444	**8.27719**	0.772825	2.16467	0.00035	0.0438849
103	ENSCAFG00000013973	GFAP	chr9:18569003-18579645	**18.1663**	11.2918	-0.685991	-2.12348	0.00035	0.0438849
104	ENSCAFG00000007047	MAL	chr17:34977297-35000278	**2.21698**	0.446887	-2.31061	-3.44302	0.00035	0.0438849
105	ENSCAFG00000019195	C16orf71	chr6:36665594-36676041	1.51007	**3.48157**	1.20512	2.5106	0.0004	0.0492505
106	ENSCAFG00000025374	DEM1	chr15:2497652-2498816	**22.4939**	12.6428	-0.831219	-2.55677	0.0004	0.0492504

a- FPKM of the gene in the control dog group
b- FPKM of the gene in the IG heterozygous dog group
c- The value of the test statistic used to compute significance of the observed change in FPKM
d- The uncorrected p-value of the test statistic
e-The Fase-discovery-rate-adjusted p-value of the test statistic
f- the default of the program when one of the values is zero

Supplement Table 3.12. Primers used to amplify mir184, and genotype results on a subset of IG dogs. **A**. primer pair used for mir184 amplification. **B**. Genotype results for the three SNPs found downstream the 3' end of the gene.

A. Primers used to amplify mir184 gene													
Forward primer		**Primer sequence**			**Reverse primer**		**Primer sequence**			**Amplicon location**			**Size**
Mir184F1		ggtccccgtaaacagagtatggtg			Mir184R1		gccagatgtccagaggagagacg			60,776,605-60,777,244			640 bp
B. Genotype results for 3 SNPs.													
#	SNP location	IG homozygous affected[a]			IG het normal[b]		IG het affected[c]						Normal Boxer
		Dog 1	Dog 2	Dog 3	Dog 1	Dog 2	Dog 1	Dog 2	Dog 3	Dog 4	Dog 5	Dog 6	
1	60,776,772	T/T	A/T	A/A	A/A	A/A	A/A	A/A	T/T	A/A	A/T	T/T	T/T
2	60,776,794	C/C	T/C	T/T	T/T	T/T	**C**/T	**C**/T	**C**/C	**C**/T	T/**C**	**C**/C	C/C
3	60,776,964	G/G	A/G	A/A	A/A	A/A	A/A	A/A	G/G	A/A	A/G	G/G	G/G

a- Dogs that are homozygous to the Col27A1 risk haplotype on CFA11 and are PRA affected.
b- Dogs that are heterozygous to the Col27A1 risk haplotype on CFA11 and are phenotypically normal.
c- Dogs that are heterozygous to the Col27A1 risk haplotype on CFA11 and are phenotypically PRA affected.

Supplement Table 3.13. Selected mir455 (A) and mir3609 (B) target genes.

Target gene	**Full Name**
A. mir455 target genes.	
TMEM212	transmembrane protein 212
ST6GAL2	ST6 beta-galactosamide alpha-2,6-sialyltranferase 2
CACNB4	calcium channel, voltage-dependent, beta 4 subunit
SALL1	sal-like 1 (Drosophila)
ADAM22	ADAM metallopeptidase domain 22
PJA2	praja ring finger 2
ARRDC3	arrestin domain containing 3
TMED2	transmembrane emp24 domain trafficking protein 2
B. mir3609 target genes.	
ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4 [Source:HGNC Symbol;Acc:34]
RHO	rhodopsin [Source:HGNC Symbol;Acc:10012]
LRAT	lecithin retinol acyltransferase (phosphatidylcholine--retinol O-acyltransferase) [Source:HGNC Symbol;Acc:6685]
CLOCK	clock homolog (mouse) [Source:HGNC Symbol;Acc:2082]
PDE6A	phosphodiesterase 6A, cGMP-specific, rod, alpha [Source:HGNC Symbol;Acc:8785]
EYA4	eyes absent homolog 4 (Drosophila) [Source:HGNC Symbol;Acc:3522]
OPN5	opsin 5 [Source:HGNC Symbol;Acc:19992]
LCA5	lebercilin; Leber congenital amaurosis 5 [Source:HGNC Symbol;Acc:31923]
EYS	eyes shut homolog (Drosophila) [Source:HGNC Symbol;Acc:21555]
ADAM9	ADAM metallopeptidase domain 9 [Source:HGNC Symbol;Acc:216]
DFNB31	deafness, autosomal recessive 31 [Source:HGNC Symbol;Acc:16361]
MOBKL2B	MOB1, Mps One Binder kinase activator-like 2B (yeast) [Source:HGNC Symbol;Acc:23825]
M6PR	mannose-6-phosphate receptor (cation dependent) [Source:HGNC Symbol;Acc:6752]
STK38L	serine/threonine kinase 38 like [Source:HGNC Symbol;Acc:17848]
PRPF8	PRP8 pre-mRNA processing factor 8 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:17340]
CHM	choroideremia (Rab escort protein 1) [Source:HGNC Symbol;Acc:1940]
ELAVL2	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)
CNGB3	cyclic nucleotide gated channel beta 3
CNGA1	cyclic nucleotide gated channel alpha 1

Supplement Figure 3.1. Sequence retrieved from DNA-seq and RNA-seq reads in the 5' end of Col27A1, where a gap was observed in the database. The red sequence was retrieved from DNA-seq reads. SNPs are highlighted in light blue- both alleles were observed in the reads. The upper case sequence was retrieved from RNA-seq reads. The black sequence is the 5' UTR, the blue sequence is first 40 bases of the coding sequence -exon 1.

gcgtggccgagcttgcatgtgagtgggtgtgagcgtgtgactgagttt
ggggtaagcgtggacgtgcgtgggagtgtgtatgcaccggcgagcaga
tt*gtgaaggtggaggtggaggtggaggtggaggtgg*cctctaaggccc
gggctgggactggggttgtggttgtgcgaacttgctgggcgctcgcgc
ctgggtcaaggtgtgcagtggcgcgctcccgcggggaccgcaggggg
tctcggtgggctcccgagggtgtgtgggtgggtgtgcgggtgggggggc
gttgagctcccgactcttaccgrgcggacagagaaggg
[nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnn]
[gap]

CGCCTGCCCCCTGGGGCGCCCCGGCGGCCCCATGGGGCGCGCCCACAC
TTGCCCCCCGGGCTCGGGAGCATGAAGTAGGGGCCCGCCATGGGAGCG
GGAGCGGCGCGGGGGGCCCGAGGCACGGCGA

[nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnn]
[gap]

tgtgctctcggttcccggctcggaagyccgtccgtgcgactctccgyt
gcctccagccccgggcttcggcttctgcccggtgcccgcgtctgtgtg
gctctgmacgtccctgctccggggagcccgtccgctcgggtctgtgcg
cctccgtgtgtccccgcgtctccgggg*ccgggccgggccgggggcgggg*
*gggggctcggcgcc*tgtcattttgcgcgtgtgagagagaaaacgcgag
agccggcgtgtcggagtagagagggcccgaggatcaccaggcattgca
gaggcgggggcgggggcggcagggaacgtccaaggagggagcgtggcc
cagccgggacccgggagctggtggcctggttgtaaagggcgagcgagt
gcaggagcgaggcactgagggtgggtgtgcgtgggagcgagcaaggat
gggagcagccctggggccctacgcgctgtggctcctgctcgctgc

Supplement Figure 3.2. Sequence across the gap at the 5' end of Col27A1 in a boxer (**A**) and in an IGPRA affected dog (**B**). Highlighted in light blue are polymorphisms between the two dogs. In blue is the coding sequence of exon 1. Highlighted in red is the donor site of intron 1.

A. A boxer DNA sequence.

CAGTGGCGCGCTCCCGCGGGGGACCGCAGGGGGTCTCGGTGGGCTCCCGAGGGTGTGTGGGTGG
GTGTGCGGGTGGGGGGCGTTGAGCTCCCGACTCTTACCGAGCGGACAGAGAAGGGGAGAGTGTG
ACCAACACTCAGCAAGCCCACCCACTGGGTGCCCGGGGAAGTCTGTTCCGGGCGGGGGGGGGGG
GCGCTCTGACCTGGGAGCAGGGCCGTGCATGTGTGTGTATGGGCGGTGGGACACTGGTCGTCGC
TGGCGTGGGAGTCTAGCCACTGGGTGCCGGTCCTGGACCTGCCCGGGAACGGGGCTCCTCTCCC
CGTAGGATGTGTCCCGTTTCCCCCTCCATCTCGGCCCCCATCTCGGGCCGCGGTGGTCCTGCGG
GGCGGCCTCGGCCGCCTCCCTCCTCCCGGTGGATCTCGGTTTCCCCGCTGGCACTGGGGTGGGG
GCGGGGGCGGGAGTCTCCGCGGGGCCGCGGGGGGCGGCCGGGGCTGTGCGCGCGGCCGCGGCCG
GCAGGGGGCGGCGTGGGGCAGCGGAGGGCGGCCGGGCGCGGGGAGGGGAGAAGGTGGAGAGCCG
CCGCGGGCCGAGACTTTAAATCGCCTTCATTTCCCCCAAATCCTTCCTTTTGCTCTCCTCCCCC
AGGCCGGCGGGGAGGCCGCTTCCACCGCCCCGCGCGCTCCCTCGCCCGCCCCGCTCGGCCTCCG
GGGGCCGCCAGCAGCCCGCCCGCAAAGTTGGGTCGTCTCGCCTGTTCTCGCCGCGTCTCCTTCC
TCGTCGGTGGAGCAGGACGAGAGCGAGAGGCAGGGGCGAGCGCGGCGCCGGACGCCGGGGACCA
TGGGCCTGGCGCGGGCGCCCGCGGGGCCCCGGCCGCGCCGCCTGCCCGCTGGGGCGCCCCGGGG
CGCGGGGCTGCGCGGGGGGCGCGGGGGCCGCGCGCTGTGAGCCGGCCTGGCGCGGCGGGGCGGG
GGGCCGGCGGCCCCATGGGGCGCGCCCACACTTGCCCCCCGGGCTCGGGAGCATGAAGTAGGGG
CCCGCCATGGGAGCGGGAGCGGCGCGGGGGGCCCGAGGCACGGCGGCGGCGGCGGAGGCGGCGG
CGGCGGCGCGCGGGGGGTGAGTACAAGCTCGGGGGCGCGCCCTCCCCGGCTTCCTGCGGCCCCG
GCCCCGGCCCCGCCCCCCGCCCCCCGCGGGCCCCCCGCGGCCACCGCGAGCTGCGCCCCCCGGC
CCGGCGAGGCCCGGCGAACTGCCTGCCCTGCCCCGCTGGCCGGGCTCCGCGGTCGCGCGGGGCG
CGGGGCGGGGGGCGCGGGGCGGGGGCCGAGCCCCGGTCGCGGCCGGGCGTGGGGGCGGGGGGCG
CCCCGGGGACGCCGGTGCCCACGGGTCCCGGCTCGGAAGTCCGTCCGTGCGACTCTCCGCTGCC
TCCAGCCCCGGGCTTCGGCTTCTGCCCGGTGCCCGCGTCTGTGTGGCTCTGCACGTCCCTGCTC
CGGGGAGCCCGTCCGCTCGGGTCTGTGCGCCTCCGTGTGTCCCCGCGTCTCCGGGGCCGGGCCG
GGCCGGGGCGGGGGGGGGCTCGGCGCCTGTCATTTTGCGCGTGTGAGAGAGAAAACGCGAGAGC
CGGCGTGTCGGAGTAGAGAGGGCCCGAGGATCACCAGGCATTGCAGAGGCGGGGGCGGGGGCGG
CAGGGAACGTCCAAGGAGGGAGCGTGG

B. An IGPRA affected dog sequence.

CAGTGGCGCGCTCCCGCGGGGGACCGCAGGGGGTCTCGGTGGGCTCCCGAGGGTGTGTGGGTGG
GTGTGCGGGTGGGGGGCGTTGAGCTCCCGACTCTTACCGAGCGGACAGAGAAGGGGAGAGTGTG
ACCAACACTCAGCAAGCCCACCCACTGGGTGCCCGGGGAAGTCTGTTCCGGGCGGGTGGGCGGG
GGGGGGGGGCGCTCTGACCTGGGAGCAGGGCCGTGCATGTGTGTGCATGGGCGGTGGGACACTG
GTCGTCGCTGGCGTGGGAGTCTAGCCACTGGGTGCCGGTCCTGGACCTGCCCGGGAACGGGGCT
CCTCTCCCCGTAGGATGTGTCCCGTTTCCCCCTCCATCTCGGCCCCCATCTCGGGCCGCGGTGG
TCCTGCGGGGCGGCCTCGGCCGCCTCCCTCCTCCCGGTGGATCTCCGTTTCCCCGCTGGCACTG
GGGTGGGGGCGGGGGCGGGGAGTCTCGGCGGGGCCGCGGGGGGCGGCCGGGGCTGTGCGCGCGG
CCGCGGCCGGCAGGGGGCGGCGTGGGGCAGCGGAGGGCGGCCGGGCGCGGGGAGGGGAGAAGGT
GGAGAGCCGCCGCGGGCCGAGACTTTAAATCGCCTTCATTTCCCCCAAATCCTTCCTTTTGCTC
TCCTCCCCCAGGCCGGCGGGGAGGCCGCTTCCACCGCCCCGCGCGCTCCCTCGCCCGCCCCGCT
CGGCCTCCGGGGGCCGCCAGCAGCCCGCCCGCAAAGTTGGGTCGTCTCGCCTGTTCTCGCCGCG

TCTCCTTCCTCGTCGGTGGAGCAGGACGAGAGCGAGAGGCAGGGGCGAGCGCGGCGCCGGACGC
CGGGGACCATGGGCCTGGCGCGGGCGCCCGCGGGGCCCCGGCCGCGCCGCCTGCCCGCTGGGGC
GCCCCGGGGCGCGGGGCTGCGCGGGGGGCGCGGGGGCCGCGCGCTGTGAGCCGGCCTGGCGCGG
CGGGGCGGGGGGCCGGCGGCCCCATGGGGCGCGCCCACACTTGCCCCCCGGGCTCGGGAGCATG
AAGTAGGGGCCCGCCATGGGAGCGGGAGCGGCGCGGGGGGCCCGAGGCACGGCGGCGGCGGCGG
AGGCGGCGGCGGCGGCGGCGGCGGCGCGCGGGGGGTGAGTACAAGCTCGGGGGCGCGCCCTCCC
CGGCTTCCTGCGGCCCCGGCCCCGCCCCCCGCCCCCCGCGGGCCCCCCGCGGCCACCGCGAGCT
GCGCCCCCCGGCCCGGCGAGGCCCGGCGAACTGCCTGCCCTGCCCCGCTGGCCGGGCTCCGCGG
TCGCGCGGGGCGCGGGGCGGGGGGCGCGGGGCGGGGGCAGAGCCCCGGTCGCGGCCGGGCGTGG
GGGCGGGGGGCGCCCCGGGGACGCCGGTGCCCACGGGTCCCGGCTCGGAAGTCCGTCCGTGCGA
CTCTCCGCTGCCTCCAGCCCCGGGCTTCGGCTTCTGCCCGGTGCCCGCGTCTGTGTGGCTCTGC
ACGTCCCTGCTCCGGGGAGCCCGTCCGCTCGGGTCTGTGCGCCTCCGTGTGTCCCCGCGTCTCC
GGGGCCGGGCCGGGCCGGGGCGGGGGGGGGCTCGGCGCCTGTCATTTTGCGCGTGTGAGAGAGA
AAACGCGAGAGCCGGCGTGTCGGAGTAGAGAGGGCCCGAGGATCACCAGGCATTGCAGAGGCGG
GGGCGGGGGCGGCAGGGAACGTCCAAGGAGGGAGCGTGGCCC

CHAPTER FOUR: FOUR AUTOSOMAL RETINAL DISEASES IN THE DOG ARE SUCCESSFULLY MAPPED BY GENOME-WIDE ASSOCIATION USING POOLED SAMPLES.

