EPIGENETIC MODIFICATIONS AND CONSERVED, NON-CODING DNA PLAY A ROLE IN REGULATION OF TYPE IV COLLAGEN GENE EXPRESSION

A Dissertation

by

JESSICA ASHLEY MOODY

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

May 2008

Major Subject: Veterinary Microbiology

EPIGENETIC MODIFICATIONS AND CONSERVED, NON-CODING DNA PLAY A ROLE IN REGULATION OF TYPE IV COLLAGEN GENE EXPRESSION

A Dissertation

by

JESSICA ASHLEY MOODY

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

Approved by:

Chair of Committee,	Keith E. Murphy
Committee Members,	Ann B. Kier
	Charles R. Long
	Weston W. Porter
	James E. Womack
Head of Department,	Gerald R. Bratton

May 2008

Major Subject: Veterinary Microbiology

ABSTRACT

Epigenetic Modifications and Conserved, Non-Coding DNA Play a Role in Regulation of Type IV Collagen Gene Expression.

(May 2008)

Jessica Ashley Moody, B.S., Texas A&M University

Chair of Advisory Committee: Dr. Keith E. Murphy

Type IV collagens are components of basement membranes throughout the body and are involved in maintenance of the structural integrity of tissues as well as cellular differentiation, growth, and adhesion. Members of this collagen family are uniquely arranged in pairs in a head-to-head orientation and share a proximal promoter region. The *COL4A5*-*COL4A6* gene pair is involved in numerous human diseases and cancer metastasis. For these reasons, defining the mechanisms that regulate collagen gene expression is of specific interest. To study type IV collagens, an *in vitro* model system was characterized. Comparative genomics was utilized to identify conserved, non-coding DNA in *COL4A5* and *COL4A6*. These sequences were transfected into cell lines differing in type IV collagen expression and tested for the ability to regulate transcription of a reporter gene. Each cell line was also treated with the epigenetic modifying agents, 5-Aza and TSA. The effects on type IV collagen expression were determined. The *COL4A5*-*COL4A6* promoter region was extensively characterized using ChIP analysis; antibodies against RNAPII, acetylated histone H3, and H3K9me2 were

used. Additionally, bisulfite sequencing was carried out on each cell line to determine the methylation status of CpG dinucleotides in the promoter.

Cell lines differing in expression of *COL4A5* and *COL4A6* were identified: 1) SCC-25 keratinocytes and HEK-293 cells transcribed both *COL4A5* and *COL4A6*, 2) HT-1080 cells selectively activated *COL4A5*, and 3) SK-N-SH neuroblastoma cells did not express either gene. In SK-N-SH cells, histone modifications were shown to facilitate formation of condensed chromatin to prevent transcription initiation; repression was independent of DNA methylation. Activation of *COL4A5* and *COL4A6* in SCC-25 and HEK-293 cells involved acetylation of histones, although differences between the two cell types were seen. In addition, conserved, non-coding sequences were shown to affect transcription of a reporter gene; these sequences may be interacting with the transcription machinery to modulate collagen expression. Finally, repression of *COL4A6* in HT-1080 cells appeared to be mediated through DNA methylation of the promoter; selective activation of *COL4A5* may involve conserved, non-coding DNA. In summary, epigenetic modifications as well as conserved sequences are intimately involved in regulation of type IV collagen gene expression.

DEDICATION

To Andrew

and

Mom & Dad

For all the years of unwavering support and selfless love

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Keith Murphy, for his continued support throughout the course of my undergraduate and graduate studies. I am also appreciative of the guidance provided by my committee members, Drs. Ann Kier, Charles Long, and James Womack. Special thanks go to the newest member of my committee, Dr. Weston Porter, for his intellectual input on the project and for providing lab space and reagents to complete experiments.

I would also like to thank my friends and colleagues at Texas A&M University. In particular, I would like to recognize former members of the Canine Genetics Laboratory including Drs. Ashley Hamlet, Stephanie Herbst, Jacquelyn Wahl, and Rebecca Bell for providing moral and intellectual support. Thanks go to Elizabeth Wellberg, Brian Laffin, and Dr. Rick Metz for allowing me to move into their lab space and for assisting with experiments.

Dr. Michelle Boggs deserves a special recognition for her support during the course of my graduate career. She was there to celebrate every achievement and offer encouragement after every failure. Undoubtedly, this project would not have been successful without her.

Finally, I would like to thank my parents, Lu and Russell Henske, and my husband, Andrew Moody. Your unconditional love, support, encouragement, and faith have carried me through all these years.

NOMENCLATURE

5-Aza	5-Aza-2′-deoxycytidine
ChIP	Chromatin immunoprecipitation
CTD	C-terminal domain
GTF	General transcription factor
DNMT	DNA methyltransferase
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
MBD	Methyl CpG binding domain
ORF	Open reading frame
PIC	Pre-initiation complex
RNAPII	RNA polymerase II
TSA	Trichostatin A
TSS	Transcription start site

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Type IV Collagen Background

Type IV collagens are major constituents of basement membranes and function to maintain the structural integrity of tissues through associations with laminins, heparan sulfate proteoglycans, and nidogen. They are involved in filtration as well as cell differentiation, proliferation, and adhesion (1-3). The importance of type IV collagens is highlighted by their involvement in various human diseases including Alport syndrome, leiomyomatosis, nail-patella syndrome, Goodpasture's disease, and benign familial hematuria (4-9). Table 1 summarizes important aspects of these conditions.

In addition to those diseases described in Table 1, type IV collagens are known to play a significant role in cancer metastasis. They have been shown to be involved in prostate cancer, colorectal epithelial tumors, adenoid cystic carcinoma, breast cancer, and oral carcinogenesis (10-14). The anti-angiogenic property of type IV collagens functions to inhibit proliferation and migration of endothelial cells (15-18). Indeed, changes in composition and/or loss of type IV collagen proteins from the basement membrane are correlated with tumor invasiveness. These proteins can be used as a diagnostic marker and may provide researchers with a novel therapeutic target for the treatment of cancer.

This dissertation follows the style of *The Journal of Biological Chemistry*.

Table 1 Summary of diseases involving the type IV collagens. Members of the type IV collagen family are involved in numerous human diseases. A brief description of several of these diseases as well as the associated clinical signs is presented here.

Disease	Description	Clinical Presentation
Alport syndrome	- caused by mutations in *COL4A3*, *COL4A4*, and *COL4A5* - characterized by absence of or aberrations in the α3.α4.α5(IV) network	- renal failure often associated with deafness and ocular lesions
Leiomyomatosis	- caused by deletions in parts of *COL4A5* and *COL4A6* - characterized by absence of the α3.α4.α5(IV) and α5.α5.α6(IV) networks	- smooth muscle tumors and renal failure
Benign familial hematuria	- caused by mutations in *COL4A3* and *COL4A4* - characterized by a reduction in the α3.α4.α5(IV) network	- hematuria and thin basement membranes
Nail-patella syndrome	- caused by mutations in *LMX1B*, a transcription factor known to regulate *COL4A3*, *COL4A4*, and nephrin genes	- skeletal abnormalities, nail hypoplasia, and nephropathy
Goodpasture's disease	- caused by autoantibodies to the NC1 domain of α3(IV)	- characterized by pulmonary hemorrhage and crescentic glomerulonephritis

Type IV Collagen Genes

Members of the type IV collagen family are designated α1(IV)- α6(IV) and are encoded by the genes *COL4A1- COL4A6*, respectively (19). These genes are unique in that they are arranged in pairs in a head-to-head orientation and share a bi-functional promoter. *COL4A1-COL4A2*, *COL4A3-COL4A4*, and *COL4A5-COL4A6* are transcribed from opposite strands on human chromosomes 13, 2, and X, respectively (20,21). Organization of collagen genes is similar across mammalian species (22,23). Type IV

collagen genes are also present in invertebrates. The head-to-head arrangement is seen in the *D. melanogaster* genome; however, the genes are unpaired in *C. elegans* (24,25).

Based on the genomic arrangement and sequence data, the evolutionary history of type IV collagen genes can be deduced. The ancestral collagen gene underwent duplication and inversion. Following divergence, the gene pair was then duplicated as a single unit. One final round of duplication gave rise to another gene pair. Based on genomic sequences, two classes exist: the alpha 1-like class includes *COL4A1*, *COL4A3*, and *COL4A5*, while *COL4A2*, *COL4A4*, and *COL4A6* are considered to be in the alpha 2-like class (26).

Type IV Collagen Proteins

Type IV collagens are not functional as individual proteins. Instead, heterotrimers are formed through the association of different α-chains into a right-handed helix. Specific interactions among the protein domains help stabilize trimer formation. The three distinct domains that characterize type IV collagens are 1) the N-terminal 7S domain, 2) the collagenous domain, and 3) the C-terminal NC1 or non-collagenous domain (27,28). The 7S domain consists of approximately 140 amino acid residues and is involved in the higher order assembly of protomers through intra- and inter-molecular disulfide bonds. The collagenous domain consists of roughly 1300 amino acids and is characterized by Gly-Xaa-Yaa repeats, where Xaa is frequently proline and Yaa can be hydroxyproline or hydroxylysine. These repeats are periodically interrupted to confer flexibility and allow for the formation of a strong network (29). The NC1 domain, like

the 7S domain, functions to form intra- and inter-molecular disulfide bonds. Approximately 230 residues make up the NC1 domain, although the number does vary among isoforms (1).

Type IV Collagen Trimers

Individual α-chains are synthesized as pro-collagen molecules consisting of non-helical amino and carboxyl-terminal ends. These proteins undergo post-translational modification, including glycosylation and hydroxylation, to aid in stabilizing formation of the heterotrimer (30). Pro-collagen chains are brought together via interactions among the NC1 domains, and formation of the triple helix proceeds in a zipper-like fashion. While individual collagen proteins are left-handed, a heterotrimer is a right-handed helix. Upon secretion into the extracellular matrix, the terminal ends of the pro-collagen molecules are cleaved to allow formation of the network (29).

Although many different combinations are possible, only three heterotrimers can be formed: 1) α1.α1.α2(IV), 2) α3.α4.α5(IV), and 3) α5.α5.α6(IV). Experiments have shown that it is the discriminatory nature of the NC1 domain that is responsible for directing trimer formation (31). Significantly, if one α-chain is unavailable, the heterotrimer with which it is associated can not be formed. Instead, the other collagen molecules are degraded within the cell (32).

Network Formation and Distribution

In the extracellular matrix, protomers (trimers) create a complex, cross-linked network through associations with other components of the basement membrane including laminins, heparan sulfate proteoglycans, and nidogen (1-3). Collagen protomers create networks through dimerization of the NC1 domain and tetramer formation at the 7S domain (27,28). Nidogen functions to link individual networks of type IV collagen and laminin. Heparan sulfate proteoglycans play less of a structural role, but rather directly interact with cells through binding receptor molecules. Matrix metalloproteinases (MMPs) and their antagonist, tissue inhibitors of metalloproteinases (TIMPs), are responsible for collagen turnover (33,34).

There are three protomer networks, and each has a tissue-specific expression pattern. The α1.α1.α2(IV) network is formed through self-associations and is present ubiquitously in basement membranes throughout the body. The α3.α4.α5(IV) network also self-associates and is localized to the glomerular basement membrane as well as the basement membranes of the lungs, eyes, and ears (35-38). The α1.α1.α2 α5.α5.α6(IV) network is found in the basement membranes of the skin, smooth muscles, and renal distal tubules (38-41).

Regulation of Type IV Collagen Genes

COL4A1-COL4A2

The genomic arrangement and/or tissue-specific expression patterns of the type IV collagen genes pose an intriguing question: how are these genes regulated? The

COL4A1-*COL4A2* genes are separated by a ~127 bp promoter region that is insufficient to drive expression of either gene alone *in vitro*. Instead, downstream activating elements are necessary for efficient transcription (42). The TATA-less promoter is characterized by short, AT-rich motifs that are essential for transcription. Motifs for CCAATT binding protein, Sp1, and CTC binding factor (CTCBF) are present in the promoter as well as in introns 1 of both *COL4A1* and *COL4A2*. Experiments have shown these proteins to bind to their respective motifs and work in concert to activate transcription (43). The CTC box, which is bound by CTCBF, is present near the TSS of several genes important in the extracellular matrix including *COL4A5*-*COL4A6*, *COL1A1*, laminin, and osteonectin. This represents a potential mechanism by which the expression of these genes may be coordinated (44,45). In addition to activating elements, a silencer in the third intron of *COL4A2* has been identified (46). This 24 bp element is sufficient to inhibit transcription of both *COL4A1* and *COL4A2* through binding of SILBF. Inhibition is independent of relative position and orientation, thus SILBF is a true silencer.

Several lines of evidence suggest that the view of a bi-directional promoter for this gene pair should be reconsidered. Researchers have shown that although CCAAT binding factor, Sp1, and CTCBF are necessary for activation, they do not affect *COL4A1* and *COL4A2* transcription equally. Specifically, mutations in binding sites for these proteins influence *COL4A2* transcription more than *COL4A1* (44). Other experiments have shown that the first introns in *COL4A1* and *COL4A2* compete for interactions with the promoter; constructs containing both introns and the promoter inhibit transcription of both genes in an *in vitro* setting (43). Taken together, these data suggest that the promoter

region should be viewed as two overlapping promoters that share *cis*-acting elements, but function in a uni-directional manner.

COL4A3-COL4A4

Much less is known about the regulation of *COL4A3* and *COL4A4*. An enhancer sequence in intron 1 of *COL4A4* binds LMX1B, a LIM homeodomain transcription factor encoded by the *LMX1B* gene, to activate expression of both *COL4A3* and *COL4A4*. Mutations in *LMX1B* are causative for nail-patella syndrome, a disease in which patients present with skeletal abnormalities, nail hypoplasia, and nephropathy (7).

COL4A5-COL4A6

Investigating regulation of *COL4A5* and *COL4A6* expression is more complicated than the other type IV collagens because these genes are not always co-expressed. The human and murine promoter regions have been characterized, although much remains unknown concerning regulation. The human *COL4A5-COL4A6* promoter region lacks the canonical TATA box and is characterized by CCAAT, CTC, and GC boxes (47,48). As has been shown for *COL4A1* and *COL4A2* these motifs are potential interaction sites with CCAAT binding protein, CTCBF, and Sp1, respectively. The relative locations of binding sites are presented in Figure 1. Experiments to determine the functional role of these proteins have yet to be carried out. The *COL4A6* gene has two alternative promoters that are functionally separate from each other as well as the *COL4A5* promoter (49).

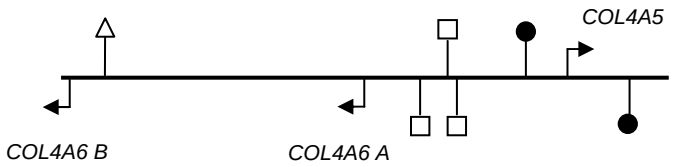

Figure 1 Relative locations of transcription factor binding sites in the collagen promoter. The relative locations of CCAAT, CTC, and GC boxes in the *COL4A5-COL4A6* promoter region are shown. Open triangle, GC box; open square, CCAAT box; filled circle, CTC box.

The murine *COL4A5-COL4A6* promoter has previously been assessed for the ability to respond to different growth factors in a cell-based assay system (47). A reporter construct containing the promoter region was transfected into three cell lines differing in type IV collagen expression and treated with various growth factors. The construct was shown to respond to four different growth factors in a cell-type specific manner, although the exact mechanism by which these factors mediate changes in gene expression remains unknown.

The human *COL4A5-COL4A6* promoter region has also been functionally characterized using a cell-based assay system (49). Constructs containing parts of the promoter were cloned upstream of the luciferase reporter gene and transfected into three cell lines differing in type IV collagen expression. As previously demonstrated, one *COL4A5* and two *COL4A6* promoters were shown to exist. All promoters were considered to be functionally separate; however, transcription was shown to be coupled through the presence of a bi-directional activator. This activator is thought to be non-functional in the cell type where *COL4A5* and *COL4A6* were not expressed. In cells

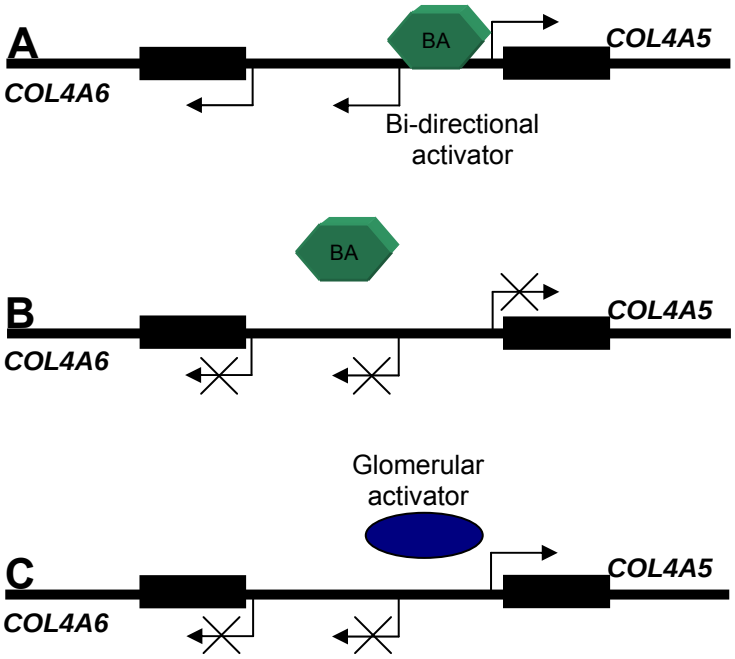

Figure 2 Proposed model for the regulation of type IV collagen genes. (*A*), Transcription of *COL4A5* and *COL4A6* is coupled through binding of the bi-directional activator. (*B*), Absence or absence of binding of the bi-directional activator contributes to gene silencing. (*C*), Selective activation of *COL4A5* is mediated through interactions between the glomerular activator and promoter. In addition, the bi-directional is absent. (Adapted from Segal, 2001)

where *COL4A5* was selectively transcribed, the promoter construct was unable to drive expression of the reporter gene. Thus, it was proposed that absence of the bi-directional activator along with unidentified distal enhancers are necessary to account for the activation of *COL4A5* without *COL4A6* (49). This distal enhancer was named the glomerular activator because of the importance of this expression pattern in the glomerular basement membrane of the kidney. Figure 2 summarizes the proposed model of collagen regulation.

In vivo experiments have also been conducted to characterize the human *COL4A5-COL4A6* promoter (50). The functionality of reporter constructs containing the entire promoter region and sequences flanking the 5'-ends of *COL4A5* and *COL4A6* were

tested in transgenic mice. Specifically, a 10.6 kb fragment upstream of *COL4A5* and 3.8 and 13.8 kb fragments upstream of *COL4A6* were assayed for the ability to turn on the *lacZ* reporter gene. These constructs were able to drive expression of the reporter gene only in the upper gastrointestinal tract. Because these genes are expressed in other tissues including the kidney, eye, and ear, additional unidentified regulatory sequences outside of the tested fragments are likely to exist (50).

Eukaryotic Gene Regulation

There exists a seemingly endless number of ways to influence gene expression in eukaryotic cells. Chromatin structure, DNA methylation, histone modification, RNAPII, and transcription factors are all involved in modulating expression at the level of transcription. Post-transcription stages of regulation exist in the form of mRNA processing, degradation, transportation, and localization. Translation and post-translational modifications represent additional layers of control. Although the end result of gene expression most definitely involves a number of different regulatory control mechanisms, only transcription and the associated players will be discussed further.

