IDENTIFICATION OF A MUTATION IN *COL4A5* CAUSATIVE FOR X-LINKED ALPORT SYNDROME IN THE DOMESTIC DOG AND ANALYSIS OF GENE EXPRESSION IN THE KIDNEYS OF AFFECTED AND NONAFFECTED SIBLINGS

A Dissertation

by

MELISSA LUANNE COX

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2003

Major Subject: Genetics

IDENTIFICATION OF A MUTATION IN *COL4A5* CAUSATIVE FOR X-LINKED ALPORT SYNDROME IN THE DOMESTIC DOG AND ANALYSIS OF GENE EXPRESSION IN THE KIDNEYS OF AFFECTED AND NONAFFECTED SIBLINGS

A Dissertation

by

MELISSA LUANNE COX

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Keith E. Murphy (Chair of Committee)	George E. Lees (Member)
James E. Womack (Member)	James E. Derr (Member)
Ann B. Kier (Head of Department)	Geoffrey M. Kapler (Chair of Genetics Faculty)

December 2003

Major Subject: Genetics

ABSTRACT

Identification of a Mutation in *COL4A5* Causative for X-linked Alport Syndrome in the Domestic Dog and Analysis of Gene Expression in the Kidneys of Affected and Nonaffected Siblings. (December 2003)

Melissa Luanne Cox, B.Sc., Dalhousie University

Chair of Advisory Committee: Dr. Keith E. Murphy

The domestic dog, *Canis lupus familiaris*, plays many roles in the lives of humans. Additionally, the dog is recognized for its potential as a model for many human hereditary diseases. Thus, the genetics and genomics of the dog are being studied extensively in order to facilitate its use as a model, as well as to help the dog for its own sake. As part of this research effort, our laboratory has added type I markers (i.e., the acidic and basic keratins, *c-kit*, type I and IV collagens, and the gene encoding uromodulin*)* to the emerging map of the canine genome. The mapping of genes, particularly those in large gene families such as the collagens, is valuable because it rapidly increases the density of gene loci on the map and provides insight regarding conservation of synteny between the dog and other mammals. The major focus of work reported here is the genetics of X-linked Alport syndrome (XLAS), a terminal renal disease that affects the human and the dog. The disease results from mutations in *COL4A5*, a type IV collagen gene. Reported here are the 1) sequencing and mapping of the canine cDNA encoding uromodulin, 2) mapping of the type I and type IV collagen

genes, 3) sequencing of the full-length cDNA of canine *COL4A5,* 4) identification of a 10 bp deletion in *COL4A5*, causative for XLAS in our colony of mixed breed dogs, 5) development of a genetic test for identification of affected and carrier dogs in the colony and 6) assessment of gene expression in the kidneys of normal and XLAS-dogs. This assessment was performed using a canine-specific oligonucleotide microarray. XLAS dogs demonstrated up-regulation of many genes involved in extracellular matrix reorganization, cell structure, and immune response, as expected in a glomerulopathy with tubulointerstitial nephritis. Trends were verified by quantitative RT-PCR. A review of the current status of canine genetics research, and current understanding of hereditary diseases in the dog, concludes this dissertation.

For Andrew
For Mum
and for all of our pups

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Keith Murphy, for the time he has invested in me and in this project. The training I have received while in his laboratory has been invaluable and I have been very fortunate to have his guidance and support in my graduate career. Dr. Murphy encouraged me to take on a challenging project, and helped me work through the stumbling blocks along the way. I have also been very lucky to have an excellent committee, comprised of Drs. James Derr, George Lees, and James Womack. Their assistance and encouragement have always been insightful and helpful.

My friends and colleagues in the Canine Genetics Laboratory have provided me with assistance beyond which one could reasonably ask. From laughs to plasmid preps, manuscript editing to dog sitting, the members of this lab have helped me in so many ways.

The colony of Navasota dogs has been established and nurtured by Dr. George Lees and a talented staff of technicians, veterinary students, and student workers. They have taught me details of the dogs' assessments and treatments, as well as general daily life with the colony. Dr. Lees has always had time to discuss ideas, impart knowledge and share his enthusiasm about the dogs and the kidney.

Our academic collaborators are numerous: Drs. Clifford Kashtan, Elaine Ostrander, Jennifer Lowe, Catherine Andre, Francis Galibert, Frode Lingaas, and their graduate students have helped with data collection, analysis, and writing of manuscripts. Our industrial collaborators in the Toxicogenomics Laboratory and Pathology Group at Eli Lilly & Co., Dr. Tim Ryan, Ms. Marnie Higgins, Mr. George Searfoss, Dr. Brian

Berridge, and Dr. Baohui Li have also helped us by providing training, microchips, and even housing for the microarray and Q-RT-PCR work I have done. It should be noted that the data published in Chapter V were generated in collaboration with this group, and Ms. Marnie Higgins contributed equally to the work that I report in this chapter.

I am so glad to have had Drs. Kelly Credille and Bob Dunstan's friendship and guidance throughout my time in grad school. Their enthusiasm for science is infectious. Also, I have been very fortunate to have an amazing group of friends to help me in grad school and in life – Gisela, the Pachucos, CWI, and so many others – graduate school can bind friendships like no other experience!

I also have an incredibly supportive family including my grandparents, my sister, and my in-laws – but two members deserve special mention. My husband, Andrew Treash, has always believed in my goals, and me, and has done everything to help me succeed. I am so lucky to have my best friend as my husband. My mother, Heather Logan, has always been my guide, my cheerleader, and my friend. Thank you, both of you, for everything.

TABLE OF CONTENTS

		Page
ABSTRACT		iii
DEDICATION		v
ACKNOWLEDGMENTS		vi
TABLE OF CONTENTS		viii
LIST OF FIGURES		x
LIST OF TABLES		xi
CHAPTER		
I	INTRODUCTION	1
	Genetics and genomics of the dog	1
	Sequencing and mapping of canine gene loci	3
	Biology of uromodulin	5
	The collagen genes	6
	Clinical aspects of Alport syndrome	8
	Cell biology and genetics of Alport syndrome	9
	Canine Alport syndrome	11
	Specific objectives	13
	Summary	14
II	RADIATION HYBRID MAPPING OF THE CANINE TYPE I AND TYPE IV COLLAGEN GENE SUBFAMILIES	15
	Overview	15
	Introduction	15
	Materials and methods	18
	Results	20
	Discussion	22

CHAPTER Page

III SEQUENCING AND RADIATION HYBRID MAPPING OF CANINE *UROMODULIN* 25

Overview 25
Introduction 26
Report 27

IV GENETIC CAUSE OF X-LINKED ALPORT SYNDROME IN A FAMILY OF DOMESTIC DOGS 38

Overview 38
Introduction 39
Materials and methods 42
Results 46
Discussion 50

V GENE EXPRESSION ANALYSIS IN A CANINE MODEL OF X-LINKED ALPORT SYNDROME 54

Overview 54
Introduction 55
Materials and methods 58
Results 59
Discussion 66

VI DIGGING UP THE CANINE GENOME – A TALE TO WAG ABOUT 70

Overview 70
Introduction 71
Current status of the canine genome 72
Canine diseases and genetic analyses 75
Future of canine genomics 82

VII SUMMARY 83

REFERENCES 88

VITA 102

LIST OF FIGURES

FIGURE		Page
1	CLUSTAL W (1.82) multiple sequence alignment (http://www.ebi.ac.uk/clustalw/)	31
2	Nucleotide and amino acid sequences for *COL4A5* exons 9 and 10	48
3	Example results of an allele-specific test for XLAS in Navasota dogs	49
4	Analysis of microarray data: Average fold change of selected genes in affected NAV dogs	63
5	Q-RT-PCR data: Gene expression (by fold change) of normal and affected NAV dogs	64
6	Q-RT-PCR and microarray data: Average fold change of up-regulated genes in affected NAV dogs	65
7	Q-RT-PCR data: Fold change of type IV collagens in normal and affected NAV dogs	65

LIST OF TABLES

TABLE		Page
1	Type IV and IV collagen gene subfamilies	16
2	Radiation hybrid mapping of collagen markers on the canine integrated 1,800-marker map	21
3	Primer sets and conditions for the amplification of canine *umod* cDNA	30
4	*COL4A5* primer sets	44
5	Primer sequences for genes amplified by Q-RT-PCR	60
6	Analysis of microarray data: Example of genes identified as up- or down- regulated in affected NAV dogs as compared with normal siblings	61
7	Analysis of microarray data: Average fold changes of 10 genes up-regulated in affected dogs	62
8	The canine genome and current resources for mapping	76

CHAPTER I

INTRODUCTION

Genetics and genomics of the dog

The domestic dog, *Canis lupus familiaris*, has contributed to our lives in many ways but most importantly as a companion. The dog is a subspecies of wolf, *Canis lupus*, and medical surveillance and treatment of it are second only to that of the human. Dogs have been bred for many purposes, with goals of both practicality and aesthetics. More specifically, nonrandom breeding programs have resulted in the creation of types and finally "purebreeds", most of which were developed in the past 250 years (Ostrander and Giniger 1997). In order to fix desirable traits, breeders have practiced various levels of inbreeding. Founder effects, population bottlenecks and surges in breed popularity have also placed intense selective pressure upon the dog, resulting in the emergence of more than 450 known hereditary diseases (OMIA 2003).

The major problem for breeders of purebred dogs is that the majority (Ostrander and Kruglyak 2000) of hereditary diseases in the dog are inherited in an autosomal recessive manner. Thus, it is not possible to prevent the production of affected animals without knowing the carrier status of the parents. Additionally, late-onset autosomal dominant diseases are also difficult to prevent because affected dogs are often used for breeding before symptoms of such diseases become evident. To improve the quality of

---

This dissertation follows the style and format of Mammalian Genome.

life for dogs and the peace of mind of their owners, it is essential to develop genetic tests that identify carriers and those dogs destined to be affected by hereditary diseases. Understanding the genetic bases of the myriad of canine hereditary diseases may also lead to preemptive treatments or better management of clinical disease.

Approximately half of the aforementioned 450 hereditary canine diseases are also found in the human and have very similar clinical presentations. Importantly, the human and canine diseases are most often caused by mutations in the same genes (Ostrander and Giniger 1997; OMIA 2003). Therefore, due to large litter sizes, short generation times, reduced genetic heterogeneity and well documented inbred lines, the dog serves as an ideal natural model for study of the genetics governing various human hereditary diseases as well as other inherited traits. It is not surprising, then, that the dog is being utilized to understand human genetics, inheritance of complex traits and in gene therapy regimens (Bartlett et al. 1996; Knapp and Waters 1997; Venta et al. 2000; Acland et al. 2001; Hungs et al. 2001; Lou et al. 2003). Of course, a great advantage that allows for the dog to be used to full potential in development of treatments and therapies applicable to the human is our extensive knowledge of canine medicine. Furthermore, because dogs usually live in the same environment as humans, external factors that may affect expression of hereditary diseases in both animals are the same. Even so, the cost of maintaining colonies of dogs is often prohibitive. However, collaborations among breeders, clubs and geneticists often allows for the collection of pedigrees and appropriate samples (blood, DNA, biopsies, etc.) for the study of these diseases without having to maintain such colonies.

Sequencing and mapping of canine gene loci

During the last decade, the major impediment to using the dog as a model has been the paucity of information about the canine genome and the lack of resources necessary to study canine genetics. The 38 small, acrocentric chromosomes of the dog complicated the standardization of the canine karyotype and prevented genomic mapping efforts typical of other model organisms such as the mouse. In 1999, the consensus karyotype was established with the use of canine whole chromosome-specific fluorescence *in situ* hybridization (FISH) paint probes (Breen et al. 1999a; Breen et al. 1999b).

While FISH is certainly a valuable tool, a critical step in the study of the canine genome was the construction of a radiation hybrid (RH) cell panel. To this end, a 5000-rad RH panel was constructed from canine fibroblasts and fused with HTK3-1 Chinese hamster cells (Vignaux et al. 1999). The most recent version of this panel, RHDF5000-2, is comprised of 118 cell lines on which 3270 markers have been mapped (Guyon et al. 2003). The integration of cytogenetic data and meiotic linkage data has resulted in a map of the canine genome with 1-Mb resolution (Guyon et al. 2003).

An important advance in canine genomics has been the identification and characterization of microsatellite markers. Briefly, characterization of 150 microsatellite markers and 30 linkage groups provided the framework for a linkage map that was rapidly expanded upon in recent years (Mellersh et al. 1997). These markers are the foundation on which early maps were built and the current version of the canine linkage map has 1596 microsatellite markers distributed across all chromosomes (Guyon et al.

2003). The utility of microsatellites is further emphasized by the fact that these markers can be used to assess linkage of markers with hereditary diseases (Flinter et al. 1989; Ron et al. 1994; Tanaka et al. 2003). However, in order for this to be effective, markers used in linkage analysis must be selected based on information content. Therefore, using the integrated linkage-radiation hybrid map developed in 1999, 172 microsatellite markers were selected and characterized for use in whole genome screens. This subset of markers provides 10 cM coverage of the canine genome and is termed the Minimal Screening Set -1 (MSS-1) (Richman et al. 2001). The MSS-1 was enhanced by multiplexing that streamlines screens of the genome (Cargill et al. 2002). Finally, construction of the most recent version of the canine map (Guyon et al. 2003) made possible the identification of a more comprehensive screening set, the MSS-2. The MSS-2 is comprised of 327 microsatellites providing 5 cM coverage and has an average spacing of 9 Mb (Guyon et al. 2003).

The importance of the dog as a model for study of human hereditary diseases is evidenced by the decision of the National Human Genome Research Institute (NHGRI) (http://www.genome.gov/11008069) to fully support sequencing the genomes of several breeds. Celera conducted the first large scale sequencing effort for the dog using a male Standard Poodle; this 1X coverage sequence is only available through collaborations with The Institute for Genome Research (TIGR). The need for a higher resolution sequence was recognized by NHGRI and sequencing of the dog that will yield a final product with 6.5X coverage will likely be completed in January 2004. The dog being sequenced is a female Boxer named Tasha. Other breeds to be sequenced are yet to be

determined but the Beagle, Doberman Pinscher and the gray wolf will be sequenced with final resolution of approximately 1X. All sequencing is being done using a whole genome shotgun approach at The Whitehead Institute. Data are made available on a daily basis through the deposition of trace sequences into Ensembl (ftp://ftp.ensembl.org/pub/traces/canis_familiaris/) and the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/trace).

Although the sequence of the dog will soon be available, mapping is still an important endeavor because markers are necessary to both localize and provide "landmarks" for sequence information. Additionally, mapping is important in comparative genomics for establishing regions of conserved synteny. As part of the effort to supplement mapping of the canine genome, our laboratory maps large gene families, particularly those encoding structural proteins of medical relevance. For example, the acidic and basic keratins (Miller et al. 1999; Miller et al. 2001) are implicated in various skin diseases, and the type I and type IV collagens are known to be involved in diseases of the connective tissue and facia. We are also interested in genes expressed in an organ-specific fashion that may potentially play roles in diseases of the human and dog. The uromodulin gene and the collagen gene superfamily are of current interest to us.

Biology of uromodulin

Uromodulin, also known as "Tamm-Horsfall glycoprotein" is a renal-specific, GPI-anchored cell membrane protein, and is the most abundant protein found in normal

urine. A definitive function for uromodulin has not yet been established but recently it was discovered that mutations in the gene are causitive for two hereditary human renal diseases: medullary cystic kidney disease and familial juvenile hyperuricemic nephropathy (Hart et al. 2002). These diseases are transmitted in autosomal dominant fashion and are typified by juvenile onset of hyperuricemia, gout, and progressive renal failure.

The collagen genes

The superfamily of collagen genes consists of more than 40 genes, grouped into 27 subfamilies. These proteins are the major components of connective tissues, extracellular matrix, and basement membranes in the body. Fibrous polypeptides, such as the type I and type IV collagens, coil into left-handed helices, and three such chains, in different combinations, wrap around each other to form a right-handed super-helix. The triple-helical domains intertwine and supercoil and the proteins usually form fibers or networks. Collagens are characterized by Gly-X-Y repeats in the collagenous domains. The presence of glycine (the smallest amino acid) at every third amino acid position is essential for the tight folding pattern to occur. Proline and hydroxyproline are often found in the X and Y positions, respectively. Globular sequences may interrupt the collageneous domains and confer greater flexibility. A short, non-collagenous domain (7S domain) exists at the amino terminus, while a long globular domain (NC1 domain) of approximately 230 amino acid residues exists at the carboxyl terminus. These non-

collagenous domains contain conserved cysteine residues that are critical for disulfide bonding (Pihlajaniemi et al. 1990).

There are two type I collagens, COL1A1 and COL1A2. These are found in bone, skin, and tendons, and mutations in each are responsible for forms of Ehlers Danlos syndrome and Osteogenesis imperfecta. There are six known type IV collagens, COL4A1-COL4A6, which form lattices upon which other basement membrane components are assembled. No known diseases are caused by mutations in the *COL4A1* or *COL4A2* genes. In contrast, mutations in the other type IV collagen genes are causative for several diseases, including Alport syndrome (AS) which is categorized as X-linked Alport syndrome (XLAS) (Hudson et al. 2003) and autosomal recessive Alport syndrome (ARAS) (Hudson et al. 2003). Autosomal dominant forms of the disease have been described, but the genetic cause is unknown (Hood et al. 1995; Ciccarese et al. 2001; Longo et al. 2002). Mutations in *COL4A4* causative for ARAS may also be responsible for Benign Familial Hematuria (Thin Basement Membrane Disease) (Lemmink et al. 1996). In addition to the mutations in *COL4A5* that cause X-linked dominant AS, contiguous and extensive gene deletions, typified by deletions of the 5' regions of *COL4A5* and *COL4A6* (including up to exon 3 of *COL4A6*), result in diffuse leiomyomatosis which is often seen in conjunction with AS (Zhou et al. 1993; Heidet et al. 1997). Although *COL4A3* has not been implicated in a disease other than ARAS, it is involved in Goodpasture syndrome, an autoimmune disorder in which the globular NC1 domain of COL4A3 is an epitope for autoantibodies (Turner et al. 1992).

Clinical aspects of Alport syndrome

Alport syndrome is a heterogeneous group of hereditary glomerular diseases that lead to progressive renal failure. While rare, AS is an important cause of human end stage renal disease (ESRD) worldwide. Information regarding the prevalence of AS, as well as other monogenic diseases that cause ESRD, is sparse (Levy and Feingold 2000). One recent estimate, based on data from Finland, was that the birth prevalence of AS is 1 in 53,000 live births (Pajari et al. 1996). Such data likely apply to large populations of European origin, but data regarding large populations of other geographic origins are lacking (Levy and Feingold 2000). In Europe, AS accounts for 1.5% of children starting renal replacement therapy (Loirat et al. 1994). In the United States, AS affects about 3% of children with chronic renal failure and about 0.3% of adults with ESRD (USRDA 1996).

Although the phenotypic spectrum of AS is diverse, the underlying structural abnormalities that are common to all forms of the disease are defective basement membranes in various organs, hematuria, and proteinuria. In the kidney, the alterations in the structure of glomerular basement membranes (GBM) are associated with progressive changes in glomerular function. Permselectivity of the glomerular filtration barrier becomes altered and proteinuria (the abnormal loss of protein in the urine) ensues. Progressive glomerular injury and scarring (glomerulosclerosis) occur such that individual glomeruli eventually cease filtration entirely. Progressive reduction of glomerular filtration rate (GFR) eventually leads to renal failure. These sequential glomerular changes are accompanied by progressive injury, inflammation and scarring

that affects the tubules and interstitium of the renal cortex. Thus, although the nephropathy associated with AS is initiated by molecular defects in the GBM, eventually the entire kidney is affected by the progressive structural damage and functional deterioration that lead to ESRD.

While it is true that different modes of genetic transmission exist for AS, the majority (more than 95%) of cases are either XLAS or ARAS. Approximately 85% of human AS cases are X-linked. In males with XLAS, the clinical symptoms generally begin with microscopic hematuria starting at around five years of age, followed by sensorineural deafness (Cohen et al. 1961; Crawfurd and Toghill 1968), hypertension by the mid-teens, and a deterioration of renal function, culminating in chronic renal failure in the mid-twenties (McCarthy and Maino 2000). Carrier females, due to lyonization, are mosaics. Since XLAS is a dominant disease, females usually show microscopic hematuria in their early twenties, and subsequent hypertension, but the remarkable reserve capacity of the kidney ensures that only 5-10% of carriers progress to chronic renal failure (McCarthy and Maino 2000). In addition to the aforementioned problems arising from AS, patients may also exhibit ocular defects (Colville and Savige 1997). At present, the only treatment options available for patients with AS-induced chronic renal failure (CRF) are dialysis and eventual renal transplantation.

Cell biology and genetics of Alport syndrome

Renal function is impaired in XLAS and ARAS because the GBM lacks the crucial type IV collagen network containing COL4A3, COL4A4, and COL4A5 that is

required for long-term stability of the GBM. This network is an assembly of heterotrimeric proteins containing each of these three peptides, and absence of one any of the three chains precludes formation of the network. Thus, a mutation in any of the genes encoding COL4A3, COL4A4, or COL4A5 results in a dominant negative effect. Consequently, the integrity of the GBM is compromised and chronic progressive renal disease inevitably ensues. XLAS and ARAS have identical clinical manifestations, and are characterized by the absence of COL4A3, COL4A4, and COL4A5 in the GBM of affected individuals. Genetic causes of the rare autosomal dominant forms of AS are obscure. XLAS results from mutations in *COL4A5* (Martin et al. 1998) that was mapped to Xq22.3 (Hostikka et al. 1990). ARAS results from mutations in either the *COL4A3* or *COL4A4* genes, which are found on chromosome 2 in humans.

Interestingly, COL4A3, COL4A4, and COL4A5 are not part of the GBM early in development. COL4A1 and COL4A2 initially predominate in the GBM of immature glomeruli, and are replaced quantitatively by COL4A3, COL4A4 and COL4A5 as glomerular maturation continues. Thus, glomerular development normally involves an isotypic switch from the fetal-type COL4A1/COL4A2 GBM network to the adult-type COL4A3/COL4A4/COL4A5 GBM network in mature glomeruli (Miner 1998). The COL4A3, COL4A4, and COL4A5 chains have more cysteine residues in their triple helical and 7S regions, as compared to COL4A1 and COL4A2. This increases disulfide bonding and increases resistance to endopeptidases as compared to COL4A1 and COL4A2 (Kalluri et al. 1997). In subjects with XLAS, however, the normal isotypic

switch cannot occur because of the absence of functional COL4A5. As a consequence, the GBM eventually becomes "leaky" because of this abnormality.

Although type IV collagen abnormalities have been identified as the inciting cause of AS, important aspects of AS remain ill defined (Kashtan 2002). For example, the complete absence of the GBM COL4A3/COL4A4/COL4A5 network does not impair glomerular development nor does it cause glomerular function to be abnormal early in life (Harvey et al. 1998). Also, mechanisms linking the type IV collagen defects to subsequent events, beginning with GBM changes and culminating with ESRD, are incompletely understood (Kashtan 2002). Moreover, treatments of proven efficacy to retard or halt progression of renal failure in AS patients have not been developed (Cosgrove et al. 1996; Lu 1999; Miner and Sanes 1996).

Canine Alport syndrome

The dog is the only naturally occurring model for XLAS and two models are under study: the Samoyed (Zheng et al. 1994) and a mixed breed (Lees et al. 1999). The progression of AS in these dogs from birth to ESRD is very similar to that observed in humans. However, unlike in human beings with XLAS who may also develop sensorineural deafness and ocular abnormalities, only renal disease has been noted in spontaneous canine models of XLAS. In dogs with XLAS, ultrastructure of the GBM is completely normal at birth despite lack of the normal developmental isotypic switch and complete absence of the COL4A3/COL4A4/COL4A5 network in the GBM (Harvey et al. 1998). Renal biopsies of Samoyeds with XLAS examined at one month of age, when

the isotypic switch would usually take place, exhibit bilaminar splitting of the GBM (Harvey et al. 1998). This suggests that COL4A1 and COL4A2 are necessary and sufficient for normal GBM development, while the COL4A3, COL4A4, and COL4A5 network is necessary for GBM stability and maintenance of glomerular function (Harvey et al. 1998).

In 1993, Dr. George E. Lees of Texas A&M University examined several mixed breed dogs from a single family in which some adolescent males developed CRF. Dogs in this family exhibited the clinical, immunohistochemical, and pathological characteristics typical of AS. The structural and functional abnormalities of their GBM, along with an X-linked dominant mode of inheritance, indicated the dogs had a disease analogous to human XLAS (Lees et al. 1999). As expected, XLAS in the dog is characterized by absence of COL4A5 in the GBM as determined by immunofluorescence using a monoclonal antibody against the NC1 region of the polypeptide. Immunofluroescence microscopy also confirmed absence of COL4A3 and COL4A4 in the GBM of affected dogs, as is observed in affected humans (Lees et al. 1999).

To further characterize the clinical and genetic aspects of the disease, Dr. Lees established a breeding colony of carrier females at TAMU-CVM (Lees et al. 1999). Normal female hound dogs were artificially inseminated with the semen of two affected male dogs that had lived to puberty. Carrier female progeny of these matings were raised and subsequently bred to a normal Labrador Retriever to produce additional affected

males, carrier females and normal dogs (males and females). Dogs in this colony are termed Navasota (NAV) dogs.