Manuscript from: Orly Goldstein, Jennifer Johnson, Julie Jordan, Peter A Schweitzer, and Gregory M. Acland. Four retinal degeneration diseases in the dog are successfully mapped by genome-wide association using pooled samples.
In preparation.

4.1 Summary

Whole-genome association studies are widely used to identify disease-causing genes, in humans as well as in animals. We had previously identified the loci responsible for three cone-rod dystrophies in the dog population, affecting the American Staffordshire breed (*crd1*), the Pit Bull Terriers (*crd2*), and the Glen of Imaal Terries (*crd3*). We also identified the locus responsible for the progressive retinal atrophy segregating in the Italian Greyhounds (IGPRA). This was done by means of genome-wide association of single nucleotide polymorphisms on individual dogs. Here we were able to identify the four loci by using three affected pooled samples, and three control pooled samples, genotyped on the Illumina canine SNP array. The pooled samples represented between 9 and 15 dogs in each group (case/control). The analysis combined an association analysis with an in house script to account for the neighborhood of each SNP, giving a higher score for supporting evidence from the SNP environment. A combination of homozygosity analysis and association analysis was also successful for the autosomal recessive diseases. The tradeoff for this approach was larger blocks, since the rare recombinants were not detectable. This proof of principle is the first one we are aware of done in the canine population, and suggests that the pooling strategy for mapping traits, particularly but not limited to autosomal recessive traits, is practical and can be beneficial for small research budgets.

4.2 Introduction

Genome-wide association studies (GWAS) were widely used in mapping clinical diseases and traits since 2005, with a significant success reported in more than 1,600 publications for more than 200 traits, as listed in *The Catalog of Published Genome- Wide Association Studies*[1]. This valuable hypothesis-free approach is not limited to human population, and was proven effective

in the dog population as well, identifying many genes responsible for inherited traits and diseases in different breeds[2-6]. Unfortunately, GWAS are very costly due to the price of high-density SNP arrays. A simple strategy of reducing the cost of a study is to replace individual genotyping with genotyping of DNA pools. To date, more than 20 pooled-based GWAS have been published, many reporting genome-wide significant associations for diseases and traits such as follicular lymphoma, otosclerosis, multiple sclerosis, Alzheimer's disease, melanoma, psorasis, and skin colour[7-13]. However, the criticism of pooled-based GWAS is that they have reduced power relative to conventional GWAS because of errors introduced by estimating allele frequency DNA pools. While this is true when studying complex traits, in simple mendelian traits this problem is not significant. Association analysis is based on the assumption that if a disease is expressed as a result of a mutation in the DNA, all affected individuals would carry that mutant chromosome, and linked SNPs would be present in the affected group in a higher frequency than in unaffected individuals, the control group. When the disease is inherited as a recessive trait, it usually means that the affected individual would have two affected alleles, and in most cases these alleles would be identical by decent (IBD). Combining this with the concept of association analysis, that would mean that an affected individual would be homozygous for the mutation, as well as for all the SNPs that are linked to that mutation, usually SNPs that are in close proximity. Therefore pooling affected dogs together should not change the genotype calling in the DNA segment carrying the mutation if all affected dogs are IBD, and they are expected to fall within the homozygous cluster for those SNPs that are linked to the disease. For any other SNPs that are not linked to the disease, genotype signal can fall either within one of the three clusters, AA, AB, BB, or outside of them (and then scored as no-call). SNPs in which the alleles proportions are deviated from 50-50% (carrier ratio) or 100-0% (homozygous ratio) would most likely get a no-

call score, since the algorithm would place the score between the clusters. As a result, the expected call rate for the pooled samples would be lower than individual samples. We would not expect to see that homozygous block in the control group, except for those areas where the segmental chromosome is fixed in the breed.

In this exercise, we examined the pooled sample strategy in four different retinal diseases, inherited independently in four separate breeds. Three of them are inherited as autosomal recessive (*crd1*, *crd2*, and *crd3*) while the fourth one is inherited as autosomal dominant with incomplete penetrance (IGPRA). Detailed description for each disease can be found in references 6, 14 and chapter 9. Briefly, *crd1* and *crd2* are cone-rod dystrophies identified in the American Staffordshire Terrier and the American Pit Bull Terrier, respectively. Experimental pedigrees were developed, breeding affected dogs to research colony dogs of known genetic and phenotypic history. The samples used in the GWAS were collected from these mixed-bred affected and unaffected dogs. In contrast, samples for the other two diseases, *crd3* and IGPRA, segregating in the Glen of Imaal Terrier and the Italian greyhound, respectively, were collected mostly from purebred dogs. The exceptions were a few dogs of mix-breed segregating the diseases (see methods for more details).

We had previously mapped the loci for all four diseases by running GWAS on individual samples. Here we report the results of GWAS on pooled samples, where we pooled five dogs into one sample, totaling six samples per disease (three samples in the case group and three in the control group), reducing the cost and showing the success of this approach.

4.3 Material and Methods

4.3.1 Samples collection: DNA samples were extracted from blood from affected and control dogs from all four different diseases: *crd1*, *crd2*, *crd3* and IGPRA. All DNAs were diluted to 50 ng/ul concentration. Three affected- and three control- pooled samples were created for each of the four diseases by pooling 5 different dogs per pool, representing between 9 and 15 dogs per affected/control group (Table 4.1; some dogs were represented in more than one pooled sample).

Table 4.1. Experimental design of pooled and individual samples used in GWAS for each of the four diseases.

		Pooled samples				Individual samples	
		Affected		Control		Affected	Control
Disease	**Experimental design**	**Pooled samples (number of dogs per pool)**	**Total number of dogs**	**Pooled samples number of dogs per pool)**	**Total number of dogs**	**Total number of dogs**	**Total number of dogs**
crd1	Pedigree-based	3 (x5)	13	3 (x5)	9	17	18
crd2	Pedigree-based	3 (x5)	13	3 (x5)	12	**14**[a]	**13**[a]
crd3	Population-based	3 (x5)	13	3 (x5)	15	**21**[a]	**22**[a]
IGPRA	Population-based	3 (x5)	9	3 (x5)	14	23	18

[a] **-** Samples run on Affymetrics SNP arrays.

4.3.2 Genome-wide association study (GWAS):

The pooled samples were genotyped using the Illumina Canine SNP chip (Illumina canineHD BeadArrays), which listed 173,662 loci, following the standard protocol. Data from the arrays were analyzed using the Illumina GenomeStudio software (v2011.1), genotyping module (v1.9.4), and Illumina Genome Viewer (v1.9.0). SNP calling was performed using cluster positions provided by Illumina for these BeadArrays and converted to A/B allele calls.

Heterozygosity plots were produced by plotting the B allele frequency for each SNP using chromosome coordinates from the CanFam2 reference genome. Call rates for all individual samples and pooled samples were retrieved from GenomeStudio software, and were calculated with the exclusion of the 1,547 markers designed as "intensity only loci", which are used for copy number variants (CNVs) investigation.

The accuracy of the genotype calls of the pooled samples was measured in the *crd1* and IGPRA studies only, since individual and pooled samples were both run on the Illumina SNP array. For each pooled sample, the genotype calls from the individual samples were retrieved and only the SNPs, where all 5 dogs had a genotype call, were analyzed (10 observed alleles, no missing data). The composition of the alleles in the pool was determined by the individual genotype calls. All SNPs were categorized to 11 categories representing the ratio of the A and B alleles in the pool: 10/10 when all 10 alleles observed in the individual genotypes were A, 9/10 to 1/10 when any number between 1 and 9 of A alleles were observed out of the 10, and 0/10, when all alleles observed were B. Within each category, we counted the number of SNPs scored as AA, AB, BB, or "no call" in the pooled sample. An error call was defined as follows: when all 10 alleles were A (all five dogs were homozygous AA), any other call of the pooled sample but AA was considered an error call (AB, BB or "no call"). When all 10 alleles were B (all five dogs were homozygous BB), any other call of the pooled sample but BB was considered an error call (AB, AA or "no call"). When both alleles were present in the pooled sample, in any ratio, an AA call or a BB call was counted as an error call, and an AB or a "no call" score were counted as a true call, since the "no call" scores were most likely a cryptic heterozygosity (see results for a support for this assumption).

Genotype calls were converted to Plink-format file and association was tested using the association command without pedigree or sex information (http://pngu.mgh.harvard.edu/purcell/plink/[15]). The genotype calls for each of the mutation-bearing chromosome for each disease were retrieved from the files and aligned by location to identify the homozygous block.

Homozygosity analysis: Runs of homozygosity were done on affected pooled samples only (n=3) using Plink, on the 172,115 SNPs (the CNVs markers were not included) with the following criteria: sliding window criteria: 1000 Kb, 50 SNPs, 0 missing calls, 0 heterozygous call, 0.05 threshold; homozygous segment criteria: length- 1000 Kb, 50 SNPs, 50 density (Kb/SNP). For *crd3* disease this run didn't give any region of homozygosity shared by all three samples in the autosomal chromosomes so more relaxed criteria were used, allowing one missing call and one heterozygous call within the sliding windows. The output was then filtered for those chromosomes where all three pooled samples shared a homozygous block.

4.3.4 "SNP neighborhood" analysis

Script in Perl was written to score SNPs by their P-values, as well as their neighboring SNPs (Supplement Figure 4.1). Each SNP received a score based on its P-value and the score increased as a function of the number of significant SNPs in the SNP's neighborhood, their P-value and their distance from the SNP (the definition of neighborhood and significance were variables and were changed per each run). The more significant neighbor SNPs existed and the closer they were to the SNP, the higher the score for that SNP was. The scores then were scaled to a 1-10 scale, with 10 given to a SNP with a significant P-value and the most supporting neighborhood.

4.4 Results

4.4.1 Call rate:

The canine HD Illumina SNPchip array insures an average call rate of 99.8% of the total 172,115 SNP loci, (http://www.illumina.com/documents//products/datasheets/datasheet_caninehd.pdf). The call rates of the pooled samples were observed between 0.752 and 0.868 with an average of 0.798 for these SNP loci (Supplement Table 4.1). Specifically to *crd1* and IGPRA, the average of call rates of the pooled samples was 0.808, or a 0.192 of "no call" rate (Table 4.2). The average call rate of the dogs that were used in the pools for *crd1* and IGPRA studies, and were run individually on the arrays, was 0.991, or a 0.009 of "no call" rate (Table 4.2). This result suggests that the relative likelihood of a SNP in a pooled sample to be scored as a "no call" due to its heterozygous allele composition (has both alleles in any ratio) is 21.33 more than due to a real no call as a result of a technical limitation (0.192/0.009).

Table 4.2. Call rates of the pooled and individual samples in the *crd1* and IGPRA studies.

Pool samples	**Pooled sample call rate**	**Pooled sample "no call" rate**	**Call rate of the five dogs used in each pooled sample**					**Average of call rate in the five individual dogs**	**Average of "no call" rate in the five individual dogs**
			Dog 1	**Dog 2**	**Dog 3**	**Dog 4**	**Dog 5**		
Pool_1_*crd1*_Affecteds	0.805	0.195	0.991	0.993	0.993	0.991	0.994	0.992	0.008
Pool_2_*crd1*_Affecteds	0.843	0.157	0.993	0.975	0.993	0.991	0.993	0.989	0.011
Pool_3_*crd1*_Affecteds	0.847	0.153	0.993	0.993	0.975	0.993	0.993	0.989	0.011
Pool_1_*crd1*_control	0.803	0.197	0.991	0.991	0.992	0.993	0.991	0.991	0.009
Pool_2_*crd1*_control	0.868	0.132	0.991	0.994	0.993	0.994	0.993	0.993	0.007
Pool_3_*crd1*_control	0.842	0.158	0.993	0.993	0.992	0.994	0.993	0.993	0.007
Pool_1_IGPRA_Affected	0.771	0.229	0.994	0.991	0.991	0.992	0.992	0.992	0.008
Pool_2_IGPRA_Affected	0.775	0.225	0.993	0.993	0.992	0.991	0.993	0.992	0.008
Pool_3_IGPRA_Affected	0.763	0.237	0.992	0.992	0.991	0.992	0.993	0.992	0.008
Pool_1_IGPRA_control	0.794	0.206	0.993	0.993	0.993	0.994	0.994	0.993	0.007
Pool_2_IGPRA_control	0.794	0.206	0.992	0.994	0.993	0.990	0.993	0.992	0.008
Pool_3_IGPRA_control	0.794	0.206	0.994	0.945	0.992	0.990	0.994	0.983	0.017
Average	0.808	**0.192**						0.991	**0.009**