Transcription

In general, transcription involves the binding of a sequence-specific activator to the promoter region. In addition to a DNA-binding domain, activators possess an activation domain necessary to assist in recruiting general transcription factors (GTFs) and proteins responsible for altering chromatin structure. DNA becomes more accessible,

and GTFs including TFIIA, B, D, E, F, and H assemble on the core promoter to form the pre-initiation complex (PIC). RNA polymerase II (RNAPII) is then positioned over the transcription start site (TSS). Core promoter elements and distal-acting enhancers work together with the PIC to coordinate transcription initiation. The C-terminal domain (CTD) of RNAPII is phosphorylated at serine 5, which recruits protein complexes that modify histones important for gene activation. Upon elongation, the CTD becomes phosphorylated at serine 2. This modification recruits proteins necessary for elongation. These factors associate directly or indirectly with RNAPII to assist in transcription (51,52).

It is important to mention that the order of recruitment is not identical for every gene. In fact, the binding of certain sequence-specific activators is dependent upon chromatin modifying activities. Nevertheless, a coordinated effort among numerous players is necessary for proper activation of transcription (51).

Chromatin and Histone Modifications

Chromatin, in its simplest definition, is the complex association of nucleosomes and non-histone proteins with DNA. Chromatin is involved in DNA packaging to allow billions of bp to exist within the confines of the nucleus. Additionally, chromatin structure is associated with numerous biological processes including transcription, replication, DNA repair, and recombination. For these reasons, much research has been dedicated to understanding the regulation of nucleosome dynamics (51).

The basic unit of the nucleosome is an octamer of two copies of each of the four core histone proteins H2A, H2B, H3 and H4. Approximately 146 bp of DNA are wrapped around the histone octamer. Histone H1 exists outside of the nucleosome core to assist with DNA packaging. Histone proteins have globular and tail domains, which are subject to an array of post-translational modifications including acetylation, methylation, phosphorylation, and ubiquitination. A diverse set of enzymes are responsible for catalyzing these modifications, the functions of which are beginning to be revealed (53,54).

The effects of covalent modifications on histone residues can be direct or indirect. Certain modifications alter charge, thereby influencing DNA-histone interactions to make DNA more or less accessible to transcription machinery. A modification can also act as a docking site for an effector protein to bind and elicit a specific response or it can actually prevent docking of non-histone proteins. Furthermore, modifications can recruit chromatin remodeling complexes, which use ATP hydrolysis to alter histone-DNA contacts (53). Table 2 summarizes post-translation modifications and their role in transcription.

Acetylation

Acetylation of lysine residues is catalyzed by histone acetyltransferases (HATs) and is almost invariably associated with an open chromatin state and active transcription. In direct contrast, histone deacetylases (HDACs) function as global transcription repressors by removing acetyl groups. Addition of acetyl groups neutralizes the basic

charge on lysine, which presumably loosens DNA contacts with histones and makes histones more mobile. Subsequently, DNA becomes more accessible for recognition by various protein complexes and GTFs. Acetylated lysine residues also act as docking sites for proteins with bromodomains. Included among these are chromatin remodeling complexes and HATs, which can modify other lysine residues in the vicinity (53).

Table 2 Summary of histone post-translational modifications. Histone residues are subject to hundreds of post-translational modifications, some of which are presented here along with their role in transcription. (Adapted from Berger, 2007)

Chromatin Modifications	**Residues**	**Effect on Transcription**
Acetylated lysine	H3 (9, 14, 18, 56), H4 (5, 8, 13, 16), H2A, H2B	Activation
Phosphorylated serine, threonine	H3 (3, 10, 28), H2A, H2B	Activation
Methylated arginine	H3 (17, 23), H4 (3)	Activation
Methylated lysine	H3 (4, 36, 79) H3 (9, 27), H4 (20)	Activation Repression
Ubiquitylated lysine	H2B (120) H2A (119)	Activation Repression

Global acetylation patterns have been assessed in yeast as well as human cells. At the genome level, higher eukaryotic organisms generally have lower levels of acetylation and more heterochromatin, presumably due to the more specialized function of a given cell type (55,56). Despite these differences, the acetylation pattern for genes actively undergoing transcription is very similar and is characterized by high amounts of acetylation at the promoter and 5′ end of the coding region. In fact, more acetylation of

histones H3 and H4 correlates with higher transcription rates (57). Acetylation of lysine residues is also associated with known regulatory elements, DNase hypersensitive sites, and evolutionarily conserved sequences (56).

Methylation

Histone methylation occurs at lysine and arginine residues and is catalyzed by histone methyltransferases (HMTs). Lysine can be mono-, di-, or tri-methylated, while arginine can exist in the mono- or di-methyl state. In contrast to acetylation, methylation does not alter charge, but rather acts as a docking site for effector proteins with chromodomains (53).

Methylation is associated with both active and repressed transcription. In general, the response to methylation is dependent upon which residue is modified, the methylation state (mono-, di-, or tri-), and other modifications on nearby histones. Methylation of one residue can enhance or encourage the modification of another residue, known as histone crosstalk (53). In addition, localization of methylation is tightly regulated. For example, methylation of histone H3K36 occurs within the open reading frame (ORF) of actively transcribed genes and helps prevent cryptic initiation of transcription. If methylation of H3K36 is misdirected to the promoter region, transcription is repressed (51). Methylation of H3K9 has long been associated with gene silencing through binding of heterochromatin protein 1 (HP1). In light of new research, however, this relationship does not appear to be so simple. More specifically, di- and tri-methylated H3K9 in the ORF are associated with the gamma isoform of HP1 and active transcription, while

methylation targeted to the promoter regions correlates with gene silencing (58). The context driven nature of histone methylation makes assigning a function to a particular modification complicated.

Methylated lysine residues have long been thought to represent a stable epigenetic mark; however, with the recent discovery of lysine demethylases, it is now known that this modification is reversible (59). Understanding the functional consequences and how this alters our current view of gene regulation are the subject of extensive research. Proteins that catalyze arginine demethylation have yet to be discovered, although arginine can be converted to citrulline resulting in the passive loss of methyl groups (60).

Phosphorylation and Ubiquitination

Histone residues are subject to phosphorylation and ubiquitination, although much less is known regarding the function of these modifications as compared to acetylation and methylation. Phosphorylation of serines or threonines results in a negative charge and is recognized by proteins containing a 14-3-3 domain. Ubiquitination is a rather large modification found on histones H2A and H2B and is associated with both active and repressed transcription (53). Undoubtedly, future research will focus on delineating the function of these modifications as they are likely to play an intimate role in regulating gene expression.

Methylation of DNA

DNA methylation represents yet another layer of regulatory control. In vertebrates, DNA methylation occurs almost exclusively in the context of CpG dinucleotides and is involved in genomic imprinting, X chromosome inactivation, carcinogenesis, and cellular differentiation. Methylation of DNA is also associated with repetitive DNA, transposons, and heterochromatin (61,62).

Early in mammalian development, the whole genome is demethylated. Around the time of implantation a wave of *de novo* methylation is carried out by the *de novo* methyltransferases (DNMTs) 3A and 3B. Methylation is propagated through subsequent cell divisions by maintenance DNMT1, which recognizes hemi-methylated DNA (62-64). Aberrations in methylation, such as hypermethylation of tumor suppressor genes, are associated with numerous cancers (65,66).

Approximately 60% of promoters are associated with CpG islands. CpG islands must 1) be 200 bp in length, 2) have a G + C content of 50%, and 3) have a ratio of observed to expected CpG frequency of 0.6 (62). The dogma has been that regardless of expression these sites are unmethylated (67). Some genes, however, do exhibit tissue-specific methylation patterns during development that inversely correlate with expression (68). Furthermore, experiments have shown that a subset of genes can be de-repressed by treating cells with agents that force global demethylation (69,70). It has also been demonstrated that gene activity can be repressed *in vitro* by methylation of promoter regions (71). These observations have lead scientists to propose that methylation of CpG dinucleotides plays a role in tissue-specific gene repression.

There are several mechanisms by which methylated DNA can mediate gene silencing. Methylated DNA can recruit and bind regulatory proteins with methyl CpG binding domains (MBDs). One such protein, MeCP2, tethers a multi-protein repressor complex that includes HDACs to remove acetyl groups. Histone residues are subsequently methylated, thereby supporting a compact chromatin state (68). In another model, proteins with a MBD and SET domain recognize methylated DNA to directly methylate histones. It has also been shown that methylated histone tails can recruit DNMTs, which in turn methylate DNA (72,73). Finally, presence of methylated DNA can preclude binding of certain transcription factors and activators (74). Many factors have been shown to be sensitive to methyl CpG including AP-2, c-Myc, E2-F, and CREB (75).

Several methods to silence genes by CpG methylation have been described; however, it is important to keep in mind that DNA methylation may also work in concert with other repressive mechanisms to keep genes in the inactive state.

Comparative Genomics

Historically, identification of *cis*-acting elements involved in gene regulation has been hindered by the limited availability of experimental approaches. Now, with access to numerous vertebrate whole genome sequences and bioinformatics analysis tools, researchers are in a position to systematically explore the biological relevance of non-coding sequences through the use of comparative genomics. Comparative genomics aligns sequences from multiple species to identify regions of homology. This technique

has widely been used to identify coding boundaries, and is now being utilized to analyze non-coding DNA. Those regions that show evolutionary conservation in non-coding DNA are selectively constrained and, therefore, thought to be functional (76-79).

To classify a sequence as conserved, non-coding, the extent of homology and the number of species that show conservation must be considered. Based on the recognized criteria, regions that show ≥ 70% homology between at least two species and are ≥ 100 bp in length are considered to be conserved (78). It has been estimated that 3% of non-coding sequence is conserved among distantly related mammals (80-82). Evidence from analyses using human genotype data suggests that these conserved, non-coding sequences are selectively constrained and do not represent mutational cold spots (83).

A number of studies have successfully identified regions of conserved, non-coding DNA that are functional and important in disease. Conserved, intronic sequences were found to regulate transcription of the breast cancer susceptibility gene, *BRCA1*, in an *in vitro* setting. *BRCA1* is often down-regulated in breast cancer patients, so determining if and how these regions are involved in the development of this disease will be informative (84). In another study 45 conserved, non-coding sequences surrounding *RET*, a tyrosine kinase gene involved in development, were systematically examined using a cell-based assay approach. This study demonstrated the biological relevance of an evolutionarily conserved enhancer associated with Hirschsprung disease. Significantly, the majority of conserved, non-coding sequences surrounding the *RET* gene exhibited activity *in vitro* in a cell-type specific manner highlighting the fact that evolutionarily

conserved regions are of biological interest (85). More studies successfully using comparative genomics will likely be seen in the future.

Summary

There are ~35,000 genes in the human genome, only a fraction of which are expressed in a given cell type; therefore, unique mechanisms to control gene expression must exist (56,86). Indeed, eukaryotic cells are highly evolved to modulate expression at numerous levels including transcription, mRNA processing, and post-translation. Regulation of the type IV collagen genes, *COL4A5* and *COL4A6*, must be tightly controlled to ensure tissue-specific patterns of expression. Previous research dedicated to understanding control of these genes has suggested that regulatory elements outside of the *COL4A5*-*COL4A6* promoter exist. Comparative genomics has been successfully utilized to identify functionally significant, conserved, non-coding regions and will likely reveal *cis*-acting elements important in controlling collagen expression. In addition, evaluation of epigenetic modifications associated with the promoter region will shed light on *COL4A5*-*COL4A6* regulation.

CHAPTER II

MATERIALS AND METHODS

Cell Culture

Cell lines were obtained from American Type Culture Collection (ATCC, Manassas, VA) and maintained according to the manufacturer's recommended protocols. HEK-293 cells (ATCC CRL-1573) were maintained in minimal essential Eagle's medium with Earle's balanced salt solution, non-essential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1500 mg/L sodium bicarbonate, supplemented with 10% horse serum. HT-1080 cells (ATCC CCL-121) and SK-N-SH brain neuroblastoma cells (ATCC HTB-11) were maintained in minimal essential Eagle's medium with Earle's balanced salt solution, non-essential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1500 mg/L sodium bicarbonate, supplemented with 10% fetal bovine serum. SCC-25 keratinocyte cells (ATCC CRL-1628) were maintained in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F12 with 2.5 mM L-glutamine, 15 mM HEPES, 0.5 mM sodium pyruvate, and 1200 mg/L sodium bicarbonate, supplemented with 10% fetal bovine serum and 0.4 µg/mL hydrocortisone.

RNA Isolation

HEK-293, HT-1080, SCC-25, and SK-N-SH cells were collected and resuspended in 750 µl RNA STAT60 (Iso-Tex Diagnostics, Friendswood, TX). Total RNA was

isolated using the *mir*Vana™ miRNA Isolation Kit (Ambion Inc., Austin, TX) following the manufacturer's recommended protocol. Samples were quantified using the NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Reverse Transcription

Reverse transcription (RT) reactions were carried out on total RNA from HEK-293, HT-1080, SCC-25, and SK-N-SH cells using the High Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems Inc., Foster City, CA). Each 20 µl RT reaction contained: 10 µl total RNA, 1X RT random primers, 20 U RNase Inhibitor, 1X RT buffer, 50 U MultiScribe™ reverse transcriptase, and 1 mM dNTPs. Reactions were run in an Eppendorf MasterCycler (Eppendorf North America Inc., New York, NY) for 10 minutes at 25°C, 120 minutes at 37°C, 5 seconds at 85°C, and then held at 4°C.

Quantitative Real Time RT-PCR

TaqMan®

TaqMan® quantitative real time RT-PCR (qRT-PCR) was used to characterize all four cell lines for collagen expression and to determine relative expression levels following treatment with epigenetic modifying agents. Primers targeting the 3′ ends of *COL4A5* and *COL4A6* were used for these experiments. In addition, TaqMan® and primers targeting the 5′, middle, and 3′ sections of both collagens were used to assess how transcripts are processed. RNA was isolated and reverse transcribed as previously

mentioned. Each 20 µl qRT-PCR contained 10 ng RT product, 1X TaqMan® Universal PCR Master Mix (No AmpErase® UNG), and 1X TaqMan® assay, which contained the forward and reverse primers as well as the TaqMan® probe (Applied Biosystems Inc.). Primer and probe sets are listed in Table 3. Reactions were run in an Applied Biosystem 7500 real time machine for 10 minutes at 95°C, followed by 40 cycles of 95°C for 15 seconds and 60°C for one minute.

Table 3 Primers for Taq Man® qRT-PCR. Pre-designed primers targeting exon-exon boundaries were used to quantify relative expression levels of *COL4A5* and *COL4A6*. *GAPDH* was used as the normalization control.

Gene	ABI Assay No.	Exon Boundary
COL4A5	Hs00166712_m1	1-2
	Hs01012435_m1	28-29
	Hs01012455_m1	50-51
COL4A6	Hs00904472_m1	1-2
	Hs00904446_g1	26-27
	Hs00904466_m1	44-45

Syber Green

Syber Green was used to determine relative expression levels of the *COL4A6* transcript variants *A* and *B* in SCC-25 and HEK-293 cells. RNA was isolated and cDNA was made following above mentioned protocols. Each 25 µl Syber Green qRT-PCR contained 10 ng RT product, 1X SYBR® GREEN PCR Master Mix (Applied Biosystems Inc.), and 500 nM of each forward and reverse primer. Primer sequences are shown in Table 4. Reactions were run using the conditions described for TaqMan® qRT-PCR. In addition, dissociation curves were used to verify primer specificity.

Table 4 Primers for Syber Green qRT-PCR. Primers targeting *COL4A6* transcript variants *A* and *B* were designed using ABI PRISM Primer Express (Applied Biosystems Inc.) and analyzed using BLAST (URL: http://www.ncbi.nlm.nih.gov/BLAST/). *TBP* was used as the normalization control.

Gene	Forward Primer	Reverse Primer
COL4A6 A	CTGCTGGTCTTCTTTACCTTCCA	CAATATTTTCACATCTTCTGCTTGTATG
COL4A6 B	CTGAGTAAGGAAACAGCCTCCAA	GCGGCCCGTGCTCAT

qRT-PCR Analysis

Relative quantification of gene expression was determined by utilizing the comparative critical threshold (C_T) method. C_T values corresponding to mRNA transcript levels were subtracted from the respective C_T values of the normalization control, resulting in ΔC_T values. *GAPDH* and *TBP* were used as normalization controls for TaqMan® and Syber Green, respectively. The largest ΔC_T value was arbitrarily used as a constant and subtracted from all other ΔC_T values to determine $\Delta\Delta C_T$ values. Fold changes were then determined by calculating $2^{-\Delta\Delta CT}$. A two-tailed t-test was used to determine the level of statistical significance.

Identification of Conserved, Non-Coding DNA

Approximately 307 kb of genomic sequence encompassing human *COL4A5* and *COL4A6* (chrX: 107,519,809-107,827,431; University of California, Santa Cruz, March 2006, hg18) was compared to the orthologous sequences of 6 non-human vertebrates (chimpanzee, rhesus, mouse, rat, dog, and cow). The human sequence was also compared to the genomes of seven non-human vertebrates (armadillo, elephant, tenrec, rabbit, chicken, opossum, and zebrafish), where *COL4A5* and *COL4A6* have not been annotated. Comparisons were conducted using the University of California, Santa Cruz Genome

Bioinformatics program (URL: http://genome.ucsc.edu/). Conservation scores were based on the phylogenetic hidden Markov model. Sequences that were ≥ 100 bp in length and exhibited ≥ 70% sequence identity between the human and at least one other vertebrate were classified as conserved. Sequences that did not overlap exons were considered to be non-coding. In addition, conserved, non-coding sequences were compared to their respective genomes using BLAST to eliminate any repetitive elements. Conserved, non-coding regions were designated according to their positions relative to the *COL4A5* TSS.

Cloning of Conserved, Non-Coding DNA

Primers were designed and analyzed using NetPrimer (URL: http://www.premierbiosoft.com/netprimer /index.html) to amplify conserved, non-coding sequences individually (Table 5). In addition, a primer set was designed to amplify a non-conserved, non-coding sequence. XhoI restriction sites were added to the 5′ ends of both primers. Each PCR contained: 50 ng human genomic DNA, 1X Buffer B, 1.5 mM $MgCl_2$, 0.5 mM dNTP mix, 1 μM of each primer, and 1 U Taq DNA Polymerase (Fisher Scientific, Pittsburg, PA). Reactions were run in an Eppendorf MasterCycler for 5 minutes at 94°C, 30 seconds at 55°C, 30 seconds at 72°C followed by 34 cycles of 30 seconds at 94°C, 30 seconds at 63°C, and 30 seconds at 72°C. PCR products were gel purified using the QIAEX II Gel Extraction kit (Qiagen, Valencia, CA) and digested with XhoI (Promega Corp., Madison, WI) following the manufacturers' protocols. Digests were then purified using the QIAquick PCR Purification kit (Qiagen). The pGL3-

Table 5 Primers to amplify conserved, non-coding sequences. Forward and reverse primer sequences to amplify each conserved region are shown. Each construct is designated by its location with respect to the *COL4A5* transcription start site (+1). The coordinates and amplicon sizes of each conserved, non-coding sequence are also shown.