Affected male NAV dogs typically develop proteinuria between four and six months of age. Chronic progressive renal disease associated with glomerular and tubulointerstitial lesions causes a reduction in GFR between 6 and 9 months of age. ESRD occurs at an average age of 11 months (Lees et al. 1999). The reserve capacity of the kidney often masks the signs and symptoms of CRF until late in the disease process but maintenance of the colony of dogs provides a model that allows study of the earliest processes leading to CRF. The dog is especially useful in this regard because its large size, as compared with many traditional (e.g., rodent) animal models, more closely mimics the physiological and environmental stresses on the human kidney. However, to use the NAV dog as a model, identification and characterization of a mutation in *COL4A5* that is causative for XLAS as well as an efficient way to diagnose dogs at the genetic level, thereby obviating the need for biopsy, is necessary.

Specific objectives

There were three objectives of this project: 1) to supplement the map of the canine genome by adding gene loci to the integrated meiotic linkage-radiation hybrid map, 2) to identify the mutation causative for XLAS in the NAV dog and develop a genetic test to identify carriers and at-risk dogs and 3) to generate renal gene expression profiles of affected and non-affected NAV dogs in order to understand the genetic changes that accompany AS and generalized ESRD. The first two goals are

straightforward but the third is far more complex. To accomplish the third objective, probing of a canine oligonucleotide microarray and quantitative real time PCR (Q-RT-PCR) are necessary. These experiments are imperative for evaluation of gene expression in kidneys with ESRD. A canine microarray (Affymetrix GeneChips) has been constructed and purchased by collaborators for our use. The six type IV collagens are represented on this array, as are many other canine genes of potential interest including those encoding other components of the GBM.

Summary

Mapping of the uromodulin gene and the type I and type IV collagens added gene loci to the map of the canine genome. Secondly, by sequencing the entire coding region of *COL4A5*, a mutation causative for XLAS in the NAV dog was identified. Subsequent to identification of the mutation, a genetic test was developed to identify carriers and dogs destined to develop XLAS. Finally, analysis of gene expression using a canine oligonucleotide microarray has revealed differential expression of genes in affected NAV dogs as compared to normal NAV dogs.

CHAPTER II

RADIATION HYBRID MAPPING OF THE CANINE TYPE I AND TYPE IV COLLAGEN GENE SUBFAMILIES*

Overview

We are interested in the collagen gene super-family and its involvement in hereditary diseases of the human and domestic dog. Presented here are radiation hybrid mapping of the type I and type IV collagen gene subfamilies on the most recent version of the canine map. The *COL1A1* gene was mapped to chromosome (Chr) 9, *COL1A2* was mapped to Chr 14, *COL4A1* and *COL4A2* were mapped to Chr 22 and *COL4A3* and *COL4A4* were mapped to Chr 25. The *COL4A5* and *COL4A6* genes, while linked to one another, are not linked in the present version of the canine map but likely are present on the X Chr. These data provide insight into molecular evolution of these subfamilies and increase the number of mapped genes in discrete regions of the canine genome.

Introduction

The super-family of collagen genes features 31 genes grouped into 19 subfamilies. Collagens are the major components of connective tissues, the extracellular matrix and basement membranes (Table 1). Types I-IV represent fibrous collagens while

*Reprinted with permission from Lowe JK, Guyon R, Cox ML, Mitchell DC, Lonkar AL et al. (2003) Radiation hybrid mapping of the canine type I and type IV collagen gene subfamilies. Funct Integr Genomics 3, 112-116

all the others are non-fibril forming. The fibrous collagens coil into left-handed helices, and three such chains, in different combinations, wrap around each other to form a right-handed super-helix. The triple-helical domains intertwine and supercoil; the proteins usually form fibers or networks. The importance of these networks is evident since mutations in genes encoding many of the collagens are causative for several human hereditary diseases (Table 1). Some of these diseases (e.g., Alport syndrome and Ehlers Danlos syndrome) have also been described in the dog and result from mutations in the canine collagen orthologs (Freeman et al. 1987; Zheng et al. 1994; Rodriguez et al. 1996; Lees et al. 1999).

Table 1. Type I and IV collagen gene subfamilies.

Gene	Location (human)	Location (mouse)	Location (dog)	Tissue-specific expression	Examples of associated diseases
COL1A1	HSA 17q21.3-q22.1	MMU 11	CFA9	Bone, skin, tendon	Ehlers Danlos syndrome
COL1A2	HSA 7q22.1	MMU 6	CFA18	Bone, skin, tendon	Osteogenesis imperfecta Ehlers Danlos syndrome
COL4A1	HSA13q34	MMU 8	CFA22	All basement membranes	No known diseases
COL4A2	HSA13q34	MMU 8	CFA22	All basement membranes	No known diseases
COL4A3	HSA 2q35-q37	MMU 1	CFA25	Basement membranes of kidney, eye, cochlea, lung, brain	Goodpasture syndrome Autosomal dominant Alport syndrome Autosomal recessive Alport syndrome Benign familial hematuria
COL4A4	HSA 2q35-q37	MMU 1	CFA25	Basement membranes of kidney, eye, cochlea, lung, brain	Autosomal dominant Alport syndrome Autosomal recessive Alport syndrome Benign familial hematuria
COL4A5	Xq22.3	X	X – unlinked	Basement membranes of kidney, eye, cochlea, lung, brain, epithelium	X-linked dominant Alport syndome Diffuse leiomyomatosis (with *COL4A6*)
COL4A6	Xq22.3	X	X – unlinked	Basement membranes of kidney, esophagus, lung, epithelium	Diffuse leiomyomatosis (with *COL4A5*)

Approximately 400 hereditary diseases have been identified in the domestic dog, *Canis familiaris* (Patterson 2001). The diseases affect every modern breed of dog and many diseases are clustered within some particularly susceptible breeds. Dissection of the genetics responsible for many of the diseases has been hindered, until recently, by the lack of a high-resolution map of the canine genome. However, the most recent version of the canine map now includes 1800 markers and spans > 90% of the genome (Breen et al. 2001).

The canine map is comprised of 1,078 microsatellite markers and 372 gene-based markers. The microsatellite markers provide the tool for genome scanning of informative pedigrees while the gene markers allow examination of syntenic relationships between the dog and other mammals. To date, however, fewer than 10 genes have been mapped per canine chromosome. In an effort to supplement this, we have undertaken the mapping of large gene families that encode structural proteins (Miller et al. 1999, 2001). As part of this effort, several members of the collagen gene super-family have been cloned and mapped.

Reported here is the expansion of the data set of mapped collagen genes. Specifically, radiation hybrid mapping of the type I and type IV collagen gene subfamilies shows that *COL1A1* and *COL1A2* are on *C. familiaris* autosome 9 (CFA9) and CFA14, respectively; *COL4A1* and *COL4A2* map to CFA22; and *COL4A3* and *COL4A4* to CFA25. The *COL4A5* and *COL4A6* genes likely reside on the X chromosome. The work presented here is part of an ongoing study to compare the

organization of the collagen gene super-family of the dog and human and to examine the roles of type I and IV collagens in hereditary diseases of the dog.

Materials and methods

Primers for RH mapping

Partial sequences of canine *COL1A1* and *COL1A2* (Campbell et al. 1998) and *COL4A1* to *COL4A6* (Thorner et al. 1996) were previously determined. Based on these sequences and additional sequencing done during this work (data not shown; see Table 2), primers were designed to produce PCR products specific for each of these genes. PCR was carried out using anonymous canine genomic DNA samples. Resultant amplification products were sequenced to verify identity. Primers used and sizes of products are listed in Table 2.

Amplification of the markers was carried out using standard conditions on dog and hamster DNA and also a mixture (1:3) of dog and hamster DNA as previously described (Priat et al. 1998). PCR for radiation hybrid (RH) typing were performed under different conditions specific for individual markers. For *COL1A1*, *COL1A2* and *COL4A6*, reactions were performed in a final volume of 15 µl using 50 ng RH DNA, 0.3 µM each primer, 0.316 mM $(NH_4)_2SO_4$, 67 mM Tris-Cl (pH 8.8 at 25°C), 0.01% Tween-20, 100 µM each dNTP, 1.5 mM $MgCl_2$, and 0.15 U Biolase Taq (Bioline) using annealing temperatures of 59°C, 58°C and 61°C for *COL1A1, COL1A2* and *COL4A6*, respectively. For *COL4A1* to *COL4A5*, amplification was performed in 10 µl reactions, containing 50 ng RH DNA, 0.3 µM each primer, 250 µM each dNTP (Pharmacia), 2

mM $MgCl_2$ (except for *COL4A2*, 1.5 mM $MgCl_2$), 1X AmpliTaq Buffer and 0.5 U AmpliTaq Gold (Perkin-Elmer). PCR reactions were carried out in PTC-200 PCR machines (MJ Research). The annealing temperatures for *COL4A1* and *COL4A2* were touchdown programs of 63°C-53°C and 67°C-57°C respectively, consisting of 7 min at 95°C, 20 cycles of 30 s at 94°C, 30 s at 63°C or 67°C (decreasing 0.5°C per cycle), 1 min at 72°C and 15 cycles of 30 s at 94°C, 30 s at 53°C or 57°C, 1 min at 72°C, with a final extension of 2 min at 72°C. For *COL4A3*, *COL4A4* and *COL4A5*, classical programs with fixed annealing temperatures of 55°C, 63°C and 60°C, respectively, were used.

Mapping analysis

The eight genes comprising the type I and type IV collagen gene subfamilies were integrated into the canine RH map (Breen et al. 2001) using the previously described canine-rodent radiation hybrid panel RHDF5000 (Vignaux et al. 1999). Since markers are constantly being added to the map, the current version is, of course, replete with more markers than the published version. Analysis was done using the two-point analysis of the computer program Multimap (Matise et al. 1994a). Markers were assigned to linkage groups using a Lod $\geq$8.0. Data concerning the canine RH map (Breen et al. 2001) are available at: http://www-recomgen.univ-rennes1.fr/doggy.html and http://www.fhcrc.org/science/dog_genome/dog.html.

Results

The *COL1A1* gene mapped to CFA9, 7.65 centiRay (cR) from GNGT2 and 27.8 cR from Ren126A15, as previously stated (Breen et al. 2001). Fluorescence *in situ* hybridization (FISH), chromosome paint studies and RH mapping of multiple genes suggest that this region of CFA9 corresponds to *Homo sapiens* autosome 17q (HSA17; Werner et al. 1997; Priat et al. 1998; Breen et al. 1999; Yang et al. 1999; Mellersh et al. 2000), consistent with the observation that human *COL1A1* maps to HSA 17q21.3-q22 (Huerre et al. 1982). Canine *COL1A2* mapped to CFA14 with strong Lod scores (>15) establishing linkage with the *CALCR* and *PON2* genes (Table 2). This region corresponds to regions of HSA7, and human *COL1A2* has been mapped to 7q21.3-q22.1 (Retief et al. 1985).

The *COL4A1* and *COL4A2* genes were mapped to CFA22 with high Lod scores (Table 2). They are co-localized on the map in a region of conserved synteny with human Chr 13, based on RH mapping of *F7*, *ATP4B* and *GPCR* gene marker loci and heterologous painting data (Yang et al. 1999; Breen et al. 2001). Their localization further confirms the conserved segment of human HSA13 on CFA22. The *COL4A3* and *COL4A4* genes were mapped to CFA25 (Table 2), and are closely linked to one another in a region corresponding to a small conserved segment of HSA2q. These data extend the coverage of this segment and further define the breakage between HSA2q and HSA8p. The *COL4A5* and *COL4A6* genes remained unlinked in the present version of the map. However, these two collagens are linked to each other with a Lod score of 5.3 (Table 2). They are also linked to microsatellite markers REN130F03 and REN143I17,

respectively, with Lod scores of 4.9 for both markers (Table 2). While these Lod score values are under the threshold of 8, they are the best Lod scores obtained for *COL4A5* and *COL4A6* with any other marker of the map. Since REN130F03 and REN143I17 are localized on the X chromosome, these data strongly suggest that *COL4A5* and *COL4A6* reside on the canine X chromosome.

Table 2. Radiation hybrid mapping of collagen markers on the canine integrated 1800-marker map.

Gene	Primers (5'–3')	Length of PCR product (bp)	PCR Tm (^{o}C)	Canine Chr (CFA)	Best Lod score with the closest markers (two-point analysis of MultiMap)	Accession number
COL1A1	F- ACCTGCCTTCTTGCTT ACACTTC R- AGGGCACAGGTGATA CAGTGATA	800	59	CFA 9	23.7 GNGT2 20.5 REN124K12	AF291995
COL1A2	F- TTGACACCCAGACAC ACCAGAGGGGTCTGA TGACACTCTTT R- TATTTAAAGCTCACTC AGCATT GCACAGGGT CTCTCGACCT	2200	58	CFA 14	15.9 CALCR 15.8 PON2	AF291996
COL4A1	F- CGGGAGGCGAAGTTA CAC R- TAGGCACAGTCAAAC AACAGA	1100	63-53	CFA 22	26.1 REN183N15 22.9 FH2538	U50933
COL4A2	F- GGCTGTCCACCACTGC CCCGCT R- TATCCAATCCAGAGAC TCCGCCATCCA	172	67-57	CFA 22	26.1 REN183N15 22.9 FH2538	U50934
COL4A3	F- TGCCATTCTTGTTCTGT AAC R- ATGGCTATGGCAATCG TAGG	600	55	CFA 25	14.6 C25.213 14.4 REN61G15	U50935
COL4A4	F- CACCCTGCCCTTTGCC TACT R- GGACACGGCGGGATG GAC	213	63	CFA 25	11.7 *COL4A3* 9.9 REN61G15	U50936
COL4A5	F- CCCTGGACCAGATGGA ATGC R- ACCGTGAGCTCTTTTA TTTCCTTGG	181	60	unlinked	4.9 REN130F03 4.8 REN307M14	U07888
COL4A6	F- GTGAAGCACAGCCAGT CAGAACAG R- CTCGCACACCGAACA ACG	1100	61	unlinked	5.3 *COL4A5* 4.9 REN143I17	U50937

Discussion

Placement of gene families or subfamilies on the canine genome map is a mechanism not only for rapidly increasing the density of the map, but, more importantly, for further establishing the evolutionary relationship between the canine, murine, and human genomes. The arrangement of gene families reveals regions of conservation as well as evolutionary breakpoints among these genomes. Of particular importance to examination of the dog is that mapping of gene families helps to increase the density of genes in discrete regions of the canine genome. Finally, sequence and mapping data of canine collagen genes facilitate examination of these genes as candidates in study of collagen-related hereditary diseases.

It was found that *COL1A1* and *COL1A2* mapped to CFA9 and 14, respectively. The regions harboring the canine genes exhibit regions of conserved synteny with HSA17 and 7, respectively, and, as expected, those are the chromosomes on which the human orthologs are found (Table 1). Mapping of *COL4A1* and *COL4A2* genes to CFA22 confirms the region of conserved synteny between CFA22 and HSA13q. The *COL4A1* and *COL4A2* genes are co-localized on the canine genome, i.e., they map to the same position on the RH map. This arrangement is also observed in human (www.ensembl.org) and mouse (www.ncbi.nlm.nih/LocusLink). In the human, these genes are located on HSA13q34 and are separated by only 50 kb. In the mouse, they are located at the same locus, which is 85 cM from the telomere on chromosome 8.

The *COL4A3* and *COL4A4* genes are closely linked on CFA25. This finding is consistent with the breakage between HSA2q and HSA8p previously detected (Yang et

al. 1999). The two genes are also organized in tandem in the human and are closely positioned on the human draft sequence (www.ensembl.org). This motif is also found in the mouse (www.ncbi.nlm.nih/LocusLink).

The RH data strongly suggest that *COL4A5* and *COL4A6* genes are linked to each other with a Lod of 4.9 by microsatellite markers REN130F03 and REN143I17, respectively. Since these two makers are on the X chromosome, it is highly likely that *COL4A5* and *COL4A6* are also on the X chromosome. The X chromosome is the largest canine chromosome but is less represented in the RH panel because of haploidy. Therefore, the X chromosome is poorly covered (as compared to other chromosomes) in the recent version of the map (Breen et al. 2001). This explains why several groups are not yet linked together and consequently why *COL4A5* and *COL4A6* are still unlinked. However, these genes have been mapped to the X chromosome in both the human and mouse and, due to the highly conserved status of chromosome X throughout mammalian evolution, the localization of these two collagens on canine chromosome X is strongly favored. With respect to linkage of *COL4A5* and *COL4A6,* in the human these two genes are also very close to one another on Xq22.3. This is additional evidence supporting the assignment of canine *COL4A5* and *COL4A6* to the X chromosome. Finally, the genomic organization of type IV collagen genes is similar with respect to their genomic exon/intron number, size and organization (Heidet et al. 2001). This feature, associated with their tandem organization at several positions on canine, human and murine genomes, strongly supports a common ancestral origin and their probable evolution through *cis* and *trans* duplications.

The canine map currently contains an average of less than ten gene loci per chromosome. The addition of more genes belonging to large gene families will further improve comparative mapping studies by allowing detection of microrearrangements between individual genes. Moreover, the mapping of such gene families also constitutes a powerful source of candidate genes for further analyses of diseases of interest. Indeed, the current map has more than 500 highly informative microsatellites and may be confidently used as a tool for selecting candidate disease regions by linkage analysis. Therefore, the detailed mapping and analyses of the collagen type I and IV subfamilies presented in this work will help in identifying responsible genes in canine pedigrees in which collagen-associated diseases segregate.

CHAPTER III

SEQUENCING AND RADIATION HYBRID MAPPING OF CANINE *UROMODULIN**

Overview

Our interest is in understanding the genetic bases for hereditary renal diseases of the domestic dog (*Canis familiaris*) and in characterizing gene loci for placement on the map of the canine genome. We report here on the cloning, sequencing and radiation hybrid mapping of the canine cDNA encoding Uromodulin (Umod), a renal-specific glycoprotein. The cDNA is 2.3 kb in length and, as expected, comparisons of nucleotide sequences reveal that canine *umod* is quite similar to *umod* of other mammals. The predicted amino acid sequence of canine uromodulin has at least 70% identity with other mammalian uromodulin proteins. Canine *umod* has been mapped on the RHDF5000 radiation hybrid panel and positioned on the most recent canine genome map. Data indicate that *umod* is linked to the marker *CZP2* (*canine zona pellucida* gene) on an RH group not yet assigned to a canine chromosome. The human *umod* and *CZP2* genes are located on chromosome 16p13.

---

*Reprinted with permission from: Sequencing and radiation hybrid mapping of canine *Uromodulin*. Cox ML, Quignon P, Galibert F, Lees GE, Murphy KE. (2003a) DNA Seq 14(1), 61-69

Introduction

Uromodulin (Umod), also known as Tamm-Horsfall glycoprotein (THP), is the most abundant protein found in normal mammalian urine. First described by Morner in 1895 as uromucoid, and characterized by Tamm and Horsfall in 1950 (Kokot and Dulawa 2000), gene cloning (Pennica et al. 1987) confirmed that the nucleotide sequences of uromodulin and THP are identical. Uromodulin is a cell membrane protein with a glycosyl phosphatidylinositol (GPI) anchor (Rindler et al. 1990) and its expression is limited to the kidney. Tubular cells in the ascending loop of Henle and distal convoluted tubule have Umod on the basal and luminal poles of their cell membranes as well as on the membranes of the Golgi body and endoplasmic reticulum (Kokot and Dulawa 2000). Depending upon the species, the 85 kDa glycoprotein is composed of between 640 and 644 amino acids. The gene is 25 kb and consists of 11 exons, the first of which is non-coding (Yu et al. 1994; Prasadan et al. 1995), resulting in a cDNA of 2.3 kb.

While a definitive function for Umod has not been established, it has been suggested that this glycoprotein may be involved in (1) renal stone and tubulointerstitial diseases, (2) host defense mechanisms and (3) function(s) of cytokines and neutrophils. Renal stone and tubulointerstitial diseases are common in the human and dog and thus are well studied. A role for Umod in these diseases has been postulated due to the fact that this protein is found in all cast nephropathies and binds to immunoglobulin light chains (Ying and Sanders 2001). Because Umod binds (with high affinity) the type I pili of *Escherichia coli* (Dulawa et al. 1988; Parkkinen et al. 1988), it is possible that the

protein has some role in protection against bacterial pathogens. Finally, Umod may play a role in host immune responses. This is based on the facts that cytokines IL-1, TNF and IL-2 all bind to carbohydrate chains found on the glycoprotein (Sherblom et al. 1988), Umod binds to neutrophils (Cavallone et al. 1999) and to T-cells (Mishra et al. 1994) and that Umod isolated from the urine of pregnant women is more immunosuppressive than that isolated from non-pregnant women (Muchmore et al. 1985). This latter finding is correlated with a reduction in the number of oligomannose chains present on Umod of pregnant women (Kokot and Dulawa 2000).

We are interested in hereditary diseases of the dog, and one of our current foci is the type IV collagen genes. Three of the type IV collagens, *COL4A3*, *COL4A4* and *COL4A5* encode proteins important in structure of glomerular basement membranes (Kashtan 1998). The *COL4A5* gene is especially interesting to us because mutations in it cause X-linked Alport syndrome (XLAS) (Kashtan 1998), a disease that occurs in the dog and human (Kashtan 1998). During the amplification of the cDNA encoding COL4A5, a fragment of the cDNA encoding Umod was fortuitously amplified through RT-PCR and RACE. Using gene-specific primers for amplification, *umod* was sequenced in its entirety and RH mapped on the latest canine RH map (Breen et al. 2001).

Report

Canine renal samples from dogs affected with XLAS were snap-frozen in liquid nitrogen at necropsy. Total RNAs were isolated using the SNAP Total RNA Isolation

Kit (Invitrogen, Carlsbad, CA). To amplify the full-length cDNA of *COL4A5*, gene specific RT-PCR primers were designed and used to generate overlapping fragments. One set of primers, *COL4A5*-C-F: 5'GTTCTATGGGAGATACTGGTTTGCCTGGATT AC3' and *COL4A5*-C-R: 5'AGGGAGTCCGTTGCGTCCTGGAT3', was used to amplify a 950 bp segment at the 3' end of *COL4A5*. A 50 µl volume reaction included 50-100 ng total RNA and the components of the Advantage One-Step RT-PCR Kit (Clontech BD Bioscience, Palo Alto, CA). Conditions for reverse transcription and amplification were as follows: 50°C for 60 min and denaturation at 94°C for 5 min, followed by 35 cycles at 94°C for 30 s, 63°C for 30 s, and 68°C for 1 min, finishing with an extension at 68°C for 2 min, on a PE 2400 thermal cycler (Perkin-Elmer, Boston, MA). Products were resolved by electrophoresis through 1.2% agarose and visualized after staining with ethidium bromide.

Upon electrophoresis, one large amplicon band was visualized. After purification using the QIAquik Gel Extraction Kit (Qiagen, Valencia, CA) and re-electrophoresis, two PCR products, one of 950 bp and one of 1050 bp, were resolved. Both products were inserted into pCR4.0-TOPO (Invitrogen, Carlsbad, CA) and the resultant constructs were used to transform chemically competent *Escherichia coli* TOP10 cells (Invitrogen, Carlsbad, CA). After transformation, bacterial colonies were randomly picked, and grown overnight in Luria Broth containing 50 ug/ml ampicillin. Plasmids were purified using the Plasmid Mini Kit (Qiagen, Valencia, CA) or the PerfectPrep Plasmid Mini Kit (Eppendorf, Westbury, NY). Sequencing reactions were carried out using the dye termination method and resolved on the 373XL and 377XL automated sequencing

machines (Applied Biosystems, Foster City, CA). As expected, the 950 bp fragment was derived from *COL4A5*. Sequence analysis revealed the 1050 bp fragment as a partial cDNA derived from *umod* that corresponds to exons three through seven of human *umod*. This gene is expressed only in the kidney.

It is interesting that the set of primers amplified two different cDNAs because when *COL4A5* primers were designed and examined using the BLAST internet tool (www.ncbi.nlm.nih.gov/blast/BLAST.cgi), only *COL4A5* sequences from various organisms were cited as sharing sequence similarity/identity. That is, *umod* sequences from human, cow, mouse and rat were not listed and thus, it was assumed that the chosen primers were highly specific. This suggested that the *umod* product did not arise from simple RT-PCR.

All mRNAs present in the total RNA isolated from canine kidney should have been reverse transcribed, using either the oligo(dT) or gene-specific primer as starting sites. In the amplification process, cDNAs for both *COL4A5* and *umod* were amplified; however, at some point during the cycling, the *umod* cDNA served as a template for a 3' RACE reaction, driven by the anti-sense primer (*COL4A5*-C-R: 5'AGGGAGTCCGTTG CGTCCTGGAT3') for *COL4A5* even though this primer and *umod* are identical at only 12 of the 23 bases. However, the conditions for amplification allowed the primer to drive synthesis of the *umod* cDNA. This amplification did not include the entire 3' portion of the *umod* cDNA. Since the elongation time during cycling was only one minute, as was appropriate for the expected 950 bp fragment of *COL4A5*, it is likely that this reduced time of elongation resulted in the truncated *umod* cDNA. Regardless, because of the

fortuitous amplification, sequencing of the 1050 bp fragment was done so that *umod*-specific RACE primers (Table 3) for capture of the 5' and 3' ends of the cDNA and subsequent sequencing of it could be designed. Both strands of the entire cDNA were sequenced. Assembly of contiguous fragments and determination of the putative amino acid sequence for Umod was accomplished using Vector NTI Suite II software (Informax, Bethesda, MD).