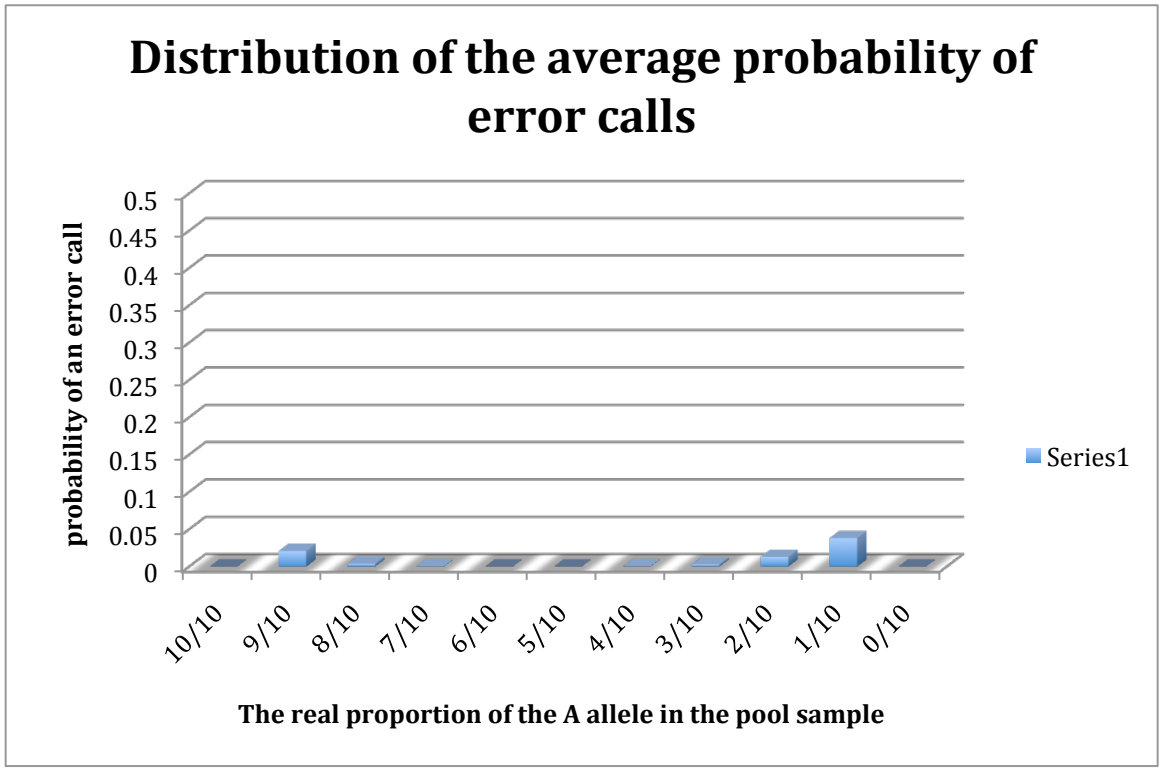
4.4.2 Call accuracy:

The accuracy of the genotype calls in the pooled samples was evaluated on the *crd1* and IGPRA pooled samples only. The analysis was done only on SNPs with no missing data (92.9% - 97.8% of all SNPs). The numbers of AA, AB, BB, and "no calls" were counted within each category of "A" allele ratio (Supplement Table 4.2). The number of error calls per sample varied from 4,801 SNPs to 20,286 SNPs (2.85%-12.06%, Supplement Table 4.3). The lowest numbers of percentage of error calls were observed in the categories where all five dogs were homozygous to either the A allele or the B allele, or when the ratios of A to B allele were between 0.4 and 0.6 (Table 4.3, Figure 4.1A). The highest numbers of error calls were observed when the allele ratio was 1/10 (either 1 "A" allele and 9 "B" alleles, or 9 "A" alleles and 1 "B" allele), with an average probability of making an error call of 0.0209 and 0.0382 (Table 4.3, Figure 4.1A). Interestingly, the distribution of error calls was asymmetric, and biased to the B allele: there were more error calls when the B allele was in a majority over the A allele (Table 4.3; Figure 4.1B and C) and the difference was significant by paired 2-tailed t-test: 8.14E-05, 5.27E-06, 5.48E-06, 5.68E-03 for 1/10, 2/10, 3/10 and 4/10 allele ratio, respectively. The difference was not significant when the animals were all homozygous AA compared to all homozygous BB (t-test=0.28). However, although we have no explanation for this bias, it does not affect the ability of the pooled assay to succeed.

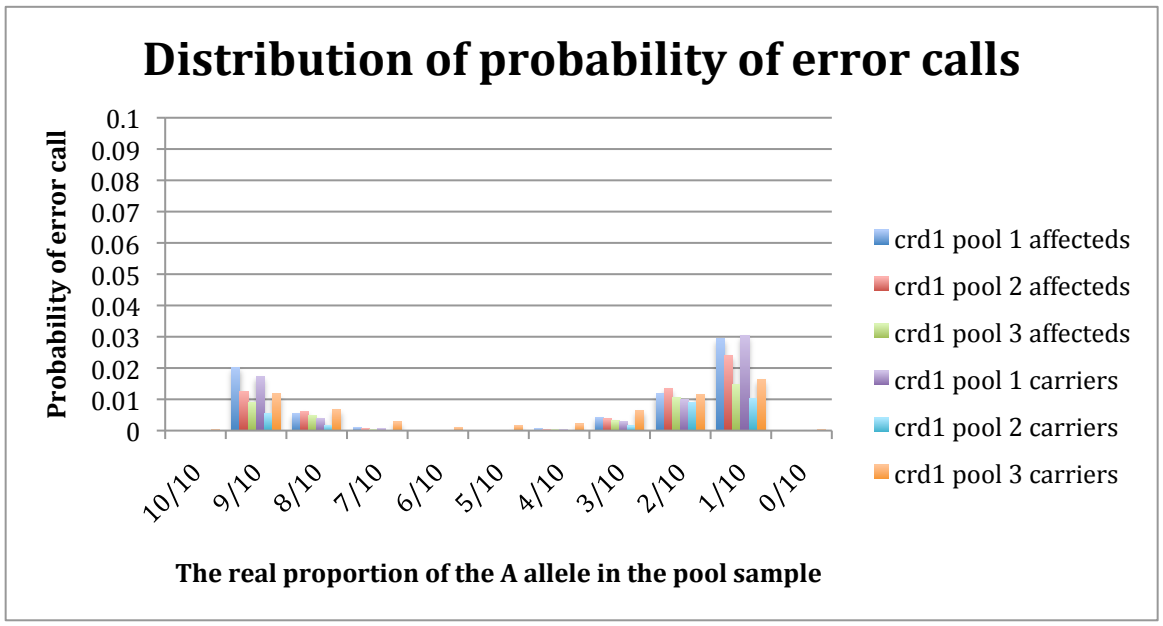
Table 4.3. The probability on an error call in each of the categories of the real portion of the A allele in the pool of 10 alleles in each of the pooled samples in crd1 and IGPRA studies.

	A/(A+B) = proportion of the "A" allele observed in the individual dogs										
	10/10	**9/10**	**8/10**	**7/10**	**6/10**	**5/10**	**4/10**	**3/10**	**2/10**	**1/10**	**0/10**
crd1 pool 1 affecteds	1.31E-04	2.02E-02	5.56E-03	1.02E-03	1.37E-04	6.56E-05	6.02E-04	4.29E-03	1.17E-02	2.93E-02	1.49E-04
crd1 pool 2 affecteds	9.09E-05	1.26E-02	6.25E-03	6.36E-04	3.64E-05	8.49E-05	4.00E-04	3.81E-03	1.35E-02	2.40E-02	3.03E-05
crd1 pool 3 affecteds	4.24E-05	9.02E-03	4.92E-03	4.17E-04	5.45E-05	2.42E-05	4.30E-04	3.12E-03	1.05E-02	1.48E-02	4.24E-05
crd1 pool 1 carriers	8.99E-05	1.73E-02	4.01E-03	7.19E-04	5.39E-05	9.59E-05	3.77E-04	3.00E-03	9.85E-03	3.03E-02	1.26E-04
crd1 pool 2 carriers	1.55E-04	5.41E-03	1.71E-03	8.92E-05	0.00E+00	1.19E-05	5.35E-05	1.53E-03	9.14E-03	1.03E-02	1.78E-04
crd1 pool 3 carriers	2.14E-04	1.17E-02	6.85E-03	2.80E-03	9.69E-04	1.65E-03	2.13E-03	6.50E-03	1.17E-02	1.65E-02	4.40E-04
IGPRA_pool1_affecteds	8.36E-05	2.62E-02	1.70E-03	1.43E-04	0.00E+00	2.99E-05	8.36E-05	1.14E-03	1.01E-02	5.30E-02	9.56E-05
IGPRA_pool2_affected	4.75E-05	3.65E-02	3.29E-03	1.96E-04	0.00E+00	0.00E+00	1.07E-04	1.65E-03	1.69E-02	6.18E-02	7.73E-05
IGPRA_pool3_affected	2.39E-05	3.11E-02	3.88E-03	3.29E-04	2.99E-05	3.58E-05	5.74E-04	3.59E-03	2.13E-02	5.83E-02	4.18E-05
IGPRA_pool1_control	1.01E-04	2.72E-02	1.93E-03	1.19E-04	5.97E-06	5.97E-06	4.18E-05	1.59E-03	1.43E-02	5.28E-02	4.18E-05
IGPRA_pool2_control	0.00E+00	3.05E-02	5.27E-03	5.62E-04	5.98E-05	1.20E-05	3.05E-04	3.68E-03	1.76E-02	5.80E-02	4.18E-05
IGPRA_pool3_control	1.06E-04	2.26E-02	1.34E-03	1.31E-04	6.25E-06	0.00E+00	4.38E-05	1.20E-03	1.09E-02	4.93E-02	1.06E-04
Average	9.05E-05	2.09E-02	3.89E-03	5.97E-04	1.13E-04	1.68E-04	4.29E-04	2.92E-03	1.31E-02	3.82E-02	1.14E-04

A.


Distribution of the average probability of error calls
probability of an error call
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
10/10
9/10
8/10
7/10
6/10
5/10
4/10
3/10
2/10
1/10
0/10
Series1
The real proportion of the A allele in the pool sample

B.


Distribution of probability of error calls
Probability of error call
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
10/10
9/10
8/10
7/10
6/10
5/10
4/10
3/10
2/10
1/10
0/10
crd1 pool 1 affecteds
crd1 pool 2 affecteds
crd1 pool 3 affecteds
crd1 pool 1 carriers
crd1 pool 2 carriers
crd1 pool 3 carriers
The real proportion of the A allele in the pool sample

C.

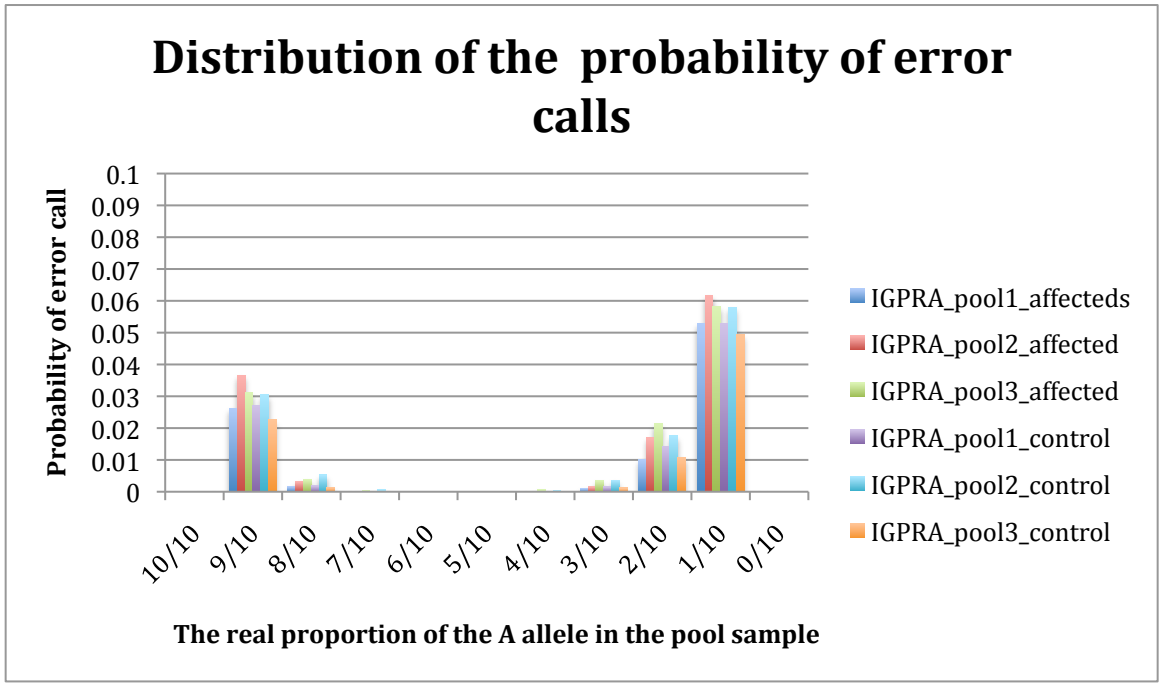


Figure 4. 1. Distribution of the probability of error calls in pooled samples in the crd1 and the IGPRA studies. **A**. The average of the probabilities of an error call in each of the real A to B allele ratios. **B**. The probability of an error call in each of the real A to B allele ratios in the crd1 pooled samples. **C**. The probability of an error call in each of the real A to B allele ratios in the IGPRA pooled samples.

4.4.3 Association analysis:

Association analysis was run on each disease separately (three cases versus three controls). The number of SNPs with P values lower than 0.05 were 174, 244, 75, and 374 for *crd1*, *crd2*, *crd3*, and IGPRA diseases, respectively (Table 4.4, Figure 4.2). In the *crd1* study, the highest $-\log_{10}P$ value was 2.33, observed in four SNPs on CFA7 (Supplement Table 4.4, Figure 4.2). In the *crd2* study, the highest $-\log_{10}P$ value was 2.01, observed in two SNPs on the X chromosome. The highest $-\log_{10}P$ value observed on an autosomal chromosome was 1.55 in 7 SNPs from varies chromosomes (CFA5, 9, 12, 14, 21, 23, and 32; Supplement Table 4.4, Figure 4.2). Therefore, since *crd2* is inherited as an autosomal recessive disease, there is no significant hit on a single autosomal chromosome, and the two- dimension Manhattan plot doesn't show a single peak if we ignore the X chromosome (Figure 4.2). In the *crd3* study, the highest $-\log_{10}P$ value was 2.47, observed in five SNPs on CFA16 (Supplement Table 4.4; Figure 4.2). In the IGPRA study, the highest $-\log_{10}P$ value was 2.33, observed in one SNP on CFA34. The second highest $-\log_{10}P$ value was 1.95 observed in two SNPs: one on CFA10, and one on CFA11 (Supplement Table 4.4). Two- dimension Manhattan plot shows three peaks on CFA10, 11 and 34 (Figure 4.2). With the exception of the *crd3* study, the association test, run on each SNP as independent test, was not sensitive enough to find the mutation- bearing locus with three affected and three control samples (CFA3, CFA33, and CFA11 for *crd1*, *crd2* and IGPRA respectively), if considering only the top hits.