Construct	*Coordinates of conserved non-coding sequences*	*Forward primer (5'-3')*	*Reverse Primer (5'-3')*	*Amplicon size (bp)*
-43K	107526375-107526477 107526212-107526321	TGTGGAAGATACGATGTGTAAGAC	AAACTGTTCACTTTGGACATTCTA	1,119
-41K	107528021-107528140	CCTTCTGCTTATTCTGCCC	GTCCTTGCTGTAGTGTTGTCC	1,193
-40K	107529361-107529585	AATAGAATGGCTTTCAGGC	GAAGGAGGCACATAAGTAGG	870
-28K	107540919-107541022	CCCTATCACAGACAAGACAA	ACATTTTCACCAAGAACTCA	561
-25K	107543904-107544042	CTCTCAACAAGGCTCTCCAG	CAAAGACTGAAGGCAATGTAGG	643
-110	107569568-107569698 107569709-107569809	TAAAGAAGACCAGCAGCCC	AGAGGGACAGTGAGGCTTG	572
16K	107644613-107644713	CCAAGTGAAGATGTAGAACG	TAGCAACAAGTAACTAAAAATAAA	722
21K	107590796-107590931	GTAGAGATGCGATTTCACCA	CCATCTGAGGGGAAAAAA	644
25K	107594560-107594741	TTTATTTCTTAGGAGGCTTCTTA	AAACGGAACTTTTTGGGG	537
76K	107646097-107646337	TTCTTCTTTGCTGCCTGTG	ACCGTAGGCTAACACTTGAAC	875
120K	107689280-107689403	TAGGTGTCACATCAACTAAAAAG	TTATGTTTCTCACTCAGGTCC	828

Table 5 Continued

Construct	*Coordinates of conserved non-coding sequences*	*Forward primer (5'-3')*	*Reverse Primer (5'-3')*	*Amplicon size (bp)*
136K	107706515-107706630	GAAGAGCAAAAGCAGGGC	GCCACAATGGTTGAGGAAA	666
179K	107748548-107748753	ATCTATGCTATGGGAGGAAGT	GCCCTAAACTGACAAAATCC	965
216K	107785542-107785654	CCAACCAGTATCACAGAGCAG	GCCTGGGCATACCTTTCT	318
230K	107800008-107800158	GAAGCCATTAGTCACAAAGAT	GTAAGGCTGTTTTCGGTACT	748
235K	107804508-107804758	TTGAAGGCTGTTGGTGAA	CAGGGCAATCTGAGAAAAC	756

Promoter vector (Promega Corp.) was digested with XhoI and treated with Shrimp Alkaline Phosphatase (Roche, Basel, Switzerland) according to the recommended protocols. Linearized vector was gel purified using the QIAEX II Gel Extraction kit (Qiagen). PCR products were cloned into the pGL3-Promoter vector using T4 DNA Ligase (Promega Corp.) and transformed into TOP10 competent cells (Invitrogen Corp., Carlsbad, CA). The GenElute™ Endotoxin-free Plasmid Midiprep kit (Sigma-Aldrich Co., St. Louis, MO) was used for large-scale isolation of plasmid DNA. Isolations were carried out following the manufacturers' protocol, and samples were quantified using the NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies). Clones were sequence verified following the protocol outlined below.

Sequencing

Sequencing reactions were performed with the BigDye Terminator kit and resolved on an ABI3100 automated sequencer (Applied Biosystems Inc.). Sequence reads were aligned with the appropriate reference sequences via CLUSTAL W (http://www.ebi.ac.uk/clustalw/).

Transient Transfections

HEK-293, HT-1080, SCC-25, and SK-N-SH cells were co-transfected with pGL3-Promoter and phRG-TK vectors (Promega Corp.). The pGL3-Promoter vector contains the complete firefly (*Photinus pyralis*) luciferase open reading frame under control of the SV40 promoter. The phRG-TK vector has the pGL3 backbone and contains

the complete *Renilla* (*Renilla reniformis* or sea pansy) luciferase open reading frame downstream of the TK promoter. The phRG-TK vector was used as the normalization control to account for differences in transfection efficiencies. Cells were cultured as previously mentioned and transfected using Lipofectamine 2000 (Invitrogen Corp.) following the recommended protocol. Approximately 2.5 X 10^5 HEK-293, 1.2 X 10^5 HT-1080, 6.0 X 10^5 SCC-25, and 8.0 X 10^4 SK-N-SH cells were plated the day before transfections. HEK-293, HT-1080, and SK-N-SH cells were transfected with 800 ng total DNA using 2.0 µl Lipofectamine 2000 (Invitrogen Corp.). SCC-25 cells were transfected with 6.0 µg total DNA using 10.0 µl Lipofectamine 2000 (Invitrogen Corp). The firefly to *Renilla* ratio was 10:1 for SK-N-SH cells, and a 20:1 ratio was used for the remaining cell types. The Dual-Glo Luciferase Assay System (Promega Corp.) was used in accordance with the manufacturer's protocol, and the Packard LumiCount™ (Packard Instrument Co., Meriden, CT) was used to measure firefly and *Renilla* luminescence 24 hours post-transfection. Each pGL3-construct was transfected in duplicate on at least three separate occasions, and relative luciferase units (RLUs) were determined.

Treatment with 5-Aza-2′-deoxycytidine and Trichostatin A

HEK-293, HT-1080, SCC-25, and SK-N-SH cells were treated with 1) 5 µM 5-aza-2′-deoxycytidine (5-Aza) for 72 hours, 2) 300 nM trichostatin A (TSA) for 24 hours, and 3) 5 µM 5-Aza for 48 hours followed by addition of 300 nM TSA for 24 hours. Media was renewed every 24 hours for the 5-Aza and 5-Aza plus TSA treatment groups. After treatment, cells were harvested and RNA was isolated and reverse transcribed.

Quantitative RT-PCR was performed using TaqMan® and primers specific for the 3′ ends of the *COL4A5* and *COL4A6* transcripts. Relative expression levels were determined using the $\Delta\Delta C_t$ method and represented as relative fold changes as compared to the control. Experiments were carried out as previously described.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were carried out on HEK-293, HT-1080, SCC-25, and SK-N-SH cell lines. Cells were grown in 10 cm plates and fixed with 1% formalin for 10 minutes at 37°C. Crosslinking was stopped by addition of 125 mM glycine. Cells were then washed with PBS and complete protease inhibitors (Roche), pelleted at 2,000 rpm for 4 minutes at 4°C, and resuspended in 200 µl SDS lysis buffer plus complete protease inhibitors for every 1 X 10^6 cells. Samples were incubated on ice for 10 minutes and sonicated. Samples were centrifuged for 10 minutes at 13,000 rpm at 4°C, and supernatants were collected. Chromatin was then diluted 5-fold with ChIP dilution buffer, and complete protease inhibitors were added. Samples were pre-cleared twice with salmon sperm DNA/protein A agarose-50% slurry and incubated overnight at 4°C with antibody. Approximately 10 µg of each of the following antibodies were used: anti-RNA polymerase II (Upstate Cell Signaling, Billerica, MA), anti-acetyl histone H3 (Upstate Cell Signaling), and anti-dimethyl lysine 9 (Abcam Inc., Cambridge, MA). Anti-HP1 was used in SK-N-SH cells (Upstate Cell Signaling). The following day, samples were incubated with salmon sperm DNA/protein A agarose-50% slurry for 1 hour at 4°C and then centrifuged at 5,000 rpm for 1 minute at 4°C. Agarose pellets were washed with

low salt wash buffer, high salt wash buffer, lithium chloride buffer, and TE buffer. Pellets were eluted with 1% SDS and 0.1 M $NaHCO_3$. Five M (5 M) NaCl and RNase A were added, and crosslinks were reversed by incubating samples at 65°C for 4 hours. DNA was precipitated through addition of 100% ethanol, and samples were incubated overnight at -20°C. Samples were pelleted, resuspended in 100 µl water, 2 µl 0.5 M EDTA, 4 µl 1 M Tris pH 6.5, and 1 µl 20 mg/mL proteinase K, and then incubated for 1.5 hours at 45°C. DNA was purified using QIAquick PCR Purification kit (Qiagen). PCR was performed with primers listed in Table 6. All reactions contained: 1X Buffer B, 1 µM of each primer, 2 µl DNA, and 1 U Taq DNA Polymerase (Fisher Scientific). Concentrations of $MgCl_2$ and dNTP mix varied: -41K, -6B, 6B, and +6B contained 1.5 mM $MgCl_2$ and 0.6 mM dNTP mix; 6A contained 2.0 mM $MgCl_2$ and 0.6 mM dNTP mix; -40K, 5A, and 76K contained 2.5 mM $MgCl_2$ and 0.4 mM dNTP mix; 5A contained 2.5 mM $MgCl_2$ and 0.8 mM dNTP mix. Reactions were run in an Eppendorf MasterCycler for 5 minutes at 94°C followed by 34 cycles of 30 seconds at 94°C, 30 seconds at primer annealing temperature (see Table 6), and 30 seconds at 72°C. A final extension was carried out at 72°C for 10 minutes. Products were resolved on agarose gels.

Bisulfite Sequencing

Genomic DNA was isolated from HEK-293, HT-1080, SCC-25, and SK-N-SH cells using Wizard DNA Isolation kit (Promega Corp.). Conversion of unmethylated cytosines was done using the EpiTect® Bisulfite kit (Qiagen) according to the

recommended protocol. MethPrimer was used to analyze the *COL4A5* and *COL4A6* promoter regions for presence of CpG islands and to design primers for bisulfite

Table 6 Primers for ChIP analysis. Forward and reverse primer sequences to amplify regions of the *COL4A5* and *COL4A6* genes for ChIP analysis are shown. Primer annealing temperatures (T_m) are included.

ChIP Product	Forward Primer (5′-3′)	Reverse Primer (5′-3′)	T_m
-6B	GGGATGGTTAGAGGGTCAA	GCAGAGCCCAGATGAGC	54°C
6B	GCTCATCTGGGCTCTGC	CCATCTGTCTTATGTGGGAATA	54°C
+6B	TATTCCCACATAAGACAGATG	ATATCTCCCATATAAAGACTTGA	50°C
6A	TATTTTCACATCTTCTGCTTGTA	CTTTTCTATTCGTTCATTTGC	51°C
5A	GCAAATGAACGAATAGAAAAG	GAGGGACAGTGAGGCTTG	52°C
+5A	CAAGCCTCACTGTCCCTC	GCAAGCAGCCCTATTTTT	56°C
-41K	AGAAGATGT CCAAAACACACAG	CACAAAGCCGTGAGCAAT	55°C
-40K	CTGGGTTTATGTTCTTAGATG	CTGCCCTGGTTAATGGT	52°C
76K	GTCAGCAGGCAACAGGGA	CTAACATCACAGATAAAAGAAGCAA	55°C

sequencing (http://www.urogene.org/methprimer/). Two primer sets were used to analyze the *COL4A5* and *COL4A6 A* promoter regions: BS1.1 (F 5′- GAAATTTTTTGTTTTGTGAGTAGTTG -3′), (R 5′- AACTCAATTCCAATCTCATTAATATAC -3′); BS1.2 (F 5′- TAATGAGATTGGAATTGAGTTTAGAAAA -3′), (R 5′- CCTCACCAACACAAAAAAATTTAA -3′). Two primer sets were used to evaluate methylation of the *COL4A6 B* promoter: BS2.1 (F 5′- AGTTATAATTTGAAATGGGAGGGAG -3′), (R 5′- AAACATCAACAAAACCCAAATAAAC -3′); BS2.2 (F 5′- GTGTTTTAAAGGGAAATAGGTTTAG -3′), (R 5′-

AAAATACCTAAAAATAAAAAAAACTACC -3′). All reactions contained: 50 ng converted DNA, 1X Buffer B, 2.5 mM $MgCl_2$, 0.6 mM dNTP mix, 1 μM of each primer, and 1 U Taq DNA Polymerase (Fisher Scientific). Reactions were run in an Eppendorf MasterCycler for 5 minutes at 94°C, 34 cycles of 30 seconds at 94°C, 30 seconds at 61°C (BS1.1, 1.2, and 2.2) or 63°C (BS2.1), 30 seconds at 72°C, followed by a final extension at 72°C for 10 minutes. PCR products were gel purified using QIAEX II Gel Extraction kit (Qiagen) and cloned into pCR4-TOPO (Invitrogen Corp.) following the manufacturer's protocol. Ten clones for each primer set were sequenced using the BigDye Terminator kit and products were resolved on an ABI3100 automated sequencer (Applied Biosystems Inc.). DNA methylation patterns were analyzed using BiQ Analyzer (http://biq-analyzer.bioinf.mpi-sb.mpg.de/).

CHAPTER III

RESULTS:

THE ROLE OF CONSERVED, NON-CODING DNA IN TYPE IV COLLAGEN GENE REGULATION

Characterization of Model Cell Lines

COL4A5 and *COL4A6* are network-forming collagens that have been implicated in cancer metastasis and numerous other diseases. They have a unique genomic arrangement and exhibit tissue-specific expression patterns, thus regulation must be tightly controlled. To gain insight into how expression of this gene pair is modulated, an *in vitro* model system must be identified. Understanding what mechanisms account for differences in collagen expression in immortalized cell lines might shed light on *in vivo* gene regulation.

Using qRT-PCR, four cell lines differing in type IV collagen gene expression were identified (see Figure 3A). *COL4A6* primers used for this part of the study detected both transcript variants. *COL4A5* and *COL4A6* were both found to be expressed in SCC-25 and HEK-293 cells, while SK-N-SH cells did not express either gene. HT-1080 cells selectively activated *COL4A5*. Relative expression of *COL4A5* was approximately four fold higher in SCC-25 cells than HEK-293 cells ($p < .0280$). In SCC-25 cells, expression of *COL4A5* was roughly 3.5 times greater than *COL4A6* ($p < .0152$), while transcript levels were approximately equal in HEK-293 cells.

In a separate set of experiments relative expression levels of *COL4A6* transcript

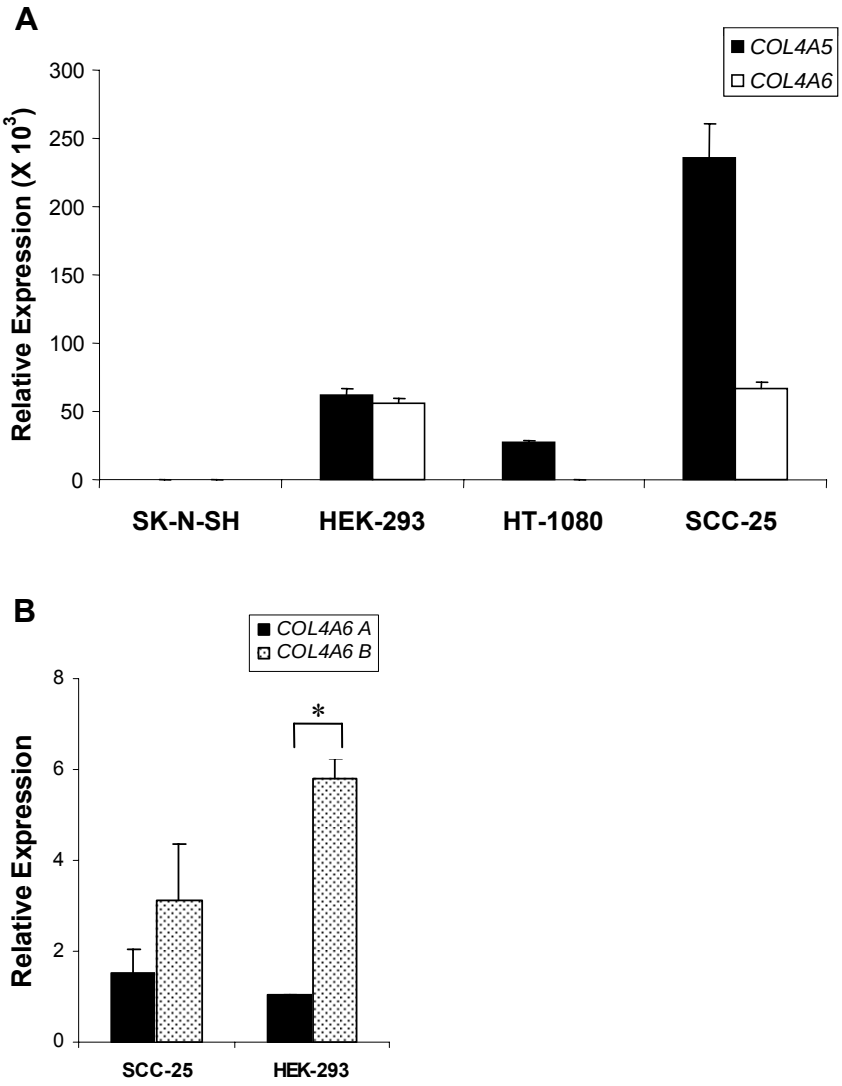

Figure 3 Characterization of type IV collagen gene expression in model cell lines. (*A*), Relative expression levels of *COL4A5* and *COL4A6* were determined for SK-N-SH, HEK-293, HT-1080, and SCC-25 cells using qRT-PCR. (*B*), Relative expression levels of *COL4A6* transcript variants *A* and *B* were determined for SCC-25 and HEK-293 cells using qRT-PCR. *$p < 0.05$

variants *A* and *B* were determined. For both HEK-293 and SCC-25 cells, *COL4A6 B* expression was higher than *COL4A6 A*. This relationship was only statistically significant in HEK-293 cells ($p < .0038$) (see Figure 3B). All four cell lines were shown to be unique

with respect to expression of *COL4A5* and *COL4A6*, thus providing an appropriate *in vitro* model system with which to study collagen gene regulation.

Evaluation of the Processing of Collagen Transcripts

There are three minimal promoters in just over a one kb segment of DNA, one for *COL4A5* and two alternative promoters to transcribe *COL4A6* (49). Because of the close proximity, recruitment of RNAPII and the transcription machinery to one promoter may affect transcription initiation at another. This is especially true for HT-1080 cells where *COL4A5* was found to be selectively activated. Full length *COL4A6* transcripts were not produced as determined by real time using primers specific to the 3′ end; however, it was unknown whether or not transcription was actually initiated at this site. Thus, primers targeting the 5′, middle, and 3′ portions of *COL4A5* and *COL4A6* transcripts were used to determine relative expression levels in all four cell lines. The results are displayed in Figure 4. *COL4A6* transcripts were never detected in HT-1080 cells, suggesting that transcription was not initiated despite the presence of RNAPII nearby at the *COL4A5* promoter. In SK-N-SH cells, transcripts were not detected for any of the primer sets used. Transcription initiation of *COL4A5* and *COL4A6*, therefore, appears to be followed by elongation and termination.