Table 3. Primer sets and conditions for the amplification of canine *Umod* cDNA.

Primer Name	Type of amplification	Fragment size (bp)	Annealing temp (˚C)	Sequence
COL4A5-C-R	5' RACE	1050	63	AGGGAGTCCGTTGCGTCCTGGAT
UMOD-RACE-2	5' RACE	740	68	TAACCTGCCCCGTACTCGGTGCTGC
UMOD-RACE-3	3' RACE	830	66	CGGATGGCACTCTTCCAGACTCCTG ACTAC
UMOD-F1	RT-PCR	740	Touchdown 64-60	F: CTTTCCTCTCTGACCTCGGTGTGG R: TGTGCGGTTCACGATGCCCT
UMOD-F2	RT-PCR	1020	Touchdown 64-60	F: CCGTTGCGTCCTGGATGAGTA R: AAGGGGTCTGTGGCGTTGCT
UMOD-F3	RT-PCR	590	Touchdown 64-60	F: TAAACATCAGCGTGGGCGGGA R: GAAACATCAGGGTCAGGGTGGC
Urom-Ex8&9	PCR for RH mapping	650	Touchdown 61-51	F: ACTGTGAAGTGTATCTCTGCGAC R: GGTTTTGGTCTATGATGCCTC

The nucleotide sequence (Figure 1) of canine *umod* cDNA has been reported to GenBank (Accession no. AF498324). The full-length cDNA is 2.3 kb; from the "ATG" start codon to the "TGA" stop codon the length is 1929 bp. Comparisons of genomic sequences of *umod* from other organisms to the canine *umod* cDNA suggest that, as in all species reported thus far, the canine gene consists of eleven exons and 10 introns; the first exon is non-coding. The exons range in size from 25 bp (exon 1) to 742 bp (exon 3). As shown in Figure 1, canine *umod* is most closely related to human and bovine *umod* with 80.7 and 79.9% identity to bovine and human *umod*, respectively, using the

GeneStream alignment program (http://vega.igh.cnrs.fr/bin/align-guess.cgi) to determine percentage identities.

```
dog             GAATTCTACTTTTCTGGCTTCAGGACTCCAGGCATCACAGACAGGGAGGAAACGCCAAGG 60
cow             --------CCTTTCTGGTTTCAGGACTTCAGGGATCAGGGACAGA-AGGAAATCTCAAGA 51
human           --------------------CAGGACACCAGACATCAGAGACAGAGAGAAAAATTCAAAG 40
mouse           ------------------------------------------------------------
rat             ------------------------------------------------------------


dog             GGTCAACCTGTCCTTGCCTTTGGGCAGGTCTTCTCCAGACCAGAAGTAACTGTGAAGAGC 120
cow             GGCCAGCCTTTCCTTGCCTTTGGGAAAGTCTTC-CCAGACCAGAAGAAGCTGTGAGGAGC 110
human           GGCCAACCCGTCTTT-CCTTTGGGCAGGTGCTATCTAGACCTGAAGTAGCGG-GAAGAGC 98
mouse           ------------------------------CCACGCGTCCGCCAGGACAGAGGGCAGAGT 30
rat             ---------------------------------------------GGAATTCCGGGCAGAGC 17
                                                                 * *  ***


dog             AGAAAAAAAATGGGGCAGCTTTCCTCTCTGACCTCGGTGTGGATGGTAGTGGTGGTAACCT 180
cow             AGAAAGGA-TGAAGTGTCTTTTCTCTCCGAACTTCATGTGGATGGCAGCGGTGGTGACCT 169
human           AGAAAGGA-TGGGGCAGC---CATCTCTGACTTGGATGCTGATGGTGGTGGTGG---CCT 151
mouse           GTAAAGGA-TGGGGA------TCCCTTTGACCTGGATGCTGCTGGTAATGATGGTAACCT 83
rat             ACAAAGGA-TGGGGCAGCTGCTCTCTTTGACCTGGCTGCTGCTGGTTATGGTGGTAACTC 76
                  ***  * **  *          **  **  *   **  * ***    * ***   *


dog             CTTGGGTCATCATAGCTGCAAACATTGATACTGTGGAAGCAAGAAGCTGCTCTGAATGTC 240
cow             CTTGGGTCATCATACCTGCAGCAACTGACACCTCATCAGCAAAAAGCTGCTCTGAATGTC 229
human           CTTGGTTCATCACAACTGCAGCCACTGACACCTCAGAAGCAAGATGGTGCTCTGAATGTC 211
mouse           CCTGGTTCACTCTGGCTGAAGCCAGTAACTCAACAGAAGCGAGACGGTGTTCTGAATGCC 143
rat             CCTGGTTCACTGTAGCTGGAGCCAATGACTCACCAGAAGCAAGAAGGTGTTCTGAATGCC 136
                * *** ***      *** *   * * *  *      *** * * * ** ******** *


dog             ACAGCAATGCCACCTGCATGGAGGATGGGATGGTCACAACATGTTCCTGCCTGGTGGGTT 300
cow             ACAGCAATGCCACTTGTACGGTGGACGGGGCTGCCACGACCTGCGCCTGCCAGGAGGGCT 289
human           ACAGCAATGCCACCTGCACGGAGGATGAGGCCGTTACGACGTGCACCTGTCAGGAGGGCT 271
mouse           ACAACAACGCCACCTGCACGGTGGATGGTGTGGTCACAACGTGCTCCTGCCAGACCGGCT 203
rat             ACGACAATGCCACCTGTGTGTTGGATGGGGTGGTCACCACATGCTCCTGCCAGGCAGGCT 196
                **  *** ***** **   *  *** *     *  ** ** **  **** * *   ** *


dog             TCACTGGCAGCGGCTTTGAGTGCGTGGACCTGGATGAATGTGCCATTCCTGGTGCCCACA 360
cow             TCACTGGCGACGGCCTCGAGTGTGTGGATCTGGACGAATGCGCCGTTCTGGGGGCGCACA 349
human           TCACCGGCGATGGCCTGACCTGCGTGGACCTGGATGAGTGCGCCATTCCTGGAGCTCACA 331
mouse           TCACTGGTGATGGGCTGGTGTGTGAGGACATGGATGAGTGTGCTACCCCATGGACTCACA 263
rat             TCACTGGAGATGGGCTGGTGTGTGAGGACATAGATGAGTGTGCCACCCCGTGGACTCACA 256
                **** **    **  *    ** * ***  * ** ** ** **    *   *  * ****


dog             ACTGCTCGGAGGGGAGCAGCTGTATGAATACGCTGGGCTCCTACCTCTGCACCTGCCCCG 420
cow             ACTGCTCCGCCACCAAGAGCTGCGTGAATACGCTGGGCTCTTACACGTGCGTCTGCCCTG 409
human           ACTGCTCCGCCAACAGCAGCTGCGTAAACACGCCAGGCTCCTTCTCCTGCGTCTGCCCCG 391
mouse           ACTGCTCC---AACAGCAGCTGTGTGAACACCCCGGGCTCGTTTAAGTGCTCCTGTCAGG 320
rat             ACTGCTCC---AACAGCATCTGCATGAACACACTGGGCTCCTACGAGTGCTCCTGTCAGG 313
                *******       *  * ***  * ** ** *  ***** *     ***  *** *  *
```

Figure 1. CLUSTAL W (1.82) multiple sequence alignment (http://www.ebi.ac.uk /clustalw/). Comparison of nucleotide similarity of canine *umod* to human (Homo sapiens, GenBank Accession no. NM_003361, 79.7%), cow (*Bos taurus*, GenBank Accession no. S75958, 80.7%), mouse (*Mus musculus*, GenBank Accession no. L33406, 72.6%), and rat (*Rattus norvegicus*, GenBank Accession no. M63510, 80.7%). Percentage identities were determined with the GeneStream align program (http://vega.igh.cnrs.fr/bin/align-guess.cgi).

```
dog             ACGGCTTCCGTCTGACACCGGGGCTGGGCTGCATCGACGTGGATGAGTGCTCGGAGCCGG 480
cow             AAGGTTTTCTCCTGAGCTCGGAGCTCGGCTGCGAGGATGTGGACGAGTGTGCAGAGCCAG 469
human           AAGGCTTCCGCCTGTCGCCCGGTCTCGGCTGCACAGACGTGGATGAGTGCGCTGAGCCTG 451
mouse           ATGGTTTTCGTCTGACGCCTGAGCTGAGCTGCACTGATGTGGATGAGTGCTCAGAGCAGG 380
rat             ATGGCTTTCGTCTGACGCCTGGGCTGGGCTGCATTGATGTGAATGAGTGCACAGAGCAGG 373
                * ** ** *  ***    * *  **  *****   ** *** * *****  * ****  *


dog             GGCTCAGCCGCTGCCATGCCCTGGCCACCTGCATCAACAACAAGGGCAATTACTCGTGCG 540
cow             GGCTCAGCCGCTGCCACGCCCTGGCCACTTGCATCAATGGCGAGGGCAACTACTCGTGCG 529
human           GGCTTAGCCACTGCCACGCCCTGGCCACATGTGTCAATGTGGTGGGCAGCTACTTGTGCG 511
mouse           GGCTCAGTAACTGTCATGCCCTGGCCACCTGTGTCAACACAGAAGGCGACTACTTGTGCG 440
rat             GGCTCAGTAACTGTCATTCCCTGGCTACCTGTGTCAACACGGAAGGCAGCTACTCATGCG 433
                **** **   *** **  ******* ** **  ****       ***   ****  ****

dog             TGTGCCCAGCGGGCTACCGAGGGGACGGGCAGCACTGTGAGTGCTCCCCGGGCTCCTGCG 600
cow             TGTGTCCTGCGGGCTACCTGGGAGACGGAAGGCACTGTGAGTGTTCCCCGGGCTCCTGTG 589
human           TATGCCCCGCGGGCTACCGGGGGGATGGATGGCACTGTGAGTGCTCCCCGGGCTCCTGCG 571
mouse           TGTGTCCCGAGGGCTTTACAGGGGATGGTTGGTACTGTGAGTGCTCCCCAGGCTCCTGTG 500
rat             TGTGTCCCAAGGGCTATAGAGGGGATGGTTGGTACTGTGAGTGCTCCCCTGGCTTCTGTG 493
                * ** **   *****     ** ** **   * ********** ***** **** *** *

dog             GCCCGGGCTTGGACTGCGTGCCCGTGGGC------GACGCGCTGGTGTGCGCGGACCCGT 654
cow             GGCCTGGGCTAGACTGCGTGCGGGAGGGC------GACGCGCTCGTGTGCGTGGACCCGT 643
human           GGCCGGGGTTGGACTGCGTGCCCGAGGGC------GACGCGCTCGTGTGCGCGGATCCGT 625
mouse           AGCCAGGACTGGACTGCTTGCCCCAGGGCCCGGATGGAAAGCTGGTGTGTCAAGACCCCT 560
rat             AGCCAGGGCTGGACTGCTTGCCTCAGGGTCCCAGTGGAAAGCTGGTGTGCCAAGACCCCT 553
                  ** **  * ****** ***    ***       *    *** *****    ** ** *

dog             GCCAGGAGCATCGCATCCTGGATGAGTACTGGCGCAGCACCGAGTACGGGGCAGGTTACA 714
cow             GCCAGGTGCACCGCATCCTGGACGAATACTGGCGCAGCACAGAGTACGGCTCCGGCTACA 703
human           GTCAGGCGCACCGCACCCTGGACGAGTACTGGCGCAGCACCGAGTACGGGGAGGGCTACG 685
mouse           GCAATACATATGAGACCCTGACTGAGTACTGGCGCAGCACAGAGTATGGTGTGGGCTACT 620
rat             GCAATGTCTATGAAACCCTGACTGAGTACTGGCGCAGCACAGACTATGGTGCCGGCTACT 613
                *  *     *    * ****   ** ************** ** ** **    ** ***

dog             CCTGCGACGTGGGCCTGAACGGCTGGTACCGCTTCACAGGGCCAGGTGGGGTGCGCCTGG 774
cow             TCTGTGATGTCAGTCTGGGCGGCTGGTACCGCTTCGTGGGCCAGGCCGGCGTGCGCCTGC 763
human           CCTGCGACACGGACCTGCGCGGCTGGTACCGCTTCGTGGGCCAGGGCGGTGCGCGCATGG 745
mouse           CCTGTGACGCGGGTCTGCACGGCTGGTACCGGTTCACAGGCCAGGGTGGCGTTCGCATGG 680
rat             CCTGTGACTCAGATATGCACGGCTGGTACCGGTTCACAGGCCAGGGTGGCGTTCGCATGG 673
                 *** **        **  ************ ***   ** *  *  ** *  *** **

dog             CGGAGACCTGCGTGCCAGTCCTGCACTGCAACACGGCCGCGCCCATGTGGCTCAATGGCA 834
cow             CCGAGACCTGCGTGCCCGTCCTGCACTGCAACACGGCCGCGCCTATGTGGCTCAACGGCA 823
human           CCGAGACCTGCGTGCCAGTCCTGCGCTGCAACACGGCCGCCCCCATGTGGCTCAATGGCA 805
mouse           CTGAGACCTGTGTGCCCGTCCTGCGATGCAACACGGCGGCACCCATGTGGCTCAATGGCT 740
rat             CTGAGACCTGTGTGCCCGTACTGAGGTGCAACACGGCTGCACCCATGTGGCTCAATGGCT 733
                * ******** ***** ** ***   *********** ** ** *********** ***

dog             CACACCCGACCAGAGACCAGGGCATCGTGAACCGCACAGCCTGTGCGCACTGGAGGGGCC 894
cow             CGCATCCATCGAGCGACGAGGGCATCGTGAACCGCGTGGCCTGTGCACACTGGAGCGGTG 883
human           CGCATCCGTCCAGCGACGAGGGCATCGTGAGCCGCAAGGCCTGCGCGCACTGGAGCGGCC 865
mouse           CTCATCCCTCGAGTAGTGAAGGCATTGTGAGCCGCACGGCCTGTGCACACTGGAGCGACC 800
rat             CGCATCCTTCGAGCAGAGAGGGCATTGTGAGCCGCACAGCCTGCGCACACTGGAGCGACC 793
                * ** **  * **     * ***** **** ****   ***** ** ******** *

dog             ACTGCTGCCTGTGGGATGCGTCCATCCAGGTGAAGGCCTGCGCCGGCGGCTACTATGTCT 954
cow             ACTGCTGCCTGTGGGACGCGCCCATCCAAGTGAAGGCCTGTGCCGGGGGCTACTACGTGT 943
human           ACTGCTGCCTGTGGGATGCGTCCGTCCAGGTGAAGGCCTGTGCCGGCGGCTACTACGTCT 925
mouse           AATGCTGCCGGTGGTCCACAGAGATCCAGGTGAAGGCTTGCCCAGGTGGCTTCTATATTT 860
rat             ACTGCTGCCTGTGGTCCACAGAGATCCAGGTGAAGGCCTGCCCTGGTGGCTTCTATGTTT 853
                * ******* ****    *     **** ******** **  * ** **** ***  * *
```

Figure 1. Continued

```
dog             ACAACCTCACGGAGACCCCTGAGTGCTACCTGGCCTACTGCACAGACCCCACCTCTGTGT 1014
cow             ACAACCTGACGGCGCCCCCTGAGTGCCATCTGGCTTACTGCACAGACCCCAGCTCTGTGG 1003
human           ACAACCTGACAGCGCCCCCCGAGTGTCACCTGGCGTACTGCACAGACCCCAGCTCCGTGG 985
mouse           ACAACTTGACAGCGCCCCCTGAGTGCAATCTGGCTTACTGCACCGATCCTAGTTCCGTGG 920
rat             ACAACTTGACAGAGCCCCCTGAGTGCAATCTGGCTTACTGCACCGATCCTAGCTCCGTGG 913
                ***** * ** * * **** *****  * ***** ******** ** ** *  ** ***

dog             TGGGGACATGTGAGGAGTGCAGTGTAGAAGAGGACTGCAAATCCCATGATGGCATGTGGA 1074
cow             AGGGGACGTGTGAGGAGTGCCGTGTGGATGAGGACTGCAAATCGGATAATGGTGAATGGC 1063
human           AGGGGACGTGTGAGGAGTGCAGTATAGACGAGGACTGCAAATCGAATAATGGCAGATGGC 1045
mouse           AGGGGACTTGCGAAGAATGCAGGGTAGATGAAGATTGCATATCGGATAACGGCAGATGGC 980
rat             AGGGGACTTGCGAAGAATGCGGGGTAGATGAAGACTGCGTATCTGATAACGGCAGATGGC 973
                 ****** ** ** ** *** *  * ** ** ** ***  ***  ** * **    ***

dog             GCTGCCAGTGCAAACAGGACTTCAATGTCACTGATCTCTTCCTTCTGGACCG---GCTGG 1131
cow             ACTGCCAGTGCAAACAGGACTTCAACGTCACCGATCTCTCCCTCCTGGAGCGCAGGCTGG 1123
human           ACTGCCAGTGCAAACAGGACTTCAACATCACTGATATCTCCCTCCTGGAGCACAGGCTGG 1105
mouse           GCTGCCAGTGTAAACAGGACTCCAACATCACAGATGTCTCCCAATTGGAGTACAGGCTGG 1040
rat             GCTGCCAGTGTAAACAGGACTTCAACGTTACAGATGTCTCCCTCCTGGAGCACAGGCTGG 1033
                 ********* ********** ***  * ** *** *** **   ****      *****

dog             AGTGTAGGCCCAATGATATCAAGGTGTCCCTGAGCAAGTGTCAGCTGAAGAGCCTGGGCT 1191
cow             AATGTGGGGTTGATGACATTAAGTTGTCCCTGAGCAAGTGCCAGCTGAAGAGTCTGGGCT 1183
human           AATGTGGGGCCAATGACATGAAGGTGTCGCTGGGCAAGTGCCAGCTGAAGAGTCTGGGCT 1165
mouse           AGTGTGGGGCCAATGACATCAAGATGTCCCTCAGAAAGTGCCAGCTACAGAGTTTGGGCT 1100
rat             AGTGTGAGGCCAATGAAATCAAGATATCCCTCAGCAAGTGCCAGCTACAGAGTTTGGGCT 1093
                * ***  *    **** ** *** * ** **  * ***** *****  ****  ******

dog             TTGAGAAGGTTTTCATGTACCTGCGTGACAGCCAGTGCTCAGGCTTCAATGAGAGGGGCG 1251
cow             TTGAGAAGGTCTTCATGTACCTGCATGACAGCCAGTGCTCAGGCTTCACTGAGAGGGGCG 1243
human           TCGACAAGGTCTTCATGTACCTGAGTGACAGCCGGTGCTCGGGCTTCAATGACAGAGACA 1225
mouse           TTATGAATGTCTTCATGTACCTGAATGACAGACAATGCTCAGGCTTCAGTGAGAGTGATG 1160
rat             TTATGAAGGTCTTCATGTACCTGAATGACAGACAGTGCTCAGGCTTCAGTGAGAGGGGTG 1153
                *    ** ** ************  ****** *  ***** ******* *** ** *

dog             ACCGGGACTGGGTATCTGTGGTGACCCCAGCCAGGGATGGTCCCTGTGGAACAGTGATGG 1311
cow             ACCGGGACTGGATGTCTGTGGTGACCCCAGCCAGGGATGGCCCCTGTGGGACAGTGATGA 1303
human           ACCGGGACTGGGTGTCTGTAGTGACCCCAGCCCGGGATGGCCCCTGTGGGACAGTGTTGA 1285
mouse           AACGAGACTGGATGTCCATAGTGACCCCTGCCAGGAATGGTCCCTGTGGGACAGTATTGA 1220
rat             AACGAGACTGGATGTCCATAGTGACTCCTGCCAGGGATGGTCCCTGTGGGACAGTATTGA 1213
                * ** ****** * **  * ***** ** *** ** **** ******** *****  **

dog             TGAGGAATGAAACTCATGCCACATACAGCAACACTCTTTACCTGGCAGATGAGATCGTCA 1371
cow             CGAGGAATGAGACCCACGCCACATACAGCAACACTCTCTACCTGGCAGATGAGATCATCA 1363
human           CGAGGAATGAAACCCATGCCACTTACAGCAACACCCTCTACCTGGCAGATGAGATCATCA 1345
mouse           GGAGAAACGAAACCCATGCCACCTACAGCAACACCCTCTACCTGGCAAATGCGATCATCA 1280
rat             GGAGAAATGAAACCCACGCCACCTACAGCAACACCCTCTACCTGGCAAGTGAGATCATCA 1273
                 *** ** ** ** ** ***** *********** ** *********  ** **** ***

dog             TCCGTGACCGCAACATCAAAATCAACTTTGAGTGTTCCTACCCCCTGGATATGAAAGTGA 1431
cow             TCCGTGACCTCAACATCAGAATCAACTTTGCGTGCTCCTATCCCCTGGACATGAAAGTCA 1423
human           TCCGTGACCTCAACATCAAAATCAACTTTGCATGCTCCTACCCCCTGGACATGAAAGTCA 1405
mouse           TTCGGGACATCATCATAAGAATGAACTTTGAATGCTCTTACCCTCTGGACATGAAAGTCA 1340
rat             TCAGGGATATCAACATCAGAATCAACTTTGAATGCTCTTACCCTCTGGACATGAAAGTCA 1333
                *  * **   ** *** * *** *******  ** ** ** ** ***** ******** *

dog             GCTTGGAGACCTCCCTGCAGCCGATAGTCAGCTCTCTAAACATCAGCGTGGGCGGGACAG 1491
cow             GCCTGAAGACCTCCCTGCAGCCAATGGTCAGTGCCCTCAACATCAGCATGGGCGGGACCG 1483
human           GCCTGAAGACCGCCCTACAGCCAATGGTCAGTGCTCTAAACATCAGAGTGGGCGGGACCG 1465
mouse           GCCTGAAGACCTCCCTACAGCCCATGGTCAGTGCCCTGAACATCAGCTTGGGTGGGACAG 1400
rat             GTCTGAAGACCTCCCTACAGCCTATGGTTAGTGCCTTGAACATCAGCTTGGGTGGGACAG 1393
                *  ** ***** **** ***** ** ** **  *  * ********  **** ***** *
```