Table 4.4. List of chromosomes carrying SNPs with P- values lower than 0.05. In bold is the chromosome bearing the mutation for each disease.

Chromosome	Number of SNPs	Total distance in Mb	Median of the distance between SNPs in Kb	The ratio between the number of SNPs and the Medain
A. crd1 disease				
2	18	60.26	482	0.0373
3	**62**	**5.98**	**62**	**1.0000**
4	10	41.51	1,473	0.0068
5	3	3.93	1,967	0.0015
6	1	-	-	-
7	17	59.29	827	0.0206
8	1	-	-	-
9	1	-	-	-
11	2	5.07	5,072	0.0004
12	15	21.98	1,234	0.0122
13	1	-	-	-
15	2	2.53	2,533	0.0008
16	3	42.77	21,383	0.0001
21	18	44.37	1,424	0.0126
26	4	2.90	882	0.0045
28	2	4.26	4,255	0.0005
29	1	-	-	-
31	1	-	-	-
33	1	-	-	-
35	2	1.68	1,678	0.0012
36	7	8.37	333	0.0210
37	1	-	-	-
38	1	-	-	-
Total number of SNPs	174			
B. crd2 disease				
1	8	92.42	1,647	0.0049
3	3	40.21	20,107	0.0001
4	1	-	-	-
5	3	19.43	9,714	0.0003
6	2	0.66	657	0.0030
7	1	-	-	-
8	2	3.31	3,307	0.0006
9	1	-	-	-
11	2	13.11	13,105	0.0002

12	50	56.77	138	0.3623
13	2	14.48	14,482	0.0001
14	9	38.40	4,893	0.0018
16	4	6.84	714	0.0056
19	9	10.57	378	0.0238
20	1	-	-	-
21	2	0.29	292	0.0068
22	9	29.41	1,433	0.0063
23	9	44.27	2,788	0.0032
24	19	32.62	294	0.0646
25	4	5.14	867	0.0046
27	2	7.59	7,591	0.0003
28	9	24.34	1,924	0.0047
31	2	0.19	190	0.0105
32	11	12.61	1,107	0.0099
33	**55**	16.69	**132**	**0.4167**
36	1	-	-	-
38	1	-	-	-
39	21	83.22	807	0.0260
41	1	-	-	-
Total number of SNPs	244			
C. crd3 disease				
1	2	15.32	15,315	0.0001
4	1	-	-	-
5	3	23.27	11,636	0.0003
8	1	-	-	-
9	1	-	-	-
15	4	2.38	539	0.0074
16	**43**	**12.38**	**91**	**0.4725**
19	1	-	-	-
21	1	-	-	-
23	2	0.05	45	0.0444
24	1	-	-	-
25	1	-	-	-
28	4	5.79	1,159	0.0035
30	2	26.88	26,882	0.0001
31	1	-	-	-
32	1	-	-	-
34	2	1.17	1,168	0.0017
36	1	-	-	-
39	3	88.71	44,356	0.0001
Total number of SNPs	75			
D. IGPRA disease				

1	64	111.77	315	0.2032
2	15	32.38	328	0.0457
3	5	26.86	1,845	0.0027
4	2	62.10	62,096	0.0000
5	16	66.71	2,666	0.0060
6	2	2.54	2,541	0.0008
7	14	45.42	928	0.0151
8	8	49.21	3,358	0.0024
9	16	36.85	1,415	0.0113
10	18	63.48	1,484	0.0121
11	**65**	**42.82**	**64**	**1.0156**
13	8	47.33	1,233	0.0065
14	17	56.13	1,705	0.0100
15	1	-	-	-
17	8	42.26	5,875	0.0014
18	6	10.11	1,409	0.0043
19	10	32.80	480	0.0208
21	16	45.34	1,602	0.0100
22	3	4.42	2,208	0.0014
23	1	-	-	-
24	9	36.01	3,839	0.0023
25	12	33.53	1,538	0.0078
26	3	30.48	15,239	0.0002
27	1	-	-	-
29	3	7.76	3,878	0.0008
30	12	34.52	527	0.0228
32	8	32.12	250	0.0320
34	14	36.78	896	0.0156
36	3	19.10	9,548	0.0003
38	1	-	-	-
39	8	108.68	1,438	0.0056
41	5	1.57	438	0.0114
Total number of SNPs	374			

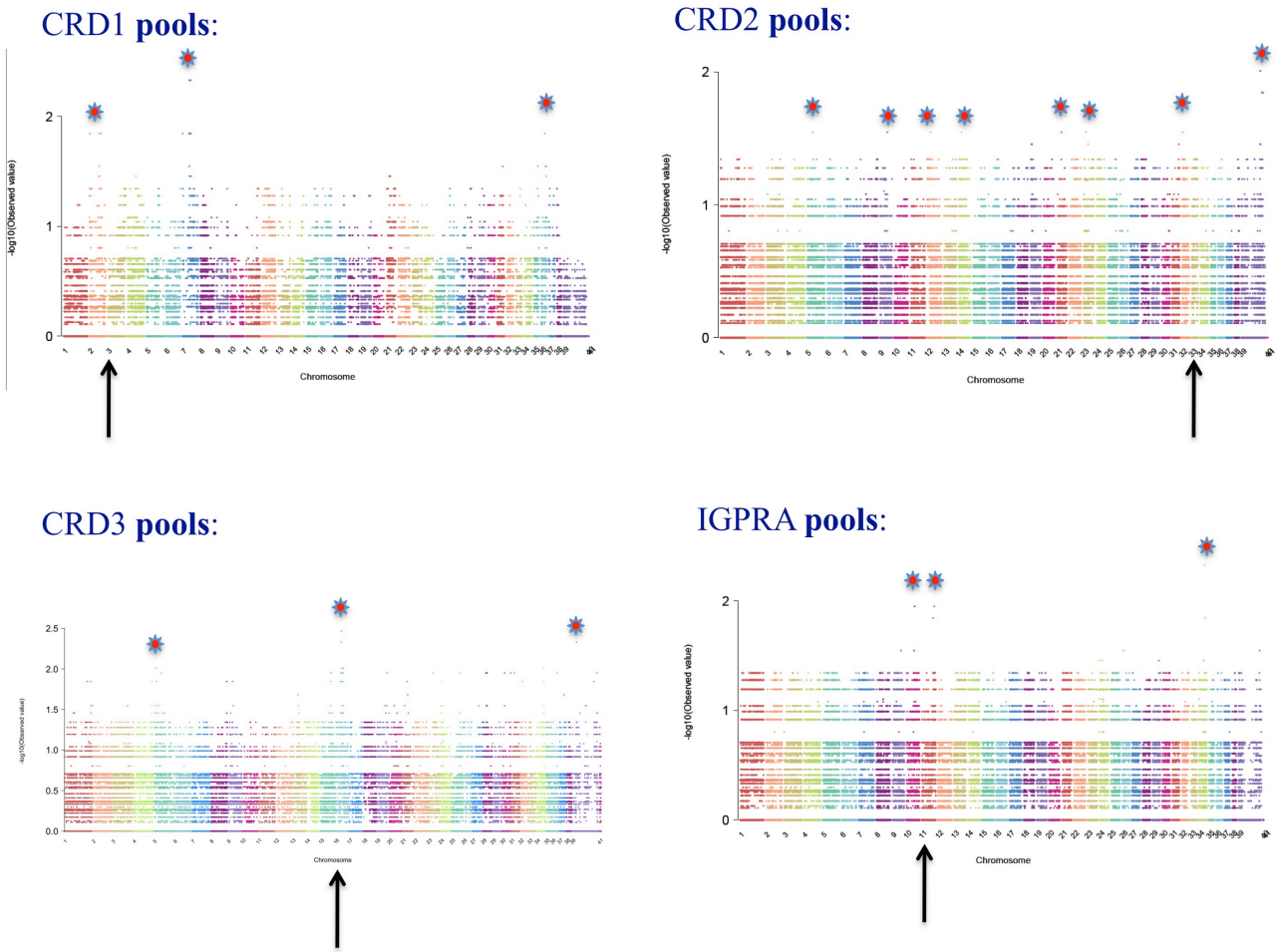


Figure 4.2. Manhattan plots summarizing results of GWAS using pooled samples in four canine hereditary retinal degenerations, *crd1*, *rcd2*, *crd3* and IGPRA. X- axis = canine chromosomes 1-38 plus the X- chromosome presented as chromosome 39 and 41. Y-axis = probability statistics ($-\log_{10}P$). Stars marks the top P-values.

4.4.4 "Neighborhood" analysis:

The association analysis run on each SNP independently does not account for the "environment" of each SNP tested, and ignores the information on the number of significant SNPs within each locus and their proximity to each other. Since we cannot run haplotype analysis due to losing this information when pooling the samples, we wanted to find a way that we will still be able to account for the "environment" of each SNP. We counted the number of SNPs within each chromosome, and calculated the median of the distances between the SNPs. Then we looked at the ratio of the number of SNPs to the median (Table 4.4). In all four diseases the highest value of the ratio was observed in the chromosome bearing the mutation, with 1.00, 0.4167, 0.4725, and 1.0156 for *crd1*, *crd2*, *crd3*, and IGPRA diseases, respectively (Table 4.4). In the case of crd2, chromosome 12 also received a high score of 0.3623. This practice mimics a haplotype analysis done in non-pooled samples.

To increase the power of the association test on pooled samples, we went one step further, and added a weight to the proximity of the significant SNPs to each other, and to the number of SNPs. We defined the neighborhood as 1 Mb interval, and the significant P-value as less than 0.1. The perl script (Supplement Figure 4.1) generated a new score for each SNP, taking under consideration the P-value of the SNP, and its neighborhood: the number of significant SNPs within the neighborhood, their P-value, and the distance between them and the SNP. The more neighboring SNPs with significant P-values there were, the higher their P-value was, and the closer they were to the SNP, the higher the score of that SNP. Then we scaled it to a 1-10 scale. The top scores for each disease were in the chromosomes bearing the mutation for each disease (Supplement Table 4.5), identifying CFA3, CFA33, CFA16 and CFA11 as the most significant chromosome for carrying the mutation for *crd1*, *crd2*, *crd3*, and IGPRA, respectively

(Figure 4.3). This approach also excluded the false positive hits within each disease: in *crd1*, the four SNPs on CFA7 with the highest P-value from the association analysis received a neighboring score of 2, suggesting no support from the SNPs' environment for real association; in *crd2*, the SNPs with the highest P-value from the association analysis on the X chromosomes, as well as the ones on the 7 autosomal chromosomes received a neighboring score between 1 and 3, suggesting no neighboring support for the hit, though CFA12 seams to have some support (bin score of 7, data not shown), but not as significant as the CFA33; in IGPRA, the associated significant SNP on CFA34 received a neighborhood score of 2, and the one on CFA10 received a score of 1.

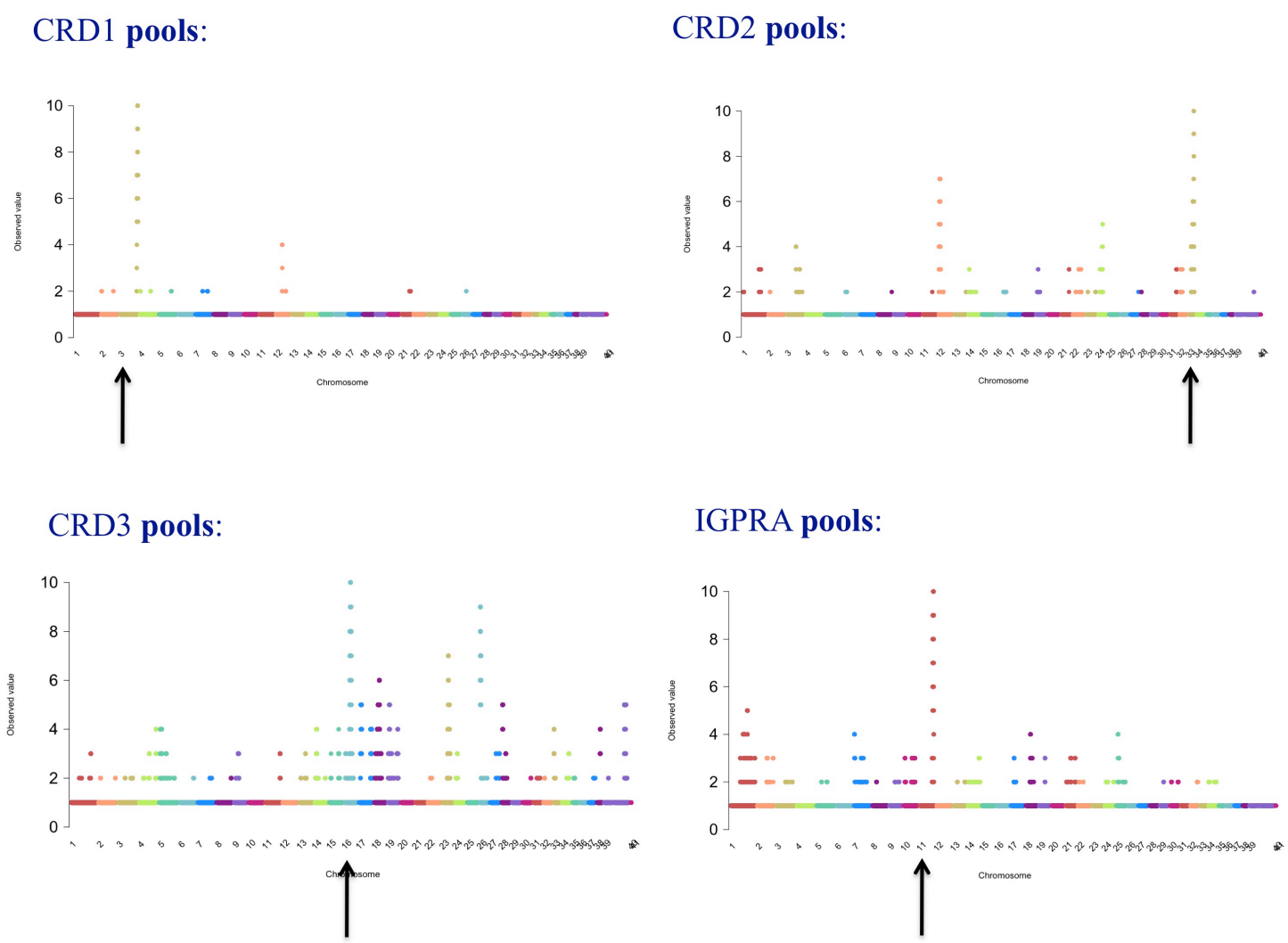


Figure 4.3. Manhattan plots summarizing results of "neighborhood GWAS" using pooled samples in four canine hereditary retinal degenerations, *crd1*, *rcd2*, *crd3* and IGPRA. X- axis = canine chromosomes 1-38 plus the X-chromosome presented as chromosome 39 and 41. Y-axis = "neighborhood score". In all four diseases the disese locus was identified (arrow).