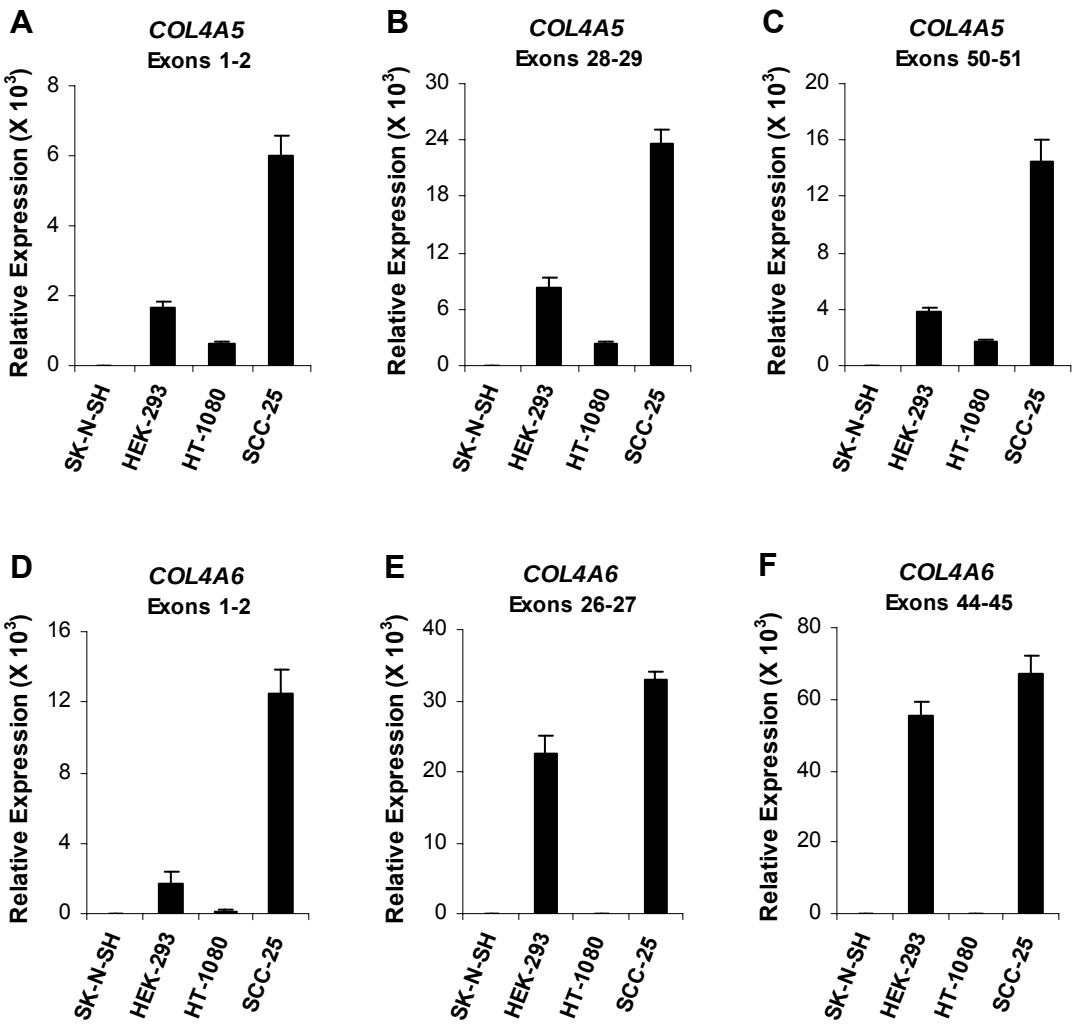

Figure 4 Evaluation of the processing of type IV collagen transcripts. qRT-PCR was used to determine relative expression levels of (*A*-*C*), *COL4A5* and (*D*-*F*), *COL4A6* for SK-N-SH, HEK-293, HT-1080, and SCC-25 cells using primer and probe sets targeting the (*A*, *D*), 5′, (*B*, *E*), middle, and (*C*, *F*), 3′ ends of the corresponding mRNA transcripts. Probes spanned exon-exon boundaries, as noted.

Identification of Conserved, Non-Coding Sequences

Previous researchers have evaluated the *COL4A5* and *COL4A6* promoter regions using *in vitro* model systems (47,49). In addition, transgenic mice bearing regions of the collagen promoter have been generated to assess tissue-specific expression patterns (50).

Based on these experiments, it was concluded that the promoter region is insufficient to drive expression of *COL4A5* and *COL4A6* in a cell-type specific manner. It has been hypothesized that unidentified, *cis*-acting elements downstream of the promoter region are necessary for activation of transcription. Indeed, it has been shown that several non-coding elements present in intronic DNA help control expression of other type IV collagen genes (42).

Each type IV collagen gene is over 200 kb in length, with the majority of the sequence representing intronic DNA. For example, the first intron in *COL4A5* is approximately 100 kb (87). Consequently, efforts to identify sequences that are involved in gene regulation have been hindered. In recent years, researchers have experienced success in utilizing comparative genomics to narrow the search for regulatory DNA. This is based on the premise that conserved DNA is selectively constrained and, therefore, thought to be functional (76-79). For the current study, the comparative genomics approach was deemed appropriate and used for the identification of putative regulatory elements involved in collagen regulation.

Human *COL4A5* and *COL4A6* genomic sequences were aligned to the genomes of 13 other species. Approximately 260 kb of *COL4A5* and the first 50 kb of *COL4A6* were evaluated. Because previous evidence exists to suggest that distal regulatory elements selectively activate *COL4A5*, the entire gene was included for analysis. The most current human genome sequence was used for alignment; however, since the time of analysis, this sequence has been updated and additional vertebrate sequences are now available.

A total of 38 sequences were identified as being conserved between the human reference and at least one other vertebrate using the established criteria of ≥ 70% homology and ≥ 100 bp. Of these, 13 conserved sequences overlapped exons of *COL4A5*; no sequences overlapped *COL4A6* exons. The remaining 25 conserved segments did not align with any cDNA sequences and are, therefore, likely to represent non-coding regions.

To prioritize a set of conserved, non-coding sequences, those regions conserved between the human and three or more non-primate vertebrates were chosen to be tested for ability to regulate gene expression *in vitro*. A total of 17 regions fit the criteria. Primers were designed to amplify each conserved, non-coding sequence individually with the exception of -43K and -110. Both primer sets were able to capture two sequences. The boundaries that define these conserved regions are not well-defined; therefore, to help ensure that all significant elements were captured, each forward and reverse primer was at least 25 bp upstream and downstream, respectively, of the conserved, non-coding sequence. Figure 5 displays the relative locations of conserved sequences.

Each region was cloned into the pGL3-promoter vector in the orientation of the gene in which it is located. Five conserved, non-coding sequences present in *COL4A6* were also cloned in the *COL4A5* direction. Additionally, a primer set was designed to amplify a non-conserved, non-coding sequence to assess how a random segment of DNA would behave. The pGL3-promoter vector was chosen so that, in addition to activators, repressor elements could be identified. Each sequence was transfected into all four cell lines and tested for the ability to regulate expression of the reporter gene in a cell-type

specific manner. Results are displayed in Figure 6. Activity of the pGL3-promoter vector was arbitrarily set at 100%, and conserved, non-coding sequences were graphed as % pGL3-promoter. Constructs with % pGL3-promoter activities above 100% were believed to have increased transcription of the reporter gene, whereas values below 100% represent inhibition of transcription. Constructs with % pGL3-promoter activities near 100% were said to have no effect on transcription.

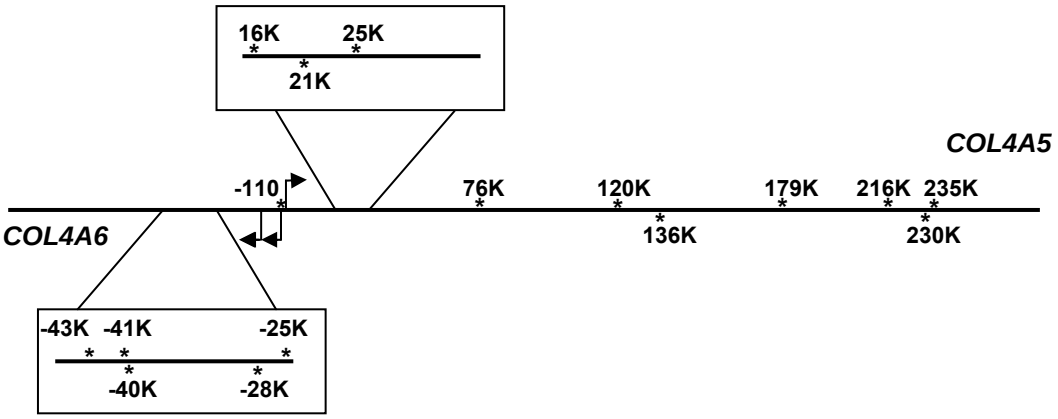

Figure 5 Relative locations of conserved, non-coding sequences in *COL4A5* and *COL4A6*. The relative locations of the 15 conserved sequences as well as the non-conserved, non-coding sequence are shown. Each sequence was named based on the relative location with respect to the *COL4A5* transcription start site.

In Vitro Analysis of the *COL4A5-COL4A6* Promoter

The pα6(-110) and pα5(-110) constructs represent conserved, non-coding DNA in the promoter region near the *COL4A5* and *COL4A6 A* TSSs. Percent pGL3-promoter activity of pα6(-110) was relatively unchanged in SK-N-SH, HT-1080, and SCC-25 cell lines. In HEK-293 cells, transcription was repressed with a % pGL3- promoter value of 51%. In the α5 direction, the *COL4A5-COL4A6* promoter activated transcription in only SK-N-SH cells with a % pGL3-promoter value reaching 191%. Expression of the

reporter gene remained relatively unchanged in HT-1080 cells and was decreased in HEK-293 and SCC-25 cells; % pGL3-promoter values were 115%, 58%, and 58%, respectively. Results are displayed in Figure 6A.

The DNA segment cloned upstream of the reporter gene in pα6(-110) and pα5(-110) included the CCAAT and CTC boxes present in the *COL4A5*-*COL4A6* promoter region. Figure 1 is a diagram of the relative locations of these motifs. Approximately 340 and 230 bp upstream and downstream, respectively, of the *COL4A5* TSS were captured. With respect to *COL4A6 A*, 520 and 50 bp upstream and downstream, respectively, were included in these constructs. Despite inclusion of TF binding sites and almost 600 bp of the promoter region, neither construct could significantly enhance transcription in HT-1080, HEK-293 or SCC-25 cells. It can be concluded, therefore, that elements outside of this region are required for efficient transcription in these cell types. In addition, distal-acting elements may be necessary as has been previously shown for cell types selectively expressing *COL4A5*. The possibility also exists that the SV40 promoter in the pGL3 vector negatively interfered with the *COL4A5*-*COL4A6* promoter. To resolve this, the collagen promoter should be cloned into a promoter-less vector and tested in transfection experiments. Finally, it is important to mention that the only cell type where the promoter exhibited activity was SK-N-SH cells in which neither gene was found to be transcribed. Although these results were unexpected, it is unlikely that the promoter is truly active when put in the context of chromatin.

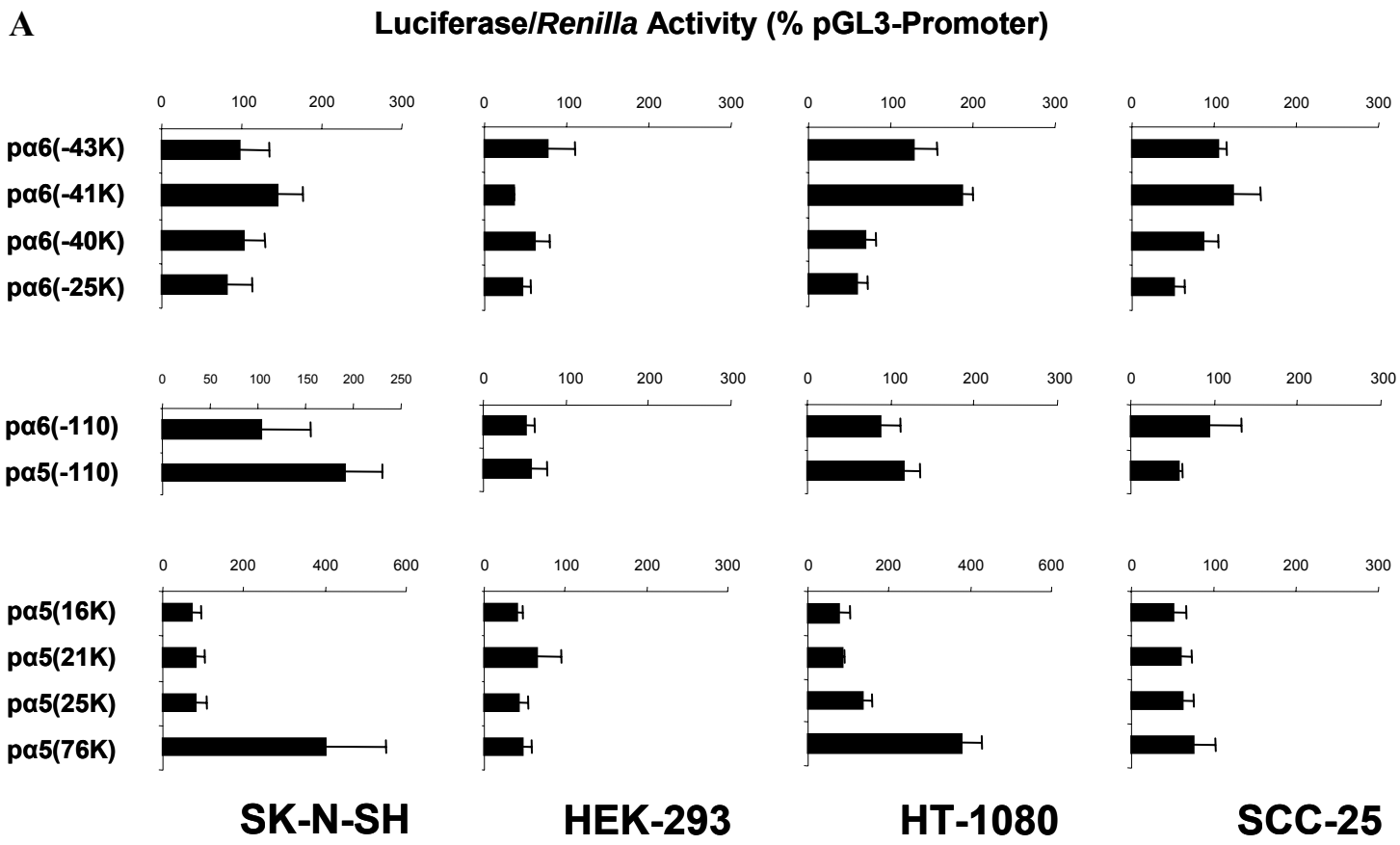

Figure 6 Percent pGL3-promoter activity of conserved, non-coding sequences. Conserved sequences were cloned into the pGL3-promoter vector in the *COL4A6* (pα6) and *COL4A5* (pα5) directions and transfected in to SK-N-SH, HEK-293, HT-1080, and SCC-25 cell lines. The % pGL3-promoter activity is shown for (*A*), conserved, non-coding sequences located in *COL4A6*, the promoter, and *COL4A5* as well as (*B*), conserved sequences located in *COL4A6*, but cloned into the vector in the α5 direction. In addition, promoter activities for six constructs containing conserved regions deep within *COL4A5* are displayed. Luciferase activities represent mean +/- S.E. (n = 3).

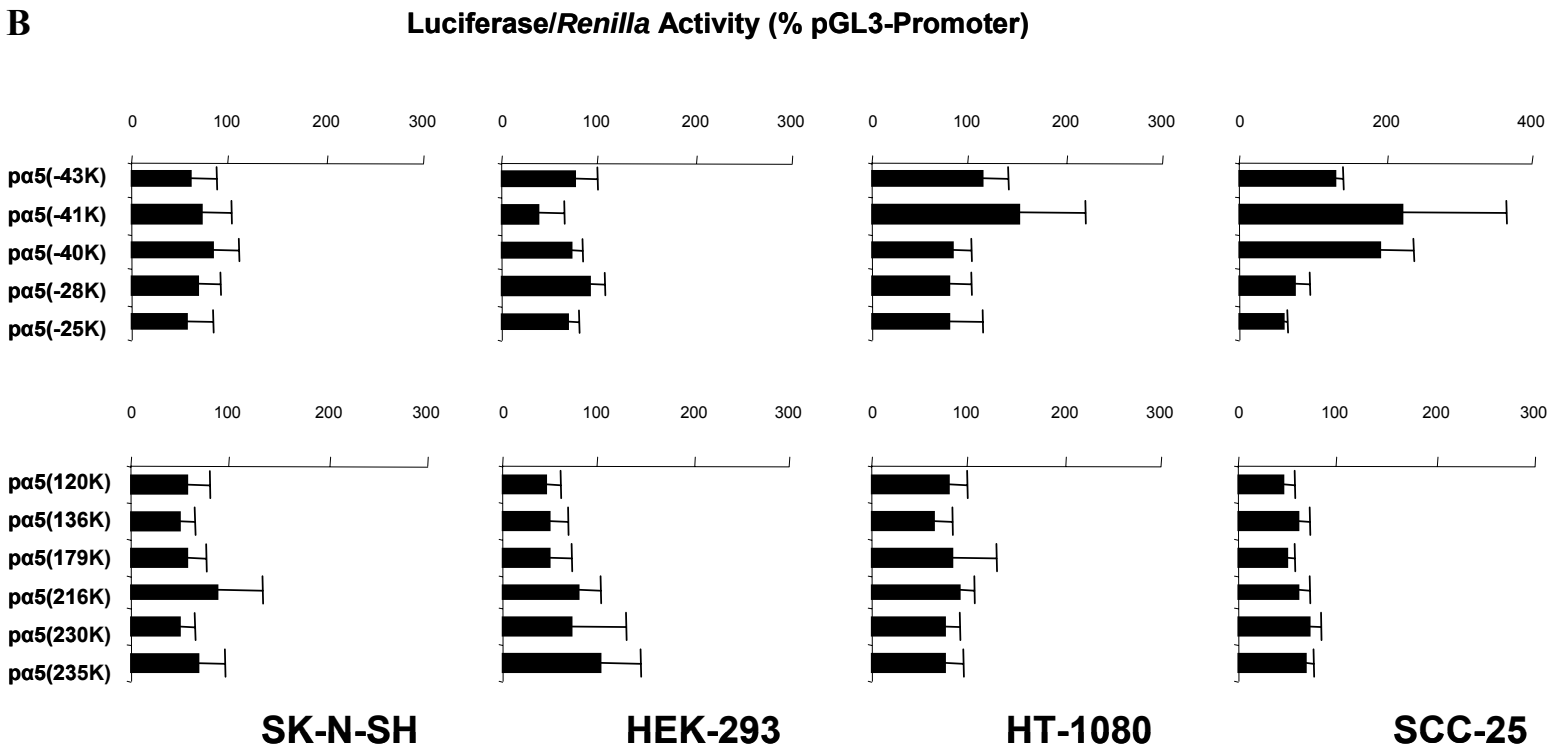

Figure 6 Continued

In Vitro Analysis of Conserved, Non-Coding Sequences

Constructs that exhibited differences in activity across cell types were identified and considered to be potentially significant in collagen regulation. Of the four pα6 constructs, only pα6(-41K) showed differences among cell types. The % pGL3-promoter activity of pα6(-41K) was decreased in HEK-293 cells (36%) and moderately increased in SCC-25 cells (123%). The pα6(-41K) construct had larger effects on transcription in SK-N-SH and HT-1080 cells with an increased activity of 144% and 187%, respectively. See Figure 6A. When pα6(-28K) was transfected into SCC-25 cells, the normalization vector did not exhibit activity levels above background, thus RLUs were unable to be calculated. For some unknown reason, these two vectors interacted to negatively impact transcription of the normalization vector; therefore, data from pα6(-28K) was not included for analysis.

The pα5(16K), pα5(21K), and pα5(25K) constructs did not exhibit cell-type specific differences. On the other hand, transcription was significantly increased when pα5(76K) was transfected into SK-N-SH and HT-1080 cells. The % pGL3-promoter activity reached 403% and 381%, respectively. Expression of the reporter gene decreased in HEK-293 and SCC-25 cells; % pGL3-promoter values were 48% and 77%, respectively. These results are shown in Figure 6A.