Figure 1. Continued

```
dog             GCATGTTCACCGTGCGGATGGCACTCTTCCAGACTCCTGACTACACACAGCCCTACCAAG 1551
cow             GCACATTCACCGTGCGAATGGCACTCTTCCAGAGCCCTGCCTACACACAGCCCTACCAAG 1543
human           GCATGTTCACCGTGCGGATGGCGCTCTTCCAGACCCCTTCCTACACGCAGCCCTACCAAG 1525
mouse           GCAAGTTCACCGTGCGGATGGCATTGTTCCAGAGCCCTACCTACACACAGCCCCACCAAG 1460
rat             GCAAGTTCACTGTGCAGATGGCACTGTTCCAGAACCCTACCTACACACAGCCCTACCAAG 1453
                ***  ***** ****  *****  * *******  ***  ****** ****** ******

dog             GCTCCTCTGTGACCCTGACTATCGAGGCCTTTCTCTATGTGGGCACCATGCTGGATGGGG 1611
cow             GCTCCTCTGTGACCCTGTCCACAGAAGCGTTTCTCTACGTCGGCACCATGCTGGATGGGG 1603
human           GCTCCTCCGTGACACTGTCCACTGAGGCTTTTCTCTATGTGGGCACCATGTTGGATGGGG 1585
mouse           GTCCTTCTGTGATGCTGTCCACTGAGGCTTTTCTGTATGTGGGCACCATGCTGGATGGGG 1520
rat             GTCCTTCTGTGATGCTCTCCACTGAGGCTTTTCTGTATGTGGGCACCATGCTGGATGGGG 1513
                *  * ** ****  **  * *  ** ** ***** ** ** ********* *********

dog             GTGATTTGTCCCGGTTTGCACTGCTGATGACTAACTGTTATGCCACACCCAGCAGCAACG 1671
cow             GTGACTTGTCCCGGTTTGTACTGCTCATGACCAACTGCTATGCCACACCCAGCAGCAATG 1663
human           GCGACCTGTCCCGATTTGCACTGCTCATGACCAACTGCTATGCCACACCCAGTAGCAATG 1645
mouse           GTGACTTGTCCCGGTTTGTACTGCTAATGACCAACTGCTATGCCACACCCAGTAGCAACT 1580
rat             GTGACTTGTCCCGGTTTGTACTGCTAATGACCAACTGCTATGCCACACCCAGTAGCAACT 1573
                * **  ******* **** ****** ***** ***** ************** *****

dog             CCACAGACCCCTTGAAGTACTTCATCATCCAGGACAGATGTCCACGCACTACGGACTCAA 1731
cow             CCACAGACCCCTTGAAATACTTCATCATCCAGGACAGATGTCCACGTGCTGCGGACTCAA 1723
human           CCACGGACCCCCTGAAGTACTTCATCATCCAGGACAGATGCCCACACACTAGAGACTCAA 1705
mouse           CCACAGACCCTGTGAAATACTTCATTATCCAGGACAGTTGTCCACGTACAGAAGATACAA 1640
rat             CCACGGACCCTGTAAAATACTTCATTATCCAGGACAGATGTCCACATACAGAAGATACAA 1633
                **** *****  * ** ******** ********** ** ****   *    **  ***

dog             CCATCCAGGTGGTGGAGAATGGGGAGTCCCCTCAGGGCCGATTTTCTGTACAGATGTTCC 1791
cow             CCATCCAAGTGGAGGAGAATGGGGAGTCCCCTCAGGGCCGGTTTTCTGTCCAGATGTTCC 1783
human           CTATCCAAGTGGTGGAGAATGGGGAGTCCTCCCAGGGCCGATTTTCCGTCCAGATGTTCC 1765
mouse           CCATTCAGGTGACAGAGAATGGCGAGTCATCTCAGGCCCGATTTTCTGTTCAGATGTTCC 1700
rat             CCATTCAGGTGACAGAGAATGGCGAGTCCTCTCAGGCCCGGTTCTCTATTCAGATGTTCC 1693
                * ** ** ***   ******** *****  * **** *** ** **  * **********

dog             GTTTTGCCGGGAACTACGACCTGGTCTACCTGCACTGTGAAGTGTATCTCTGCGACATCA 1851
cow             GTTTTGCTGGAAACTACGACCTGGTGTACCTGCATTGTGAAGTGTATCTCTGTGACACCG 1843
human           GGTTTGCTGGAAACTATGACCTAGTCTACCTGGACTGTGAAGTCTATCTCTGTGACACCA 1825
mouse           GGTTTGCAGGAAACTACGACCTTGTCTACCTTCACTGCGAGGTGTACCTATGTGACTCTA 1760
rat             GGTTTGCAGGAAACTCCGACCTTGTCTACCTTCACTGCGAGGTGTACCTGTGTGACACTA 1753
                * ***** ** ****  ***** ** *****  * ** ** ** ** ** ** ** ***

dog             TTAATGAAAAATGCAAACCTACCTGCTCTGGGACCAGATTCCGCAGTGGAGGCATCATAG 1911
cow             TGAATGAAAAGTGCAGACCTACCTGCCCTGAGACCAGATTCCGCAGTGGGAGCATCATAG 1903
human           TGAATGAAAAGTGCAAGCCTACCTGCTCTGGGACCAGATTCCGAAGTGGGAGTGTCATAG 1885
mouse           CGAGTGAACAGTGTAAACCTACCTGCTCTGGTACTAGATTTCGAAGTGGGAACTTCATAG 1820
rat             TGAGTGAGCAGTGTAAACCTACCTGTTCTGGTACTAGATATCGAAGTGGGAACTTCATAG 1813
                  * ***  * ** *  ********  ***  ** ****  ** *****     ******

dog             ACCAAAGCCGCGTCCTGAACTTGGGTCCCATCACACGGAAAAATGTCCAGGCAGTGGTCT 1971
cow             ACCAAACCCGTGTCCTGAACTTGGGTCCCATCACACGGAAGGGGGGCCAGGCTGCAATGT 1963
human           ATCAATCCCGTGTCCTGAACTTGGGTCCCATCACACGGAAAGGTGTCCAGGCCACAGTCT 1945
mouse           ATCAGACCCGTGTCCTGAACTTGGGTCCCATAACACGACAAGGTGTCCAGGCCTCAGTGT 1880
rat             ATCAGACCCGTGTCCTGAACTTGGGTCCCATCACACGACAAGGTGTCCAGGCCTCAGTGT 1873
                * **   *** ******************** *****  *    * ******     * *

dog             CAAGGGCTGCTTCCAGCAGCTTGGGGTTCCTGAAGGTCTGCCTGCCTCTGCTTCTATCGG 2031
cow             CAAGGGCTGCTCCCAGTAGCTTGGGGCTTCTGCAGGTCTGGCTGCCTCTGCTTCTGTCGG 2023
human           CAAGGGCT---TTTAGCAGCTTGGGGCTCCTGAAAGTCTGGCTGCCTCTGCTTCTCTCGG 2002
mouse           CCAAGGCTGCTTCCAGCAACTTGAGGCTCCTGAGCATCTGGCTGCTGTTGTTTCCCTCAG 1940
rat             CCAAGGCTGCTTCCAGCAACTTGGGGTTCCTGAGCATCTGGCTGCTGCTGTTTCTCTCAG 1933
                * * ****      ** * **** ** * ***    **** ****   ** ***  ** *
```

Figure 1. Continued

```
dog             CCACCCTGACCCTGATGTTTCAGTGATTGGCAGCTGGAAACCCTGTGCTCTGTGGCT--- 2088
cow             CCACTTTGACCCTGATGTCTCCGTGACTG-TGGCCGGAAATCCTGTACTCTGTGGCTACC 2082
human           CCACCTTGACCCTGACTTTTCAGTGACTGACAGCGGAAAGCCCTGTGCTCCATGGCTGCC 2062
mouse           CCACTTTGATCTTCATGGTTCAATGATGGAAAGCAGAAAACCTGGT-----GTGGCTCCC 1995
rat             CTACTTTGACCCTGATGGTTCATTGATGGAAAACAGAAAACCTGGT-----GTGGCTCCC 1988
                * **  *** * * *    **  ***  *    * * **  *  **      *****

dog             ----TCA-TTCCTGCTGGCTG-GGGAT--GATGCCGCTTAGTGCTCCAGCCACAGAAAAG 2140
cow             AAACTCACTTTCTACTGACTGCGAGGC--CATGCAGGCTAGCACTCCACCCATAGAGAAA 2140
human           AT-CTCACCTCCTGCTGGGCAGGGGGCATGATGCGGGCCAGTGCTCCAGCCACAGAAAAG 2121
mouse           AG-CTCACTTCCTGCTGGCCAGGAGTGGGGATGCAGGCTGGTTATC-AGCCAGGAGAGAG 2053
rat             AG-TTCACTTCCTGCTGGCCAGGAGTGGGGATGCAGGCTGGTAAGC-AGCCAGGAGAGAG 2046
                    ***  * ** ***     * *     **** *    *    * * ***      *

dog             GGAACTCACACTTCAGTCAC-CTGCTCCT-CTTTCCCCACCTTTAATCCTCATTATCACA 2198
cow             GGAGCCCACACTTCAGTCACTCTGCTCCTATTTTTCTCCCCTTTAATCCTTGTTACCAAA 2200
human           AAAGTTCATGCTTTGTTCAGCCTG--CCTTCTTTTCTCCCTTTTAATCCTGGCTGTCGAG 2179
mouse           GGAGCTCACACTGCCTCCAGCTTGCACGTACTCTTT----CTTTAATCCTCACCATCA-A 2108
rat             GGAGCTCACACTGCCTCCAGCTTGCACTTACCCTTTT--CCTTTAACCCTCACCATCA-A 2103
                  *   **  **     **   **  * *    *       ***** ***      *

dog             AAATATCTTAATATCTTTAAATTCTGCCTTCTTTCAAAATGGGACTTGTGACATGTGTGC 2258
cow             AAAATAGCCTGTGTCTTTAAATGCTGCTTTCTCTCAAAATGGGTCTTGTGATATGTCTGT 2260
human           AAACA-GCCTGTGTCTTTAAATGCTGCTTTTTCTCAAAATGGGACTTGTGACG---GTGT 2235
mouse           AAACAAGTTGGCATCTTTCAGTGCTGTTCTTCCTTCAAATG-------TGATATGT---- 2157
rat             AACCGAGTTGATATCTTTCAATGCTGCTGATTCTTCAACTGGAACTTGTAATGTGTCTGT 2163
                **           ***** * * ***       *  ** **       * *

dog             ACATGAGGCCCCATGTCTCCTTAAAGATTGTGGCAAAATAATAAT-----AATTTTAAGA 2313
cow             TCATGAGGCCCC-TATCTCCTTAAAGGGTATTGCAAAATAATAATTAAAAAAAGCTCAGT 2319
human           ACCTGAGGCCCC-CATCTCCTTAAAGAGTGTGGCAAAATAATGATTTTTAAATCTC---- 2290
mouse           --ATGAAGCCCC-TGTCTCCTGAAGGAATGTGGCAAAATAATAAGTTTAAAGTATCAAAA 2214
rat             ACATGAAGCCCC-TGTCTCTTTAAGGAATGTATCAAAATAATAAGTTTAAAACAATCCGG 2222
                   *** *****   **** * ** *  * *  ********* *      *

dog             AAAAAAAAAAAAAAAAAAAAAA 2335
cow             CTAAAAAAAAAAAAAAAAAAA- 2340
human           ----------------------
mouse           AAAAAAAAAAAAAA------- 2229
rat             AATTC----------------- 2227
```

Figure 1. Continued

The predicted canine Umod contains 642 amino acids and has an estimated molecular weight (not including associated carbohydrate moieties) of 72039.96 kDa (Peptide Molecular Weight Calculator: http://elmo.ucsc.edu/MWform.html). The putative canine Umod has 81.8 and 82.3% similarity at the amino acid level to bovine and human Umod, respectively. As expected, nucleotide and amino acid sequences of uromodulin of the mouse and rat are slightly more divergent; there is 72.6% identity and

72.0% similarity of nucleotide and amino acid sequences between dog and mouse, and 70.9% identity and 75.2% similarity of nucleotide and amino acid sequences between dog and rat.

Umod has been localized to human HSA16p13.11 (Jeanpierre et al. 1993; Pook et al. 1993), murine MMU7F1-F2 (Fukuoka and Matsuda 1997), and rat RNO1q36-q37 (Fukuoka and Matsuda 1997). In order to add type I loci to the map of the canine genome and add data on synteny among species, primers for radiation hybrid mapping of canine *umod* were designed: Umod-Ex8/9: 5'ACTGTGAAGTGTATCTCTGCGAC3' and 5'GGTTTTGGTCTATGATGCCTC3'. Amplification using these primers generated a PCR product of 650 bp from a stock canine genomic DNA sample. Amplifications were performed in 10 µl reactions containing 50 ng DNA of each of the RHDF5000 hybrid cell lines, 0.3 µM of each primer, 250 µM of each dNTP (Pharmacia, Peapack, NJ), 2 mM $MgCl_2$, 1X AmpliTaq Buffer and 0.5 U AmpliTaq Gold (Perkin-Elmer, Boston, MA). PCR reactions were carried out in PTC-200 thermal cyclers (MJ Research, Waltham, MA) with the following program: 8 min 94°C, followed by 20 cycles of 30 s 94°C, 30 s 61°C (decreasing of 0.5°C per cycle), 1 min 72°C and 15 cycles of 30 s 94°C, 30 s 51°C, 1 min 72°C and a final extension of 3 min 72°C. PCR products were separated by electrophoresis through 2% agarose and visualized after staining with ethidium bromide. Results were scored in terms of present, absent or ambiguous in each of 126 hybrid cell lines composing the radiation hybrid panel (Vignaux et al. 1999). These data were computed based on the most recent version of the canine map (Breen et al. 2001) using the Multimap package (Matise et al. 1994b). In this study, a two-point

analysis was performed, i.e., the program computed the *umod* RH data against the 1500 RH data of the previous map, and gave the marker which has the best Lod score among all of the markers.

RH mapping revealed that canine *umod* is linked to the zona pellucida gene, *CZP2*, with a Lod score of 13.8, in a small RH group not yet assigned to a canine chromosome. Due to tight linkage with *CZP2*, the fragments amplified using RH primers for *CZP2* and *umod* were sequenced to confirm that they were two separate genes (data not shown). The *umod* and *CZP2* genes are also linked in human, and are located on chromosome HSA16p13.11 and HSA16p12, respectively. These regions of HSA16 (16p13.11 and 16p12) have been shown through fluorescence *in situ* hybridization to have conservation of synteny with the central part of canine chromosome 6 (Breen et al. 2001), suggesting that *umod* and *CZP2* will probably be located in this region of CFA6 in future RH maps. Interestingly, *umod* is one of a tentative gene family that includes the zona pellucida genes *ZP2* and *ZP3*, the exocrine pancreas zymogen granule membrane glycoprotein gene *GP2*, and the beta-glycan (TGF-β receptor III) gene.

CHAPTER IV

GENETIC CAUSE OF X-LINKED ALPORT SYNDROME IN A FAMILY OF DOMESTIC DOGS*

Overview

Alport syndrome is a hereditary disease of type IV (basement membrane) collagens that occurs spontaneously in humans and dogs. In the human XLAS is caused by mutations in *COL4A5*, resulting in absence of type IV collagen α5 chains from the GBM of affected individuals. The consequence of this defect is progressive renal failure for which the only available treatments are dialysis and transplantation. Recent studies support the prospect of gene transfer therapy for Alport syndrome, but further development of required technologies and demonstration of safety and efficacy must be accomplished in a suitable animal model. We previously identified and have propagated a family of mixed-breed dogs with an inherited nephropathy that exhibits the clinical, immunohistochemical, pathological, and ultrastructural features of human XLAS. To identify the causative mutation, *COL4A5* cDNAs from normal and affected dogs were sequenced in their entirety. Sequence analyses revealed a 10 bp deletion in exon 9 of affected dogs. This deletion causes a frame-shift that results in a premature stop codon in exon 10. Characterization of the causative mutation was followed by development of an allele-specific test for identification of dogs in this kindred that are destined to develop XLAS.

---

*Reprinted with permission from Cox ML, Lees GE, Kashtan, Murphy KE (2003b) Genetic cause of X-linked Alport syndrome in a family of domestic dogs. Mamm Genome 14, 396-403

Introduction

Alport syndrome is a hereditary progressive renal disease that is often associated with hearing loss, ocular abnormalities, or both, and rarely with leiomyomatosis (Kashtan 1998; Tryggvason and Martin 2001). The disease is caused by mutations in genes that encode certain members of the type IV collagen family of peptides. Type IV collagen is an important structural component of BM, the thin, sheet-like extracellular matrices that separate cells from their underlying stroma in tissues throughout the body (Hudson et al. 1993; Miner 1999). Similar to other collagens, type IV collagen is a triple-helical protein consisting of three α chains. The collagen α chains have (Gly-Xaa-Yaa)$_n$ repeats, with glycine being the only amino acid small enough to fit into the center of the triple helix. Type IV collagen α chains have many interruptions in their Gly-Xaa-Yaa repeats, allowing for flexible kinks in the triple helix. In addition to the collagenous domain, type IV collagen α chains have a carboxyterminal noncollagenous (NC1) domain and an aminoterminal noncollagenous (7S) domain that are important for their assembly into a network structure. Six genetically distinct type IV collagen α chains have been described. The α1(IV) and α2(IV) chains are ubiquitous in BM and are present in triple-helical molecules in a 2:1 ratio. The other α chains have variable and more restricted BM distributions (Miner 1999).

Alport syndrome is an uncommon, but nonetheless important cause of human ESRD worldwide. One recent estimate, based on data from Finland, was that the birth prevalence of AS was about 1 in 53,000 live births (Pajari et al. 1996). Such data likely apply to large populations of European origin, but data regarding large populations of

other geographic origins are lacking (Levy and Feingold 2000). In Europe, AS accounts for 1.5% of children starting renal replacement therapy (Loirat et al. 1994). In the United States, AS affects about 3% of children with chronic renal failure and about 0.3% of adults with ESRD (United States Renal Data System 1996).

Alport nephropathy develops as a consequence of absence from the GBM of a crucial assembly of type IV collagen heterotrimers containing α3, α4, and α5 peptide chains (Gunwar et al. 1998; Tryggvason and Martin 2001). A defect in any one of the three genes (*COL4A3*, *COL4A4*, and *COL4A5*) encoding these chains can disrupt proper assembly of the α3/α4/α5 chain network, which is required for maintenance of adult GBM structure and function (Harvey et al. 1998). In 80-85% of families, AS is X-linked (i.e., XLAS) and caused by mutations in the *COL4A5* gene which encodes the α5(IV) collagen chain (Barker et al. 1990; Hostikka et al. 1990). Since the first report of a causative mutation in *COL4A5* (Barker et al. 1990), more than 350 other mutations have been identified in XLAS patients (Martin et al. 1998; Inoue et al. 1999; Jais et al. 2000). About 15% of AS cases are autosomal recessive, and autosomal dominant inheritance of AS occurs rarely (Kashtan 1998). Autosomal forms of AS are caused by mutations in the *COL4A3* or *COL4A4* genes (Lemmink et al. 1994; Mochizuki et al. 1994; Boye et al. 1998; Heidet et al. 2001; Longo et al. 2002), which are located on human Chromosome (Chr) 2. Almost all AS mutations are private; most are point mutations, small deletions or insertions (Tryggvason and Martin 2001). Those causing substitutions of a glycine residue within the collagenous domain may produce a less severe disease phenotype; however, nonsense mutations, splice site mutations, and large rearrangements usually

produce more severe phenotypes with earlier onset of disease (Gross et al. 2002). Additionally, large deletions involving certain portions of *COL4A5* and *COL4A6*, which is adjacent to *COL4A5* on the X Chr, result in Alport syndrome and diffuse leiomyomatosis (Tryggvason and Martin 2001).

Although type IV collagen abnormalities have been identified as the inciting cause of AS, important aspects of AS remain ill defined (Kashtan 2002). For example, the complete absence of the GBM α3/α4/α5 chain network does not impair glomerular development nor does it cause glomerular function to be abnormal early in life (Harvey et al. 1998). Also, mechanisms linking the type IV collagen defects to subsequent events, beginning with GBM changes and culminating with ESRD, are incompletely understood (Kashtan 2002). Moreover, treatments of proven efficacy to retard or halt progression of renal failure in AS patients are not available. Thus, dialysis and eventual renal transplantation currently are the mainstays of treatment for patients with AS (Kashtan 1998; Tryggvason and Martin 2001).

Availability of suitable animal models provides needed opportunities for investigation of the pathogenesis and treatment of AS (Kashtan 2002). Several transgenic murine models of autosomal recessive AS have been developed (Cosgrove et al. 1996; Lu 1999; Miner and Sanes 1996), but a murine model of XLAS has not (Kashtan 2002). However, several forms of AS have been identified as spontaneous diseases in dogs (Jansen et al. 1984; Hood et al. 1995; Lees et al. 1998; Lees et al. 1999), including two forms of XLAS (Jansen et al. 1984; Lees et al. 1999). The first form was described in a family of Samoyed dogs, initially reported in 1977 (Bernard and Valli

1977), in which partial sequencing of the canine *COL4A5* cDNA (GenBank Accession no. U07888) subsequently revealed a G-T substitution in exon 35, creating a premature stop codon (Zheng et al. 1994). Recently, a second family of mixed-breed dogs with XLAS was described (Lees et al. 1999). These Navasota (NAV) dogs, so called because of the name of the town from which they originated, exhibited typical clinical, histological, immunochemical, and genetic features of XLAS; however, the sequence of exon 35 of *COL4A5* was normal in affected dogs (Lees et al. 1999).

Since 1998, a colony of NAV dogs has been propagated at Texas A&M University for studies of the pathogenesis and treatment of XLAS. Reported here are the full-length coding sequence of canine *COL4A5* (GenBank Accession no. AF470624) and identification of the frame-shift that is causative for XLAS in NAV dogs. This mutation results in the production of a truncated protein lacking most of the collagenous region and NC1 domain essential for trimerization. Determination of the sequence of the coding region and characterization of the NAV dog mutation will permit construction of a full-length, wild-type cDNA for *COL4A5*. This is necessary for utilization of our canine model to investigate possible gene transfer therapy for AS.

Materials and methods

Dogs used in this study were produced within a colony of NAV dogs that is maintained at Texas A&M University. Methods of clinical and pathological evaluations, including immunofluorescence microscopy to assess α5(IV) chain expression in epidermal BM, were as previously described (Lees et al. 1999). Samples of renal cortex

obtained at necropsy from normal and affected NAV dogs were snap frozen in liquid nitrogen. Tissue was disrupted using Dounce homogenizers (Kontes, Vineland, NJ) and the Poly(A)Pure Kit (Ambion, Austin, TX) was used to extract mRNA.

Gene-specific primers were designed for reverse transcription by using Netprimer (Premiere Biosoft, Palo Alto, CA), and gene specificity was assessed using the BLAST internet tool (http://www.ncbi.nlm.nih.gov/BLAST/). An enhanced AMV-RT (Sigma, St.Louis, MO) was used to synthesize cDNA according to the manufacturer's instructions. Both one-step (Clontech BD Bioscience's Advantage One-Step RT-PCR kit, Palo Alto, CA) and two-step (Sigma eAMV RT-PCR kit, St. Louis, MO and Clontech BD Bioscience Advantage GC2 PCR kit) RT-PCR kits were used to amplify cDNAs.

The sequence for 40% of the 3' end of canine *COL4A5* was previously published (GenBank Accession no. U07888) and was, therefore, used to design the initial reverse transcription primer (*COL4A5*-NC2-R). All other gene-specific reverse transcription primers were designed using the sequence determined for NAV dogs. RT-PCR primers *COL4A5*–E, C, and F were also designed using the published canine *COL4A5* sequence. For amplification of the 5' portion of the cDNAs, the published sequence of the human cDNA (GenBank Accession no. NM_000495) was used to design sense primers. Data derived from sequencing of the 3' fragments were used to design canine-specific antisense primers. Each subsequent primer pair contained one human and one canine-specific primer (Table 4). In order to maximize both gene specificity and cross-species amplification, touchdown protocols were used for most PCRs. In this method a high

temperature is used for the initial annealing step, and the temperatures is decreased one degree per cycle to a target temperature at which the remainder of the cycles take place. Standard thermocycling protocols were used for amplification of two fragments (*COL4A5*-J and *COL4A5*-Q+). All annealing temperatures are reported in Table 4.

Amplification products of expected sizes were cloned into pCR2.1 (Invitrogen, Carlsbad, CA), and were transformed into chemically competent *E. coli* JM109 (Promega, Madison, WI) or DH5αMCR cells (GibcoBRL Life Technologies, Carlsbad, CA). Plasmid DNA was isolated and sequencing reactions were carried out with the dye termination method and resolved on the 373XL and 377XL automated sequencing machines (PE Biosystems, Foster City, CA). Assembly of contiguous fragments, alignment, and sequence analysis were carried out using Vector NTI Suite II (Informax, Bethesda, MD) software.

Total RNA was isolated from canine peripheral leukocytes using the RNA Blood Module, RNAqueous, and Poly(A)Purist Kits (Ambion). Because the large introns of canine *COL4A5* have not been sequenced, the primeres were designed to amplify the cDNA on either side of the mutation site. Nested RT-PCR was performed on mRNA reverse transcribed using *COL4A5*-RevT-5, by using the primer set *COL4A5*-Q+ and amplified with the Advantage GC-2 RT-PCR kit (Clontech BD Bioscience). An aliquot of that product was subsequently reamplified, with the *COL4A5*-AS1 set of fluorescently labeled primers, in which the sense primer was labeled with 6-FAM (Sigma Genosys, The Woodlands, TX). Fragments were resolved on the ABI 310 capillary-based Genetic

Analyzer (PE Biosystems) and analyzed using the Genotyper 2.0 software (PE Biosystems).

Table 4. COL4A5 primer sets.