4.4.5 Homozygosity analysis

We wanted to see if the loci for all four diseases can be identified by homozygosity analysis run on only three samples- the affected pooled samples. Results suggest that this is possible with a trade off of some false positive hits, and larger regions to investigate (Table 4.5). In the *crd1* disease, four chromosomes carried homozygous blocks larger than 1.0 Mb with a total of 10.43 Mb candidate regions for carrying the mutation (Table 4.5A). In the *crd2* disease, only two chromosomes were identified: CFA24 with a 1.29 Mb block, and CFA33 with a 2.31 Mb block (Table 4.5B). In *crd3*, six chromosomes carried homozygous blocks larger than 1.0 Mb with a total of 9.66 Mb candidate regions for carrying the mutation (Table 4.5C). In IGPRA disease, only one hit was observed, CFA11 with a 2.79 Mb block interval (Table 4.5D). Combining a homozygosity analysis with the P-value from the association analysis reduced the candidate region to the disease locus only (Table 4.5). These homozygosity regions can be visualized on the heterozygosity plot generated by GenomeStudio (Supplement Figure 4.2).

Table 4.5. Blocks of homozygosity shared among all three affected pooled samples across autosomal chromosomes. In Bold are the chromosomes bearing the mutation for each disease.

Number	CFA	Beginning	End	Size of block (in MB)	Association hit
A. *crd1* disease					
1	1	3,068,620	4,842,799	1.77	No
2	2	69,867,312	71,256,781	1.39	No
3	**3**	89,216,567	90,861,550	1.64	Yes
4	**3**	91,802,222	94,469,321	2.67	Yes
5	18	5,374,587	8,333,065	2.96	No
			Total	**10.43**	
B. *crd2* disease					
1	24	4,919,402	6,254,584	1.29	No
2	**33**	26,970,019	29,283,214	2.31	Yes
			Total	**3.6**	
C. *crd3* disease					
1	12	26,118,989	27,451,074	1.33	No
2	14	15,757,142	16,986,373	1.23	No
3	**16**	27,739,723	29,367,229	1.63	Yes
4	**16**	29,451,384	30,580,074	1.13	Yes
5	17	13,008,182	14,815,215	1.81	No
6	18	6,956,527	8,210,900	1.25	No
7	26	6,568,721	7,846,291	1.28	No
			Total	**9.66**	
D. IGPRA disease					
1	**11**	68,908,950	71,695,361	2.79	Yes
			Total	**2.79**	

4.5 Discussion

There are numbers of factors that influence the ability of GWA studies to detect genetic association. Among them are phenotype accuracy, the linkage disequilibrium (LD) between the causal variant and the probed SNPs, the number of individuals in each group, and the analysis approach taken. Specifically to pooling, there are additional factors. These include precision of allele frequency measurements made by the SNP genotyping microarray, the accuracy of pool construction by pipetting, the number of individuals pooled and the number of pooled samples. Moreover, in pooling, much more caution is required in making sure the pooled individuals do indeed have the same phenotype, since we lose the ability to compare subphenotype, and we cannot directly measure genotype, only allelotype. However, we were able to show in four different independent cases, that if all the factors are being carefully considered, the necessary measurements are taken, and the correct analysis used, pooled-samples GWAS can be successful, saving funds on the arrays, and allowing subsequent replication, fine-mapping, and sequencing of associated genomic regions.

The Illumina array has 172,115 SNPs. Our individual samples arrays showed 99.1% average call rate. That is 170,565 SNPs that technically work well and give a genotype call output when using the clustering measures of the array. These SNPs are evenly spaced to give an average of ~70 SNPs per Mb. We had shown, that for SNPs that worked on all the individual samples (average of 96.8% of all SNPs), the accuracy of the pooled sample genotype call using the Illumina software default algorithm for clustering, was more than 95%, and most of the error calls were for SNPs where the real allele ratio

in the pool was 1/10 (118,104 error calls across all 12 pooled samples out of a total of 161,035 errors; 73.3%). The direction of the error was as such that a cryptic heterozygous pooled-sample was scored as a homozygous sample. The opposite direction, a homozygous pooled sample scored as a heterozygous or "no-call", was very low (<0.04%). This would suggest that we are more likely to get false- positive homozygous blocks, but we will not miss the homozygous block carrying the mutation.

We were also able to show that the "no-call" scores in the pool were more likely a result of a cryptic heterozygosity in the pooled samples, due to the presence of the two alleles, in any ratio. When the disease is inherited as autosomal recessive this is very helpful, since the exact ratio is not relevant, as long as we distinguish between homozygous and cryptic heterozygous SNPs.

Our results showed that except for one disease (*crd3*), the association test on the pooled samples for each disease did not yield the correct hit. Since segments of DNA are inherited together, looking at haplotypes is usually more powerful than testing each SNP in an independent manner. Haplotype analysis in pooled samples is not possible, due to the lack of specific genotypes for the individual dogs. When we evaluated the supporting or unsupporting neighborhood of each SNP, we mimic haplotype analysis and increased the power enough to identify the correct locus for each disease. In *crd3* disease, the initial association test identified the mutation- bearing locus, maybe due to the fact that the 6 pooled samples represented 28 different dogs, compared to 22, 23 and 25 in the other three studies, and that the control group had mostly normal dogs and not carriers.

Homozygosity analysis is a valid approach when searching for mutations causing autosomal recessive diseases. In theory, if all the animals that are in a pooled sample are

homozygous for the same allele, we expect the pooled sample to be scored as homozygous as well. It is hard to predict how many homozygous blocks one would find in any given set of animals. It depends on the breed, the samples (purebred or mixed), the size of the block set forth by the analysis, and the stringency of the analysis. *crd1* and *crd2* are mixed-bred dogs generated by crossing the affected proband dog to a nonaffected colony dog, and therefore are expected to have a higher heterozygous genome compared to purebred dogs. The number of homozygous blocks in *crd1* and *crd2* were 5 (on 4 chromosomes) and 2, respectively. In *crd3*, Glen of Imaal Terrier dogs, we found 7 blocks (on 6 chromosomes), but only one block in the Italian Greyhound dogs. In all four studies, combination of homozygosity analysis with basic association analysis identified the correct locus.

That said, a major limitation of our pooled sample approach was the call error rate in cases where the allele ratio was 1/10. That is translated in action to missing the rare recombinant animals that would reduce the homozygous block. Another problem with pooling samples is that when the score is a "no-call", there is still a chance that it is not due to cryptic heterozygosity, which means that sometimes the observed homozygous block is smaller than it really is. Those two problems result in having different homozygous blocks in size and location when comparing the homozygous blocks observed in the individual runs to the pooled run in *crd1* and IGPRA. In the *crd1* study, the homozygous block on CFA3 was 93,432,486-94,693,816 (1.26 Mb) when individual arrays were used for each dog, while when using the pooled samples, two homozygous blocks were observed: one of 1.699 Mb size (89,186,028-90,885,261) and one of 2.684 Mb (91-785,257-94,469,321). The mutation is located distal to the second homozygous

block. Although the pooled approach did map the disease to the correct location, the potential segment bearing the mutation was not accurately identified, and a further genotyping of individual dogs for specific critical SNPs would have been necessary in order to establish the correct region. This, economically, would not be a problem, since the saving on the arrays leaves more funds for subsequent SNPs genotyping on individual dogs. In the IGPRA study, the homozygous block on CFA11 was 69,975,843-71,720,598 (1.74 Mb) when individual arrays were used for each dog, while when using the pooled samples, one homozygous block was observed of 2.83 Mb size (68,891,921-71,720,598). The rare recombinant from the proximal end of the block was missed in the pool, resulting in a larger block.

4.6 References

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, and Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009, 106(23): 9362-9367.

2. Downs LM, Bell JS, Freeman J, Hartley C, Hayward LJ, Mellersh CS. Late-onset progressive retinal atrophy in the Gordon and Irish Setter breeds is associated with a frameshift mutation in C2orf71. Anim Genet. 2012 Jun 12. doi: 10.1111/j.1365-2052.2012.02379.x. [Epub ahead of print].

3. Shearin AL, Hedan B, Cadieu E, Erich SA, Schmidt EV, Faden DL, Cullen J, Abadie J, Kwon EM, Gröne A, Devauchelle P, Rimbault M, Karyadi DM, Lynch M, Galibert F, Breen M, Rutteman GR, André C, Parker HG, Ostrander EA. The MTAP-CDKN2A Locus Confers Susceptibility to a Naturally Occurring Canine Cancer. Cancer Epidemiol Biomarkers Prev. 2012 Jul;21(7):1019-27. Epub 2012 May 23.

4. Grall A, Guaguère E, Planchais S, Grond S, Bourrat E, Hausser I, Hitte C, Le Gallo M, Derbois C, Kim GJ, Lagoutte L, Degorce-Rubiales F, Radner FP, Thomas A, Küry S, Bensignor E, Fontaine J, Pin D, Zimmermann R, Zechner R, Lathrop M, Galibert F, André C, Fischer J. PNPLA1 mutations cause autosomal recessive congenital ichthyosis in golden retriever dogs and humans. Nat Genet. 2012 Jan 15;44(2):140-7. doi: 10.1038/ng.1056.

5. Kuchtey J, Olson LM, Rinkoski T, Mackay EO, Iverson TM, Gelatt KN, Haines JL, Kuchtey RW. Mapping of the disease locus and identification of ADAMTS10 as a candidate gene in a canine model of primary open angle glaucoma. PLoS Genet. 2011 Feb;7(2):e1001306. Epub 2011 Feb 17.

6. Goldstein O, Mezey JG, Boyko AR, Gao C, Wang W, Bustamante CD, Anguish LJ, Jordan JA, Pearce-Kelling SE, Aguirre GD, Acland GM. An ADAM9 mutation in canine cone-rod dystrophy 3 establishes homology with human cone-rod dystrophy 9. Mol Vis. 2010 Aug 11;16:1549-69.

7. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, Iyadurai K, Becker N, Brooks-Wilson A, Curry JD, Spinelli JJ, Holly EA, Riby J, Zhang L, Nieters A, Smith MT, Brown KM. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009 Aug;41(8):873-5. Epub

2009 Jul 20.

8. Schrauwen I, Ealy M, Huentelman MJ, Thys M, Homer N, Vanderstraeten K, Fransen E, Corneveaux JJ, Craig DW, Claustres M, Cremers CW, Dhooge I, Van de Heyning P, Vincent R, Offeciers E, Smith RJ, Van Camp G. A genome-wide analysis identifies genetic variants in the RELN gene associated with otosclerosis. Am J Hum Genet. 2009 Mar;84(3):328-38. Epub 2009 Feb 19.

9. Comabella M, Craig DW, Camiña-Tato M, Morcillo C, Lopez C, Navarro A, Rio J; BiomarkerMS Study Group, Montalban X, Martin R. Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. PLoS One. 2008;3(10):e3490. Epub 2008 Oct 22.

10. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics. 2008 Sep 29;1:44.

11. Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward NK. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet. 2008 Jul;40(7):838-40. Epub 2008 May 18.

12. Capon F, Bijlmakers MJ, Wolf N, Quaranta M, Huffmeier U, Allen M, Timms K, Abkevich V, Gutin A, Smith R, Warren RB, Young HS, Worthington J, Burden AD, Griffiths CE, Hayday A, Nestle FO, Reis A, Lanchbury J, Barker JN, Trembath RC. Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene. Hum Mol Genet. 2008 Jul 1;17(13):1938-45. Epub 2008 Mar 25.

13. Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, van der Ouderaa FJ, Cox DR. A genomewide association study of skin pigmentation in a South Asian population. Am J Hum Genet. 2007 Dec;81(6):1119-32. Epub 2007 Oct 15.

14. Goldstein O, Mezey JG, Schweitzer PA, Boyko AR, Gao C, Bustamante CD, Jordan JA, Aguirre GD, Acland GM. IQCB1 and PDE6B Mutations Cause Similar Early Onset Retinal Degenerations in Two Closely Related Terrier Dog

Breeds. Invest Ophthalmol Vis Sci. 2013 Oct 25;54(10):7005-19. doi: 10.1167/iovs.13-12915.

15. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J Hum Genet. 2007; 81(3):559-575.

Supplemental material:

Supplement Tables 4.1-4.5.

Supplement Figures 4.1-4.2

Supplement Table 4.1. Call rates of pooled samples:

Pooled sample	**Call rate**
crd1 Affected samples- pool 1	0.805
crd1 Affected samples- pool 2	0.843
crd1 Affected samples- pool 3	0.847
crd1 obligated carriers samples- pool 1	0.803
crd1 obligated carriers samples- pool 2	0.868
crd1 obligated carriers samples- pool 3	0.842
crd2 Affected samples- pool 1	0.768
crd2 Affected samples- pool 2	0.792
crd2 Affected samples- pool 3	0.803
crd2 obligated carriers samples- pool 1	0.762
crd2 obligated carriers samples- pool 2	0.752
crd2 obligated carriers samples- pool 3	0.794
crd3 Affected samples- pool 1	0.812
crd3 Affected samples- pool 2	0.786
crd3 Affected samples- pool 3	0.788
crd3 obligated carriers samples- pool 1	0.807
crd3 control samples- pool 2	0.800
crd3 control samples- pool 3	0.792
IGPRA Affected samples- pool 1	0.771
IGPRA Affected samples- pool 2	0.775
IGPRA Affected samples- pool 3	0.763
IGPRA control samples- pool 1	0.794
IGPRA control samples- pool 2	0.794
IGPRA control samples- pool 3	0.794
Average	**0.798**

Supplement Table 4.2. The number of SNPs with AA, AB, BB or "no call" genotype calls in the pooled samples, within each category of "A" to "B" allele composition.