Conserved, non-coding elements in *COL4A6* were also cloned into the reporter vector in the *COL4A5* orientation; *cis*-acting DNA in *COL4A6* can also interact with the *COL4A5* promoter. Of the five constructs, two exhibited cell-type specific differences. In HT-1080 cells, the pα5(-41K) construct increased transcription by 50% over the empty

pGL3-promoter vector. The same results were seen for this construct in SCC-25 cells with a % pGL3-promoter activity of 225%. Transcription of the reporter gene was repressed in SK-N-SH and HEK-293 cells; % pGL3-promoter activity was 70% and 39%, respectively. The pα5(-40K) construct moderately decreased expression in all cell types except for SCC-25 cells where % pGL3-promoter activity reached 193% (Figure 6B).

An additional six constructs containing conserved elements in *COL4A5* were assessed for the ability to affect transcription. Surprisingly, none of these constructs displayed differences among cell types. See Figure 6B.

In addition to determining which constructs exhibited cell-type specific differences, the activity of all constructs within a cell type was assessed. For each cell line, % pGL3-promoter values for the majority of constructs fell within a particular range. For constructs within this range, therefore, the effects on transcription appear to be independent of the conserved, non-coding DNA transfected. The range for SK-N-SH and SCC-25 cells was approximately 50-90%, while the ranges for HT-1080 and HEK-293 cells were roughly 60-90% and 35-75%, respectively.

Constructs within these ranges were not investigated at this time, which included pα6(-41K), pα5(76K), pα5(-41K), and pα5(-40K) in HEK-293 cells and pα5(-41K) and pα5(-40K) in SK-N-SH cells. In HT1080 cells, pα5(-40K) was excluded for further analysis. The pα6(-41K) and pα5(76K) constructs did not appear to be important in SCC-25 cell lines. Significantly, the non-coding, non-conserved DNA segment, pα5(-16K), fit into this category, thus providing support for this method of analysis. It is important to

mention that this is not a fail-proof way to identify putative regulatory elements. The criteria used in this study to narrow the search are only a starting point, and the remaining conserved, non-coding elements should be investigated in the future.

The constructs that may contain functional DNA elements are pα6(-41K), pα5(76K), pα5(-41K), and pα5(-40K). The pα6(-41K) and pα5(76K) constructs enhanced transcription of the reporter gene in both SK-N-SH and HT-1080 cell lines. These elements could potentially activate transcription of *COL4A5* in HT1080 cells; however, their role in SK-N-SH cells is less clear. Significantly, in HT-1080 cells -41K also operated in the *COL4A5* direction as pα5(-41K) was shown to increase reporter gene activity. In SCC-25 cells, pα5(-41K) and pα5(-40K) were shown to enhance transcription, thus providing evidence to suggest that these elements could potentially activate *COL4A5* and/or *COL4A6*.

General conclusions can be drawn from this data. First, these functional constructs contain transcription factor binding sites. In the cell types in which the constructs function, transcription factors are available to bind to their recognition sequences and interact with the SV40 promoter to facilitate transcription. It can not be determined, however, whether or not these elements are functional when in their native chromatin environment. Additional studies will have to be carried out to assess this.

Surprisingly, only a few of the 21 different constructs evaluated in this study appear to be potentially functional. Previous experiments done using this type of assay system showed that the majority of conserved, non-coding sequences tested were able to exhibit cell-type specific differences (84,85). There are several possibilities that exist to

help explain the observed inconsistencies between the current and previous studies. In the current study, amplicon sizes of conserved, non-coding sequences were much smaller than those used in other experiments. Perhaps other regulatory elements nearby the conserved regions exist and were not captured in this study. Alternatively, the effects seen on transcription in the previous studies could be non-specific, i.e. resulted from random transcription factor binding sites that are likely to exist over several kb of DNA. In addition, the utilization of different vector constructs may affect results. This study used a vector with a promoter, while other studies have used the native promoter or promoter-less vectors.

As mentioned previously, constructs excluded from further analysis should be re-evaluated in the future. It is possible that these sequences act in concert with one another to activate or repress transcription of *COL4A5* and/or *COL4A6*. More insight can also be gained from placing these conserved, non-coding elements next to the native promoter and assessing how transcription is affected in different cell types. Nevertheless, initial analyses have provided a basis for future studies.

Evaluation of Histone Acetylation of Conserved, Non-Coding Sequences

Genome-wide acetylation patterns were recently mapped in human T cells. Gene-rich regions were found to be associated with the highest levels of histone acetylation, although acetylation across euchromatin domains was not uniform. Instead, hyperacetylated regions co-localized with active promoters as well as conserved, non-coding sequences thought to be biologically significant. In addition, these “acetylation

islands" co-localized with regulatory elements known to function in T cells. The authors concluded that acetylation of *cis*-acting elements prevents the spreading of heterochromatin and promotes an open chromatin structure to allow transcription factors to bind (56).

Based on the data summarized above, ChIP analysis was used to determine the acetylation state of the conserved, non-coding sequences identified in *COL4A5* and *COL4A6*. Antibodies directed against acetylated histone H3 were used to immunoprecipitate SCC-25, SK-N-SH, HEK-293, and HT-1080 chromatin. Primers were designed to amplify regions spanning -41K, -40K, and 76K; acetylation patterns were assessed in all four cell lines, and the results are displayed in Figure 7.

Both pα5(-41K) and pα5(-40K) constructs significantly enhanced transcription of the reporter gene in SCC-25 cells. ChIP analysis revealed the association of acetylated histone H3 with -41K and -40K as evidenced by the ability to amplify these DNA regions. In HT-1080 cells, acetylated histones were also associated with 76K and -41K, which was shown to function in both directions. According to this data, these conserved sequences are correlated with acetylated histones, thus providing additional evidence to support their role in activation of *COL4A5* and/or *COL4A6*.

In SK-N-SH cells, -41K and 76K were associated with acetylated histones. Because neither *COL4A5* nor *COL4A6* was transcribed in this cell line, it is unlikely that these conserved sequences are functional with regards to gene activation. Each cell type is independent of the other, thus these data do not negate the potential functional role of -41K and 76K in SCC-25 and HT-1080 cells.

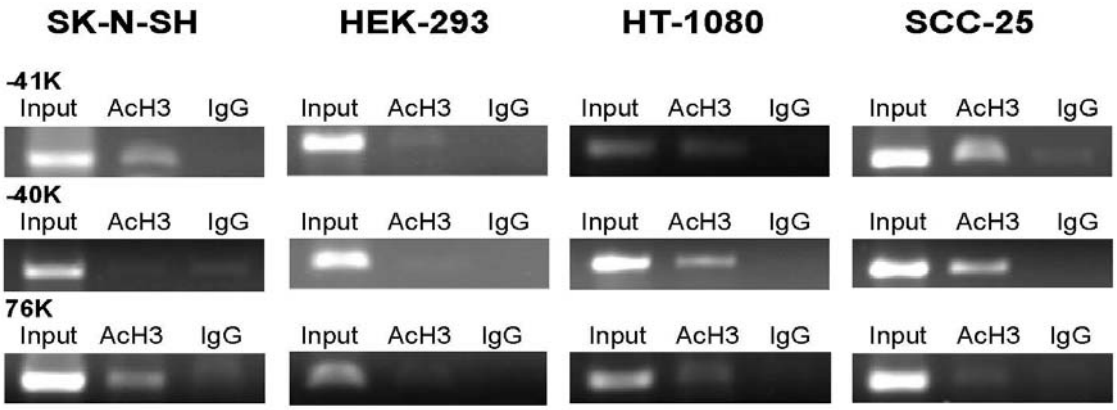

Figure 7 Chromatin immunoprecipitation analysis of conserved, non-coding sequences. ChIP analysis was used to determine the histone acetylation state of the conserved, non-coding sequences -41K, -40K, and 76K. Acetylation patterns were assessed in SK-N-SH, HEK-293, HT-1080, and SCC-25 cell lines.

Acetylated histones were also associated with other conserved sequences that did not appear to be functional in the cell-based assay including -41K in HEK-293, -40K in HT-1080, and 76K in SCC-25 cells. Consequently, transfection data was re-evaluated. In HEK-293 cells, -41K appears to be repressing transcription of the reporter gene more than what was considered to be average in this cell type. The % pGL3-promoter activities for pα6(-41K) and pα5(-41K) were 36% and 39%, respectively. Significantly, these were the two lowest values obtained in HEK-293 transfections. It can be argued, therefore, that acetylation patterns can aid in identifying functionally significant elements that may be missed otherwise. In addition, the association of acetylated histones with -40K and 76K in HT-1080 and SCC-25 cells, respectively, merits further investigation.

In summary, strong evidence exists to support the functional role of -41K and 76K in HT-1080 cells, -41K and -40K in SCC-25 cells, and -41K in HEK-293 cells. These sequences were shown to 1) be highly conserved across numerous species, 2) affect transcription of a reporter gene in a cell-based assay, and 3) be associated with

acetylated histone H3. Further experiments should be executed to further delineate their role in transcription of *COL4A5* and *COL4A6*.

CHAPTER IV

RESULTS:

THE ROLE OF EPIGENTIC MODIFICATIONS IN TYPE IV COLLAGEN GENE REGULATION

Eukaryotic cells mediate changes in gene expression by altering chromatin structure and interactions between the transcription machinery and DNA. Epigenetic marks including histone modifications and DNA methylation are intimately involved in this process. To determine the role epigenetics play in modulating expression of collagens, three sets of experiments on all four cell lines were carried out: 1) cells were treated with epigenetic modifying agents and the effect on expression of *COL4A5* and *COL4A6* was measured, 2) ChIP analysis was used to assess the chromatin structure surrounding the collagen promoter, and 3) the methylation status of CpG dinucleotides in the collagen promoter region was determined using bisulfite sequencing.

The *COL4A5*-*COL4A6* gene pair is located on the X chromosome. Consequently, the mechanism of X chromosome inactivation and the number of X chromosomes in each cell line must be considered when analyzing data from these experiments. Random X chromosome inactivation involves association of Xist, an untranslated RNA molecule, with the inactive X. Coating of the inactive X chromosome with Xist transcripts is then followed by histone modifications that facilitate heterochromatin formation and DNA methylation. The inactive X is also associated with late replication timing and is localized to the nuclear periphery to ensure a silenced state (88).

Based on cytogenetic analyses, SCC-25 cells were shown to have one X chromosome, HT-1080 cells consisted of XY, SK-N-SH cells contained two X chromosomes, and HEK-293 cells were associated with three X chromosomes (http://www.atcc.org). Because of dosage compensation, it can be assumed that in each cell only one X chromosome was active; this is true for all four cell lines.

Type IV Collagen Response to Epigenetic Modifying Agents

SK-N-SH, HEK-293, HT-1080, and SCC-25 cells were treated with TSA, 5-Aza, and a combination treatment of TSA plus 5-Aza. TSA inhibits HDAC activity, thereby forcing acetylation of histones, while treatment with 5-Aza results in global demethylation of CpG dinucleotides. Response to treatment was assessed using qRT-PCR and primers targeting the *COL4A5* or *COL4A6* transcripts; the *COL4A6* primers used did not discriminate between the transcript variants.

Any changes in gene expression in HT-1080 and SCC-25 cells were the result of epigenetic modifying agents acting on the single active X chromosome. The same is true for SK-N-SH and HEK-293 cells despite the fact that more than one X chromosome was present. In other words, it is unlikely that treatment with TSA or 5-Aza alone will affect expression of the collagen genes present on the inactive X chromosome. As outlined above, several layers of repression are responsible for keeping the inactive X in a silenced state. Results for each cell line are displayed in Figure 8.

SK-N-SH

Neither collagen gene was transcribed under normal conditions in SK-N-SH cell lines. Treatment with TSA dramatically increased expression of *COL4A5* and *COL4A6* approximately 16- and 7.7-fold, respectively. Relative expression of both genes slightly increased upon treatment with 5-Aza, although the effects were certainly not as dramatic as what was seen after TSA treatment; fold changes for *COL4A5* and *COL4A6* were 2.9 and 1.6, respectively. Finally, expression of both genes increased upon combination treatment; a 34-fold increase for *COL4A5* and a 9.9-fold increase for *COL4A6* were seen. This increase appears to be synergistic, especially for *COL4A5*; perhaps the combined treatment was enough to relieve repression of the collagen genes on the inactive X chromosome. Results are displayed in Figure 8A.

Based on these observations, repression of *COL4A5* and *COL4A6* appears to be mediated through chromatin structure. More specifically, forced acetylation was enough to de-repress both genes, suggesting that an unidentified repressive complex operates to support a tightly packed chromatin structure that involves the removal of acetyl groups from histones. Moreover, repression appears to be independent of DNA methylation as treatment with 5-Aza had little effect on expression levels.

SCC-25

Although *COL4A5* and *COL4A6* were shown to be actively transcribed in SCC-25 cells, expression of both transcripts increased slightly upon treatment with 5-Aza. The relative fold changes as compared to the control were approximately 2.5 and 3.8 for

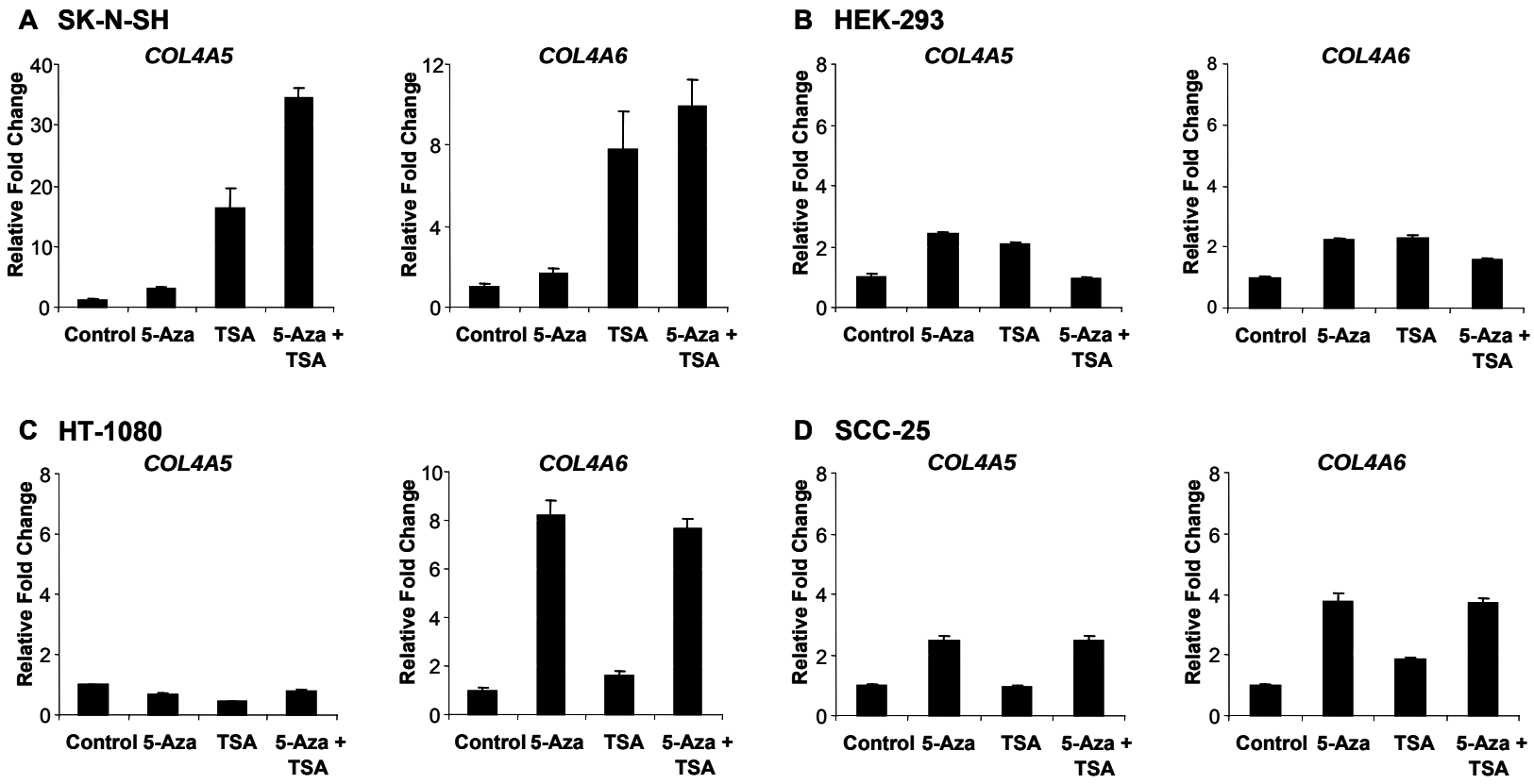

Figure 8 Type IV collagen gene expression following treatment with epigenetic modifying agents. (*A*), SK-N-SH, (*B*), HEK-293, (*C*), HT-1080, and (*D*), SCC-25 cell lines were treated with 5-Aza, TSA, and 5-Aza plus TSA. Relative expression levels of *COL4A5* and *COL4A6* were determined using qRT-PCR.

COL4A5 and *COL4A6*, respectively. Treatment with TSA resulted in a slight increase in fold change of *COL4A6* (1.8), which may reflect complete de-repression. Expression of *COL4A5* was unchanged following treatment with TSA suggesting that histones in this area may already be fully acetylated. Combined treatment resulted in a moderate increase in fold changes for both genes; *COL4A5* and *COL4A6* fold changes were 2.5 and 3.7, respectively. Results are presented in Figure 8D.

These data taken together suggest that under normal growth conditions in SCC-25 cells, histones associated with the collagen promoter are in the acetylated state to promote transcription. DNA methylation may act to repress collagen expression in a sub-population of cells, although it is possible that the slight increase in expression following treatment resulted from an indirect effect, *i.e*. change in expression at another locus. More experiments are necessary to determine the role, if any, of DNA methylation in this cell type.

HEK-293

HEK-293 cells were shown to produce both transcripts, thus the effects on expression following treatment with epigenetic modifying agents were only moderate. In fact, the general trends resembled those seen for SCC-25 cells. Relative fold changes of *COL4A5* and *COL4A6* increased approximately 2.4- and 2.2-fold, respectively, following treatment with 5-Aza. Whether or not this slight increase reflects actual methylation of DNA at the collagen promoter under normal conditions can not be determined from this data; however, further investigation is warranted. TSA treatment resulted in a roughly 2-

fold increase in expression of both *COL4A5* and *COL4A6*, which is, again, most likely due to complete de-repression.

Relative fold change of *COL4A6* increased approximately 1.6 upon combination treatment, while *COL4A5* was unchanged. Based on the trends observed for treatment with 5-Aza or TSA alone, increased expression would be expected following combination treatment. This lack of increase may be a result of toxicity leading to cell death. See Figure 8B for results.

These data support the hypothesis that histone acetylation or DNA demethylation alone is not enough to activate collagens on the inactive X chromosome. Because there are two inactive X chromosomes in HEK-293 cells, a much larger increase in expression following treatment would be expected.

HT-1080

In HT-1080 cells, relative fold changes of *COL4A5* compared to the control did not increase after treatment with 5-Aza, TSA, or 5-Aza plus TSA. It does not appear, therefore, that DNA methylation affects expression of *COL4A5*; the lack of increase in expression following treatment with TSA may reflect fully acetylated histones in this area. Significantly, combination treatment and 5-Aza alone resulted in a 7.6- and 8.2-fold increase, respectively, in *COL4A6* expression. No change was seen following treatment with TSA. These results suggest that DNA methylation plays a role in repression of *COL4A6* that is independent of histone acetylation. Results are shown in Figure 8C.