Name	Size (bp)	Location (to human: NM_ 033380)	Annealing Temp. ('C)	for	Sequences
COL4A5-S-F/R	316	F: 43 R: 359	55-52	RT-PCR	S-F:TTCTTCCACTCTTAAAAAGC S-R:GAAATCCTCTCTCCCCCT
COL4A5-Q-F/R	695	F: 207 R: 902	63-60	RT-PCR	Q-F:AACTGCGTGGAGTCAGCCTGG Q-R:GACCTTGCTCACCTTTTTCACCTTTG
COL4A5-P-F/R	470	F: 696 R: 1166	63-57	RT-PCR	P-F:CACTGCCAGGACCAAAGGGTAATC P-R:CAGGGTAACCAGGATCACCAGGTAAAC
COL4A5-M-F/R	488	F: 1074 R: 1562	63-59	RT-PCR	M-F:AAGGAGAGCCAGGCAAAAGAGGTA M-R:CTGTTTCACATATCTCATCACTAGGAGAGA
COL4A5-5'-F5/R4/5	629	F: 1303 R: 1932	65-58	RT-PCR	F5:TGGAGAAAAAGGAGAGCGAGGATT R4/5:GTGCCAGGTAAACCAGGAAGCC
COL4A5-J-F/R	378	F: 1756 R: 2132	65	RT-PCR	J-F:ACATGCTGGTCTAACAGGTCCCAAAGGATTA J-R:CTGGATTTCCTGCCACACCTTGTATGCC
COL4A5-H-F/R	547	F: 2099 R: 2646	71-64	RT-PCR	H-F:GGAAAAAGGCATACAAGGTGTGGCAGGA AATC H-R:AGCCCAGGAGGTCCCATTGGTCCAGGTT
COL4A5-G-F/R	754	F: 2515 R: 3269	65-58	RT-PCR	G-F:TGGTCCAAAAGGTGATCGTGGTTTCC G-R:GAGGTCCTGTAAGCCCTGGTTTCCC
COL4A5-F-F/R	316	F: 3211 R: 3527	66-63	RT-PCR	F-F:ACCATCAGGTCCCAAAGGTAACCCCG F-R:TAATCCAGGCAAACCAGTATCTCCCATA GAAC
COL4A5-C-F/R	968	F: 3494 R: 4462	66-63	RT-PCR	C-F:GTTCTATGGGAGATACTGGTTTGCCT GGATTAC C-R AGGGAGTCCGTTGCGTCCTGGAT
COL4A5-E-F/R	572	F: 4199 R: 4771	68-63	RT-PCR	E-F:GGTGTTCCAGGATTTCCAGGTATGAAGGGAC E-R:AGGCATGGTACTGAAGCGACGAAGGCAACT
COL4A5-NC2-F/R	571	F: 4670 R: 5241	63-57	RT-PCR	NC2-F:ATTTATGAAGGCTTTTCTCTCTTATATGTCC NC2-R:CGTGTCCTCAAGTCTCCTGCTTTC
COL4A5-T-F/R	500	F: 5189 R: 5689	64-59	RT-PCR	T-F:ATATGTTCAGCAAACCTCAGTCAGAAAC T-R:GGATACAGCAGGATTAGTAGCACCG
COL4A5-Q+-F/R	592	F: 573 R: 941	60	RT-PCR	Q+-F:GCAATGGAACCAAGGGAGAACG Q+-R:AGGGTAACCAGGATCACCAGGTAAAC
COL4A5-AS1-F/R	252/ 242	F: 687 R: 940	58-56	RT-PCR	AS1-F:ATTATGTCATCACTGCCAGGA AS1-R:GTTCACTGATCTGCCCAGG
COL4A5-RevT-1		3809-3838		GSP-RT	AATGCCAGGTAATCCTCCATCACCCTTTTGAC
COL4A5-RevT-2		3737-3770		GSP-RT	CAAAGCCTGGTTGACCTGGTTCACCCTTCTGTCC
COL4A5-RevT-3		2660-2689		GSP-RT	AATCCCTGGTGGTCCTGGTGGTCCCTGAAC
COL4A5-RevT-4		2768-2795		GSP-RT	GACTGCCTCTTTCACCTGGGGGTCCTGG
COL4A5-RevT-5		1607-1641		GSP-RT	GTGTCACCTTTGTCACCTTTCACTCCTT GTTCTCC

Results

Sequencing of canine COL4A5 cDNA and determination of the causative mutation

Previous work demonstrated an X-linked pattern of disease transmission and absence of type IV collagen α5 chains in the GBM of affected NAV dogs (Lees et al. 1999). Thus, an examination of *COL4A5* as a candidate gene was warranted. To do this, cDNA of *COL4A5* was synthesized from normal and affected dogs. Amplification of the entire coding sequence was difficult because the collagenous region has an unusually high GC level. This promotes formation of secondary structures during reverse transcription (RT), PCR, and subcloning. For example, rearrangements and deletions of amplicons occurred in recombination-deficient *Escherichia coli* (i.e., INVα, TOP10) strains. Therefore, JM109 and DH5*α*MCR strains of *E. coli* were used for propagation of constructs harboring segments of *COL4A5*. A second complicating factor is that the conservation of nucleotide sequences among members of the collagen gene super-family is very high, particularly in the collagenous region. This necessitated the use of gene-specific RT and PCR primers. Therefore, in order to capture the entire coding region of *COL4A5*, 13 overlapping fragments were amplified (Table 4). The smallest fragment amplified is 316 bp (*COL4A5*-F, *COL4A5*-S) while the largest is 968 bp (*COL4A5*-C) (Table 4). This sequential approach allowed determination of the complete sequences of the *COL4A5* cDNAs from multiple normal (GenBank Accession no. AF470624), and affected NAV dogs and includes the 5' and 3' UTRs. The BLAST internet tool was used to compare each of the fragments with the published partial canine *COL4A5* (GenBank Accession No. U07888) and full-length human *COL4A5* (GenBank Accession no.

NM_000495) sequences to verify that the amplicons were indeed *COL4A5*. Additionally, there was sequence consensus between areas of overlap for each fragment. The characteristic motifs of type IV collagens are present, including the globular 7S amino terminus, a long, conserved collagenous region with multiple interruptions (Pihlajaniemi et al. 1990), and the NC1 region. The sequence of the normal canine *COL4A5* has 93% identity to human, and the deduced open reading frame has 94% identity.

Comparison of the sequences of the normal and affected dogs revealed a 10 bp deletion in the putative exon 9 (as deduced from sequence information for human *COL4A5*) of *COL4A5* from dogs with XLAS (Figure 2). To confirm this mutation, this primer set (*COL4A5-Q*) was also used to amplify the same region from a total of ten normal and ten affected dogs, as well as an unrelated dog (a female Dalmatian). Sequence analysis revealed that the mutation segregated with XLAS affected dogs. As expected, when this region was amplified and sequenced from five carrier females, analysis revealed that approximately half of the transcripts harbored the mutation, while half were normal (data not shown).

This frame-shift mutation disrupts the amino acid Gly-X-Y collagenous domain repeat and results in a premature stop codon in putative exon 10 (Figure 2). This mutation is different from that described for Samoyed dogs (Zheng et al. 1994) and causes elimination of 85% of the carboxyl terminus. This results in the absence of the majority of the collagenous region and the entire NC1 domain. Because these are

necessary for trimerization of the collagen peptides, it is not surprising that the 10 bp mutation is so deleterious.

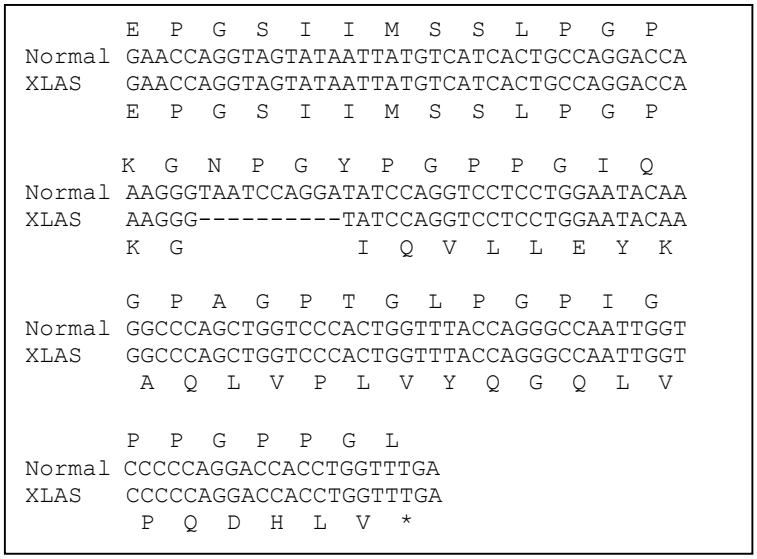

Figure 2. Nucleotide and amino acid sequences for COL4A5 exons 9 and 10. This shows the deletion, resulting frame-shift and premature stop codon, as determined in normal and affected male Navasota dogs with XLAS. A 10 bp deletion in exon 9 causes XLAS in NAV dogs.

Genetic test

Prior to work reported here, identification of dogs destined to develop XLAS was accomplished using immunofluorescence microscopy of skin biopsy specimens to assess expression of α5(IV) chains in the epidermal BM of one-week-old puppies. Immunostaining of α5 chains in epidermal BM has a uniform linear pattern in normal males and females, has a discontinuous linear (mosaic) pattern in carrier females, and is absent in affected males (Lees et al. 1999). Characterization of the mutation responsible for canine XLAS allowed for the development of a PCR-based test of *COL4A5* cDNA

that distinguishes between the two alleles (normal and mutant) in NAV dogs. A test using genomic DNA has not yet been developed because the small *COL4A5* exons (e.g., exon 9 is 84 bp) are separated by very large introns, which have not yet been sequenced, and are sufficiently different from those in human to prohibit successful primer design for amplification. Therefore, peripheral leukocytes were collected for isolation of mRNA for subsequent RT-PCR. Although only illegitimate transcripts of *COL4A5* are found in the leukocytes, sufficient quantities of mRNAs for use in RT-PCR exist. Two sets of nested primers were used in this experiment. The product from primer set *COL4A5*-Q+ is used in a subsequent reaction with the primer set *COL4A5*-AS1 (Table 4).

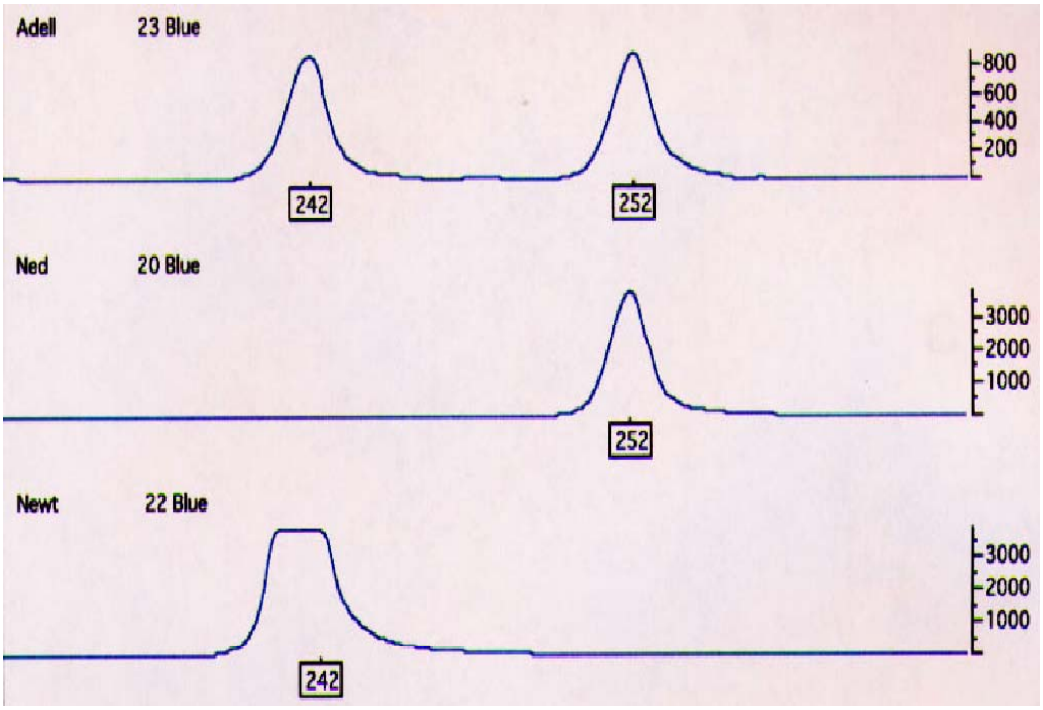

Figure 3. Example results of an allele specific test for XLAS in Navasota dogs. The test discriminates carrier females (Adell), normal males (Ned) or females (not shown), and affected males (Newt).

The allele present in normal males and females is 252 bp, while affected males have a 242 bp allele that results from the 10 bp mutation; carrier females are heterozygous (Figure 3). To verify the accuracy of the test, three normal male dogs, three affected males, two carrier females, and one unrelated normal dog (a Dalmatian) were genotyped.

Discussion

The entire canine *COL4A5* cDNA was assembled and sequenced. This analysis showed that canine *COL4A5* has 93% nucleotide identity to human *COL4A5* (GenBank Accession no. NM_033380), and an amino acid similarity of 94%. The level of conservation was less for murine *COL4A5* cDNA (GenBank Accession no. AB041350) (i.e., nucleotide identity of 88% and amino acid similarity of 90%). Nucleotide identity to the published partial canine 3' *COL4A5* cDNA (GenBank Accession no. U07888) was 99%.

Interestingly, comparisons of sequences revealed different splicing for the previously published canine *COL4A5* cDNA (31) and the sequence we report. Two transcript variants were originally noted in human *COL4A5* cDNAs from kidney and leukocytes (Guo et al. 1993). Other studies revealed a total of three transcript variants with different tissue-specific expression patterns (Martin and Tryggvason 2001). Two previously uncharacterized 9 bp exons were identified in intron 41 and were designated exons 41A and 41B (Martin and Tryggvason 2001); each codes for a Gly-X-Y repeat (Nakanishi et al. 1994). Human isoform I (GenBank Accession no. NM_000495) is found in many tissues, including the kidney, liver, placenta, spleen, and epithelium, and

is 18 bp shorter than the full-length transcript, because exons 41A and 41B are skipped. The previously published partial canine *COL4A5* cDNA (GenBank Accession no. U07888), sequenced from renal tissue (Zheng et al. 1994), is orthologous to human isoform I. Human isoform II (GenBank Accession no. NM_033380) is the predominant form found in the kidney and is also found in the liver, placenta, and spleen. This variant uses both exons 41A and 41B, and is most similar to the NAV canine *COL4A5* renal cDNA sequence reported herein. Human isoform III (GenBank Accession no. NM_033381) is expressed solely in epithelium and has only the sequence from exon 41B. Thus, isoform III is 9 bp different from the aforementioned transcripts.

The predicted effect of the identified mutation is the production of a truncated peptide chain that is unable to participate in post-translational, extracellular assembly in type IV collagen networks. In XLAS, the α3 and α4 chains are usually absent from the GBM (as determined by immunostaining), even though their genes are intact (Nakanishi et al. 1994; Knebelmann et al. 1996; Kalluri et al. 1997; Heidet et al. 2000). This is thought to be due to failure of the proper assembly of stable α3/α4/α5 trimers. Not all mutations that result in the production of a defective α5 chain prevent formation of an α3/α4/α5 trimer (Kalluri et al. 1997); although such trimers are not fully functioning proteins. Upon secretion, these defective trimers are subject to extracellular degradation, since they are unable to form a stable supramolecular network (Kalluri et al. 1997). Mutations that produce truncated α5 chains lacking the noncollagenous 1 (NC1) domain prevent trimerization, resulting in the intracellular degradation of the α3 and α4 chains (Kalluri et al. 1997). Production of truncated α5(IV) chains that cannot form stable

networks is consistent with negative immunostaining for α5(IV), α3(IV) and α4(IV) chains in the GBM of affected NAV dogs (Zheng et al. 1994), as well as the absence of α5(IV) and α6(IV) chains in the epidermal BM of affected dogs (Lees et al. 1999).

Both canine models of XLAS (NAV and Samoyed) have premature stop codons (Zheng et al. 1994) and fall into the "truncated protein" group of mutations, which includes premature stop codons, frame-shifts, and large rearrangements. The production of truncated α5(IV) chains preventing formation of the COL4A3/COL4A4/COL4A5 network in GBM is consistent with the severe disease phenotype and early onset of renal failure observed in affected dogs (Lees et al. 1999), as well as humans (Gross et al. 2002). The clinical presentation of AS in these dogs is analogous to Type-S (severe) designation (Gross et al. 2002) in which juvenile onset of ESRD is found in 92% of human patients. Similarly, both canine models have early onset of proteinuria (3-6 months in the NAV, 4 months in the Samoyed) and juvenile ESRD (6-18 months in the NAV, 8-10 months in the Samoyed) (Kashtan 2002).

Previously, identification of dogs destined to develop XLAS was accomplished with immunofluorescence microscopy of skin biopsy specimens to assess expressino of α5(IV) chains in the epidermal BM of 1-week-old puppies. Immunostaining of α5 chains in epidermal BM has a uniform linear pattern in normal males and females, has a discontinuous linear (mosaic) pattern in carrier females, and is absent in affectecd males (Lees et al. 1999). However, although accurate (100%) in males, it can be inaccurate in heterozygous females owing to Lyonization, thus potentially leading to misclassification of some carrier females as genetically normal individuals. The allele-specific test

overcomes this problem by examining *COL4A5* at the genetic, rather than the protein levels. RNA extracted from peripheral leukocytes may be used for genetic tests on NAV pups, facilitating the identification of XLAS affected and carrier dogs.

In summary, the sequence of the canine *COL4A5* cDNA is reported, and the 10 bp deletion in *COL4A5* causative for XLAS in NAV dogs has been identified. Characterization of this mutation has allowed development of an allele-specific PCR test to identify dogs in this kindred destined to develop XLAS. This will (1) allow the rapid identification of NAV dogs destined to develop XLAS and (2) complement the use of skin biopsies for diagnosis. Availability of the complete cDNA sequence for *COL4A5* facilitates production of a full-length, wild-type cDNA suitable for gene transfer therapy trials, which might ultimately lead to gene therapy-based treatments for AS in humans.

CHAPTER V

GENE EXPRESSION ANALYSIS IN A CANINE MODEL OF X-LINKED ALPORT SYNDROME

Overview

A kindred of mixed-breed dogs with an inherited nephropathy that is clinically, morphologically, and genetically similar to human XLAS has been identified. A nonsense mutation in *COL4A5* causes XLAS in this kindred (Cox et al. 2003). Kidneys from unaffected dogs within the kindred are histologically normal, whereas renal sections from dogs with XLAS have abnormalities typical of chronic progressive renal disease secondary to protein-losing glomerulopathy. Lesions in affected kidneys include progressive glomerular sclerosis with capsular fibrosis, interstitial fibrosis with mixed mononuclear cell inflammation, tubular atrophy, tubular dilation, and tubular proteinosis with intraluminal casts. Transcriptional profiling of renal cortex from dogs with advanced renal dysfunction resulting from XLAS and related, normal dogs of similar ages was performed using a canine oligonucleotide array, and the results revealed marked differences in gene expression that correlated with the morphologic changes present in kidneys from XLAS-affected dogs. At the transcriptional level, approximately 1,200 genes were differentially expressed (i.e., differences greater than 2-fold) in kidneys of affected dogs compared with controls. Quantitative RT-PCR (Q-RT-PCR) confirmed the expression profiles of 14 clinically interesting genes. Differences in levels of expression

were consistent with observed histological findings of chronic inflammation and fibrosis, as evidenced by the marked up-regulation of inflammatory mediators (e.g., *gp80*, *MCP-1*), interstitial collagens (*COL1A1* and *COL1A2*) and matrix remodeling proteins (e.g., *TIMP-1*). The strong correlation between pathological changes and differences in gene expression further characterizes this canine model of XLAS and enhances its utility for studies of disease progression and therapy.

Introduction

Alport syndrome is an hereditary progressive renal disease first described in the human (Alport 1927). AS is caused by mutations in genes that encode three of the type IV collagens, which are essential structural components of basement membranes found in the glomerulus of the kidney, the cochlea, and the cornea. Perturbation of these different basement membranes occasionally leads to hearing loss and ocular abnormalities along with renal disease (Kashtan 1998; Tryggvason and Martin 2001). Type IV collagens are members of the collagen superfamily, which at latest count includes more than 40 members in 27 subfamilies (Pace et al. 2003). Collagens are triple-helical proteins consisting of three alpha chains that contain (Gly-X-Y)n repeats, with glycine being the only amino acid small enough to fit into the center of a triple helix. Interruptions in the collagenous domains of the type IV collagens allow flexibility of the triple helix, permitting the heterotrimers to form a characteristic lattice-like network. Also important for assembly of the network are the carboxyterminal noncollagenous (NC1) domain and the aminoterminal noncollagenous (7S) domain. Of the six type IV collagen chains, the

COL4A1 and COL4A2 chains are present in all basement membranes, and the remaining alpha chains have varying distributions (Miner 1999). The interactions of collagens and other proteins are vital, because mutations in genes encoding collagens are responsible for many hereditary diseases. Some of these diseases (e.g., Alport syndrome and Ehlers Danlos syndrome) have also been described in the dog and result from mutations in the canine collagen orthologs (Freeman et al. 1987; Zheng et al. 1994; Rodriguez et al. 1996; Lees et al. 1999).

XLAS in the dog was first described in a Samoyed kindred (Bernard and Valli 1977). Partial sequencing of the cDNA for *COL4A5* (GenBank Accession Number U07888) in these dogs revealed a G-T substitution in exon 35, creating a premature stop codon (Zheng et al. 1994). A family of mixed-breed dogs (termed Navasota, abbreviated NAV) with XLAS was also described (Lees et al. 1999). NAV dogs exhibit typical clinical, histological, and immunochemical features of XLAS, and the disease is transmitted in X-linked fashion. When molecular analysis was performed, it was found that NAV dogs with XLAS have a normal sequence in exon 35 of *COL4A5* (Lees et al. 1999). Upon further analysis, it was determined that these dogs have a 10 bp deletion in exon 9 (of 51), causing a premature stop codon in exon 10 (Cox et al. 2003). The collagen protein is therefore truncated, resulting in a severe disease phenotype of ESRD at an average age of 11 months (Cox et al. 2003).

Our long-range goal is to understand the genetic factors underlying the nephropathy that accompanys cellular and molecular events in AS and generalized ESRD. Chronic renal failure is a common cause of death in the dog, and few models

(Brown et al. 2003) exist to allow the identification of biomarkers and mechanisms of secondary renal disease. Since all ESRD is characterized by the loss of cells and their replacement by inflammatory cells, fibroblasts, and extracellular matrix (Ying et al. 2001), the NAV dogs can be used to evaluate generalized progressive renal failure at the molecular level.

Many techniques have been developed to examine levels of gene expression, and DNA microarrays permit comprehensive study of global gene expression in many disease models. Microarrays are of great use in the study of nephrology (Hsiao et al. 2000; Hayden et al. 2003), including the opportunity to clarify molecular mechanisms of normal or disturbed organ or cellular function, and studies of these types have been carried out in the human and the mouse (Sarwal et al. 2003; Yuan et al. 2003). For a robust study of the dog, however, a canine-specific microarray platform is necessary. A proprietary, canine-specific Affymetrix GeneChip oligonucleotide array (liCanine1a) with 22,774 probe sets (13,729 derived from canine sequences and 9,045 derived from human sequences) has been developed and validated by Eli Lilly and Company (Higgins et al. 2003). This array will facilitate the identification of genetic events that underlie disease mechanisms and will provide insight regarding the progression of disease in our NAV model of XLAS. Use of the microarray will also allow identification of candidate genes of interest as biomarkers, genes expressed in attempted compensation by the kidney, and genes involved in tubulointerstitial scarring that are shared by end-stage XLAS nephropathy and generalized ESRD due to other causes.

Materials and methods

The NAV dogs are maintained at Texas A&M University, and studied according to protocols approved by the Texas A&M University Laboratory Animal Care Committee protocols. Seven related males, four affected dogs and three age-matched normal dogs were chosen for this study. Routine clinical and pathological evaluations were performed and tabulated as previously described (Lees et al. 1999). Affected dogs were sacrificed when a standardized end-point was reached, as determined by clinical parameters. Renal cortex samples obtained at necropsy were snap frozen and total RNA was isolated from 100 mg of each sample using RNA Stat60 (TelTest, Friendswood, TX) according to manufacturer's protocol. The isolated total RNA was further purified with Qiagen's RNeasy system (Valencia, CA). For reverse transcription, SuperScript II (Invitrogen, Carlsbad, CA) together with an oligo T-7-(dT)24 primer (Operon, Valencia, CA) was used to reverse transcribe 12 μg total RNA for each sample. The cDNA was purified and the BioArray T-7 polymerase labeling kit (Enzo, Farmingdale, NY) was used to synthesize biotinylated cRNA, according to manufacturer's instructions. Biotinylated cRNA was purified and fragmented using the GeneChip Sample Cleanup Module (Affymetrix, Santa Clara, CA) prior to overnight hybridization with the canine microarray (liCanine1a). Arrays were washed and stained according to the Affymetrix eukaryotic expression sample protocol (Affymetrix GeneChip Expression Analysis Technical Manual). Microarray Suite (MAS) 5.0 and Data Mining Tool (DMT), Affymetrix-based software packages, were used to perform data analysis.

To confirm transcriptional changes detected by the canine microarray, a subset of 14 clinically interesting transcripts were selected for quantitative real-time PCR. Gene specific primers, as listed in Table 5, were designed using Primer Express 1.5 (Applied Biosystems, Foster City, CA). Total RNA was isolated from kidney cortex of normal (n=3) and XLAS affected dogs (n=4) as previously mentioned. The RNA (3.2 μg) was DNase treated (DNA-free, Ambion) and reverse transcribed using Omniscript reverse transcriptase (Qiagen). Primer specificity and amplicon size was confirmed by electrophoresis through 4% agarose prior to quantitative fluorescence-based detection using the Applied Biosystems Prism 7900HT® Sequence Detection System. Reactions contained 100 μM forward and reverse primer, SybrGreen™ PCR Master Mix (Applied Biosystems), and cDNA template. Each assay was performed in triplicate and the PCR parameters were: 50ºC, 2 minutes followed by 95ºC, 10 minutes. These pre-amplification temperatures were followed by 40 cycles of 95ºC, 15 seconds and 60ºC, 1 minute. Data were analyzed by the comparative threshold (Ct) method of relative quantitation. All values are expressed as fold change (calculated from the signal log ratio) compared with the control group.

Results

Screening of the 22,774 probe sets (not all are unique due to overlap among proprietary, canine, and human sequences represented on the array), revealed approximately 1,200 genes that are differentially expressed (i.e., greater than 2-fold).

Obviously, only a subset of these can be shown herein (Table 6). To reduce the proportional error between random noise and signal intensity, the signal log ratio (SLR),

Table 5. Primer sequences for genes amplified by Q-RT-PCR.