	A/(A+B) = proportion of the "A" allele observed in the individual dogs										
	Homozygous AA	Heterozygous									Homozygous BB
Pool genotype call	10/10	9/10	8/10	7/10	6/10	5/10	4/10	3/10	2/10	1/10	0/10
A. *crd1* pool 1- affected dogs											
AA	39,090	3,393	932	171	23	5	1	1	0	0	0
AB	0	119	1,733	4,333	7,488	7,502	6,848	3,150	1,009	49	1
BB	0	0	0	0	0	6	100	719	1,965	4,920	52,442
No call	22	4,090	4,363	4,448	1,609	892	2,500	5583	4,654	3,550	24
Total	39,112	7,602	7,028	8,952	9,120	8,405	9,449	9,453	7,628	8,519	52,467
B. *crd1* pool 2- affected dogs											
AA	40,208	2,071	1,031	105	5	1	0	0	0	0	0
AB	0	82	3,346	5,350	8,269	7,699	6,784	3,912	1,367	7	0
BB	0	0	0	0	1	13	66	628	2,220	3,962	52,812
No call	15	3,322	3,990	2,356	749	785	2,187	3,973	5,461	2,197	5
Total	40,223	5,475	8,367	7,811	9,024	8,498	9,037	8,513	9,048	6,166	52,817
C. *crd1* pool 3- affected dogs											
AA	42,791	1,491	813	69	9	0	0	0	1	0	0
AB	1	1,449	1,164	5,081	8,835	6,864	7,621	3,217	578	411	0
BB	0	0	0	0	0	4	71	515	1,733	2,453	55,535
No call	6	3,330	2,599	3,095	936	897	1,984	5,050	2,428	4,241	7
Total	42,798	6,270	4,576	8,245	9,780	7,765	9,676	8,782	4,740	7,105	55,542

D. *crd1* pool 1- control dogs											
AA	34,879	2,896	669	120	9	5	0	0	0	0	0
AB	0	216	3,103	5,945	8,609	9,978	7,636	4,590	1,548	49	2
BB	0	0	0	0	0	11	63	501	1,644	5,062	47,557
No call	15	4,887	5,081	2,451	1,031	874	2,400	4,132	6,889	4,050	19
Total	34,894	7,999	8,853	8,516	9,649	10,868	10,099	9,223	10,081	9,161	47,578
E. *crd1* pool 2- control dogs											
AA	48,583	910	287	15	0	0	0	0	0	0	0
AB	0	20	1,305	5,627	6,443	8,520	5,469	2,968	333	4	1
BB	0	0	0	0	0	2	9	257	1,537	1,728	63,000
No call	26	1,621	5,264	2,026	201	158	1,020	4,762	5,178	945	29
Total	48,609	2,551	6,856	7,668	6,644	8,680	6,498	7,987	7,048	2,677	63,030
F. *crd1* pool 3- control dogs											
AA	42,574	1,977	1,153	469	130	56	0	1	0	0	0
AB	0	3,085	1,963	4,032	6,487	6,770	5,783	3,124	1,059	1,039	1
BB	0	0	0	2	33	222	359	1,093	1,966	2,771	56,514
No call	36	3,435	2,016	2,546	1,691	2,670	2,110	2,978	2,400	5,643	73
Total	42,610	8,497	5,132	7,049	8,341	9,718	8,252	7,196	5,425	9,453	56,588
G. IGPRA pool 1- affected dogs											
AA	29,464	4,394	285	24	0	2	0	0	0	0	0
AB	0	40	1,902	7,777	9,771	9,923	9,375	4,590	515	7	0
BB	0	0	0	0	0	3	14	191	1,688	8,867	41,077
No call	14	5,662	7,271	2,928	332	146	1,494	7,289	8,779	3,554	16
Total	29,478	10,096	9,458	10,729	10,103	10,074	10,883	12,070	10,982	12,428	41,093
H. IGPRA pool 2- affected dogs											
AA	32,236	6,136	553	33	0	0	1	0	0	0	0
AB	1	19	1,304	6,459	9,000	9,607	7,952	2,923	225	0	0
BB	0	0	0	0	0	0	17	277	2,850	10,398	44,473
No call	7	4,473	7,564	2,651	314	240	1,841	6,628	7,689	2,379	13
Total	32,244	10,628	9,421	9,143	9,314	9,847	9,811	9,828	10,764	12,777	44,486

I. IGPRA pool 3- affected dogs											
AA	27,422	5,207	649	55	5	0	0	0	0	0	0
AB	0	138	2,631	7,180	10,161	10,832	8,645	4,445	909	30	0
BB	0	0	0	0	0	6	96	601	3,567	9,759	38,216
No call	4	4,626	7,367	3,040	698	626	2,577	6,422	8,287	3,185	7
Total	27,426	9,971	10,647	10,275	10,864	11,464	11,318	11,468	12,763	12,974	38,223
J. IGPRA pool 1- control dogs											
AA	32,478	4,549	324	20	1	0	1	0	0	0	0
AB	1	18	1,626	7,197	9,499	9,937	8,390	3,437	278	7	0
BB	0	0	0	0	0	1	6	267	2,388	8,853	44,748
No call	16	4,876	7,342	2,223	244	184	1,715	6,474	7,845	2,573	7
Total	32,495	9,443	9,292	9,440	9,744	10,122	10,112	10,178	10,511	11,433	44,755
K. IGPRA pool 2- control dogs											
AA	27,312	5,097	882	94	10	2	0	0	0	0	0
AB	0	44	2,861	6,929	9,447	10,113	8,516	4,325	972	10	0
BB	0	0	0	0	0	0	51	616	2,945	9,702	38,493
No call	0	5,704	6,902	3,345	849	594	2,552	6,101	8,668	4,143	7
Total	27,312	10,845	10,645	10,368	10,306	10,709	11,119	11,042	12,585	13,855	38,500
L. IGPRA pool 3- control dogs											
AA	32,727	3,615	214	21	1	0	0	0	0	0	0
AB	0	22	1,679	6,371	8,243	8,916	7,546	3,412	370	6	0
BB	0	0	0	0	0	0	7	192	1,740	7,893	45,801
No call	17	5,077	6,351	1,824	213	140	1,505	5,513	7,712	2,820	17
Total	32,744	8,714	8,244	8,216	8,457	9,056	9,058	9,117	9,822	10,719	45,818

Supplement Table 4.3. Autosomal SNPs with the highest -log10(P)-value for each disease.

CFA	SNP	Position (CanFam 2)	-log10(P)
crd1 disease:			
7	BICF2G630557683	37,960,556	2.33
7	BICF2G630557718	38,042,775	2.33
7	BICF2G630557829	38,251,057	2.33
7	BICF2P215143	45,627,305	2.33
***crd2* disease**			
5	BICF2G63035639	32,845,055	1.55
9	BICF2P952701	40,598,113	1.55
12	BICF2P1388209	19,782,720	1.55
14	BICF2P163164	18,361,055	1.55
21	BICF2G630650017	26,825,762	1.55
23	BICF2G630386022	16,896,724	1.55
32	BICF2G630600742	7,749,679	1.55
***crd3* disease**			
16	BICF2P1369072	28,070,386	2.47
16	BICF2S23248096	28,592,790	2.47
16	BICF2P207917	28,600,467	2.47
16	BICF2P355003	28,926,606	2.47
16	BICF2G630110934	31,510,907	2.47
IGPRA disease			
34	TIGRP2P398192_rs8631647	31600218	2.33

Supplement Table 4.5. Position of SNPs with the max Perl bin score of 10.

#	CFA	position
crd1		
1	3	92,888,906
2	3	92,898,131
3	3	92,918,288
4	3	92,920,698
5	3	92,934,377
6	3	92,948,610
7	3	92,959,822
8	3	93,002,840
9	3	93,049,536
10	3	93,359,777
11	3	93,432,486
12	3	93,435,448
13	3	93,443,600
14	3	93,478,160
15	3	93,520,425
16	3	93,539,230
17	3	93,542,118
18	3	93,610,097
19	3	93,618,479
20	3	93,639,445
21	3	93,668,058
22	3	93,684,805
23	3	93,697,691
24	3	93,700,678
25	3	93,731,523
26	3	93,737,251
crd2		
1	33	28,044,027
2	33	28,058,136
3	33	28,191,052
4	33	28,206,993
5	33	28,213,885
6	33	28,287,462
7	33	28,299,073
8	33	28,329,085
9	33	28,400,636
10	33	28,426,671

11	33	28,457,866
12	33	28,566,620
13	33	28,616,373
14	33	28,633,163
crd3		
1	16	30,348,569
2	16	30,390,824
3	16	30,401,873
4	16	30,406,905
5	16	30,434,408
6	16	30,451,050
7	16	30,455,364
8	16	30,473,480
9	16	30,480,010
10	16	30,496,748
11	16	30,509,498
12	16	30,539,551
13	16	30,554,749
14	16	30,569,541
15	16	30,580,074
16	16	30,609,221
17	16	30,636,048
18	16	30,671,681
19	16	30,695,399
IGPRA		
1	11	70,253,263
2	11	70,268,434
3	11	70,279,675
4	11	70,288,934
5	11	70,304,641
6	11	70,319,662
7	11	70,322,734
8	11	70,385,348
9	11	70,401,578
10	11	70,419,090
11	11	70,438,792
12	11	70,495,965
13	11	70,578,183
14	11	70,588,982
15	11	70,700,222
16	11	70,702,225

17	11	70,760,338
18	11	70,872,927
19	11	70,907,305
20	11	70,913,592
21	11	70,950,276
22	11	70,969,609
23	11	71,013,919
24	11	71,025,991
25	11	71,091,329
26	11	71,166,278
27	11	71,192,456

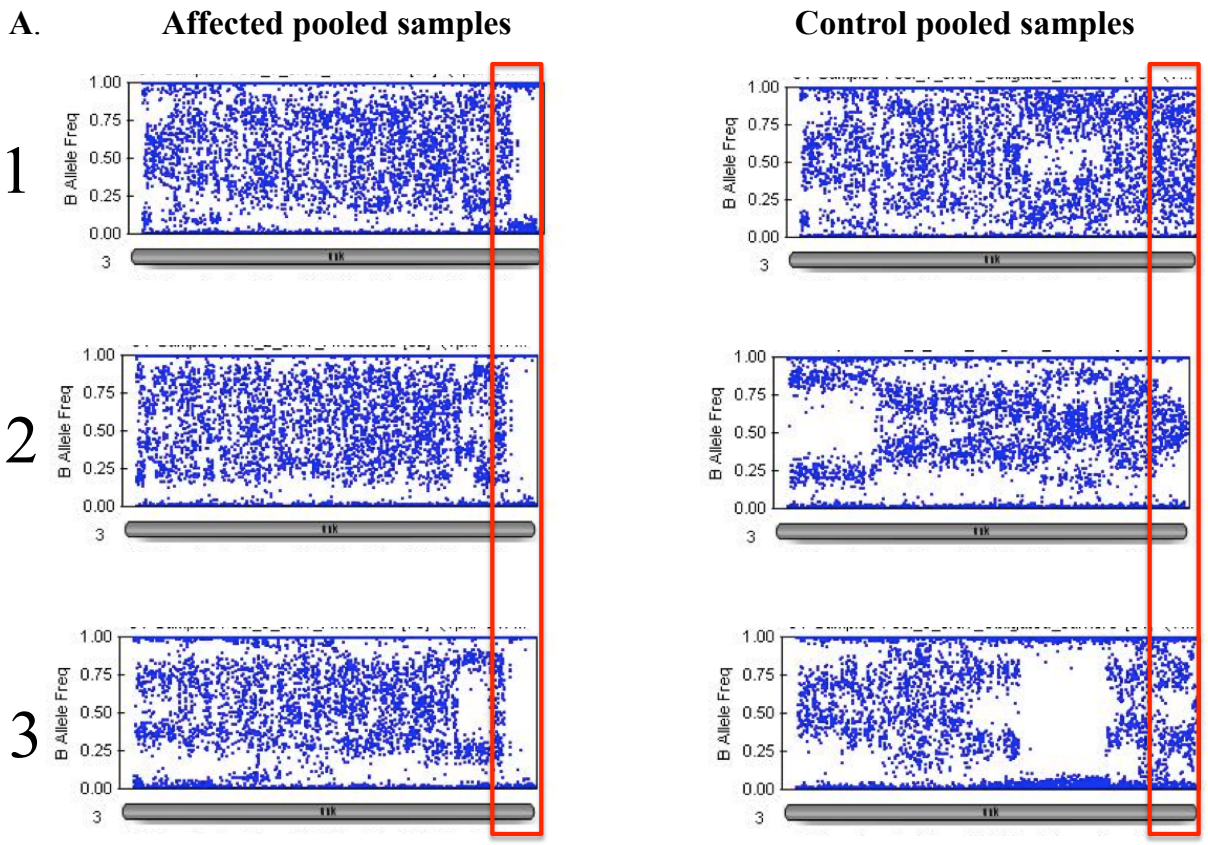
Supplement Figure 4.1. Perl script written to give weight to the SNPs based on their p value and their "neighborhood" SNPs.

```
#!/usr/bin/perl
use strict;

#read in file with snp#, Chromosome, position and p into a set of arrays
print "What is the location of your data file, with 4 columns, data starting in second row, headers in first, first column is snp name, second is snp chromosome, third is position and fourth is p value:\n";
chomp (my $filename = <STDIN>);

my @snpnum;
my @chr;
my @pos;
my @p;
my $line=0;

open FILE, $filename or die$!;
while (<FILE>)
{
chomp;
($snpnum[$line], $chr[$line], $pos[$line], $p[$line]) = split("\t",$_);
$line++;
}
close FILE;

#ask what distance from current snp is
print "What is the distance, in bp, from the current snp where you want to find other significant SNP's?\n";
chomp (my $dist = <STDIN>);

#ask what is the upper level for p to be considered for significance
print "At what p value is a SNP significant?\n";
chomp (my $sigp = <STDIN>);

#some debuging code:
print "You want to see if there are any snps that are within $dist bp at a better then $sigp p value.\n";
#print $dist - $sigp,"\n";
#print "$snpnum[0] \n";
#print "$snpnum[1] \n";
#print "$snpnum[-1] \n";
#print "$pos[$dist] \n";
```