Chromatin Immunoprecipitation Analysis of *COL4A5-COL4A6* Promoter

ChIP analysis allows the relationship between a particular protein and DNA to be determined. Histone modifications as well as DNA-bound proteins can be assessed using this technique. Antibodies directed against RNAPII, acetylated histone H3, and dimethyl histone H3K9 (H3K9me2) were used in the current study to look at DNA surrounding the *COL4A5* and *COL4A6* promoters. In addition, antibodies against HP1 were used in SK-N-SH cells because of the established association between HP1 and the silencing of genes in euchromatin. Acetylated histone H3 correlates with open chromatin that is favorable for transcription, while H3K9me2 targeted to the promoter region of genes is associated with repression. These two antibodies, therefore, were deemed to be appropriate markers to assess the role of chromatin in modulating collagen expression. Significantly, in SK-N-SH and HEK-293 cells ChIP analysis does not distinguish between the active and inactive X chromosomes; this fact must be considered when analyzing results.

ChIP analysis can determine the position of a protein along a stretch of DNA based on relative amounts of enrichment seen for overlapping primer sets. The most accurate method to quantify enrichment is qRT-PCR. Standard PCR can also be used; however, this technique is only semi-quantitative. The current study used PCR, thus determination of the precise location of a modification or protein binding was not attempted. Locations of primer sets used in this study are displayed in Figures 9A and 10A. ChIP results can be found in Figures 9B, C and 10B, C and Table 7 summarizes these findings.

Table 7 Summary of chromatin immunoprecipitation analysis of the *COL4A5-COL4A6* promoter. Results of ChIP analysis for SK-N-SH, HEK-293, HT-1080, and SCC-25 cell lines are summarized. Positive and negative associations are designated as + and -, respectively.

SK-N-SH	+6B	6B	-6B	6A	5A	+5A
H3K9me2	+	+	+	+	+	+
RNAPII	-	-	-	-	-	-
AcH3	-	-	-	-	-	-

HEK-293	+6B	6B	-6B	6A	5A	+5A
H3K9me2	+	+	+	+	+	+
RNAPII	+	+	+	+	+	+
AcH3	+	+	+	+	+	+

HT-1080	+6B	6B	-6B	6A	5A	+5A
H3K9me2	+	+	+	+	-	+
RNAPII	-	-	-	+	+	+
AcH3	+	+	-	+	+	+

SCC-25	+6B	6B	-6B	6A	5A	+5A
H3K9me2	+	+	+	-	-	-
RNAPII	+	+	+	+	+	+
AcH3	+	+	+	+	+	+

SK-N-SH

The entire promoter region was found to be associated with H3K9me2 in SK-N-SH cells as evidenced by the presence of PCR bands for all primer sets. Enrichment of RNAPII and acetylated histone H3 was not seen. Results are presented in Figures 9B and 10B. In addition, PCR bands for HP1 were present using primers 6B and +5A, whereas no products were produced with the remaining primer sets (see Figures 9C and 10C). This suggests HP1 may be involved in silencing both genes, perhaps through associations with H3K9me2 at distinct locations; the possibility also exists that enrichment resulted from the association of HP1 with the inactive X. Indeed, HP1 has been shown to be involved in heterochromatin formation of the inactive X chromosome through binding methyl H3K9 (89). Nonetheless, these findings suggest that the collagen promoter is in a

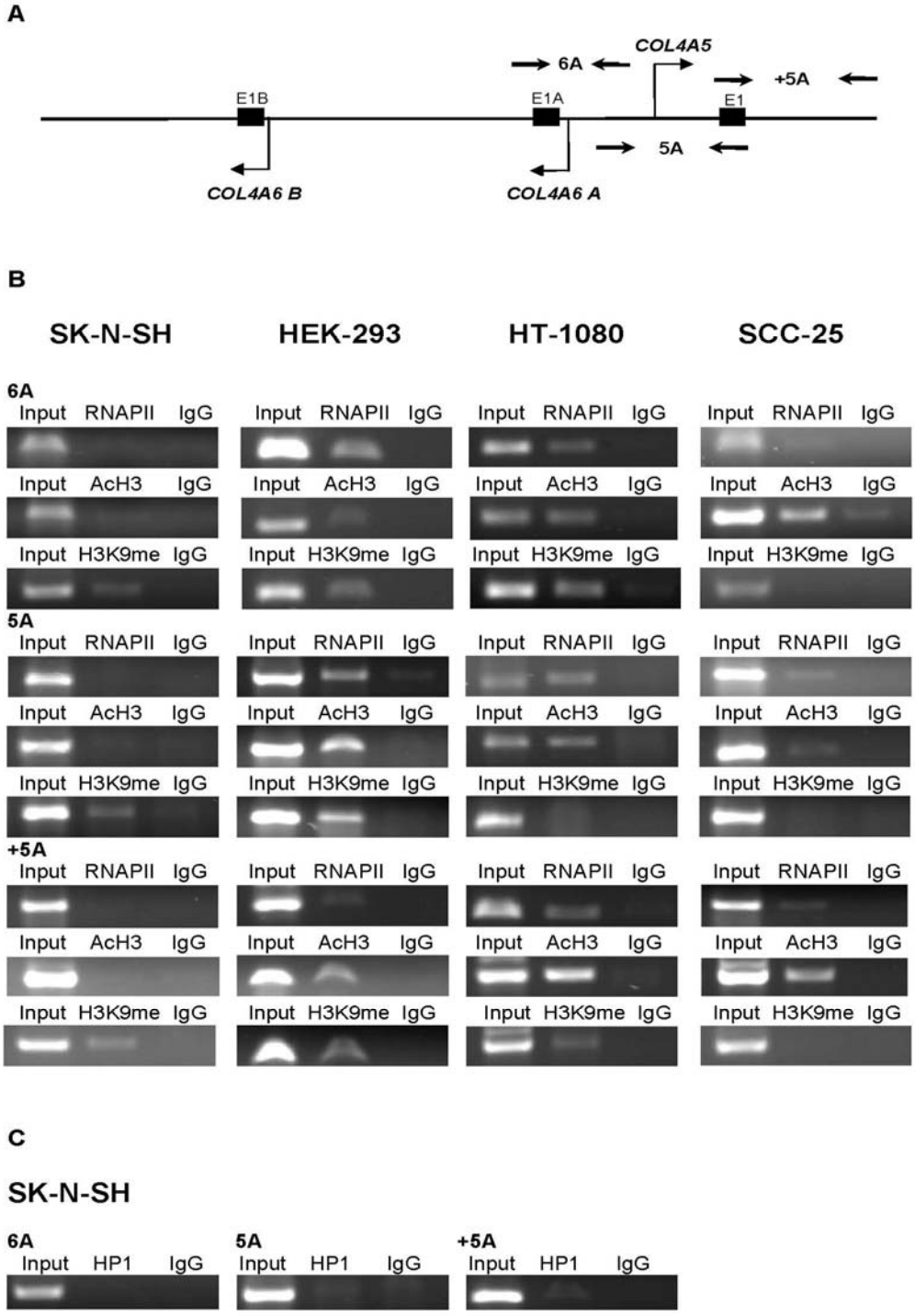

Figure 9 Chromatin immunoprecipitation analysis of *COL4A5* and *COL4A6 A* promoters. (*A*), Relative locations of primer sets used to analyze *COL4A5* and *COL4A6 A* regions are shown. ChIP analysis using antibodies against (*B*), RNAPII, H3K9me, and acetylated histone H3 (AcH3) was carried out in SK-N-SH, HEK-293, HT-1080, and SCC-25 cell lines. In addition, (*C*), HP1 was used in SK-N-SH cells.

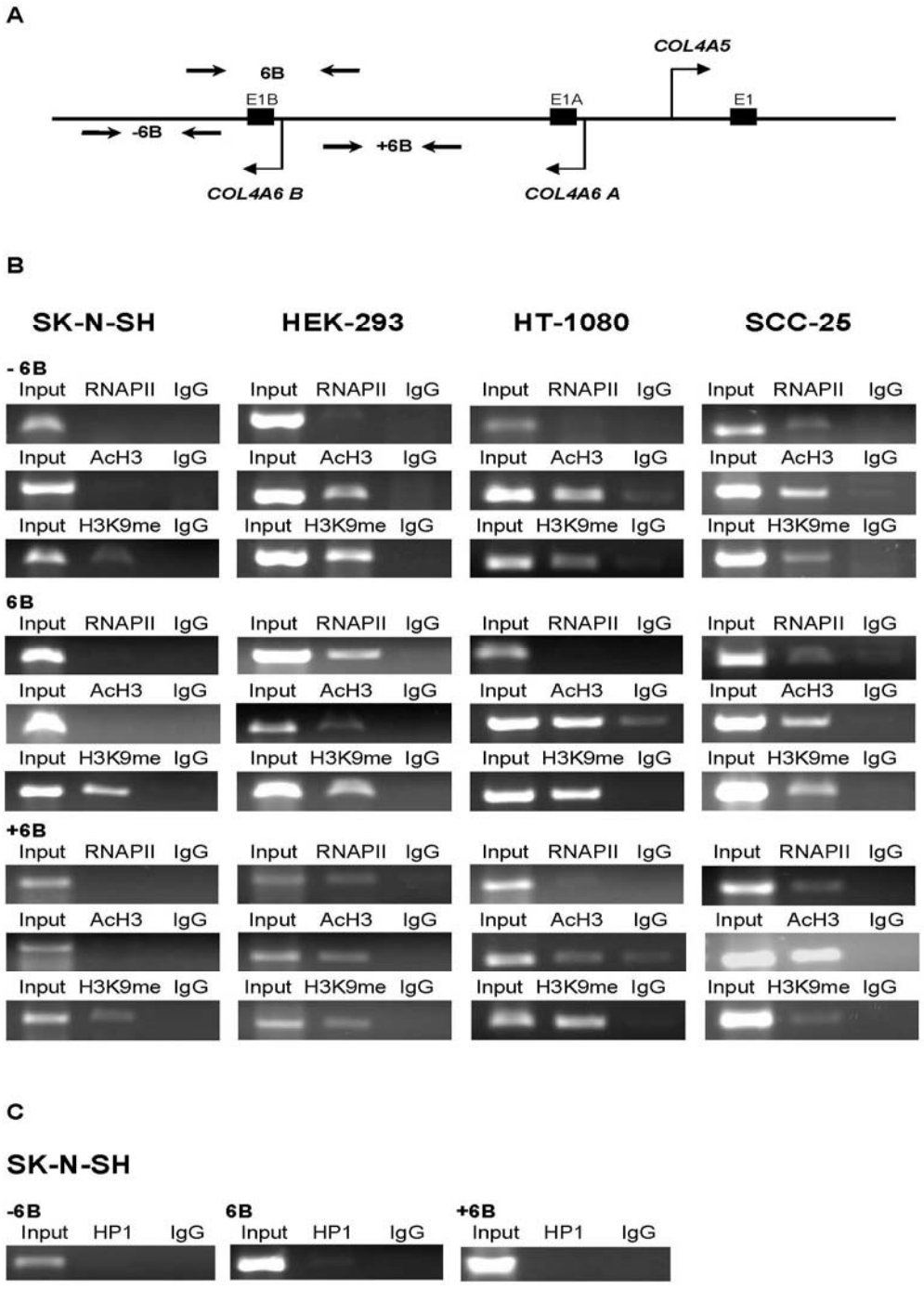

Figure 10 Chromatin immunoprecipitation analysis of *COL4A6 B* promoter. (*A*), Relative locations of primer sets used to analyze the *COL4A6 B* region are shown. ChIP analysis using antibodies against (*B*), RNAPII, H3K9me, and acetylated histone H3 (AcH3) was carried out in SK-N-SH, HEK-293, HT-1080, and SCC-25 cell lines. In addition, (*C*), HP1 was used in SK-N-SH cells.

closed chromatin state that is important in preventing RNAPII and the transcription machinery from binding to initiate transcription.

HEK-293

In HEK-293 cells, the *COL4A5* and *COL4A6* promoter region was found to be associated with RNAPII, acetylated histone H3, and H3K9me2; all six primer sets produced PCR products for each antibody. Results are displayed in Figures 9B and 10B.

RNAPII and acetylated histone H3 were most likely only associated with the active X chromosome during transcription of both collagen genes. Histone H3K9me2 was probably associated with the inactive X chromosomes, but this modification could also be present on histones surrounding the collagen promoter on the active X. In the population as a whole at any given moment, it is unlikely that every single cell was actively transcribing *COL4A5* and *COL4A6*. It has been previously shown that when transcription is not re-initiated, histones are modified to facilitate formation of closed chromatin. This prevents cryptic initiation and provides a mechanism to control the rate of transcription (65). Therefore, enrichment of H3K9me2 may be correlated with the inactive and well as the active X chromosomes; association with the active X reflects cells in different stages of transcription.

HT-1080

HT-1080 cells were characterized by a unique pattern of enrichment, most likely reflecting activation of only *COL4A5*. Results are shown in Figures 9B and 10B. RNAPII

was found to be associated with *COL4A5* and *COL4A6 A* promoters, as evidenced by PCR products obtained using 6A, 5A, and +5A primer sets. *COL4A6* transcripts were not produced in this cell type, thus association of RNAPII with the *COL4A6 A* promoter may be a result of limited resolution. When chromatin was immunoprecipitated with H3K9me2, all primer sets generated products except for 5A, which spans the *COL4A5* TSS. Acetylated histone H3 was found to be associated with the entire promoter region except for the area captured by the -6B primer set; the intensity of the product was equal to that produced by IgG, which is suggestive of non-specific pull-down. Lack of histone acetylation at DNA surrounding -6B suggests that this region is important in preventing transcription of *COL4A6 A* and *B*, whereas lack of H3K9 methylation at *COL4A5* TSS may be necessary to keep chromatin open for transcription initiation.

SCC-25

Following immunoprecipitation of SCC-25 chromatin, PCR revealed the presence of RNAPII across the *COL4A5* and *COL4A6* promoters. Acetylated histone H3 was also found to be associated with this region as PCR products were obtained for all primer sets. These results are indicative of active transcription of both genes. Significantly, products were present only using primer sets +6B, 6B, and -6B following immunoprecipitation with H3K9me2. Enrichment of H3K9me2 was not found surrounding *COL4A5* and *COL4A6 A* promoters, suggesting that differences in transcription rates exist. See Figures 9B and 10B for results.

Analysis of DNA Methylation in *COL4A5-COL4A6* Promoter

Treatment of cell lines with the epigenetic modifying agent 5-Aza provided evidence to suggest that DNA methylation may play a role in repression of collagen expression. To investigate this further, bisulfite sequencing of the entire promoter region was carried out to determine the methylation status of CpG dinucleotides. The results are presented in Figure 11.

Based on the established criteria, MethPrimer failed to detect CpG islands in the *COL4A5* and *COL4A6* promoter regions. Nevertheless, bisulfite sequencing was carried out. Four primer sets were used to capture the majority of CpGs in the area. BS1.1, BS1.2, BS2.1, and BS2.2 captured 4, 7, 15, and 3 CpG dinucleotides, respectively. CpGs are concentrated in the area encompassed by the BS2.1 primer set. See Figure 11A.

In SCC-25 cells, methyl CpGs were sparse making it very unlikely that DNA methylation plays a functionally significant role. A relatively fair amount of methyl CpGs were present across the entire promoter region in HEK-293 cells. DNA was also methylated in SK-N-SH cells, although not as many methyl CpGs were detected as compared to HEK-293 cells; complete absence of DNA methylation was seen for the region captured by the BS1.1 primer set. These findings most likely reflect methylation of the promoter region on the inactive X chromosomes. The presence of methyl CpGs for each amplicon would be expected if both the active and inactive chromosomes were associated with DNA methylation. Amplification of the entire promoter region in fewer primer sets would be helpful as this would allow the methylation profile of a contiguous piece of DNA to be assessed.

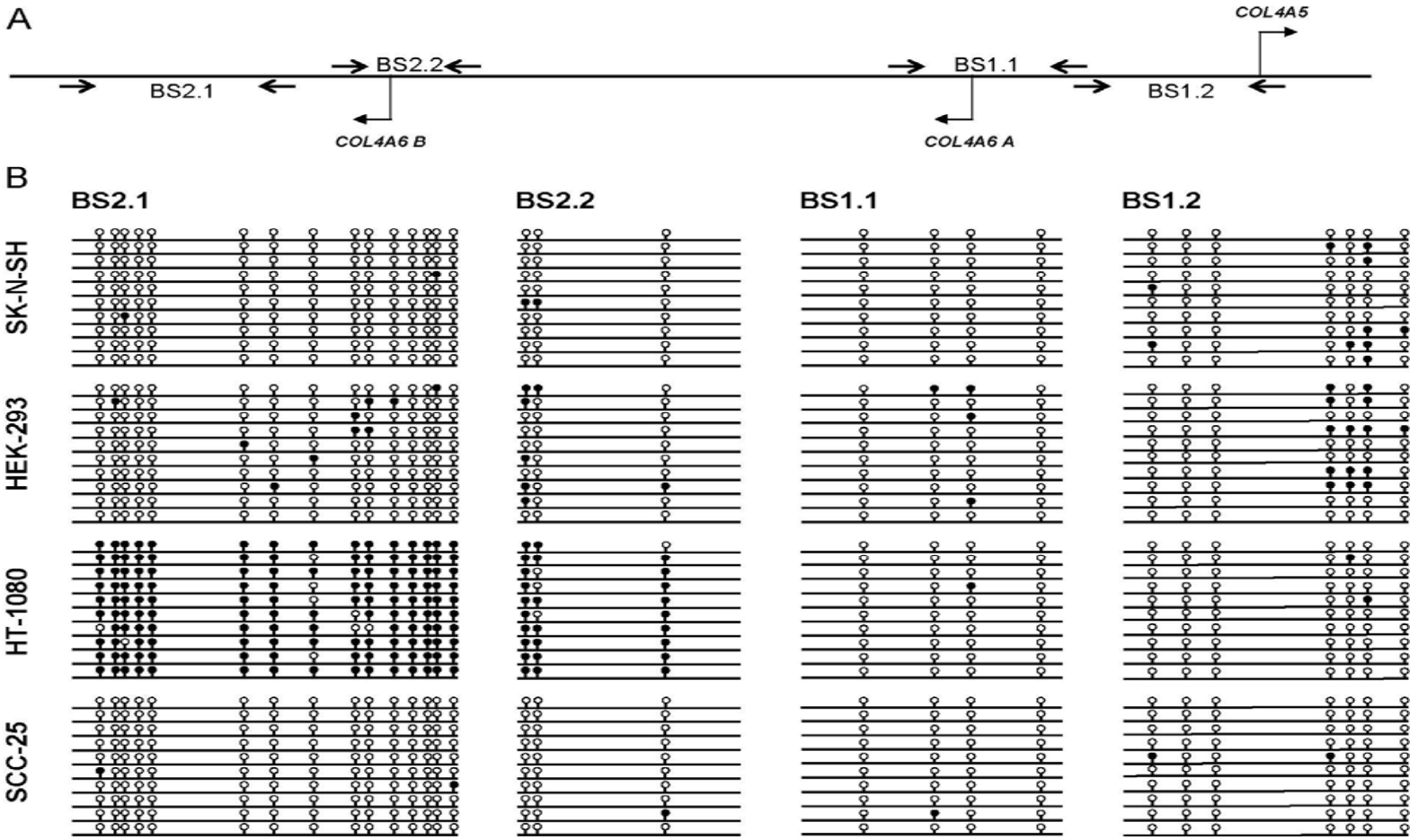

Figure 11 DNA methylation profiles of the *COL4A5-COL4A6* promoter as determined by bisulfite sequencing. (*A*), Relative locations of primer sets used to determine the methylation status of CpG dinucleotides in (*B*), SK-N-SH, HEK-293, HT-1080, and SCC-25 cell lines; open circle, CpG; filled circle, methyl CpG.