Gene	**Accession Number**	**Primer Sequence (5'-3')**
MMP-9	AF169244	F: CCCTTGAACACGCATGACATC
		R: GTCCACCTGGTTCACCTCATTC
GP80	BI395936	F: CCTCCAGATTGCGGAGAAGTT
		R: CAGGGAGGACGTGTTGAACAT
VCAM-1	CFU32086	F: ATTCTTGGAAAGAAATGGCTTCCT
		R: GCCTATTTTTCAGGCAGAAAGTTTT
TIMP-1	AF077817	F: ACTGCAGAGTGACACTCACTGCTT
		R: TGGCGGCTCTGGAAACC
EGF	AB049597	F: GGATCCACTGTCTGGATATTAATGAG
		R: TGCAGGTATAGTTTCCCTCCATATT
COL1A1	AF153062	F: GCCGCTTCACCTACAGTGTCA
		R: GAGGTCTTGGTGGTTTTGTATTCG
COL1A2	AF035120	F: CATGCTCAGCTTTGTGGATACG
		R: GCAGTTGCCTCCTGTAAAGATTG
APOA2	Proprietary	F: TGTTCCCCCCATGAAGCTACT
		R: GCACAAAAGCCCCTTCGA
OSTEO PONTIN	Proprietary	F: TTGGTAGCCAGGAAAATTCTGAA
		R: ACAATTAACTCCTTCTCATGGCTTTG
COL4A1	CFU50933	F: TGCAACATCAACAACGTGTGTAA
		R: GGCTCCGGCGTGGATAG
COL4A2	CFU50934	F: TGTACTGCAACCCTGGTGATG
		R: AGTGGTGGACAGCCAGTAGGA
COL4A3	CFU50935	F: TGGAGCCTTATATTAGCAGATGCA
		R: AATGTCAGTGGTTTGGCTGTGA
COL4A4	CFU50936	F: TTGGTCTGGCAGGGTCTTG
		R: TGGCACACTTGGTGGATGTT
COL4A5	AF470624	F: CCCAGGATGCAACGTCTCA
		R: TCTCTTGGTCATATATAGTAGCACCA TTG
COL4A6	CFU50937	F: GCCCCGCTTTAGCACCAT
		R: TCATTGCGTCTGGCGTAGTG
GAPDH	AB038240	F: CGGATTTGGCCGTATTGG
		R: GATGGCGACAATATCCACTTTG

Table 6. Analysis of microarray data: Example of genes identified as up- or down- regulated in affected NAV dogs as compared with normal siblings.

Genes up-regulated in affected dogs					
Extracellular Matrix	**Inflammatory cell Response**	**Cellular Matrix/ Structural**	**Metabolism/ Signalling**	**Protein Transport**	**Stress Response**
MMP-9	*MHC* class II genes	*dynamitin*	*selenophosphate synthetase*	*CK2 interacting protein 1*	*MAPKAPK3*
decorin	*IgG heavy chains*	*NADPH*	*phospholemman*	*integral type I protein (P24B)*	
TIMP-1	*ApoE*	*actin cross-linking factor*	*L-asparaginase*	*AP3B1*	
COL1A1	*complement c3 precursor*	*claudin-2*	*HMGCL*		
COL1A2	*IgA heavy chain*	*tubulin*	*DLST*		
COL4A1	*A3 adenosine receptor*	*SH3 domain protein*	*UBE2M*		
osteopontin	*midkine precursor*	*Rho-1*	*dipeptidyl peptidase I*		
	IL-6		*S-100 protein*		
	MCP-1				
Genes down-regulated in affected dogs					
Cell proliferation	**Metabolism/ Signalling**				
EGF	*isovaleryl CoenzymeA dehydrogenase*				
IGFBP2	*kallikrein*				
	PDEA				

Table 7. Analysis of microarray data: Average fold changes of 10 genes up-regulated in affected dogs.

Gene	Fold change	Standard deviation
gp80	25.21	1.4363
MCP-1	11.18	1.7545
COL1A1	10.88	1.4353
Osteop	10.1	1.5685
COL1A2	5.82	1.4833
Rho-1	3.54	1.3066
Decorin	3.44	1.8017
Tubulin	2.97	1.9485
SH3	2.5	1.2805
COL4A1	2.17	1.7296

was used to calculate the average of all possible combinations of normal and affected dogs. Fold change (FC) was calculated by taking the arithmetic mean of all of the individual SLRs and raising the resulting number to the power of 2, effectively reversing the log transformation for reporting purposes. Standard deviations were calculated to determine how widely the values varied (Table 7). If it was not possible to calculate FC due to lack of signals in either all normal or all affected dogs, genes were categorized as increased or decreased in affected dogs as compared to normal siblings (Table 6). A conservative estimate of genes was used to ensure statistical significance. That is, only those genes that were expressed in all 3 normal dogs, and/or all 4 affected dogs, were tabulated. Relative fold changes of several of these genes are shown in Figure 4.

A subset of 14 clinically interesting genes was chosen for validation by quantitative RT-PCR. Signal intensity values were first normalized to the GAPDH standard and then down-scaled to normalize the values against a normal control dog. Following this, FC was calculated by raising the negative of the resulting number to the power of 2. The 14 genes chosen for Q-RT-PCR analysis included several categorized as

up-regulated (*gp80, VCAM-1, TIMP-1, COL1A2, COL1A1, Osteopontin*, and *COL4A1*), one that was down-regulated (*EGF*), one that was unchanged (*ApoA2*), and five that were non-hybridizing, but of special interest (*COL4A2-COL4A6*).

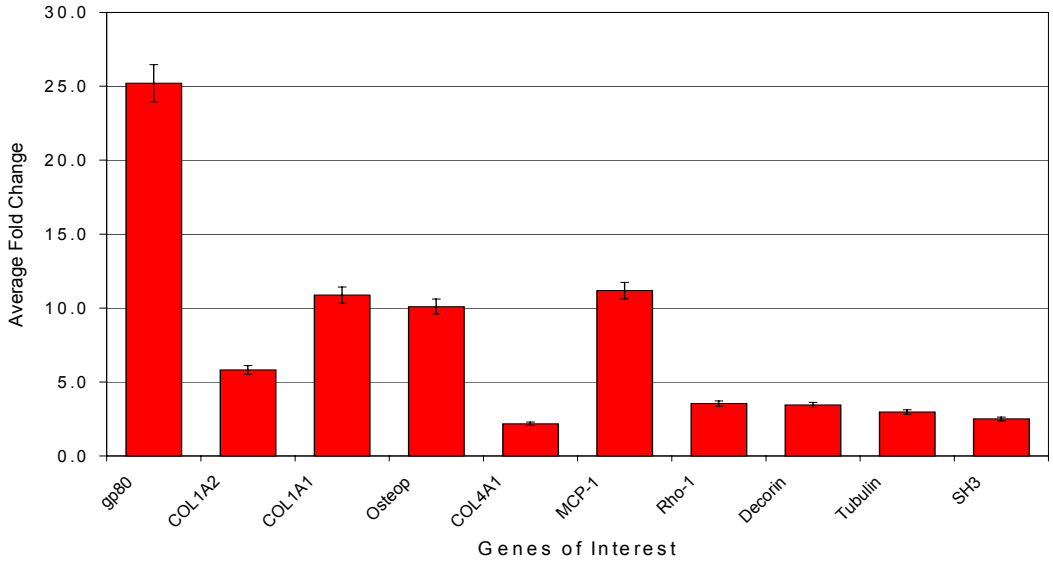

Figure 4. Analysis of microarray data: Average fold change of selected genes in affected NAV dogs.

The 2^{-ddCt} method of quantification was used to analyze the Q-RT-PCR results (Livak and Schmittgen 2001). In order to establish an internal control, both *18S* rRNA and *GAPDH* genes were amplified and compared to the genes of interest over a serial dilution of cDNA. The dCt varies with target dilution, and if the efficiencies of the targets and reference genes are similar, the reference gene may be used as a standard for those genes (Livak and Schmittgen 2001). In this case, *GAPDH* most closely matched the efficiencies of the genes of interest (data not shown), and was used as the internal control. Samples were run in triplicate, and after normalization to *GAPDH* and

calibration to an untreated control sample (a normal sib), the mean, standard deviation and fold change were calculated for all genes of interest.

All 14 genes assessed by Q-RT-PCR amplified well (Figure 5).

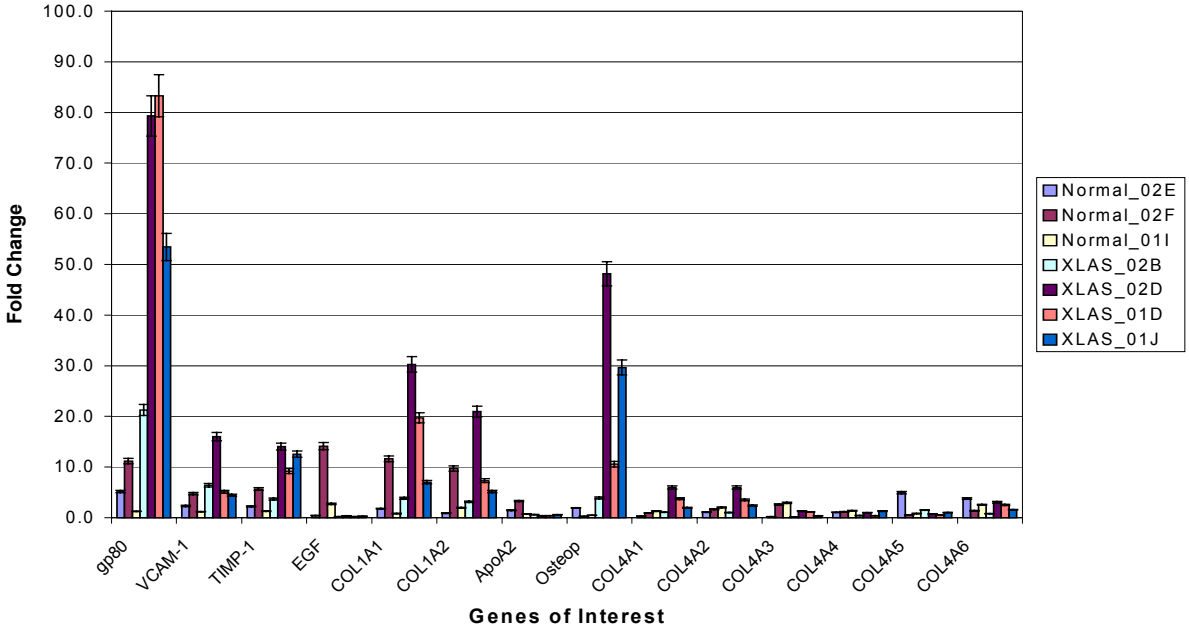

Figure 5. Q-RT-PCR data: Gene expression (by fold change) of normal and affected NAV dogs.

The 9 genes represented by working probe sets on the microarray confirmed the trends observed in hybridization to the microarray. Five of these transcripts are compared in Figure 6. Since the type IV collagens generally did not hybridize well to the chip, and they are of great interest in this disease model, all type IV collagens were subjected to Q-RT-PCR. *COL4A1* and *COL4A2* showed a very modest up-regulation in affected dogs, but the expression levels of *COL4A3*, *COL4A4, COL4A5* and *COL4A6* were not statistically significantly different between normal and affected dogs (Figure 7).

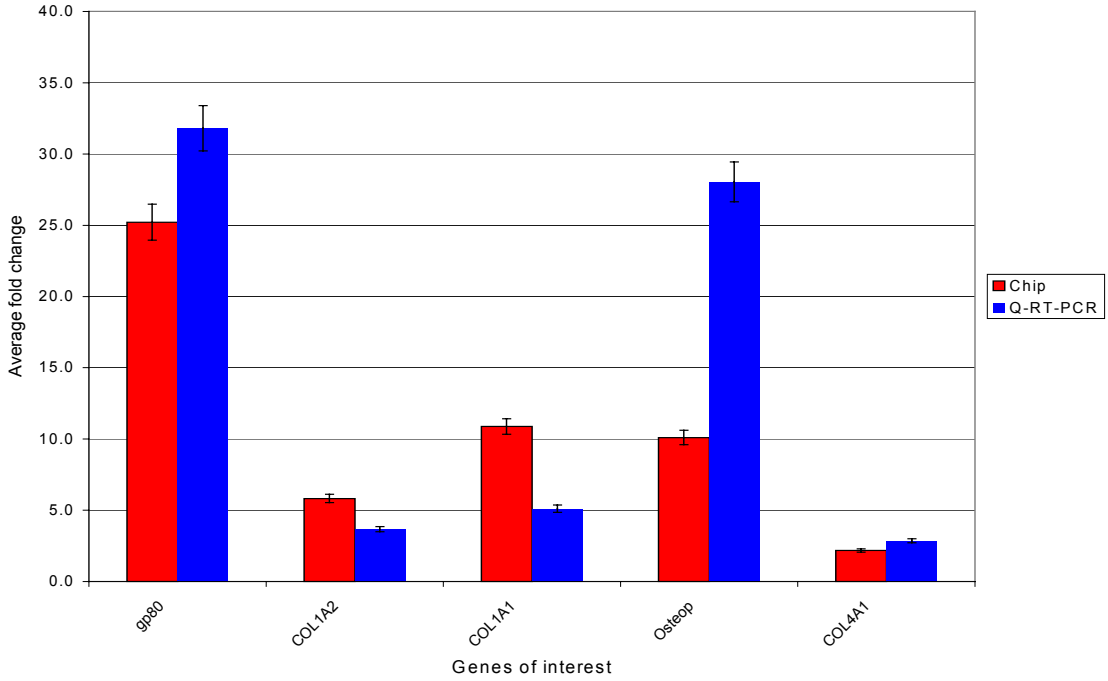

Figure 6. Q-PT-PCR and microarray data: Average fold change of up-regulated genes in affected NAV dogs.

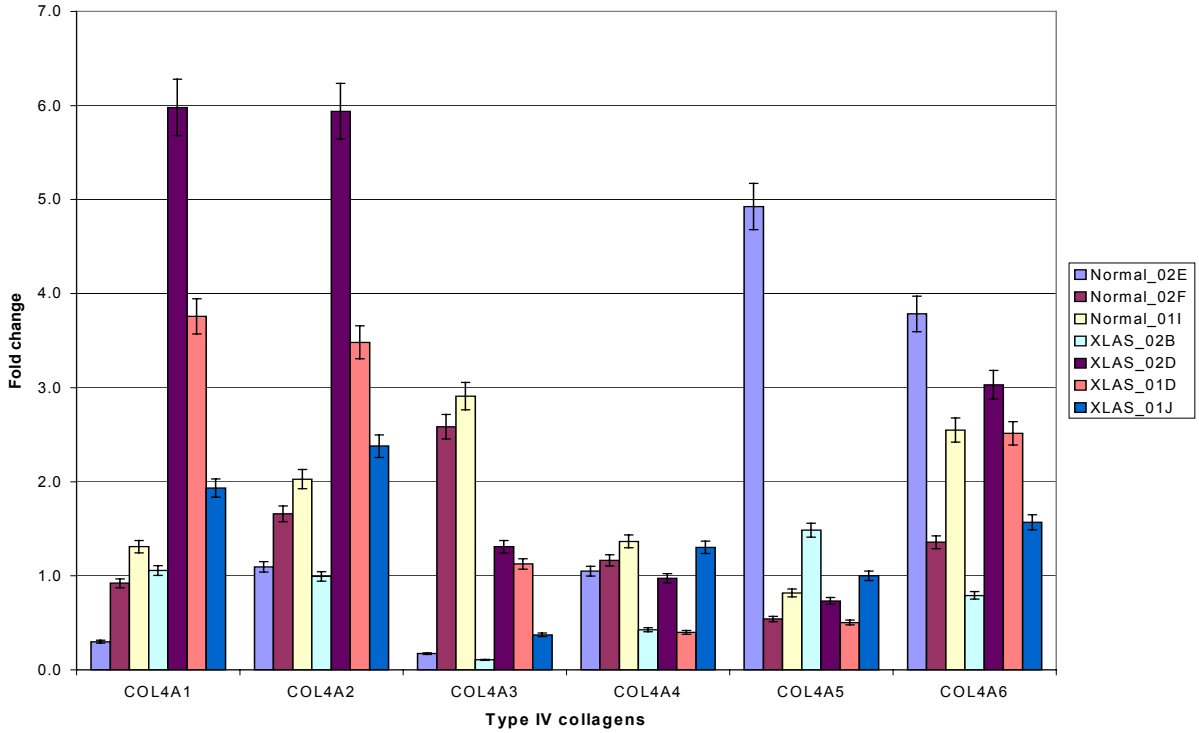

Figure 7. Q-RT-PCR data: Fold change of type IV collagens in normal and affected NAV dogs.

Discussion

Profiling the genome using microarray analysis allows examination of the molecular, pathologic and physiologic pathways of XLAS and hopefully those pathways common to ESRD regardless of the original inciting cause. Hence, this work will help reveal the involvement of known and unknown regulatory pathways of end-stage XLAS, and how these pathways differ in the kidney of normal and affected dogs. While the clinical features and histopathology of this disease are well described, these pathways are poorly understood, and can best be elucidated by such a platform.

Background noise in microarray data is a well-documented problem, particularly when samples are genetically heterogeneous; however, a sufficiently large sample size, sound experimental approach, and statistical analysis can offset these variables. Another difficulty in interpreting such data is that signal intensity may vary depending on sample preparation and individual chips. To minimize such variables, sample preparation was kept consistent, and microchips from the same batch were used on all samples.

The microarray used was previously described and validated in an analysis of acute phase response to lipopolysacchride (LPS) challenge in canine kidney (Higgins et al. 2003). Not surprisingly, more genes were up-regulated in the affected NAV dogs as compared to normal siblings than were down-regulated. This reflects the disease state, in which genes that are not normally expressed are turned on (e.g., as part of the inflammatory cell response), and those that are generally at a low working level (e.g., genes encoding proteins that are part of extracellular and cellular matrix remodelling) are up-regulated. The degree of tubulointerstitial nephritis in this canine model is

confirmed by the inflammatory cell response of up-regulated cytokine, chemokine, and MHC genes activated in the renal cortex. This inflammatory response creates further damage in the kidney (Ikezumi et al. 2003; Matsuo et al. 2003). Histologically, XLAS-affected kidneys in this study showed mixed mononuclear cell inflammation (mainly in the tubulointerstitium) along with other markers typical of chronic inflammation and fibrosis.

Expression of the type IV collagens as assessed by Q-RT-PCR confirm those reported from studies of the human. In contrast, an earlier study of the Samoyed XLAS dog, (Thorner et al. 1996) it was suggested that expression of *COL4A3*, *COL4A4* and *COL4A5* was down-regulated, relative to normal controls. It was hypothesized this may indicate pretranscriptional co-regulation of all 3 genes. This has not been found in the human (Nakanishi et al. 1996), and we see no evidence of such co-regulation in the NAV dog (Figure 8), at least in the end-stage kidney. Finally, other than *COL4A1* and *COL4A2*, none of the type IV collagens is significantly differentially expressed in XLAS affected dogs. COL4A1 and COL4A2 make up the embryonic GBM before the isotypic switch occurs in glomerular development (Miner 1998). It is logical that they are somewhat upregulated in the absence of the normal COL4A3/COL4A4/COL4A5 network, in an attempt to compensate for this missing structure (Kalluri et al 1997; Harvey et al. 1998), This is also consistent with the observations that expression of these genes as assessed by immunofluorescence in both the glomerulus and tubulointerstitial compartments is up-regulated (data not shown).

Extracellular matrix proteins are highly up-regulated in XLAS affected dogs. Seven known extracellular matrix genes are over-expressed, while none are called as under-expressed, consistent with the role of cortical fibrosis in the progression of disease. Extracellular proteoglycans are not differentially expressed, but the synthesis and turnover of extracellular matrix are demonstrated by the up-regulation of the type I collagens (*COL1A1* and *COL1A2*), two type IV collagens (*COL4A1* and *COL4A2*), as well as matrix metalloproteinases (*MMP-9*), *TIMP-1*, *decorin*, and *osteopontin*. A previous study assessing changes in matrix metalloproteinases showed *MMP-9* to be up-regulated in affected NAV dogs as compared to their normal sibs (Rao et al. 2003), which is corroborated by the data presented herein.

Monocyte chemoattractant protein-1 (*MCP-1*) and osteopontin (*SPP-1*) are chemokines that have been implicated in macrophage-mediated renal injury (Pichler et al. 1994; Hao et al. 2003; Ikezumi et al. 2003), and their up-regulation in affected NAV dogs is not unexpected. Surprisingly, neither *TGF-β* nor *endothelin* (*ET-1*) are differentially expressed, as assessed by multiple probe sets on the microarray. This is unexpected because *TGF-β1* is known to induce the synthesis of extracellular matrix proteins such as fibronectin and collagen (Noh et al. 1993; Goumenos et al. 2001; Strutz et al. 2002) and to play a role in glomerular damage and tubulointerstitial fibrosis (Sayers et al. 1999; Abbate et al. 2002).

ET-1 is a potent vasoconstrictor normally produced by glomerular epithelial cells and monocytes. This protein stimulates monocyte proliferation and increased expression of the collagens and fibronectins. ET-1 stimulates collagen matrix reorganization in

monocyte α1β1-integrin-mediated migration and MMP-2 activity (Boffa et al. 2001), causing collagen remodelling in progressive glomerulonephritis. *ET-1* is not significantly differentially expressed according to microarray analysis, although the type I collagens, which it is known to activate, are significantly up-regulated. Additionally, there was only one probe set for *MMP-2* present on the chip, to which only one of the 7 samples hybridized; therefore, we conclude that this probeset was not adequate to give a true representation of the status of *MMP-2* in the affected NAV dog. Additionally *MMP-2* mRNA expression as assessed by semiquantitative RT-PCR in another study in these dogs showed an average two-fold increase in affected dogs as compared to normal sibs (Rao et al. 2003), while *MMP-9*, which was significantly up-regulated on our chip, showed an average increase of 19-fold (Rao et al. 2003). Also, in some models of glomerulonephritis, *MMP-2* has been found to be of much less importance in the development of glomerular injury than *MMP-9* (Urushihara et al. 2002). Also, in a murine model of ARAS, *MMP-9*, but not *MMP-2*, was up-regulated (Miner and Sanes 1996). Obviously, the inconclusive evidence regarding these genes in the NAV model warrants further investigation. It is likely that by investigating changes in the XLAS kidney at stages prior to ESRD (by microarray and Q-RT-PCR) will provide more conclusive data on the roles of these genes in the disease process.

CHAPTER VI

DIGGING UP THE CANINE GENOME---A TALE TO WAG ABOUT*

Overview

There is incredible morphological and behavioral diversity among the hundreds of breeds of the domestic dog, *Canis familiaris*. Many of these breeds have come into existence within the last few hundred years. While there are obvious phenotypic differences among breeds, there is marked interbreed genetic homogeneity. Thus, study of canine genetics and genomics is of importance to comparative genomics, evolutionary biology and study of human hereditary diseases. The most recent version of the map of the canine genome is comprised of 3,270 markers mapped to 3,021 unique positions with an average intermarker distance of ~1 Mb. The markers include approximately 1,600 microsatellite markers, about 1,000 gene based markers, and almost 700 bacterial artificial chromosome-end markers. Importantly, integration of radiation hybrid and linkage maps has greatly enhanced the utility of the map. Additionally, mapping the genome has led directly to characterization of microsatellite markers ideal for whole genome linkage scans. Thus, workers are now able to exploit the canine genome for a wide variety of genetic studies. Finally, the decision to sequence the canine genome highlights the dog's evolutionary and physiologic position between man and mouse and

---

*Reprinted with permission from Greer KG, Cargill EJ, Clark LA, Cox ML, Tsai KL et al. (2003) Digging up the canine genome – A tale to wag about. Cytogen Genome Res (in press)

its importance as a model for the study of mammalian genetics and human hereditary diseases.

Introduction

The domestic dog has occupied a unique position in human lives for many centuries, serving many roles, including: warrior, shepherd, guide, retriever, hunter, and companion. Clearly, the dog has many notable behavioral and physical characteristics that have been valued over thousands of years. However, the greatest benefit of the dog may yet to be realized: its contribution to genetics.

Recent data suggest the domestic dog's origin was in East Asia (Savolainen et al. 2002). Specifically, the common origin of New and Old World dogs from gray wolves is demonstrated by mtDNA evidence dating the event to approximately 15,000 to 40,000 years ago (Leonard et al. 2002). Since its domestication, specific breeds have been developed with distinctly different appearances and behavioral traits. The rapid origin of breeds has been based on inbreeding a genetic pool composed of a limited founder number (Ostrander and Kruglyak 2000). However, the price of limited genetic heterogeneity has been the emergence of more than 450 naturally occurring hereditary diseases. The small founding populations, high levels of inbreeding and creation/maintenance of multigenerational pedigrees have resulted in a population ideal for genetic studies because often only a single mutated allele is responsible for these diseases (Ostrander and Kruglyak 2000). Furthermore, almost half of the hereditary

diseases of the dog have an analogous disease in the human (OMIA 2003). Thus, the dog serves as a natural model for study of these diseases and obviates the need to construct knockout models for study of such diseases.

Current Status of the Canine Genome

Mammalian genome comparisons

Information regarding the canine genome is significantly less than that for the murine and human genomes. Determination of the sequences of the human and murine genomes confirmed syntenic conservation of genetic loci by pairwise alignment of nearly 13,000 orthologous gene pairs. However, the median amino acid sequence identity of mouse to human is only 78.5% (Boguski 2002). Thus, human and murine comparisons are insufficient to complete our understanding of mammalian gene function and evolution. The Mouse Genome Consortium recently noted that many additional mammalian genome sequences would be necessary to fully decipher information from the human genome sequence. This acknowledgement results from the recognition that gene family changes between the murine and human genomes may represent physiological divergence from the common ancestor of the mouse and human (Waterston 2002). Therefore, because the dog is more closely related, evolutionarily and physiologically, to the human than is the mouse, it is an ideal model for comparative genetics and will complement information gleaned from sequences of other genomes.

Mapping the canine genome

The dog has 38 autosomes, plus the sex chromosomes. Unfortunately, acrocentricity and similarity in the sizes of autosomes complicated standardization of the canine karyotype. An international collaboration, DogMap, was established in 1993 to create a low-resolution marker map, and to standardize the karyotype. In 1997, sixteen linkage groups were derived from the analysis of 94 polymorphic loci (Lingaas et al. 1997). Shortly thereafter, characterization of 150 microsatellite markers and 30 additional linkage groups provided the first meiotic linkage map (Mellersh et al. 1997), and a framework for development of the 10-centimorgan (cM) map (Neff et al. 1999). A map of 341 markers distributed at an average of 9 cM soon followed (Werner et al. 1999).