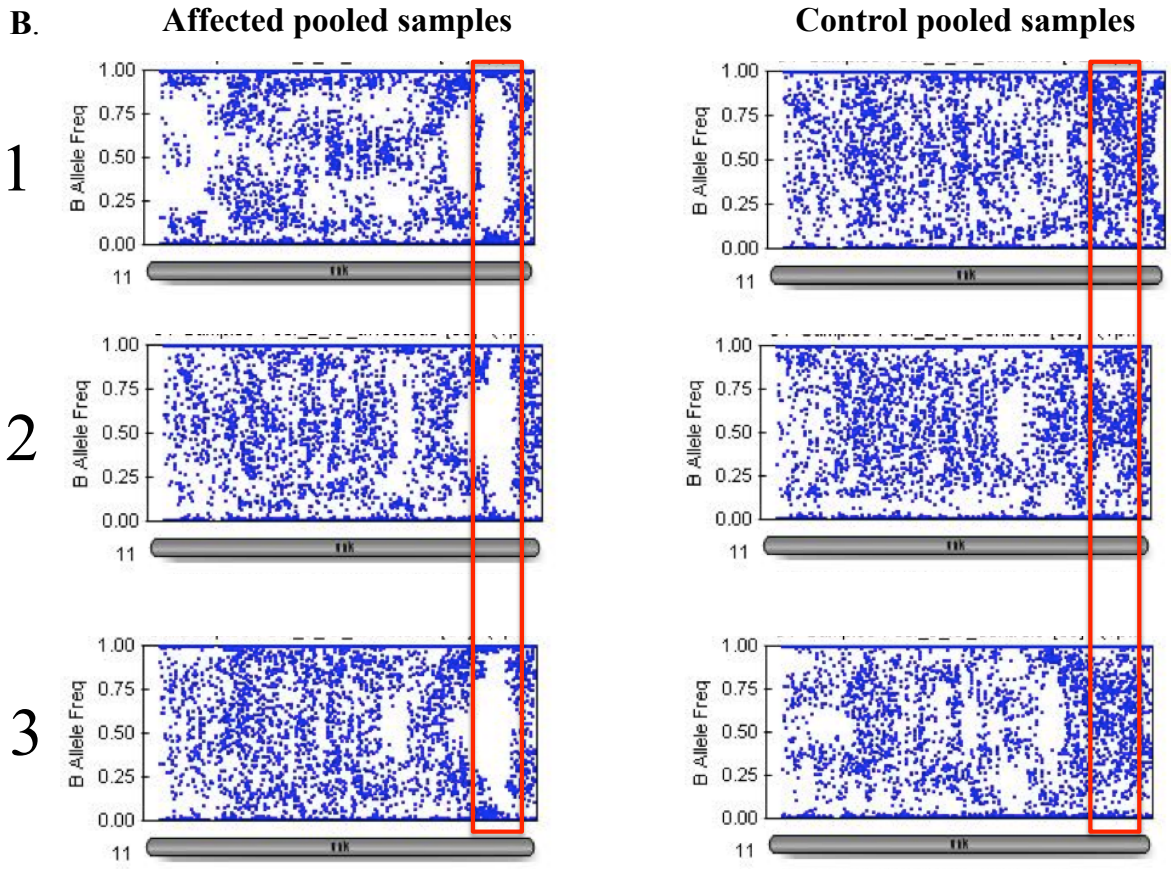
```
#for SNP (same number in each array) set size as 1, then check if same chromosome, within distance and under or equal to the set p value, if so set make the size variable larger for decreasing p and distance
my @size;
my $snp = 0;
my $checkindex = 0;

foreach (@chr)
{
	$size[$snp] = 1;
	$checkindex = 0;
	if (($sigp >= $p[$snp]) && ($chr[$snp] =~ /^[+-]?\d+$/))
		{
			$size[$snp] = $size[$snp]+  (100-(($p[$snp]/$sigp)*100)) ;
				foreach  (@snpnum)
				{

					if ($chr[$checkindex] == $chr[$snp])
					{
						if ( abs(($pos[$checkindex])-($pos[$snp])) <= $dist)
						{
							if ($sigp >= $p[$checkindex])
							{
								if ($checkindex != $snp)
								{
									$size[$snp] = ($size[$snp]
		+ (100-(($p[$checkindex]/$sigp)*100))
		+ (100-((abs($pos[$checkindex]-$pos[$snp])/$dist)*100)));
								}
							}
						}
					}

					$checkindex++;
				}
		}
	$snp++;
	if ($snp % 1000 == 0)
	{
		print "processing line $snp\n";
```

```
        }
}

#find the maximum value for the size variable so it can be used to create a weighted size
variable.
my $maxsize = 1;
$snp = 1;

foreach (@size)
{
if ($size[$snp] > $maxsize)
        {
        $maxsize = $size[$snp];
        }
$snp++;
}
print "The largest size is $maxsize.\n";


#create a weightes size variable, from 1 thru 10, as a percent of the max size.
my @wsize;
$snp = 0;
foreach (@chr)
{
        $wsize[$snp] = int((($size[$snp]/$maxsize)*10)+.5);
        if($wsize[$snp] == 0)
        {
                $wsize[$snp] = 1;
        }
        $snp++;
}

#write a new file, with size as a fourth column and weighted size as a fifth
$size[0] = "size";
$wsize[0] = "wsize";

open OUTFILE, ">$filename.output.txt";
my $linenew = 0;
foreach (@pos)
{
print OUTFILE
"$snpnum[$linenew]","\t","$chr[$linenew]","\t","$pos[$linenew]","\t","$p[$linenew]","\t
","$size[$linenew]","\t","$wsize[$linenew]\n";
$linenew++;
}
close OUTFILE;
```

Supplement Figure 4.2. Heterozygosity plots of the SNPs in the three affected and three control pooled sample on CFA3 (A) and on CFA 11 (B).



B.


Affected pooled samples
Control pooled samples
1
2
3
B Allele Freq
1.00
0.75
0.50
0.25
0.00
11

CHAPTER FIVE: DISCUSSION- SUMMARY OF RESULTS AND FUTURE PLANS.

5.1 Summary of findings

In this work we have touched upon the surface of the heterogeneity of retinal degeneration in the canine population, similar to the landscape observed in human retinitis pigmentosa. The retina, one might suggest, is the "busiest" organ in our body: to deliver vision, an act we do most of the day, and in our modern life- also at night, the phototransduction cascade must be tightly regulated to adjust to a wide variety of stimuli, including different levels of light and circadian rhythm[1-3]. In addition, the retina must rapidly cope with toxicity of bi-products, and maintain viability of cells under highly oxidizing conditions, and extremely rapid regeneration of the visual chromophore[1-3]. This execution would require a complex gene regulation network, which would simultaneously implement the many levels of regulations. Therefore, it is easy to imagine that any interruption in the balance of this network, might result in a domino effect that would eventually (and sometimes immediately) affect the normal function of the retina. Many times these imbalances occur due to mutations in the DNA, and if in germ-lines, the disease is then inherited from generation to generation. It is apparent that mutations in different genes can result in a very similar phenotype, and the same mutation on a different genetic background, even a very similar one as in siblings, can result in a different phenotype. Deciphering this heterogeneity and understanding the effect of each mutation can begin to clarify the complexity of vision impairment.

Dogs suffer from a broad spectrum of vision impairments, some unique to a specific breed, and some shared by many. Since each breed is genetically homogenous for a large proportion of its genome, complex traits can become less complex when presented on a unified genetic background. The dog model therefore is a great research tool to not only

find disease causative genes, but also to understand the many modifiers affecting the final outcome of vision.

This concept was presented, I hope, in this work. We were able to identify the mutations in several diseases, some shared by many breeds, such as *prcd*, and some uniquely affecting one breed, such as *erd* (Figure 5.1).

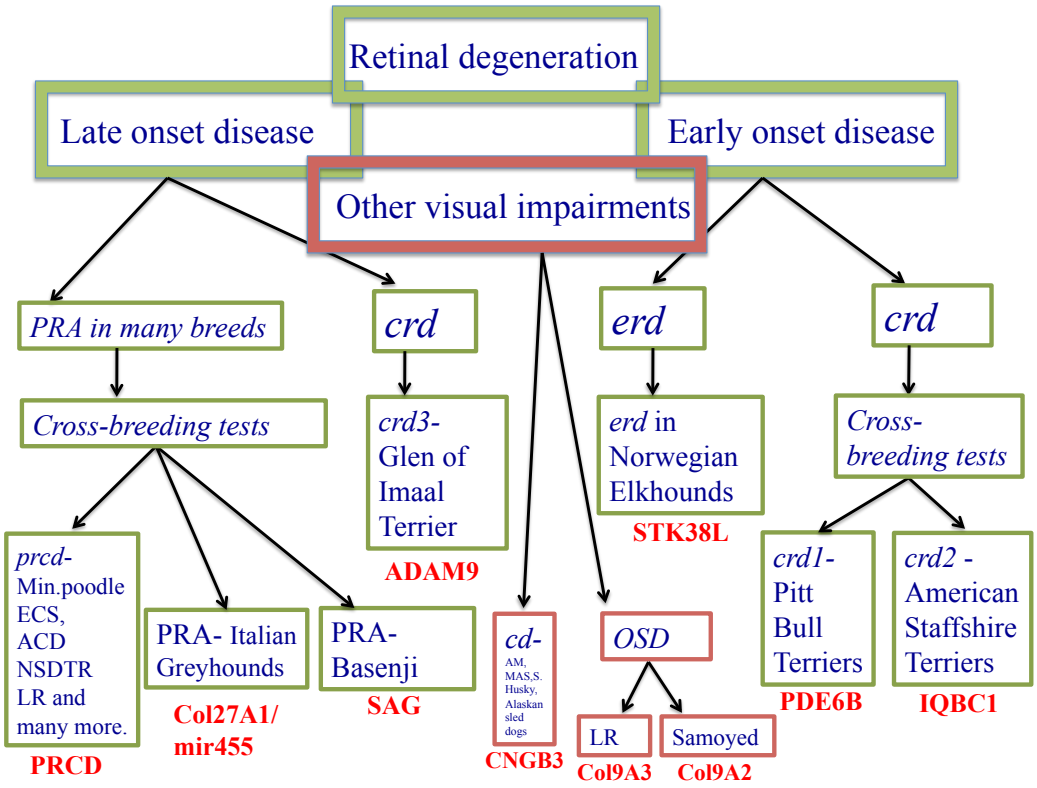


Figure 5.1. Summary of the research results. The locus, gene and mutation were identified in all diseases. Three novel Genes causing PRA were identified: PRCD, STK38L, and mir455, suggesting new candidate genes for human inherited RP, and new pathways important for photoreceptor function. The DNA alteration in IGPRA is yet to be identified.

5.2 From Linkage to Association and everything in between

The domestication of the dog from its grey wolf ancestor (*c. lupus*), unintentionally created a valuable genetic model for studying both simple mendelian traits and more complex ones. Among its advantages as a research model are: the striking similarities of canine diseases to human diseases; accurate breeding records; large families; large LDs; and an excellent large animal therapy model. Through the journey I took exploring retinal diseases in dogs, I also experienced the evolution of gene-mapping approaches, as data emerging from newly developed technologies enabled us to adapt experimental design and methods of analysis. Linkage mapping, unlike the candidate gene approach, interrogates the entire genome, including unknown genes, using informative families segregating the disease. The search is for genetic markers that co-segregate with the disease: If no recombination is observed between a marker and the disease phenotype, most likely the genomic mutation is physically close to the marker. In practice, access to family samples in the "real world" is often extremely difficult when it comes to blood collection and phenotype evaluation. For those reasons, purpose-bred research colonies were developed. By breeding affected dogs to beagles, and subsequent development of three-generation-colony pedigrees segregating the disease under investigation, with many littermates of dogs per breeding, and many meioses to look at, linkage analysis was a powerful tool to use (*prcd*, *crd1*, *crd2*, *cd*, OSD, and *erd*). The outcross to a different breed, introduced a high level of heterozygosity and informativeness that was lacking in the purebred dog population. However, these experiments were extremely expensive and time consuming, and with the late onset diseases in particular sometimes ascertainment of phenotype at an early age becomes challenging, resulting in the need to keep these dogs

for many years, further inflating the cost. Moreover, as the number of meioses observed per breeding in canine pedigrees is limited, multiple breedings are required to generate sufficiently large numbers of informative progeny (over 50 dogs on average, Table 5.1). The identification of over 2.5 million SNPs, densely and broadly placed across the whole genome enabled us to move away from linkage analysis and into association studies. Based on the long within-breed LD[4,5] (between 0.5 and 5 Mb, depending on the breed) 10,000-30,000 SNPs are considered sufficient to map a Mendelian trait with high penetrance and no phenocopies. Lindblad-Toh and her group suggested that 20 cases and 20 controls are sufficient to map autosomal recessive traits and 50 of each group to map autosomal dominant trait[6]. We showed the efficiency of this approach by mapping and identifying the mutations in five diseases, using between 10 and 20 dogs per each group (Table 5.1). In the Basenji PRA disease we were able to map the gene with less than 10 dogs total (6 affecteds and 3 controls). The average minimal LD was 1.8 Mb, though in regions that are gene-rich this interval can still be challenging in hunting for the causative mutation. Nevertheless, the reduction in the number of needed dogs, and in time and in cost obtained by SNP map genotyping and genome-wide association study (GWAS) has been very valuable and powerful.

Table 5.1. Study designs, methods of analysis, tools and results in 10 different retinal diseases in the canine population represented in this work.