DNA surrounding the *COL4A6 B* promoter region was found to be heavily methylated in HT-1080 cells (see results for primer sets BS2.1 and BS2.2 presented in Figure 11B). This observation, along with absence of methylation at the *COL4A5* promoter, provides strong evidence to suggest that methylation of CpGs plays a functional role in silencing expression of the *COL4A6 B* transcript. Significantly, the *COL4A6 A* promoter region was free of methyl CpGs (see results for primer set BS1.1). The question then becomes how do HT-1080 cells repress transcription of *COL4A6 A*? The possibility exists that methylation of the *B* promoter indirectly affects expression at the *A* promoter; however, other mechanisms may also be at work in this cell line.

CHAPTER V

CONCLUSION

Regulation of Type IV Collagen Genes Occurs at Transcription Initiation

In a recent study scientists reported that, contrary to what has previously been known, the majority of protein-coding genes in human cells experience transcription initiation; the promoter region of this class of genes is characterized by histone modifications associated with active transcription and the presence of RNAPII, although full length mRNA transcripts are undetectable. Additionally, these genes lack markers correlated with elongation, suggesting that regulation occurs post-transcription initiation (90).

The study reported herein provides evidence that, although the majority of genes may experience transcription initiation, type IV collagens do not fit into this category. Presence of RNAPII and full length transcripts were detected only when *COL4A5* and/or *COL4A6* were actively being transcribed. In SK-N-SH cells RNAPII was absent across the entire promoter region, and neither collagen transcripts were detected even when primers targeting the 5′ ends were used. The same results were seen for HT-1080 cells with respect to *COL4A6*. These findings corroborate previous research (49). Moreover, epigenetic modifications discussed in the following sections maintain the view that regulation of collagens occurs at transcription initiation.

Activation of *COL4A5* and *COL4A6* in SCC-25 and HEK-293 Cells

According to real time data, *COL4A5* and *COL4A6 A* and *B* transcripts were expressed in SCC-25 and HEK-293 cell lines. As expected for genes being actively transcribed, ChIP assays revealed the presence of RNAPII and acetylated histone H3 across the promoter region. A model is proposed for regulation of this gene pair in SCC-25 and HEK-293 cells based on these observations and what is currently known about mechanisms for gene activation (see Figure 12) (51,52). Briefly, in the absence of binding of the bi-directional activator, which has been shown to be necessary for bi-directional transcription of *COL4A5* and *COL4A6*, both genes are silent (49). The bi-directional activator binds to the promoter region in response to a stimulus and functions to recruit chromatin remodelers and HATs. As a result, nucleosomes become more mobile and contacts between DNA and histones are loosened, thus allowing for components of the PIC to assemble. Finally, RNAPII is recruited and transcription is initiated.

This is a simplified view, as it is likely that other transcription factors bind upstream of both promoters and interact with the bi-directional activator and PIC components to recruit necessary players. Two conserved, non-coding sequences, -41K and -40K, were shown to activate transcription of a reporter gene in SCC-25 cells and were also found to be associated with acetylated histone H3. These sequences may contain functional elements necessary for expression of *COL4A5* and/or *COL4A6* in SCC-25 cells. In addition to activators, repressor complexes may act to alter expression

of collagen genes. The -41K construct repressed transcription in HEK-293 cells and may be functional *in vivo*.

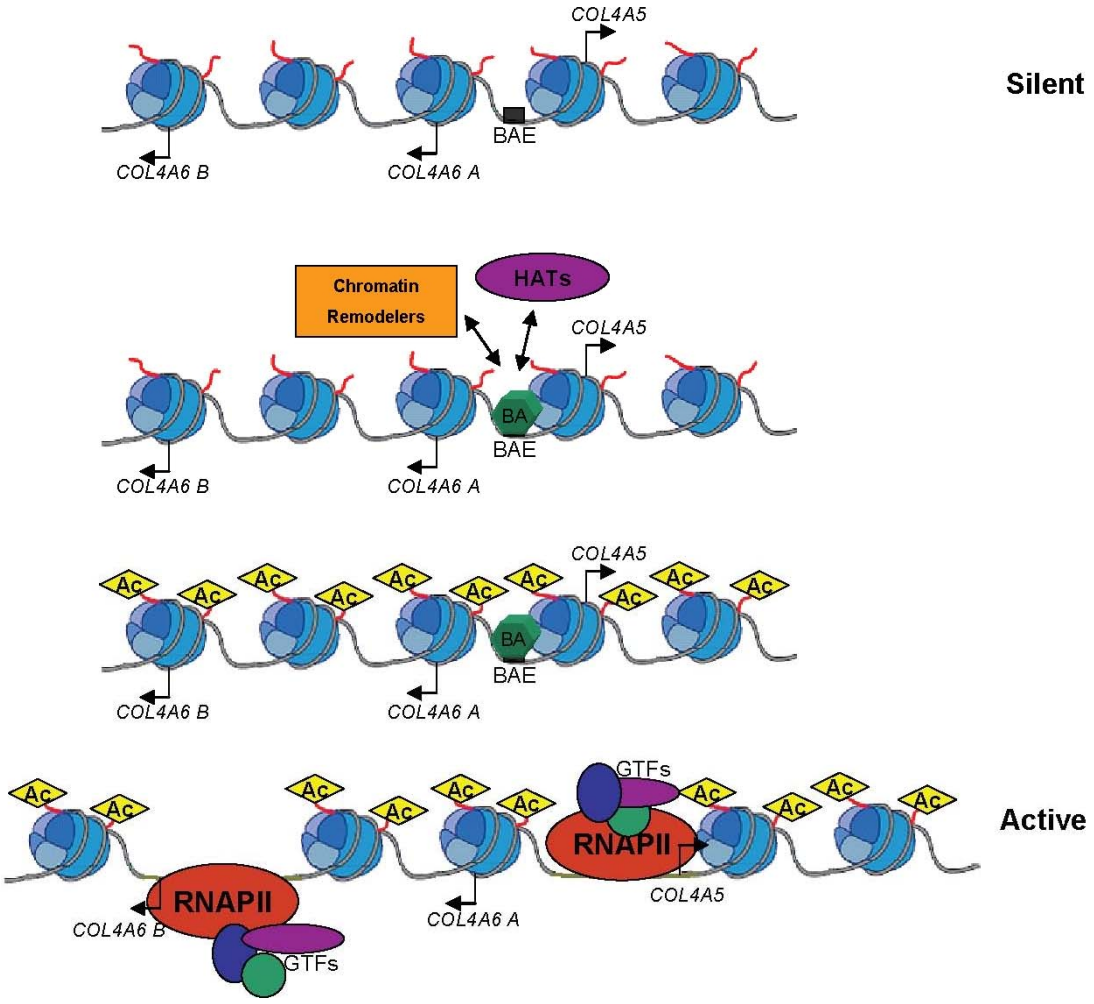

Figure 12 Proposed model for activation of collagen genes in SCC-25 and HEK-293 cells. *COL4A5* and *COL4A6* are silent in the absence of the bi-directional activator. Upon binding, the activator functions to recruit chromatin remodelers and HATs. Histone tails become acetylated allowing for formation of PIC and subsequent transcription initiation. See text for more details. BA, bi-directional activator; BAE, bi-directional activator element; Ac, acetyl group.

The model for activation presented above does not depict the apparent differences seen in expression of collagens as determined by real time. Differences in *COL4A5* will be addressed here, and the *COL4A6* transcript variants will be discussed in the following section.

Relative expression of *COL4A5* in SCC-25 cells was approximately four times that which was seen in HEK-293 cells. This finding may be a result of differences in transcription rates; evidence to support this comes from ChIP analysis and treatment data. As mentioned previously, acetylation of histone tails is directly correlated with the amount of transcription, whereas methylation of histone H3K9 is associated with closed chromatin. In HEK-293 and SCC-25 cells, the only region that was found to not be associated with H3K9me2 was the *COL4A5* promoter in SCC-25 cells. This suggests that transcription is continuously initiated and explains the higher expression levels seen in SCC-25 cells. Moreover, upon treatment with the HDAC inhibitor, TSA, *COL4A5* expression levels were unchanged in SCC-25 cells, thus supporting the view of fully acetylated histones. In contrast, forced acetylation did increase expression of both collagens in HEK-293 cells and *COL4A6* in SCC-25 cells approximately two-fold as compared to the control. These data taken together support the idea that expression levels correlate with histone modifications and transcription initiation. Although these data correlate nicely, post-transcription mechanisms of regulation may still contribute to steady state transcript levels.

It is important to mention that gene activation is not necessarily a direct result of histone modification. Instead, it is more likely that an upstream stimulus, i.e. binding of

the bi-directional activator, mediates its effects through changes in chromatin structure. So, what upstream signal accounts for differences in expression of *COL4A5* and what is the functional consequence of this? This will certainly be of interest when studying collagen expression in the future.

COL4A6 *Transcript Variants*

The *COL4A6* transcript variants *A* and *B* differ only in their signal peptide sequences, the functional significance of which is currently unknown (48). In both SCC-25 and HEK-293 cells the relative expression of *COL4A6 B* was higher than *COL4A6 A*, although only statistically significant in HEK-293 cells. The chromatin environment surrounding the *COL4A6 A* promoter in both cell lines appears to be favorable for transcription. In fact, enrichment for acetylated histone H3 was greater for the *COL4A6 A* promoter than the *COL4A6 B* promoter in SCC-25 cells; this region was not associated with methylated H3K9. Therefore, some other mechanism that is not dependent upon histone modifications may exist to account for differences in relative expression. In one scenario, the megadalton size PIC assembled at the *COL4A5* promoter could simply be too large to allow another complex to assemble only 292 bp away. Alternatively, PIC assembly at the *COL4A5* promoter could result in a conformational change in DNA that is unsuitable for protein binding at the *COL4A6 A* promoter.

Silencing of *COL4A5* and *COL4A6* in SK-N-SH Cells

In SK-N-SH cells, transcription initiation is prevented by histone modifications that facilitate heterochromatin formation. The *COL4A5* and *COL4A6* promoter regions were characterized by: 1) absence of RNAPII, 2) absence of acetylated histone H3, and 3) presence of H3K9me2. These findings correlate well with real time data, which failed to detect either transcript even using primers targeted to the 5′ end. Forced acetylation relieved repression of both genes as determined by treatment of cells with TSA, while 5-Aza had little or no effect on expression. This suggests that DNA methylation does not play a role in silencing collagens in this cell line and is supported by bisulfite data, which revealed very few methylated CpGs. Those CpGs that were methylated were probably associated with the inactive X chromosome.

These data allow for a working model of repression to be proposed (see Figure 13). An, as yet, unidentified trans-acting repressor binds to specific sequences in the promoter region and functions to recruit HDACs. It is only after deacetylation of histones that HMTs can methylate histone H3K9; presumably, other repressive marks work in conjunction with H3K9me2. The resulting closed chromatin creates an environment that is unfavorable for binding of the transcription machinery. Constitutive expression of the repressor complex ensures stable repression of *COL4A5* and *COL4A6*.

ChIP analysis showed enrichment of HP1 in the promoter region, suggesting that HP1 may be involved in repression of collagens in SK-N-SH cells. HP1 is known to silence genes through binding of H3K9me2. In contrast to the “spreading” mechanism important in silencing constitutive heterochromatin, HP1 associates with one nucleosome

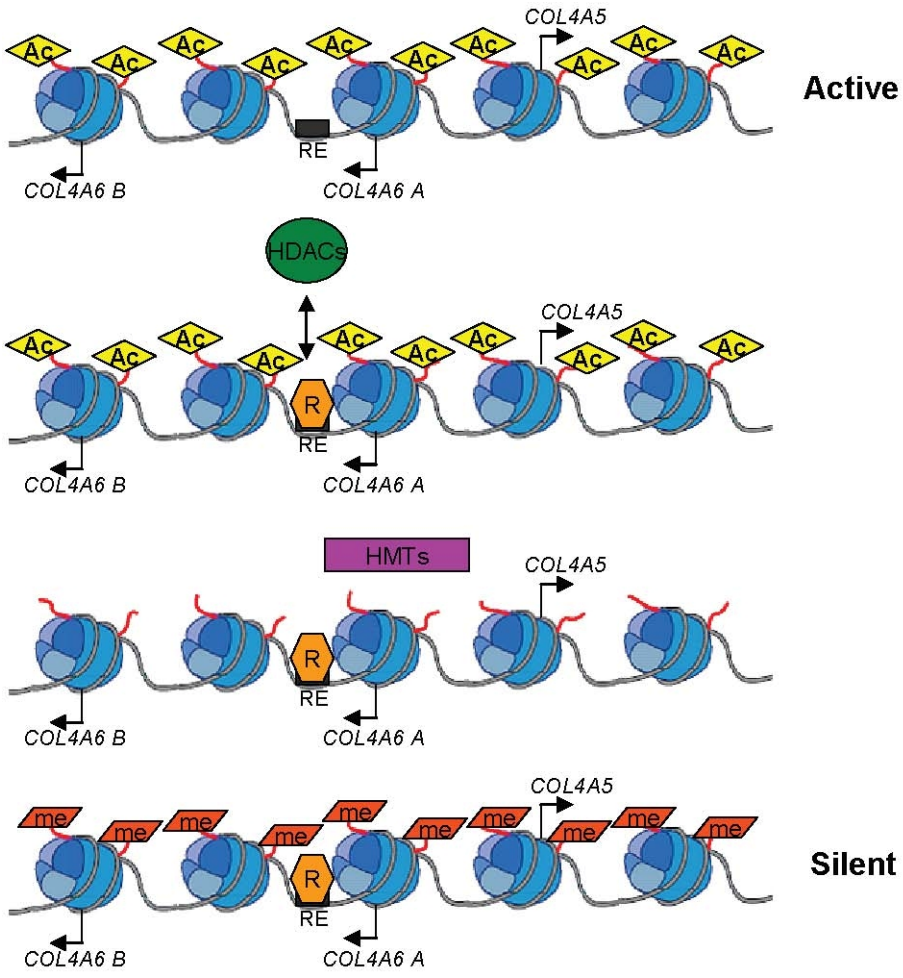

Figure 13 Proposed model for silencing of collagen genes in SK-N-SH cells. The repressor complex binds to the promoter region and recruits HDACs. Histones become deacetylated allowing for methylation of histone H3K9 by HMTs. *COL4A5* and *COL4A6* are stably silenced through constitutive expression of the repressor complex. See text for more details. R, repressor; RE, repressor element; Ac, acetyl group; me, methyl group.

near the promoter region to silence genes in euchromatin (73,91). This is evidenced in the current study by enrichment of HP1 at two discrete locations, one near the TSS of *COL4A5* and another near the TSS of *COL4A6 B*. The precise mechanism by which HP1 silences genes is not well understood, although HP1 has been shown to interact with repressors as well as components of the transcription machinery (92,93). In addition, HP1

has recently been shown to mediate interactions between DNMT1 and the HMT, G9a (94). Although HP1 may play a role in silencing collagens in SK-N-SH cells, enrichment of HP1 could be due to associations with the silent X chromosome. ChIP analysis should be carried out on the other three cell lines before more conclusions can be drawn.

In addition to those discussed here, other mechanisms may be involved in the stable silencing of *COL4A5* and *COL4A6*. For example, replication timing has been shown to be an important layer of repression. Tissue-specific genes replicate early in the tissues in which they are expressed and late in the tissues where they are turned off. Furthermore, DNA replicated early in S phase gets re-packaged in acetylated histones, whereas late replicating DNA is associated with HDACs (70,95).

Selective Activation of *COL4A5* in HT-1080 Cells

HT-1080 cells present a unique opportunity to study two genes whose promoter regions are intimately associated, but are not co-expressed. The *COL4A5* gene is transcribed in this cell type, and, as expected, the promoter region is associated with RNAPII and acetylated histones. It has been proposed that a glomerular activator exists to promote selective transcription of *COL4A5* (49). Two conserved, non-coding sequences identified in this study, -41K and 76K, may function as this activator as they were shown to enhance transcription of a reporter gene in HT-1080 cells. They were also found to be associated with acetylated histone H3. Consequently, these sequences may bind *trans*-acting factors that interact with the *COL4A5* promoter region, chromatin remodelers, and

other transcription factors to promote gene activation. Figure 14 displays the proposed model of transcription in HT-1080 cells.

The first evidence to suggest that DNA methylation is involved in regulating collagen expression comes from treatment of HT-1080 cells with the demethylating agent, 5-Aza. Expression of *COL4A6* was dramatically increased upon treatment; however, no change was seen for *COL4A5*. This suggests CpG methylation is involved in the selective silencing of *COL4A6*. Indeed, bisulfite sequencing data revealed that the *COL4A6 B* promoter was heavily methylated, while the *COL4A5* region remained relatively free of this modification.

DNA in the *COL4A6 A* promoter region was unmethylated and histones were acetylated; however, no transcripts were detected by real time. Perhaps DNA methylation at the *COL4A6 B* promoter can also act to silence transcription of *COL4A6 A*. Alternatively, the mechanisms discussed earlier for SCC-25 and HEK-293 cells may also be at work in HT-1080 cells.