Distinguishing the canine chromosomes with the use of whole chromosome-specific fluorescent *in situ* hybridization paint probes (Breen et al. 1999), the construction of canine bacterial artificial chromosome (BAC) libraries (Li et al. 1999; Schelling et al. 2002), and construction of a radiation hybrid (RH) panel accelerated development of a high-resolution map of the canine genome. An initial RH panel was constructed (Priatt et al. 1998), and updated by a whole genome RH panel (Vignaux et al. 1999). Mellersh and colleagues (2000) integrated the linkage and RH maps, a major advance for the study of the canine genome. Maps were integrated through duplicate typing of 217 markers, generating an integrated map of 724 unique markers suitable for both linkage analysis and comparative mapping. Importantly, integration allowed for identification of markers useful in whole genome linkage scans. This subset of markers, termed the Minimal Screening Set 1 (MSS-1) (Richman et al. 2001), has been

multiplexed to streamline linkage analyses (Cargill et al. 2002). Estimates of genome coverage suggest that 77% of the genome is within 1 Mb of at least one marker in the MSS-1 (Richman et al. 2001). The most recent RH panel, RHDF5000-2, was used to map 3,270 markers to 3,021 unique positions with an average intermarker distance corresponding to ~1 Mb (Guyon et al. 2003). The 1 Mb RH map facilitated characterization of a more comprehensive screening set, the MSS-2, composed of 325 microsatellite markers with an average spacing of 9 Mb (Guyon et al. 2003).

Sequencing the canine genome

The importance of the dog as a model for many human hereditary diseases is a major impetus for sequencing the canine genome. Celera (Rockville, MD) conducted the first large scale canine sequencing effort with DNA from a male Standard Poodle. Access to the sequence (coverage is 1X) is available through collaborations with The Institute for Genome Research. Work using this resource is currently underway and one aspect of this work is directed towards mapping independent gene sequences (Ostrander et al. 2002). This 1X sequence has considerable gaps and this was a factor in the decision by the National Institutes of Health to designate sequencing of the canine genome as a high priority (http://www.nih.gov/news/pr/sep2002/nhgri-12.htm). Sequencing began in June 2003 using DNA isolated from a female Boxer (http://www.genome.gov/11007358) and is being done at the Whitehead Genome Center in a manner similar to that utilized in sequencing of the murine genome. By utilizing different library sizes cloning bias will be minimized, thereby allowing a hierarchical approach for sequence assembly. This method will provide approximately 6-fold

sequence coverage and 50-fold physical coverage. Other sequencing efforts include end-sequencing of clones retrieved from BAC libraries. More specifically, sequences have been obtained from approximately 1,500 BAC clones and 668 of these are mapped, yielding an initial framework for future mapping (Guyon et al. 2003).

A second valuable resource for study of the canine genome is the development of a single nucleotide polymorphism (SNP) map. Recent results indicate that interbreed sequence comparisons are a reasonable strategy for identifying useful SNPs within many breeds (Brouillette and Venta 2002). Even so, it must be noted that a SNP map which will facilitate high throughput screening of the genome (10 Kb spacing) will require the identification of 250,000 SNPs (Ostrander et al. 2002). Additionally, efforts are being made to construct normalized or subtracted cDNA libraries, and to characterize expressed sequence tags (ESTs). McCombie and others (Table 8) maintain an interactive website for access to current canine EST projects (http://www.cshl.org/genbin/cgi-bin/golden_retriever.cgi), listing several canine cDNA resources. Lastly, the canine genome sequencing efforts will be complemented by expression profiling in canine models of human diseases and such work is underway through various collaborations (http://crisp.cit.nih.gov/crisp).

Canine diseases and genetic analyses

As progress is made in constructing maps and sequencing of the canine genome, more mutation based and linkage based genetic tests will be developed to allow detection of deleterious alleles. While mutation based tests allow guaranteed results

because the specific mutation has been defined, linkage based tests do not provide such certainty due to recombination events that may cause false negative and positive results.

Table 8. The canine genome and current resources for mapping.

Chromosome number (2N)	78
Genome size	2.8 x 10^9 bp
Meiotic Linkage Map	
Microsatellites	354
[1]Radiation Hybrid Map	
Microsatellites	1596
gene sequences + ESTs	900
BAC end sequences	668
Sequence tag sites (STS)	106
[2]Integrated Linkage and RH Map	
Common to cytogenetic and RH	102
Common to RH and meiotic linkage	251
Common to cytogenetic and meiotic linkage	52
BAC library	8.1-fold coverage
[1]BAC end sequencing	1,504
[1]BAC end sequence containing microsatellites	39
[3]ESTs	19465
[2]SNPs	78
[2]STS	200

[1]Numbers reflect those reported in 2003 (Guyon et al.), and discovery of SNPs, STSs, and BAC end sequences continues in numerous laboratories.
[2]As reported by FHCRC, http://www.fhcrc.org/science/dog_genome/dog.html.
[3]EST development is ongoing and referenced number is current as of submission date, with R. McCombie and A. George having submitted the majority of EST sequences to GenBank.

Candidate gene analyses

The candidate gene approach uses analysis of pedigrees and phenotypes together with comparisons of analogous diseases in other species to identify candidate genes that may be responsible for a given disease. An excellent example of this is the identification of the gene causative for autosomal dominant progressive retinal atrophy in the English

Mastiff. After determining the mode of inheritance by controlled outcross matings, Kijas and coworkers (2003) selected three genes responsible for approximately half of all human autosomal dominant retinal disease (http://www.sph.uth.tmc.edu/Retnet) for evaluation in the dog. Their analysis revealed that the rhodopsin gene contained a transversion altering the protein's amino acid sequence. Association testing followed, and confirmed that the rhodopsin mutation was responsible for this form of progressive retinal atrophy in the English Mastiff.

Another recent candidate gene investigation revealed the mutation responsible for XLAS in a group of mixed-breed dogs. Alport syndrome is a chronic, progressive glomerulonephritis characterized by ultrastructural changes in the glomerular basement membrane. The most common form of AS in the human is XLAS and is due to mutations in *COL4A5* (Jais et al. 2000). A colony of mixed-breed dogs segregating naturally occurring XLAS was studied and found to exhibit clinical, immunohistological, and ultrastructural characteristics virtually identical to that seen in human XLAS (Lees et al. 1999). Therefore, full-length *COL4A5* cDNAs from normal and affected dogs were sequenced. This work revealed a 10 bp deletion in exon 9, resulting in a frameshift which disrupts the collagenous region of the protein, causing a premature stop codon within exon 10. This study also led to development of a mutation based genetic test for XLAS (Cox et al. 2003b).

Other candidate gene analyses include the study of dilated cardiomyopathy (DCM) in the Doberman Pinscher. DCM provides an example of the difficulties geneticists sometimes face in studying canine hereditary diseases. That is, DCM is a

lethal, late onset disease (average age of onset is approximately 6 years) and this complicates obtaining DNA samples for linkage analysis. The possibility of phenocopies adds to the difficulties in analysis. Furthermore, the mode of inheritance is uncertain although the likely presence of a founder effect suggests a major gene is necessary for a dog to be affected. Several DCM-causing genes have been identified in human families and are candidate genes for canine DCM. In the human, every gene identified to date accounts for a small proportion of heritable DCM, suggesting that any number of genes could be responsible for canine DCM (Schönberger and Seidman 2001). Progress is being made by eliminating many of these candidate genes as causative (e.g., Meurs et al. 2001; Venta et al. unpublished results), but new approaches to identify genes for complex diseases are needed.

Linkage analyses

Linkage studies rely upon informative pedigrees that have affected and non-affected individuals available for sampling. Generally, such extended pedigrees are not available or accessible for human studies, but are common in the dog. Another advantage of using the dog in genetic studies is that the problem of locus heterogeneity inherent to human linkage studies is reduced. This is due to founder effects and minimal genetic diversity in breeds that have arisen from small populations.

Some conditions (e.g., deafness) are not amenable to the candidate gene approach since too many candidates exist. Thus, pedigree assembly for linkage analysis is necessary. While it is known that a relationship exists between hereditary hearing loss and pigmentation in the Dalmatian, the nature of this association is unclear. Even so,

candidate gene studies have eliminated a few genes such as *PAX3* (Brenig et al. 2003), but the number of remaining candidate genes is quite large. Therefore, linkage analysis is being performed by Muhle et al. (2002) who have assembled a kindred of Swiss Dalmatians and Cargill et al. (personal communication), who have assembled a kindred of US Dalmatians.

The utility of linkage analysis is exemplified by the study of hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis (RCND), a rare cancer in the German Shepherd Dog. RCND is a spontaneous, inherited disease with several human correlates, albeit none that match all facets of the disease exactly. Jonasdottir and colleagues (2000) propagated a colony segregating RCND and used the MSS-1 screening panel for linkage analysis. One microsatellite marker yielded a lod score of 16.7 and the RCND locus was mapped to canine chromosome 5. Within the targeted region are several attractive candidate genes and future work will likely identify the gene responsible for RCND.

Linkage analysis is also being used to identify loci contributing to canine hip dysplasia (CHD). CHD is a degenerative disease, primarily in large breeds, that results in part, from increased joint laxity. Functional subluxation of the hip, poor hip joint congruity, and debilitating secondary osteoarthritis are hallmarks of CHD. Although the mode of inheritance is unknown, recent investigations offer heritability estimates ranging from 0.11 to 0.68 (Bliss et al. 2002). An outcrossed pedigree of Labrador Retrievers and Greyhounds has been constructed for use in linkage studies of CHD and is currently being analyzed by a genome-wide screen (Todhunter et al. 2003).

Linkage disequilibrium

Linkage disequilibrium (LD) mapping relies upon population based gene and marker frequencies considered to be in disequilibrium if the marker and allele do not assort independently within the population. In LD, the marker and allele segregate together at a greater frequency than would be expected by their individual frequencies. This type of association would be expected for genes and markers in a population with common ancestral origin. Many breeds are derived from very small founding populations and gene flow between breeds is restricted by registration requirements. Thus, LD is a valuable tool for identification of genes responsible for hereditary diseases and other traits of interest.

The dog is particularly well suited for LD analysis. For example, Ostrander and Kruglyak (2000) presented evidence for the utility of LD through analysis of cancers in the Golden Retriever. The estimated risk of malignancy for the Golden Retriever is significantly higher than that for other breeds (Priester and McKay 1980). If the genes for malignancy originated in a popular sire, LD would be expected over large distances (Ostrander et al. 2000). This phenomenon is not restricted to cancers or a specific breed but rather applies to other diseases with hereditary components and to those breeds for which founding history, litter sizes, etc. are well documented. For those populations in which such information can be gathered, it is estimated that LD will extend 5-10 cM from loci causative for diseases (Ostrander et al. 2001). Furthermore, calculations

estimate that screening of 10 affected individuals of a given breed has adequate power for detection of regions with genetic homogeneity through identity by descent, as would be expected within a single line of dogs. If, however, dogs from various breeding lines are needed, it is estimated that only 40 dogs would be necessary for screening based upon conditions of moderate heterogeneity (Houwen et al. 1994). In this case, it would be reasonable to assume dogs within a breed carry the same mutation due to a common ancestral origin; therefore, the overlap of genetic regions revealed by LD would narrow the genomic region containing the causative gene (Ostrander et al. 2000).

A current study using LD concerns canine sebaceous adenitis (SA), an inflammatory disease of the skin characterized by hair loss and unpleasant odor, making affected animals far from desirable pets. SA is archetypical of a number of heritable canine dermatological diseases in that 1) there are a few breeds with high prevalence but many breeds have sporadic occurrence, 2) there often is an inability to definitively call a dog "normal" based on phenotype, 3) onset often occurs during maturity and 4) there are no plausible candidate genes. These findings preclude a candidate gene approach or traditional linkage studies. SA has been recognized in over 55 breeds and mongrels but the Standard Poodle and Akita have the highest prevalences. Analysis of both breeds suggests SA is an autosomal recessive trait. However, there are more normal dogs than expected for an autosomal recessive disease due to subclinically affected dogs and dogs with late-onset disease (Rosser et al. 1987; Dunstan and Hargis 1995; Reichler et al. 2001; Scott et al. 2002). The Akita (rather than the Standard Poodle) is currently being studied by several groups pursuing association studies to define LD because SA in this

breed fits the criteria established in the human for using association studies (Kruglyak 1997; Houwen et al. 1994; Elston 1998; De La Chapelle and Wright 1998).

Future of Canine Genomics

Clearly there has been an explosion in knowledge of canine genetics and the canine genome. This new information is augmenting studies of human hereditary diseases, comparative genomics and clinical veterinary medicine, yet the full impact of the new information remains to be seen. The primary benefits to be derived from study of canine genetics are 1) the addition of comparative genomic information regarding a species evolutionarily and physiologically closely related to the human, 2) a better understanding of canine hereditary diseases and subsequent elimination of many disease alleles from breeding populations and, 3) insight into analogous human hereditary diseases and implementation of therapies, including gene therapy protocols. The importance of the latter is illustrated by success in gene therapy for two different diseases. Firstly, gene therapy prevents clinical manifestations of a lysosomal storage disease in a canine model of the disease (Ponder et al. 2002). Secondly, a form of early childhood blindness, Leber congenital amaurosis, has been corrected by gene therapy in a spontaneous canine model of the disease (Acland et al. 2001). These examples and others are indicative of the power of the dog as a large animal model of human hereditary diseases. The similarity between orthologous canine and human genes ensures that analogous human disease genes will be identified. In conclusion, research in canine genetics and genomics is benefiting the human and the dog, our companion for many centuries.

CHAPTER VII

SUMMARY

Our laboratory studies the dog because of its intrinsic value to humans. While the direction of the laboratory has evolved and matured over the last four years, the goal is the same: we want to improve the health and quality of life for all dogs through increased understanding of canine genetics. The specific aims of our work are to 1) understand the genetics underlying hereditary diseases of the ear, kidney, pancreas and musculoskeleton, 2) contribute to increasing the density of the canine map by characterizing and mapping gene loci that are members of gene families and 3) to develop tools for more efficient analysis of the canine genome.

Success in addressing these objectives is shown by the facts that our laboratory has identified the mutation causative for canine XLAS, the first discovery at Texas A&M University of a gene responsible for a disease of a companion animal (Cox et al. 2003b), mapped many gene loci (Miller et al. 1999; Miller et al. 2001; Cox et al. 2003a; Lowe et al. 2003; Tsai et al. 2004) and improved our ability to carry out genome-wide analyses of the dog (Yang et al. 1999; Cargill et al. 2002; Clark et al. 2004), and been involved in the efforts to sequence the canine genome (Ostrander et al. 2002).

The impetus for work reported here is the diagnosis of AS in a family of mixed breed dogs (Lees et al. 1999). Renal failure is common in the dog and is a major cause of death. Therefore, we are interested in understanding the genetics and pathophysiology of this disease in order to develop better diagnostic and treatment

regimens. Importantly, regardless of etiology, renal diseases reach a certain common point in the disease process (secondary renal disease). Thus, the NAV model of XLAS provides insight into ESRD in general. That is, due to the reserve capacity of the kidney, most dogs with ESRD are not diagnosed until very late in the disease process (whatever its primary cause), when the majority of the renal capacity has been lost and damage to the kidney is irreversible. Our colony is an excellent model in which to follow the progression of ESRD from the very beginnings of disease. This being so, future studies in the laboratory will include a temporal study of global gene expression by microarray analysis and Q-T-PCR at various stages of the lives of NAV dogs. It is hoped that this will allow for the identification of potential biomarkers and targets for therapeutic intervention.

Chapter II describes the mapping of the type I and IV collagens. This increases the density of the map of the canine genome, and provides information pertaining to synteny and the conservation of sequence through evolution. *COL1A1* maps to CFA9, and *COL1A2* maps to CFA14, which match locations with their human orthologs. The type IV collagens, as expected, form 3 separate linkage groups. *COL4A1* and *COL4A2* map to CFA8, *COL4A3* and *COL4A4* map to CFA25, and *COL4A5* and *COL4A6* form their own linkage group with two microsattelite markers mapped to CFAX. These results confirm the evolutionary relationship of type IV collagens in other species. It is believed that *COL4A1* was the original type IV collagen, which by a duplication and inversion event, spawned *COL4A2*, in a head-to-head position. The other four type IV collagens were duplications of *COL4A1* and *COL4A2*. *COL4A1*, *COL4A3*, and

COL4A5 are most similar to each other, and *COL4A2*, *COL4A4*, and *COL4A6* are most similar.

Chapter III reports the fortuitous amplification of a cDNA encoding the renal-specific glycoprotein uromodulin. The *umod* gene is linked to *CZP2* in an RH group not yet assigned to a chromosome, but is likely CFA6. This conclusion is based on the synteny between CFA6 and HSA16, to which these genes have been mapped in the human. The function of *umod* is unknown, but sequencing of it is timely, given the recent discovery that mutations in the gene cause two hereditary diseases in the human.

Chapter IV reports the main focus of my doctoral work. The clinical study of XLAS in NAV dogs has taken place over a number of years, but the sequencing of *COL4A5*, and identification of the mutation causative for the disease was necessary for the use of this colony as a true disease model for human XLAS. The full-length sequence was generated in 13 overlapping fragments, and a 10 bp deletion was found in exon 9, resulting in a frameshift and premature stop codon in exon 10. The severe clinical phenotype of the NAV dog is partially explained by the nature of this mutation (Gross et al. 2003). After sequencing and identification of a mutation, an allele-specific genetic test was also developed to allow easy identification of carrier females and of dogs destined to develop XLAS.

There is no satisfactory treatment for human or canine XLAS and the only alternatives are dialysis and/or renal transplantation. Thus, our laboratory and others are exploring the possibility of gene transfer therapy as a treatment of AS. Introduction (and subsequent expression in the kidney) of the wild-type coding sequence to compensate for a defective type IV collagen might allow the GBM to form properly

thereby stabilizing or restoring GBM function. Theoretically, if gene transfer therapy were applied early in the disease process, progressive renal damage and failure might be prevented entirely. However, a prerequisite to trials in human patients is the development of the technology and demonstration of the safety and efficacy of gene transfer therapy in a suitable animal model (e.g., the dog) of AS. Because of this, part of the future focus of work in our laboratory is on the generation of a full-length clone of *COL4A5* for gene therapy studies, and the testing of this clone in the NAV dog as a proof-of-principle experiment for the development of the therapy for human use.

The differential gene expression of kidneys from normal and affected XLAS NAV dogs is the focus of Chapter V. Four affected dogs were used to assess gene expression in the end stage XLAS kidney as compared to three normal siblings. Very little is known about the global gene expression differences at end-stage, and probing of canine-specific microarray revealed approximately 1,200 differentially expressed genes. Trends of a subset of clinically interesting genes were confirmed by Q-RT-PCR. As expected, many genes playing important roles in maintenance and functioning of the extracellular matrix, cell structure, immune system, and signaling pathways were up-regulated in XLAS dogs. Unfortunately, the type IV collagens did not hybridize well on the chip, and were therefore scrutinized by Q-RT-PCR as well. *COL4A1* and *COL4A2* were found to be moderately up-regulated (but not as much as *COL1A1* and *COL1A2*), while the rest of the type IV collagens showed no statistically significant difference between normal and XLAS dogs. This finding is in agreement with the literature for human XLAS but contradicts a study on the Samoyed model of XLAS, in which a down-regulation in *COL4A3*, *COL4A4*, and *COL4A5* is reported

(Thorner et al. 1996) leading the authors to conclude that these genes may be co-regulated pre-transcriptionally. We see no evidence of such a finding or mechanism in our dogs.

Chapter VI is a comprehensive review of the current status of canine genetics and genomics. It highlights recent advances in the integration of a RH and linkage maps, allowing the development of a new microsatellite marker set, the MSS-2, for whole genome screens. The decision by the NIH to fund the sequencing of the dog demonstrates the importance of canine genetics to human medicine. Equally important is that the sequence and the emerging comprehensive SNP map will eventually foster development of screening and testing for various hereditary diseases of the dog.

In conclusion, this work adds to knowledge of canine hereditary diseases and canine genomics by providing the foundation for gene therapy for AS. Also, the NAV dog is an excellent model for study of end stage renal failure in the dog and human.

REFERENCES

1) Abbate M, Zoja C, Morigi M, Rottoli D, Angioletti S et al. (2002) Transforming growth factor-β1 is up-regulated by podocytes in response to excess intraglomerular passage of proteins. Am J Pathol 161(6), 2179-2193

2) Acland GM, Aguirre GD, Ray J, Zhang Q, Aleman TS et al. (2001) Gene therapy restores vision in a canine model of childhood blindness. Nat Genet 28, 92-95

3) Alport AC (1926) Hereditary familial congenital haemorrhagic nephritis. Br Med J 1, 504-506

4) Barker DF, Hostikka SL, Zhou J, Chow LT, Oliphant AR et al. (1990) Identification of mutations in the *COL4A5* collagen gene in Alport syndrome. Science 248, 1224-1226

5) Bartlett RJ, Winand NJ, Secore SL, Singer JT, Fletcher S et al. (1996) Mutation segregation and rapid carrier identification of X-linked muscular dystrophy in dogs. Am J Vet Res 57(5), 650-654

6) Bernard MA, Valli VE (1977) Familial renal disease in Samoyed dogs. Can Vet J 18, 181-189

7) Bliss S, Todhunter RH, Quaas R, Casella G, Wu R et al. (2002) Quantitative genetics of traits associated with hip dysplasia in a canine pedigree constructed by mating dysplastic Labrador Retrievers with unaffected Greyhounds. Am J Vet Res 63(7), 1029-1035

8) Boffa J-J, Tharaux P-L, Dussaule J-C, Chatziantoniou C. (2001) Regression of renal vascular fibrosis by endothelin receptor antagonism. Hypertension 37(2), 490-502

9) Boguski MS (2002) The mouse that roared. Nature 420, 515-516

10) Boye E, Mollet G, Forestier L, Cohen-Solal L, Heidet L et al. (1998) Determination of the genomic structure of the *COL4A4* gene and of novel mutations causing autosomal recessive Alport syndrome. Am J Hum Genet 63, 1329-1340

11) Breen M, Bullerdiek J, Langford CF (1999a) The DAPI banded karyotype of the domestic dog (Canis familiaris) generated using chromosome-specific paint probes. Chromosome Res 7(5), 401-6

12) Breen M, Thomas R, Binns MM, Carter NP, Langford CF (1999b) Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (*Canis familiaris*) and human. Genomics 61, 145-155

13) Breen M, Jouquand S, Renier C, Mellersh CS, Hitte et al. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes, Genome Res 11(10), 1784-95

14) Brenig B, Pfeiffer I, Jaggy A, Kathmann I, Balzari M et al. (2003) Analysis of the 5' region of the canine PAX3 gene and exclusion as a candidate for Dalmatian deafness. Animal Genet 34, 47-50 (2003)

15) Brouillette JA, Venta PJ (2002) Within-breed heterozygosity of canine single nucleotide polymorphisms identified by across-breed comparison. Animal Genet 33, 464-467

16) Brown SA, Finco DR, Brown CA, Crowell WA, Alva R (2003) Evaluation of the effects of inhibition of angiotensin convertin enzyme with enalapril in dogs with induced chronic renal insufficiency. Am J Vet Res 64(3), 321-7

17) Campbell BG, Wootton JA, MacLeod JN, Minor RR (1998) Sequence of canine *COL1A2* cDNA: Nucleotide substitutions affecting the cyanogen bromide peptide map of the alpha 2(I) chain. Arch Biochem Biophys 357, 67-75

18) Cargill EJ, Clark LA, Steiner JM, Murphy KE (2002) Multiplexing of canine microsatellite markers for whole-genome screens. Genomics 80, 250-253

19) Cavallone D, Malagolini N, Serafini-Cessi F (1999) Binding of human neutrophils to cell-surface anchored Tamm-Horsfall glycoprotein in tubulointerstitial nephritis. Kidney Int 55, 1987-99

20) Ciccarese M, Casu D, Ki Wong F, Faedda R, Arvidsson S et al. (2001) Identification of a new mutation in the alpha4(IV) collagen gene in a family with autosomal dominant Alport syndrome and hypercholesterolaemia. Nephrol Dial Transplant 16(10), 2008-12

21) Clark LA, Famula TR, Murphy KE (2004) A rapid single multiplex microsatellite-based assay for use in canine forensics. Am J Vet Res (in press)

22) Cohen MM, Cassady G, Hanna BL (1961) A genetic study of hereditary renal dysfunction with associated nerve deafness. Am J Hum Genet 13, 379-389

23) Colville DJ, Savige J (1997) Alport syndrome: A review of the ocular manifestations. Ophthalmic Genet 18(4) 161-173

24) Cosgrove D, Meehan DT, Grunkemeyer JA, Kornak JM, Sayers et al. (1996) Collagen *COL4A3* knockout: A mouse model for autosomal Alport syndrome. Genes Dev 10, 2981-2992

25) Cox ML, Quignon P, Galibert F, Lees GE, Murphy KM (2003a) Sequencing and radiation hybrid mapping of canine uromodulin. DNA Sequence 14(1), 61-69

26) Cox ML, Lees GE, Kashtan, Murphy KE (2003b) Genetic Cause of X-linked Alport Syndrome in a family of domestic dogs. Mamm Genome 14, 396-403

27) Crawfurd MD, Toghill PJ (1968) Alport's syndrome of hereditary nephritis and deafness. Q J Med 37(148), 563-576