Disease	Breeds affected	Number of affected dogs used	Number of unaffected dogs used	Polymorphism tool	CFA	Mapped interval (Mb)	Gene
Linkage studies							
OSD - *drd1*	Labrador retriever	30	50	Microsatellite (genome-wide 249 markers)	24	11.5	Col9A3
OSD - *drd2*	Samoyed	23	47	Microsatellite (genome-wide 249 markers)	15	21	Col9A2
erd	elkhound	>50	>50	microsattelitle+SNPs	27	1.6	STK38L
cd	Alaskan Malamude	>50	>50	microsattelitle+SNPs	29	1.18	CNGB3
prcd	Miniature poodle	>50	>50	microsattelitle+SNPs	9	1.5	PRCD
Association studies (colony-derived dogs)							
crd1	American Staffshire Terrier	17	18	Illumina SNP array (173,662 SNPs)	3	1.05	PDE6B
crd2	Pit bull Terrier	15	13	Affymetrix SNP array (~60,000)	33	2.67	IQCB1
Association studies (purebred dogs)							
crd3	Glen of Imaal Terrier	21	22	Affymetrix SNP array (~60,000)	16	2.74	ADAM9
Basenji PRA	Basenji	6	3	Illumina SNP array (173,662 SNPs)	25	2.09	SAG
IG PRA	Italian Greyhound	21	18	Illumina SNP array (173,662 SNPs)	11	0.413	Col27A1/ mir455
Linkage Disequilibrium							
prcd	>23 breeds	10 (from 10 different breeds)	6	Fine-mapping SNPs	9	0.106	
cd	Alaskan Malamute, S, Husky, Miniautre Australian Shepherd, Alaskan sled-dogs	5 (from 3 different breeds)*	4 (from 4 different breeds)**	Fine-mapping SNPs	29	0.5-1.04	

*- Miniature Australian Shepherd, Alaskan Malamute, Husky carrying one affected chromosome

**- Boxer, Miniature Australian Shepherd, Alaskan Malamute, Husky carrying one normal chromosome

A synergistic combination of both population study (GWAS) and family-based study (Linkage mapping) can be very beneficial, as seen in *crd2* disease. Two significant GWAS hits were observed: one on CFA12 and one on CFA33. The locus bearing the mutation showed a higher p-value, however the exclusion of CFA12 was more convincing when one SNP was genotyped on a large number of dogs in an informative perdigree, and yielded a non-significant lod score, with theta close to 0.5.

While the extensive within-in breed LD facilitates initial mapping of a chromosomal region, identifying the mutation can be challenging when no obvious candidate gene is available. When mapping diseases expressed uniquely in one breed, LD also depends on the location of the mutation as we saw in our diseases: closer to the centromer, where recombination is suppressed, LD is larger than when the mutation is on the telomeric end of the chromosome (PDE6B and Col27A1 are on the telomeric ends of their chromosomes, with LD of 0.413-1.05 Mb compare to ADAM9, SAG and IQCB1 which are located more proximal with 2.09-2.74 Mb interval; Table 5.1).

PRCD is located on the centromeric end of CFA9. As such, the number of recombinants within each available pedigree was very small, making it difficult to narrow the linkage interval. LD across unrelated dogs within the same breed in most cases did not reduce the interval. Cross-breed test matings among several breeds suggested that their diseases were allelic to each other. Under the assumption that all the breeds affected with *prcd* are identical by descent (IBD), however, we were able to show that comparing fine-map-haplotypes across breeds was extremely efficient in reducing the LD, and the number of candidate genes from 50 to 3. The initial locus was mapped by linkage analysis to a 1.9 Mb interval and the identification of an ancestral haplotype common to all affected dogs from 14 different breeds reduced the LD to 106

Kb. That this approach can sometimes work even within one breed, if dogs are taken from different geographic isolates, was shown in for *prcd* and in the toy poodle breed.
Researchers have to bear in mind though that running association studies across many breeds would only work if the disease allele is inherited IBD, which sometimes is difficult to determine. We were able to show that mapping genes in the canine population by means of association can be done on pooled samples. Careful steps have to be taken to assure success. The major tradeoff for the reduced cost of the experiment is a larger LD interval, as rare recombinants can be missed.

5.3 Novel genes, novel pathways

In this work we were able to identify three novel genes and pathways responsible for photoreceptor function and vision: PRCD, STK38L and mir455.
Reducing the *prcd*-LD interval to a 106 Kb region encompassing only three known genes was critical for the identification of the *prcd* novel gene. Excluding these three positional candidates showed, without a doubt, that the mutation causing *prcd* had to be a novel gene, and likely one of the conserved sequences in the interval. Conservation analysis, together with the analysis of a normalized canine EST library[7], suggested the presence of a novel gene, with a predicted protein of 54 amino acids[8]. The first 4 amino acids coded by the first exon are highly conserved and *prcd*-affected dogs were homozygous for a G to A transition at nucleotide 5 of the coding sequence, resulting in a cysteine to tyrosine change (C2Y)[8]. This discovery set the ground for screening human RP patients for this gene, and identifying two unrelated families with mutations in PRCD: one family from Bangladesh carrying the same mutation as observed in the dog[8], and an Israeli Arab family, with a novel mutation in PRCD causing RP in homozygous recessive

individuals[9]. More recently, a third case of mutated PRCD was identified in a Turkish family affected with RP[10]. These findings establish its etiology in RP. Our protein model suggested a transmembrane domain in the C-terminus of the protein, with an alpha-helix structure[8]. Recently, proteomic analysis of the rod and cones discs identified PRCD as one of the main proteins expressed, and at levels of molar ratio of about 1:290 between PRCD and rhodopsin[11].

We showed that an STK38L mutation causes early retinal degeneration in Norwegian elkhounds, and thus, for the first time, implicated a critical involvement of STK38L in photoreceptor development and disease. As of today, no human patients with mutations in STK38L or any of the proteins that interact with it (S100B, Mob1, Mob2) have been identified as mutated in RP patients. However, we anticipate that one or more of these proteins will probably prove to be mutated in some portion of the RP patients yet to be resolved. We followed the investigation on the *erd*- affected retinas and showed that during postnatal weeks 7 to 14 some photoreceptor cells become committed to apoptosis and some are recruited differentiate such that the ONL thickness does not change[12]. Coincident with this photoreceptor cell death and proliferation phase is a change in the visual cell population: both cone types of cells remain (L/M and S-cones) but the rods that presumably were generated after proliferation now become hybrid photoreceptors that express both rod-opsin and cone opsin. These results suggest a role for STK38L in the control of cell division and morphogenesis in photoreceptors and possibly other retinal neurons. Recently, a group of researchers in Japan provided evidence that STK38L phosphorylates Rabin8 at Ser-272 and that this phosphorylation is crucial for ciliogenesis[13]. They propose that NDR2-mediated Rabin8 phosphorylation triggers the switch in binding specificity from PS to Sec15 and thereby promotes local activation of Rab8 and ciliary membrane

formation. A mutation in this gene therefore would affect ciliogenesis. Defects in primary cilium formation impair cell sensitivity to external signals in various tissues and can cause diverse genetic disorders[14,15]. It is yet to be determined if *erd* is indeed a ciliopathy disease. Our data suggest that *erd*-affected cells have at least some capability to transport proteins to the OS, as the outer segment renewal experiments show normal incorporation of labeled opsin into the outer segment (although it is redistributed subsequently throughout the rods due to formation of hybrid rod/cones which have a cone outer segment disc morphology).

In the IGPRA dogs we were able to associate a disease in a large animal model with an alteration in the level of a microRNA gene. Mouse models for RP exist that show alterations in the level of expression of several mir genes. Our findings support these observations and suggest that some RP patients might be carrying a mutation in one of the mir genes expressed in the retina.
The overall mir gene network is not fully understood. In mammals, many miRNAs exist as duplicates or highly similar genes, which raises the question of both functional redundancy and cooperation[16-18]. Many computational target-prediction tools predict up to a couple of hundred targets per miRNA[19]. However, it was also observed that some miRNAs can control a specific gene regulation program through a few major targets, suggesting they may function as switches[20,21]. Although proteomics approach usually demonstrate changes in many genes in miRNAs-knockout mice, only a few of those may be critical for a particular phenotype. The downstream consequences of low mir455 expression in the retina remain unclear at this point. miRNA genes are also known to act as modifiers in pathogenesis[22]. Our hit to mir184 when comparing IG heterozygous dogs affected with PRA to IG heterozygous dogs showing no symptoms of PRA suggests that miRNA interaction and cooperation might influence penetrance

in dominant diseases with incomplete penetrance. The observation that a reduction in mir455 expression causes a significant change in the expression of mir3609 also suggest that miRNAs cooperate in a network, with multiple pathways and feedbacks, and changes in expression of one mir may result in changes in expression of other mir genes and maybe also of SNORD genes.

5.4 Future plans

As in any research, the newly discovered data raises more questions.

What does PRCD do in the retina? Which genes does it interacts with? What are the modifiers of the disease? Why does the gene have such a long 3'UTR? Can PRCD be rescued with gene-therapy?

What is the function of STK38L in the retina? Are there any human patients with mutations in STK38L or any of its interacting proteins (S100B, Mob1, Mob2)?

How does low expression of mir455 cause retinal degeneration? What are the target genes of mir455? What is its relationship with mir3609 and what role does the latter have in the maintenance of the retina? Can mir genes be a therapeutic approach for blindness?

Is gene therapy possible for IQCB1 in cone-rod dystrophy 2 in the Pit bull terrier?

Some of these questions are very broad and may take years to answer. Some of the questions are easier to answer in a mouse model, and some are better answered in a large animal model such as the dog. One thing is definite: there is much more to be done!

5.5 References

1. Rodieck RW (1998) The first steps in seeing. Sinauer Associates, Sunderland.

2. Palczewski K (2011) Focus on vision: 3 decades of remarkable contributions to biology and medicine. FASEB J 25(2):439–443.

3. Ridge KD, Palczewski K (2007) Visual rhodopsin sees the light: structure and mechanism of G protein signaling. J Biol Chem 282(13):9297–9301.

4. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK. Linkage disequilibrium and demographic history of wild and domestic canids. Genetics. 2009 Apr;181(4):1493-505.

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

6. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshatsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolen O, Elong R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M, Foley K, Foley C, Franke A, Friedrich D, Gage D, Garber M, Gearin G, Giannoukos G, Goode T, Goyette A, Graham J, Grandbois E, Gyaltsen K, Hafez N, Hagopian D, Hagos B, Hall J, Healy C, Hegarty R, Honan T, Horn A, Houde N, Hughes L, Hunnicutt L, Husby M, Jester B, Jones C, Kamat A, Kanga B, Kells C, Khazanovich D, Kieu AC, Kisner P, Kumar M, Lance K, Landers T, Lara M, Lee W, Leger JP, Lennon N, Leuper L, LeVine S, Liu J, Liu X, Lokyitsang Y, Lokyitsang T, Lui A, Macdonald J, Major J, Marabella R, Maru K, Matthews C, McDonough S, Mehta T, Meldrim J, Melnikov A, Meneus L, Mihalev A, Mihova T, Miller K, Mittelman R, Mlenga V, Mulrain L, Munson G, Navidi A, Naylor J, Nguyen T, Nguyen N, Nguyen C, Nguyen T, Nicol R, Norbu N, Norbu C, Novod N, Nyima T, Olandt P, O'Neill B, O'Neill K, Osman S, Oyono L, Patti C, Perrin D, Phunkhang P, Pierre F, Priest M, Rachupka A, Raghuraman S, Rameau R, Ray V, Raymond C, Rege F, Rise C, Rogers J, Rogov P, Sahalie J, Settipalli S, Sharpe T, Shea T, Sheehan M, Sherpa N, Shi J, Shih D, Sloan J, Smith C, Sparrow T, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Stone S,

Sykes S, Tchuinga P, Tenzing P, Tesfaye S, Thoulutsang D, Thoulutsang Y, Topham K, Topping I, Tsamla T, Vassiliev H, Venkataraman V, Vo A, Wangchuk T, Wangdi T, Weiand M, Wilkinson J, Wilson A, Yadav S, Yang S, Yang X, Young G, Yu Q, Zainoun J, Zembek L, Zimmer A, Lander ES. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005 Dec 8;438(7069):803-19.

7. Zangerl B, Sun Q, Pillardy J, Johnson JL, Schweitzer PA, Hernandez AG, Liu L, Acland GM, Aguirre GD. Development and characterization of a normalized canine retinal cDNA library for genomic and expression studies. Invest Ophthalmol Vis Sci. 2006 Jun;47(6):2632-8.

8. 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. Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. Genomics. 2006 Nov;88(5):551-63.

9. Nevet MJ, Shalev SA, Zlotogora J, Mazzawi N, Ben-Yosef T. Identification of a prevalent founder mutation in an Israeli Muslim Arab village confirms the role of PRCD in the aetiology of retinitis pigmentosa in humans. J Med Genet. 2010 Aug;47(8):533-7.

10. Pach J, Kohl S, Gekeler F, Zobor D. Identification of a novel mutation in the PRCD gene causing autosomal recessive retinitis pigmentosa in a Turkish family. Mol Vis. 2013 Jun 13;19:1350-5.

11. Skiba NP, Spencer WJ, Salinas RY, Lieu EC, Thompson JW, Arshavsky VY. Proteomic identification of unique photoreceptor disc components reveals the presence of PRCD, a protein linked to retinal degeneration. J Proteome Res. 2013 Jun 7;12(6):3010-8.

12. Berta ÁI, Boesze-Battaglia K, Genini S, Goldstein O, O'Brien PJ, Szél Á, Acland GM, Beltran WA, Aguirre GD. Photoreceptor cell death, proliferation and formation of hybrid rod/S-cone photoreceptors in the degenerating STK38L mutant retina. PLoS One. 2011;6(9):e24074.

13. Chiba S, Amagai Y, Homma Y, Fukuda M, Mizuno K. NDR2-mediated Rabin8 phosphorylation is crucial for ciliogenesis by switching binding specificity from phosphatidylserine to Sec15. EMBO J. 2013 Mar 20;32(6):874-85.

14. Fliegauf M, Benzing T, Omran H (2007) When cilia go bad: cilia defects and ciliopathies. Nat Rev Mol Cell Biol 8: 880–893.

15. Gerdes JM, Davis EE, Katsanis N (2009) The vertebrate primary cilium in development, homeostasis, and disease. Cell 137: 32–45.

16. Liu N., Bezprozvannaya S., Williams A.H., Qi X., Richardson J.A., Bassel-Duby R., Olson E.N. microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. Genes Dev. 2008;22:3242-3254.

17. van Rooij E., Sutherland L.B., Qi X., Richardson J.A., Hill J., Olson E.N. Control of stress-dependent cardiac growth and gene expression by a microRNA. Science 2007;316:575-579.

18. van Rooij E., Quiat D., Johnson B.A., Sutherland L.B., Qi X., Richardson J.A., Kelm R.J. Jr., Olson E.N. A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. Dev Cell. 2009;17:662-673.

19. Bartel D.P. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215-233.

20. Xiao C., Calado D.P., Galler G., Thai T.H., Patterson H.C., Wang J., Rajewsky N., Bender T.P., Rajewsky K. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell 2007;131:146-159.

21. Johnnidis J.B., Harris M.H., Wheeler R.T., Stehling-Sun S., Lam M.H., Kirak O., Brummelkamp T.R., Fleming M.D., Camargo F.D. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature 2008;451:1125-1129.

22. Bandiera S, Hatem E, Lyonnet S, Henrion-Caude A. microRNAs in diseases: from candidate to modifier genes. Clin Genet. 2010 Apr;77(4):306-13.