28) De La Chapelle A, Wright FA (1998) Linkage disequilibrium mapping in isolated populations: The example of Finland revisited. Proc Natl Acad Sci, USA 95, 12416-12423

29) Dulawa J, Jann K, Thomsen M, Rambausek M, Ritz E (1988) Tamm Horsfall glycoprotein interferes with bacterial adherence to human kidney cells. Eur J Clin Invest 18(1), 87-91

30) Dunstan RW, Hargis AM (1995) The diagnosis of sebaceous adenitis in standard poodle dogs. In. *Current Veterinary Therapy XII*. Kirk RW (ed.) WB Saunders Co, Philadelphia, PA, pp.619-622

31) Elston RC (1998) Linkage and association. Genet Epidem 15, 565-576

32) Flinter FA, Abbs S, Bobrow M (1989) Localization of the gene for classic Alport syndrome. Genomics 4(3), 335-338

33) Freeman LJ, Hegreberg GA, Robinette JD (1987) Ehlers-Danlos syndrome in dogs and cats. Semin Vet Med Surg (Small Anim) 2, 221-227

34) Fukuoka S-I, Matsuda Y (1997) Assignment of the Tamm-Horsfall protein/uromodulin gene (*Umod*) to mouse chromosome bands 7F1-F2 and rat chromosome bands 1q36-q37 by *in situ* hybridization. Cytogenet Cell Genet 79, 241-242

35) Goumenos DS, Tsamandas AC, Oldroyd S, Sotsiou F, Tsakas S (2001) Transforming growth factor-β1 and myofibroblasts: A potential pathway towards renal scarring in human glomerular disease. Nephron 87, 240-248

36) Gross O, Netzer K-O, Lambrecht R, Seibold S, Weber M (2002) Meta-analysis of genotype-phenotype correlation in X-linked Alport syndrome: Impact on clinical counseling. Nephrol Dial Transplant 17, 1218-1227

37) Gunwar S, Ballester F, Noelken ME, Sado Y, Ninomiya Y et al. (1998) Glomerular basement membrane; identification of a novel disulfide-cross-linked network of α3, α4, and α5 chains of type IV collagen and its implications for the pathogenesis of Alport syndrome. J Biol Chem 273, 8767-8775

38) Guo C, Van Damme B, Van Damme-Lombaerts R, Van de Berghe H, Cassiman JJ et al. (1993) Differential splicing of COL4A5 mRNA in kidney and white blood cells: Complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain. Kidney Int 44, 1316-1321

39) Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E et al. (2003) A 1-Mb resolution radiation hybrid map of the canine genome. Proc Natl Acad Sci, USA 100, 5296-5301

40) Hao L, Okada H, Kano Y, Inoue T, Kobayashi et al. (2003) Direct contact between human peripheral blood mononuclear cells and renal fibroblasts facilitates the expression of monocyte chemattractant protein-1. Am J Nephrol 23(4), 208-213

41) Hart TC, Gorry MC, Hart PS, Woodard AS, Shihabi Z et al. (2002) Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. J Med Genet. 39(12), 882-92

42) Harvey SJ, Zheng K, Sado Y, Naito I, Ninomiya Y et al. (1998) Role of distinct type IV collagen networks in glomerular development and function. Kidney Int 54, 1857-1866

43) Hayden PS, El-Meanawy A, Schelling JR, Sedor JR (2003) DNA expression analysis: Serial analysis of gene expression, microarrays and kidney disease. Curr Opin Nephrol Hypertens 12(4), 407-414

44) Heidet L, Cohen-Solan L, Boye E, Thorner P, Kemper MJ et al. (1997) Novel COL4A5/COL4A6 deletions and further characterication of the diffuse leiomyomatosks-Alport syndrome (DL-AS) locus define the DL critical region. Ctogenet Cell Genet 78(3), 240-246

45) Heidet L, Yi C, Guicharnaud L, Antignac C, Gubler M-C (2000) Glomerular expression of type IV collagen chains in normal and X-linked Alport syndrome kidneys. Am J Pathol 156, 1901-1910

46) Heidet L, Arrondel C, Forestier L, Cohen-Solal L, Mollet G et al. (2001) Structure of the human type IV collagen gene *COL4A3* and mutations in autosomal Alport syndrome. J Am Soc Nephrol 12, 97-106

47) Higgins MA, Berridge BR, Mills BJ, Schultze AE, Gao H et al. (2003) Gene expression analysis of the acute phase response using a canine microarray. Tox Sci 74(2), 470-484

48) Hood JC, Savige J, Hendtlass A, Kleppel MM, Huxtable CR et al. (1995) Bull terrier hereditary nephritis: A model for autosomal dominant Alport syndrome. Kidney Int 47, 758-765

49) Hostikka SL, Eddy RL, Byers MG, Höyhtyä M, Shows TB et al. (1990) Identification of a distinct type IV collagen α chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. Proc Natl Acad Sci 87, 1606-1610

50) Houwen RH, Baharloo S, Blankenship K, Raeymaekers P, Juyn J et al. (1994) Genome screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8, 480-486

51) Hsiao L-L, Stears RL, Hong RL, Gullans SR (2000) Prospective use of DNA microarrays for evaluating renal function and disease. Current Op Nephrol Hypert 9, 253-258

52) Hudson BG, Reeders ST, Tryggvason K (1993) Type IV collagen: Structure, gene organization, and role in human diseases. J Biol Chem 268, 26033-26036

53) Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG (2003) Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med 348, 2543-2556

54) Huerre C, Junien C, Weil D, Chu ML, Morabito M et al. (1982) Human type I procollagen genes are located on different chromosomes. Proc Natl Acad Sci USA 79, 6627-6630

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

56) Ikezumi Y, Hurst, LA, Masaki T, Atkins RC, Nikolic-Paterson DJ (2003) Adoptive transfer studies demonstrate that macrophages can induce proteinuria and mesangal cell proliferation. Kidney Int 63, 83-95.

57) Inoue Y, Nishio H, Shirakawa T, Nakanishi K, Nakamura H et al. (1999) Detection of mutations in the *COL4A5* gene in over 90% of male patients with X-linked Alport's syndrome by RT-PCR and direct sequencing. Am J Kid Dis 34, 854-862

58) Jais JP, Knebelmann B, Giatras I, de Marchi M, Rizzoni G et al. (2000) X-linked Alport syndrome: Natural history in 195 families and genotype-phenotype correlations in males. J Am Soc Nephrol 11, 649-657

59) Jansen B, Thorner PS, Singh A, Patterson JM, Lumsden JH et al. (1984) Hereditary nephritis in Samoyed dogs. Am J Pathol 116, 175-178

60) Jeanpierre C, Whitmore SA, Austruy E, Cohen-Salmon M, Callen DF et al. (1993) Chromosomal assignment of the uromodulin gene (UMOD) to 16p13.11. Cytogenet Cell Genet 62(4), 185-187

61) Jonasdottir TJ, Mellersh CS, Moe L, Heggebo R, Gamlem H et al. (2000) Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. Proc Natl Acad Sci USA 97, 4132-4137

62) Kalluri R, Shield CF, Todd P, Hudson BG, Neilson EG (1997) Isoform switching of type IV collagen is developmentally arrested in X-linked Alport syndrome leading to increased susceptibility of renal basement membranes to endoproteolysis. J Clin Invest 99, 2470-2478

63) Kashtan CE (1998) Disease of the month: Alport syndrome and thin glomerular basement membrane disease. J Am Soc Nephrol 9, 1736-1750

64) Kashtan CE (2002) Animal models of Alport syndrome. Nephrol Dial Transplant 17, 1359-1361

65) Kijas JW, Miller BJ, Pearce-Kelling SE, Aguirre GD, Acland GM (2003) Canine models of ocular disease: Outcross breedings define a dominant disorder present in the English mastiff and Bull mastiff dog breeds. Journal Hered 94, 27-30

66) Knapp DW, Water DJ (1997) Naturally occuring cancer in pet dogs: Important model for developing improved cancer therapy for humans. Mol Med Today 3(1), 8-11

67) Knebelmann B, Breillat C, Forestier L, Arrondel C, Jacassier D et al. (1996) Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. Am J Hum Genet 59, 1221-1232

68) Kokot F, Dulawa J (2000) Tamm-Horsfall protein updated. Nephron 85(2), 97-102

69) Kruglyak L (1997) What is significance in whole-genome linkage disequilibrium studies? Am J Hum Genet 61, 810-812

70) Lees GE, Helman RG, Kashtan CE, Michael AF, Homco LD et al. (1998) A model of autosomal recessive Alport syndrome in English cocker spaniel dogs. Kidney Int 54, 706-719

71) Lees GE, Helman RG, Kashtan CE, Michael AF, Homco LD et al. (1999) New form of X-linked dominant hereditary nephritis in dogs. Am J Vet Res 60, 373-383

72) Lemmink HH, Mochizuki T, van den Heuvel LPWJ, Schroder CH, Barrientos A et al. (1994) Mutations in the type IV collagen α3 (*COL4A3*) gene in autosomal recessive Alport syndrome. Hum Mol Genet 3, 1269-1273

73) Lemmink HH, Nillesen WN, Mochizuki T, Schroder CH, Brunner HG et al. (1996) Benign familial hematuria due to a mutation of the type IV collagen alpha 4 gene. J Clin Invest 98(5), 1114-1118

74) Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillen S et al. (2002) Ancient DNA evidence for old world origin of new world dogs. Science 298, 1613-1616

75) Levy M, Feingold J (2000) Estimating prevalence in single-gene kidney diseases progressing to renal failure. Kidney Int 58, 925-943

76) Li R, Mignot E, Faraco J, Dadotani H, Cantanese J et al. (1999) Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library. Genomics 58, 9-17

77) Lingaas F, Sorensen A, Juneja RK, Johansson S, Fredholm M et al. (1997) Towards construction of a canine linkage map- establishment of 16 linkage groups. Mamm Genome 8, 218-221

78) Livak KJ, Schmittgen TD. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ddCt} method. Methods 25, 402-408

79) Loirat C, Ehrich JHH, Geerlings W, Jones EHP, Landais P et al. (1994) Report on management of renal failure in children in Europe, XXIII, 1992. Nephrol Dial Transplant, Suppl. 1, 26-40

80) Longo I, Porcedda P, Mari F, Giachino D, Meloni I (2002) *COL4A3/COL4A4* mutations: From familial hematuria to autosomal-dominant or recessive Alport syndrome. Kidney Int 61, 1947-1956

81) Lou XY, Todhunter RJ, Lin M, Lu Q, Liu T et al. (2003) The extent and distribution of linkage disequilibrium in a multi-hierarchic outbred canine pedigree. Mamm Genome 14(8), 555-564

82) Lowe JK, Guyon R, Cox ML, Mitchell DC, Lonkar AL et al. (2003) Radiation hybrid mapping of the canine type I and type IV collagen gene subfamilies. Funct Integr Genomics 3, 112-116

83) Lu W, Phillips CL, Killen PD, Hlaing T, Harrison WR et al. (1999) Insertional mutation of the collagen genes *COL4A3* and *COL4A4* and in a mouse model of Alport syndrome. Genomics 61, 113-124

84) Martin P, Heiskari N, Zhou J, Leinonen A, Tumelius T et al. (1998) High mutation detection rate in the *COL4A5* collagen gene in suspected Alport syndrome using PCR and direct DNA sequencing. J Am SocNephrol 9, 2291-2301

85) Martin PH, Tryggvason K (2001) Two novel alternatively spliced 9 bp exons in the COL4A5 gene. Pediatr Nephrol 16, 41-44

86) Matise TC, Perlin M, Chakravarti A (1994a) MultiMap: An expert system for automated genetic linkage mapping. Proc Int Conf Intell Syst Mol Biol 1, 260-5

87) Matise TC, Perlin M, Chakravarti A (1994b) Automated construction of genetic linkage maps using an expert system (Multimap): A human genome linkage map. Nat Genet 6, 384-390

88) Matsuo S, Morita Y, Maruyama S, Marchay L, Yuzawa Y (2003) Proteinuria and tubulointerstitial injury: The causative factors for the progression of renal disease. Contrib Nephrol 139, 20-31

89) McCarthy PA, Maino DM (2000) Alport syndrome: A review. Clin Eye Vis Care 12(3-4) 139-150

90) Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K et al. (1997) A linkage map of the canine genome. Genomics 46, 329-336

91) Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C et al. (2000) An integrated linkage-radiation hybrid map of the canine genome. Mamm Genome 11,120-130

92) Meurs KM, Magnon AL, Spier AW, Miller MW, Lehmkuhl LB et al. (2001) Evaluation of the cardiac actin gene in Doberman Pinschers with dilated cardiomyopathy. Am J Vet Res 62, 33-36

93) Miller AB, Breen M, Murphy KE (1999) Chromosomal localization of acidic and basic keratin gene clusters of *Canis lupus familiaris*. Mamm Genome 10(4), 371-375

94) Miller AB, Lowe JK, Ostrander EA, Murphy KE (2001) Cloning, sequence analysis and radiation hybrid mapping of a mammalian *krt2p* gene. Funct Integr Genomics 1, 305-311

95) Miner JH, Sanes JR (1996) Molecular and functional defects in kidneys of mice lacking collagen α3(IV): Implications for Alport syndrome. J Cell Biol 135, 1403-1413

96) Miner JH (1998) Developmental biology of glomerular basement membrane components. Curr Op Nephrol Hyper 7, 13-19

97) Miner JH (1999) Perspectives in basic science: Renal basement membrane components. Kidney Int 56, 2016-2024.

98) Mishra BB, Fernandes AM, Blaese RM, Muchmore AV (1994) Characterization of T cell ligands for uromodulin: A possible role in costimulation, Cell Immunol 159, 113-23

99) Mochizuki T, Lemmink HH, Mariyama M, Antignac C, Gubler MC et al. (1994) Identification of mutations in the alpha-3(IV) and alpha-4(IV) collagen genes in autosomal recessive Alport syndrome. Nat Genet 8, 77-82

100) Muchmore AV, Decker JM (1985) Uromodulin: A unique 85-kilodalton immunosuppressive glycoprotein isolated from urine of pregnant women, Science 229(4712), 479-81

101) Muhle AC, Jaggy A, Stricker C, Steffen F, Dolf G et al. (2002) Further contributions to the genetic aspect of congenital sensorineural deafness in Dalmatians. Vet J 163, 311-318

102) Nakanishi K, Yoshikawa N, Iijima K, Kitagawa K, Nakamura H et al. (1994) Immunohistochemical study of α1-5 chains of type IV collagen in hereditary nephritis. Kidney Int 46, 1413-1421

103) Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM et al. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics 151, 803-823

104) Noh JW, Wiggins R, Phan SH (1993) Urine transforming growth factor-β is related to the degree of scarring in crescentric nephritis in the rabbit. Nephron 63, 73-78

105) Online Mendelian Inheritance in Animals (OMIA). http://www.ncbi.nlm.nih.gov /OMIM. (Accessed 15 Sept. 2003)

106) Ostrander EA, Giniger E (1997) Semper fidelis: What man's best friend can teach us about human biology and disease. Am J Hum Genet 61(3), 475-80

107) Ostrander EA, Galibert F, Patterson DF (2000) Canine genetics comes of age. Trends in Genet 16, 117-124

108) Ostrander EA, Kruglyak L (2000) Unleashing the canine genome. Genome Res 10, 1271-1274

109) Ostrander, EA, Galibert F, Mellersh CS (2001) Linkage and radiation hybrid mapping in the canine genome. In. *The Genetics of the Dog*. Ruvinsky A, Sampson J (eds.), CABI Publishing, New York, pp. 329-369

110) Ostrander EA, Lindblad-Toh K, Lander ES and the Canine Genome Mapping Community (2002) White paper proposal for sequencing the canine genome: Sequencing the genome of the domestic dog, *Canis familaris* http://www.genome.gov/10002154

111) Pace JM, Corrado M, Missero C, Byers PH (2003) Identification, characterization and expression analysis of a new fibrillar collagen gene COL27A1. Matrix Biol 22(1), 3-14

112) Pajari H, Kääriäinen H, Muhonen T, Koskimies O (1996) Alport's syndrome in 78 patients: Epidemiological and clinical study. Acta Paediatr 85, 1300-1306

113) Parkkinen J, Virkola R, Korhonen TK (1988) Identification of factors in human urine that inhibit the binding of *Escherichia coli* adhesins. Infect Immunol 56(10), 2623-30

114) Patterson DF (2001) Canine genetic disease information system: A computerized knowledge base of genetic diseases in the dog. St. Louis, Mosby-Harcourt.

115) Pennica D, Kohr WJ, Kuang WJ, Glaister D, Aggarwal BB et al. (1987) Identification of human uromodulin as the Tamm-Horsfall urinary glycoprotein. Science 236(4797), 83-8

116) Pichler R, Giachelli CM, Lombardi D, Pippin J, Gordon K et al. (1994) Tubulointerstitial disease in glomerulonephritis: Potential role for osteopontin (uropontin). Am J Pathol 144(5), 915-926

117) Pihlajaniemi T, Pohjolainen ER, Myers JC (1990) Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, α5(IV). J Biol Chem 265(23), 13758-13766

118) Ponder KP, Melniczek JR, Xu L, Weil MA, O'Malley TM et al. (2002) Therapeutic neonatal hepatic gene therapy in mucopolysaccharidosis VII dogs. Proc Natl Acad Sci USA 99, 13102-13107

119) Pook MA, Jeremiah S, Scheinman SJ, Povey S, Thakker RV (1993) Localization of the Tamm-Horsfall glcoprotein (uromodulin) gene to chromosome 16p12.3-16p13.11. Ann Hum Genet 57(pt4), 285-290

120) Prasadan K, Bates J, Badgett A, Dell M, Sukhatme V et al. (1995) Nucleotide sequence and peptide motifs of mouse uromodulin (Tamm-Horsfall protein) – the most abundant protein in mammalian urine. Biochim Biophys Acta 1260, 328-32

121) Priat C, Hitte C, Vignaux F, Renier C, Jiang Z et al. (1998) A whole-genome radiation hybrid map of the dog genome. Genomics 54, 361-378

122) Priester WA, McKay FW (1980) The occurrence of tumors in domestic animals. Natl Cancer Inst Monogr 54, 1-210

123) Rao VH, Lees GE, Kashtan CE, Nemori R, Singh RK et al. (2003) Increased expression of MMP-2, MMP-9 (type IV collagenases/gelatinases), and MT1-MMP in canine X-linked Alport syndrome (XLAS). Kidney Int 63, 1736-1748.

124) Reichler IM, Hauser B, Schiller I, Dunstan RW, Credille KM et al. (2001) Sebaceous adenitis in the Akita: Clinical observations, histopathology and heredity. Vet Dermatol 12, 243-53

125) Retief E, Parker MI, Retief AE (1985) Regional chromosome mapping of human collagen genes alpha 2(I) and alpha 1(I) (*col1A2* and *colIA1*). Hum Genet 69, 304-308

126) Richman M, Mellersh CS, Andre C, Galibert F, Ostrander EA (2001) Characterization of a minimal screening set of 172 microsatellite markers for genome-wide screens of the canine genome. J Biochem Biophys Met 47, 137-149

127) Rindler MJ, Naik SS, Nanxin L, Hoops TC, Peraldi M-N (1990) Uromodulin (Tamm-Horsfall glycoprotein/uromucoid) is a phosphatidylinositol-linked membrane protein. J Biol Chem 265(34), 20784-89

128) Rodriguez F, Herraez P, Espinosa de los Monteros A, Calabuig P, Rodriguez JL (1996) Collagen dysplasia in a litter of Garafiano shepherd dogs. Zentralbl Veterinarmed A 43, 509-512

129) Ron M, Band M, Yanai A, Weller JI (1994) Mapping quantitative trait loci with DNA microsatellites in a commercial dairy cattle population. Anim Genet 25(4), 259-64

130) Rosser EJ, Dunstan RW, Breen PT (1987) Sebaceous adenitis in the Standard Poodle: A discussion of 10 cases. J Am An Hosp Assoc 23, 341-345

131) Sarwal M, Chua MS, Kambham N, Hsieh SC, Satterwhite T et al. (2003) Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling. N Engl J Med 349(2), 125-38

132) Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an east asian origin of domestic dogs. Science 298, 1610-1613

133) Sayers R, Kalluri R, Rodgers KD, Shield III CF, Meehan DT et al. (1999) Role for transforming growth factor-β1 in Alport renal disease progression. Kidney Int 56, 1662-1673

134) Schelling C, Schläpfer J, Billault A, Guziewicz K, Gmür A et al. (2002) Construction of a canine bacterial artificial chromosome library for screening with PCR. J An Breed and Genet 119, 400-401

135) Schönberger J, Seidman CE (2001) Many roads lead to a broken heart: The genetics of dilated cardiomyopathy. Am J Hum Genet 69, 249–260

136) Scott DW, Miller WH Jr, Senter DA, Cook CP, Kirker J et al. (2002) Treatment of canine atopic dermatitis with a commercial homeopathic remedy: A single-blinded placebo-controlled study. Can Vet J 43(8), 601-603

137) Sherblom AP, Decker JM, Muchmore AV (1988) The lectin-like interaction between recombinant tumor necrosis factor and uromodulin. J Biol Chem 263(11), 5418-24

138) Strutz F, Zeisberg M, Ziyadeh FN, Yang C-Q, Kalluri R et al. (2002) Role of basic fibroblast growth factor-2 in epithelial-mesenchymal transformation. Kidney Int 61, 1714-1728

139) Tanaka T, Ikari K, Furushima K, Okada A, Tanaka et al. (2003) Genome-wide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus for ossification of the posterior longitudinal ligament of the spine. Am J Hum Genet Sept 4 (Epub ahead of print)

140) Thorner PS, Zheng K, Kalluri R, Jacobs R, Hudson BG (1996) Coordinate gene expression of the alpha3, alpha4, and alpha5 chains of collagen type IV: Evidence from a canine model of X-linked nephritis with a *COL4A5* gene mutation. J Biol Chem 271(23), 13821-13728

141) Todhunter RJ, Bliss SP, Casella G, Wu R, Lust G et al. (2003) Genetic structure of susceptibility traits for hip dysplasia and microsatellite informativeness of an outcrossed canine pedigree. J Hered 94, 39-48

142) Tryggvason K, Martin P (2001) Alport syndrome and basement membrane collagen. Chap.214 In. *The Metabolic and Molecular Basis of Inherited Disease*, Schriver CR, Beaudet AL, Sly WS, Valle D, Childs B et al. (eds.) McGraw-Hill: Philadelphia PA, pp. 5453-5466

143) Tsai KL, Guyon R, Murphy KE (2004) Identification of isoforms and RH mapping of canine c-kit. Cyto Genom Res (in press)

144) Turner N, Mason PJ, Brown R, Fox M, Povey S et al. (1992) Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha 3 chain of type IV collagen. J Clin Invest 89(2) 592-601

145) United States Renal Data System (USRDA) (1996) Incidence and prevalence of ESRD. Am J Kid Dis 28 (Suppl. 2), S34-S47

146) Urushihara M, Kagami S, Kahara T, Tamaki T, Kuroda Y. (2002) Glomerular distribution and gelatinolytic activity of matrix metalloproteinases in human glomerulonephritis. Nephrol Dial Transplant 17, 1189-1196

147) Venta PJ, Li J, Yuzbasiyan-Gurkan V, Brewer GJ, Schall WD (2000) Mutation causing von Willebrand's disease in Scottish Terriers. J Vet Intern Med 14(1), 10-19

148) Vignaux F, Hitte C, Priat C, Chuat JC, André C et al. (1999) Construction and optimization of a dog whole-genome radiation hybrid panel. Mamm Genome 10, 888-894

149) Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520-562

150) Werner P, Raducha MG, Prociuk U, Henthorn PS, Patterson DF (1997) Physical and linkage mapping of human chromosome 17 loci to dog chromosomes 9 and 5. Genomics 42, 74-82

151) Yang F, O'Brien PC, Milne BS, Graphodatsky AS, Solanky N et al. (1999) A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. Genomics 62, 189-202

152) Ying WZ, Sanders PW (2001) Mapping the binding domain of immunoglobulin light chains for Tamm-Horsfall protein. Am J Pathol 158(5), 1859-66

153) Yu H, Papa F, Sukhatme VP (1994) Bovine and rodent Tamm-Horsfall protein (THP) genes: Cloning, structural analysis, and promoter identification. Gene Expr 4, 63-75

154) Yuan B, Liang M, Yang Z, Rute E, Taylor N et al. (2003) Gene expression reveals vulnerability to oxidative stress and interstitial fibrosis of renal outer medulla to nonhypertensive elevations of ANG II. Am J Physiol Regul Integr Comp Physiol 284(5), R1219-30

155) Zheng K, Thorner PS, Marrano P, Baumal R, McInnes RR (1994) Canine X chromosome-linked hereditary nephritis: A genetic model for human X-linked hereditary nephritis resulting from a single base mutation in the gene encoding the α5 chain of collagen type IV. Proc Natl Acad Sci 91, 3989-3993

156) Zhou J, Mochizuki T, Smeets H, Antignac C, Laurila P et al. (1993) Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors. Science 261(5125), 1167-1169

VITA

Name:	Melissa Luanne Cox
Permanent Address:	91 Otterbrook Rd. PO Box 281 Brookfield, Nova Scotia Canada B0N 1C0
Education:	Dalhousie University University of King's College Halifax, Nova Scotia, Canada B.Sc., 1999, Combined Honours: Biology and History
	Texas A&M University College Station, TX, USA Ph.D., 2003, Genetics