The question then becomes what is the mechanism by which CpG methylation exerts its effects. Proteins with MBDs have been shown to bind methylated DNA and recruit HDACs to modify chromatin. Based on the data presented here, it does not appear that CpG methylation functions in this manner as acetylation was not sufficient for gene activation. ChIP analysis showed the *COL4A6 B* promoter was acetylated under normal conditions, and even treatment with TSA to fully acetylate histones had no effect on expression. Histone acetylation, therefore, is most likely necessary, but not sufficient for gene activation. Instead it is possible that methyl CpG functions directly by inhibiting the

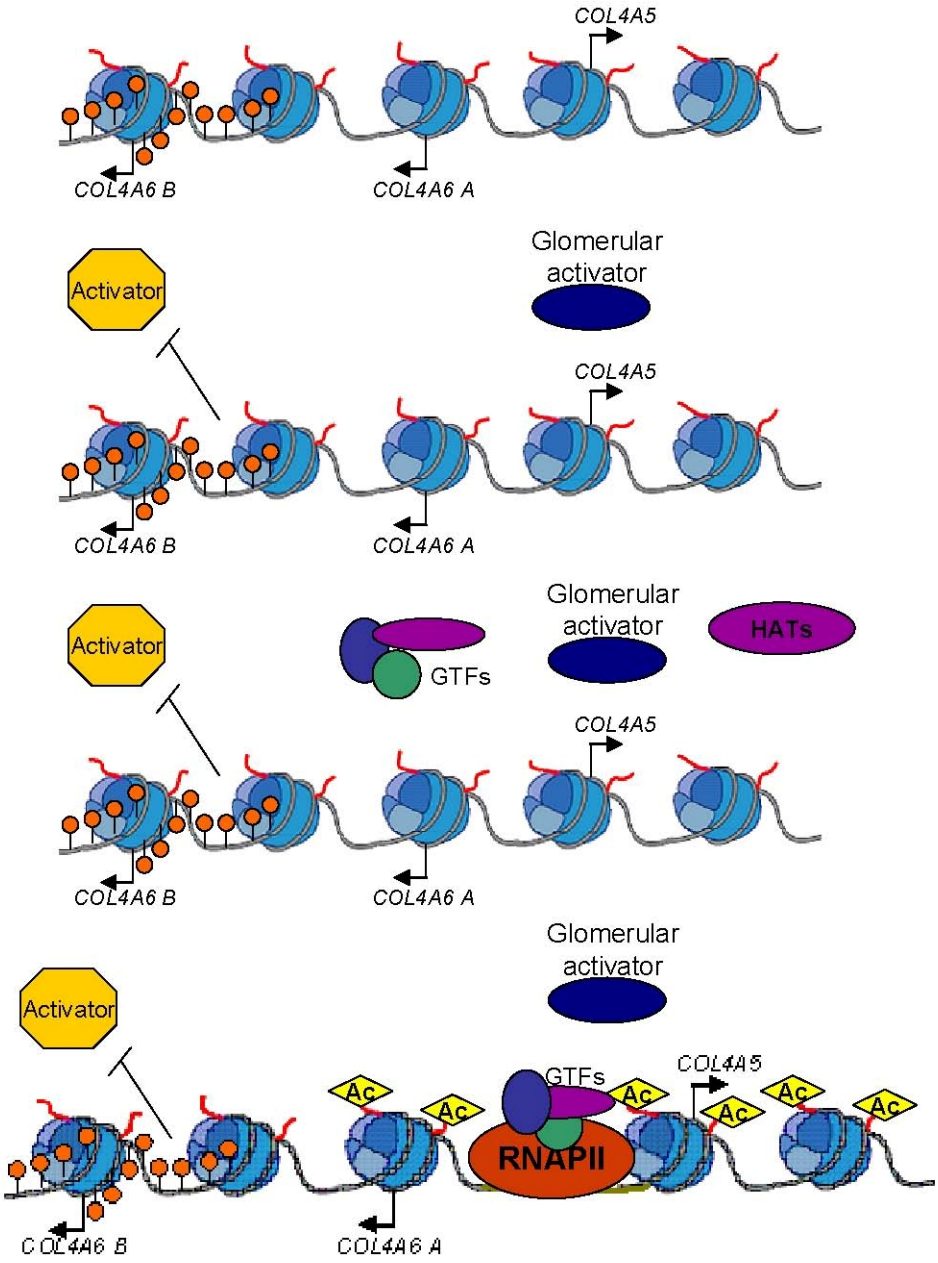

Figure 14 Proposed model for selective activation of *COL4A5* in HT-1080 cells. Methylation of CpG dinucleotides in the *COL4A6 B* promoter functions to prevent binding of an unidentified activator, thereby inhibiting transcription. Activation of *COL4A5* is mediated through interactions among HATs, GTF, and an unidentified glomerular activator. RNAPII is recruited and transcription is initiated.

binding of an activator to the promoter region to prevent transcription initiation (see Figure 14). This has been shown to be the case for several transcription factors (74,75). On the other hand, methyl CpG can actually bind repressor complexes to negatively impact transcription. In both models, the repressive mark is stable through successive cell divisions by maintenance DNMTs. More experiments are certainly necessary to determine the exact role of DNA methylation in this cell line.

Other experiments have shown expression of collagens to be inversely correlated with DNA methylation. One recent study evaluated the methylation profiles of the *COL4A5* and *COL4A6* promoters in colorectal cancers and found hypermethylation to be associated with loss of expression of both collagen alpha chains (96). Methylation of the *COL4A6 B* promoter, however, was not assessed making the current study the first to look at this region. It is important to mention that, by definition, the collagen promoter is not a CpG island. Nonetheless, based on the data presented here and the previous study, this should be re-considered. While computational analyses are undeniably useful, experimental data outweigh any method of prediction.

Summary and Future Directions

Cell type-specific epigenetic marks, including histone modifications and DNA methylation, contributed to regulation of the type IV collagen genes, *COL4A5* and *COL4A6*. Regulation occurred at transcription initiation, although mRNA processing and other post-transcription mechanisms of control may also contribute to steady state transcript levels. In addition, evidence to suggest that conserved, non-coding sequences

modulate collagen expression was provided. More experiments must be carried out to determine whether or not these findings function *in vivo*.

Although these data shed light on collagen gene regulation, several fundamental questions remain unanswered. For example, in cell lines expressing *COL4A6*, transcript levels of *B* were higher than *A*. What is the functional significance of this and how are these differences mediated? Are these differences occurring *in vivo*? In previous experiments using RT-PCR, both transcripts were found to be expressed in kidney, lung, and placenta, although relative transcript levels were not determined (48). Only *COL4A6 B* was shown to be expressed in the bladder suggesting that functionally significant differences do indeed exist (97). Determining how *COL4A6 A* and *B* operate should be the focus of future research.

Questions regarding transcription rates of *COL4A5* and *COL4A6* also exist. Specifically, do transcription rates reflect trimer composition, *i.e*. 2:1 ratio of *COL4A5* to *COL4A6*? The ratio of steady state levels of *COL4A5* and *COL4A6* has been shown to be different for different tissues; however, these studies did not determine the actual rates of transcription. Significantly, transcription rates do reflect trimer composition for *COL4A1* and *COL4A2* in cell lines, although steady state transcript levels do not. This suggests transcription and post-transcription mechanisms of regulation exist. Perhaps the same will hold true for *COL4A5* and *COL4A6*.

Although the *COL4A5-COL4A6* promoter region has been extensively characterized, the functional significance of CTCBF, Sp1, and CCAAT binding protein remains unknown. Binding motifs for these proteins are located in the *COL4A5-COL4A6*

promoter as well as the *COL4A1*-*COL4A2* promoter and introns 1 of both genes. All three proteins have been shown to bind to their respective motifs and facilitate transcription of *COL4A1* and *COL4A2*. With regards to *COL4A5* and *COL4A6*, deletion constructs tested *in vitro* suggested that CCAAT binding protein and CTCBF do not contribute to gene activation (25). The possibility exists that different experimental conditions may reveal the importance of these proteins in transcription, thus studies aimed at determining their role should be conducted in the future.

Finally, do mechanisms exist to coordinate type IV collagen expression? For example, *COL4A5* and *COL4A6* expression appears to be coupled through the presence of a bi-directional activator; the resulting protein products form the α5.α5.α 6(IV) trimer. Transcription can be uncoupled through absence of this activator and presence of additional elements to selectively transcribe *COL4A5*. The *COL4A5* gene is unique in that its protein product can also partner with α3(IV) and α4(IV). What mechanisms exist to account for expression of *COL4A3*, *COL4A4*, and *COL4A5*? This question has been addressed by studying Alport syndrome, which is caused by mutations in any one of these three genes. The results, however, are conflicting. Studies using human fibroblasts and canine models of the disease suggest that expression is not coordinated, while another canine model of Alport syndrome provided evidence to suggest that expression of *COL4A3*, *COL4A4*, and *COL4A5* is coordinated (32,98,99). The particular mutation involved in each case may contribute to differences in transcript expression, thus studies in wild-type cells should be conducted.

In addition to these fundamental questions regarding type IV collagens, more research should be dedicated to delineating the role of *COL4A5* and *COL4A6* in disease. For example, mutations in *COL4A5* are causative for the X-linked form of Alport syndrome (XLAS). In approximately 50% of human cases, the underlying mutation can not be identified in the coding region, promoter, or splice sites (100). It is possible that mutations in intronic, cis-acting elements, such as those identified in this study, can contribute to the disease phenotype. In addition, epigenetic changes may be causative for XLAS. Some patients presenting with the disease have normal trimer formation, but decreased amounts in the extracellular matrix as determined by immunohistochemistry. Perhaps these cases are characterized by abnormal mRNA expression that results from changes in histone modifications or partial DNA methylation of the promoter. These issues should be further explored as efficient and accurate diagnosis is imperative for proper treatment.

In addition to Alport syndrome, the role of type IV collagens in cancer has been widely studied. Cancer cells have been shown to produce MMPs that function in remodeling the extracellular matrix. Disruption of the type IV collagen network occurs, allowing for cancer cell migration. Indeed, numerous studies have shown that changes in composition and/or loss of the collagen network correlates with tumor invasiveness. The majority of studies to date have focused on protein levels; however, disturbances in the type IV collagen network have also been shown to result from down-regulation at the mRNA level. As mentioned previously, hypermethylation of the *COL4A5-COL4A6* promoter resulted in absence of the corresponding transcripts and proteins in colorectal

cancer. These changes provide a unique opportunity to reverse aberrant expression levels and potentially halt cancer metastasis. It is likely that epigenetic changes in type IV collagens will be described for other types of cancer as well. Findings from the current study regarding the role of epigenetic modifications in gene regulation will undoubtedly aid in future research.

In conclusion, data generated from this study have provided insight into regulation of *COL4A5* and *COL4A6* and will certainly aid in answering the fundamental questions outlined above.

REFERENCES

1. Timpl, R. (1989) *Eur J Biochem* **180**(3), 487-502
2. Yurchenco, P. D., Smirnov, S., and Mathus, T. (2002) *Methods Cell Biol* **69**, 111-144
3. Aumailley, M., and Gayraud, B. (1998) *J Mol Med* **76**(3-4), 253-265
4. Reeders, S. T. (1992) *Kidney Int* **42**(3), 783-792
5. Lemmink, H. H., Mochizuki, T., van den Heuvel, L. P., Schroder, C. H., Barrientos, A., Monnens, L. A., van Oost, B. A., Brunner, H. G., Reeders, S. T., and Smeets, H. J. (1994) *Hum Mol Genet* **3**(8), 1269-1273
6. Zhou, J., Mochizuki, T., Smeets, H., Antignac, C., Laurila, P., de Paepe, A., Tryggvason, K., and Reeders, S. T. (1993) *Science* **261**(5125), 1167-1169
7. Morello, R., Zhou, G., Dreyer, S. D., Harvey, S. J., Ninomiya, Y., Thorner, P. S., Miner, J. H., Cole, W., Winterpacht, A., Zabel, B., Oberg, K. C., and Lee, B. (2001) *Nat Genet* **27**(2), 205-208
8. Wilson, C. B., and Dixon, F. J. (1973) *Kidney Int* **3**(2), 74-89
9. Lemmink, H. H., Nillesen, W. N., Mochizuki, T., Schroder, C. H., Brunner, H. G., van Oost, B. A., Monnens, L. A., and Smeets, H. J. (1996) *J Clin Invest* **98**(5), 1114-1118
10. Dehan, P., Waltregny, D., Beschin, A., Noel, A., Castronovo, V., Tryggvason, K., De Leval, J., and Foidart, J. M. (1997) *Am J Pathol* **151**(4), 1097-1104
11. Hiki, Y., Iyama, K., Tsuruta, J., Egami, H., Kamio, T., Suko, S., Naito, I., Sado, Y., Ninomiya, Y., and Ogawa, M. (2002) *Pathol Int* **52**(3), 224-233

12. Misumi, S., Iyama, K., Honda, Y., Kitano, T., Sado, Y., Ninomiya, Y., and Shinohara, M. (2004) *Virchows Arch* **445**(1), 54-62

13. Nakano, S., Iyama, K., Ogawa, M., Yoshioka, H., Sado, Y., Oohashi, T., and Ninomiya, Y. (1999) *Lab Invest* **79**(3), 281-292

14. Tamamura, R., Nagatsuka, H., Siar, C. H., Katase, N., Naito, I., Sado, Y., and Nagai, N. (2006) *Virchows Arch* **449**(3), 358-366

15. Floquet, N., Pasco, S., Ramont, L., Derreumaux, P., Laronze, J. Y., Nuzillard, J. M., Maquart, F. X., Alix, A. J., and Monboisse, J. C. (2004) *J Biol Chem* **279**(3), 2091-2100

16. Kamphaus, G. D., Colorado, P. C., Panka, D. J., Hopfer, H., Ramchandran, R., Torre, A., Maeshima, Y., Mier, J. W., Sukhatme, V. P., and Kalluri, R. (2000) *J Biol Chem* **275**(2), 1209-1215

17. Karagiannis, E. D., and Popel, A. S. (2007) *Biochem Biophys Res Commun* **354**(2), 434-439

18. Sudhakar, A., Nyberg, P., Keshamouni, V. G., Mannam, A. P., Li, J., Sugimoto, H., Cosgrove, D., and Kalluri, R. (2005) *J Clin Invest* **115**(10), 2801-2810

19. Hudson, B. G., Reeders, S. T., and Tryggvason, K. (1993) *J Biol Chem* **268**(35), 26033-26036

20. Hostikka, S. L., Eddy, R. L., Byers, M. G., Hoyhtya, M., Shows, T. B., and Tryggvason, K. (1990) *Proc Natl Acad Sci U S A* **87**(4), 1606-1610

21. Breen, M., Jouquand, S., Renier, C., Mellersh, C. S., Hitte, C., Holmes, N. G., Cheron, A., Suter, N., Vignaux, F., Bristow, A. E., Priat, C., McCann, E., Andre,

C., Boundy, S., Gitsham, P., Thomas, R., Bridge, W. L., Spriggs, H. F., Ryder, E. J., Curson, A., Sampson, J., Ostrander, E. A., Binns, M. M., and Galibert, F. (2001) *Genome Res* **11**(10), 1784-1795

22. Filie, J. D., Burbelo, P. D., and Kozak, C. A. (1995) *Mamm Genome* **6**(7), 487
23. Lowe, J. K., Guyon, R., Cox, M. L., Mitchell, D. C., Lonkar, A. L., Lingaas, F., Andre, C., Galibert, F., Ostrander, E. A., and Murphy, K. E. (2003) *Funct Integr Genomics* **3**(3), 112-116
24. Blumberg, B., MacKrell, A. J., and Fessler, J. H. (1988) *J Biol Chem* **263**(34), 18328-18337
25. Guo, X. D., and Kramer, J. M. (1989) *J Biol Chem* **264**(29), 17574-17582
26. Zhou, J., Ding, M., Zhao, Z., and Reeders, S. T. (1994) *J Biol Chem* **269**(18), 13193-13199
27. Timpl, R., Wiedemann, H., van Delden, V., Furthmayr, H., and Kuhn, K. (1981) *Eur J Biochem* **120**(2), 203-211
28. Yurchenco, P. D., and Furthmayr, H. (1984) *Biochemistry* **23**(8), 1839-1850
29. Tsilibary, E. C., and Charonis, A. S. (1986) *J Cell Biol* **103**(6 Pt 1), 2467-2473
30. Berg, R. A., and Prockop, D. J. (1973) *Biochem Biophys Res Commun* **52**(1), 115-120
31. Boutaud, A., Borza, D. B., Bondar, O., Gunwar, S., Netzer, K. O., Singh, N., Ninomiya, Y., Sado, Y., Noelken, M. E., and Hudson, B. G. (2000) *J Biol Chem* **275**(39), 30716-30724

32. Thorner, P. S., Zheng, K., Kalluri, R., Jacobs, R., and Hudson, B. G. (1996) *J Biol Chem* **271**(23), 13821-13828

33. Gomez, D. E., Alonso, D. F., Yoshiji, H., and Thorgeirsson, U. P. (1997) *Eur J Cell Biol* **74**(2), 111-122

34. Rao, V. H., Lees, G. E., Kashtan, C. E., Delimont, D. C., Singh, R., Meehan, D. T., Bhattacharya, G., Berridge, B. R., and Cosgrove, D. (2005) *Pediatr Nephrol* **20**(6), 732-739

35. Butkowski, R. J., Wieslander, J., Kleppel, M., Michael, A. F., and Fish, A. J. (1989) *Kidney Int* **35**(5), 1195-1202

36. Kleppel, M. M., Santi, P. A., Cameron, J. D., Wieslander, J., and Michael, A. F. (1989) *Am J Pathol* **134**(4), 813-825

37. Sanes, J. R., Engvall, E., Butkowski, R., and Hunter, D. D. (1990) *J Cell Biol* **111**(4), 1685-1699

38. Miner, J. H., and Sanes, J. R. (1994) *J Cell Biol* **127**(3), 879-891

39. Yoshioka, K., Hino, S., Takemura, T., Maki, S., Wieslander, J., Takekoshi, Y., Makino, H., Kagawa, M., Sado, Y., and Kashtan, C. E. (1994) *Am J Pathol* **144**(5), 986-996

40. Peissel, B., Geng, L., Kalluri, R., Kashtan, C., Rennke, H. G., Gallo, G. R., Yoshioka, K., Sun, M. J., Hudson, B. G., Neilson, E. G., et al. (1995) *J Clin Invest* **96**(4), 1948-1957

41. Ninomiya, Y., Kagawa, M., Iyama, K., Naito, I., Kishiro, Y., Seyer, J. M., Sugimoto, M., Oohashi, T., and Sado, Y. (1995) *J Cell Biol* **130**(5), 1219-1229

42. Poschl, E., Pollner, R., and Kuhn, K. (1988) *EMBO J* **7**(9), 2687-2695

43. Pollner, R., Schmidt, C., Fischer, G., Kuhn, K., and Poschl, E. (1997) *FEBS Lett* **405**(1), 31-36

44. Fischer, G., Schmidt, C., Opitz, J., Cully, Z., Kuhn, K., and Poschl, E. (1993) *Biochem J* **292 (Pt 3)**, 687-695

45. Genersch, E., Eckerskorn, C., Lottspeich, F., Herzog, C., Kuhn, K., and Poschl, E. (1995) *EMBO J* **14**(4), 791-800

46. Haniel, A., Welge-Lussen, U., Kuhn, K., and Poschl, E. (1995) *J Biol Chem* **270**(19), 11209-11215

47. Sund, M., Maeshima, Y., and Kalluri, R. (2005) *Biochem J* **387**(Pt 3), 755-761

48. Sugimoto, M., Oohashi, T., and Ninomiya, Y. (1994) *Proc Natl Acad Sci U S A* **91**(24), 11679-11683

49. Segal, Y., Zhuang, L., Rondeau, E., Sraer, J. D., and Zhou, J. (2001) *J Biol Chem* **276**(15), 11791-11797

50. Herzog, C., Zhuang, L., Gorgan, L., Segal, Y., and Zhou, J. (2003) *Biochem Biophys Res Commun* **311**(2), 553-560

51. Li, B., Carey, M., and Workman, J. L. (2007) *Cell* **128**(4), 707-719

52. Roeder, R. G. (2005) *FEBS Lett* **579**(4), 909-915

53. Kouzarides, T. (2007) *Cell* **128**(4), 693-705

54. Workman, J. L., and Kingston, R. E. (1998) *Annu Rev Biochem* **67**, 545-579

55. Liu, C. L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S. L., Friedman, N., and Rando, O. J. (2005) *PLoS Biol* **3**(10), e328

56. Roh, T. Y., Cuddapah, S., and Zhao, K. (2005) *Genes Dev* **19**(5), 542-552

57. Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., and Young, R. A. (2005) *Cell* **122**(4), 517-527

58. Vakoc, C. R., Mandat, S. A., Olenchock, B. A., and Blobel, G. A. (2005) *Mol Cell* **19**(3), 381-391

59. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) *Cell* **119**(7), 941-953

60. Bannister, A. J., and Kouzarides, T. (2005) *Nature* **436**(7054), 1103-1106

61. Bernstein, B. E., Meissner, A., and Lander, E. S. (2007) *Cell* **128**(4), 669-681

62. Bird, A. (2002) *Genes Dev* **16**(1), 6-21

63. Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992) *Genes Dev* **6**(5), 705-714

64. Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002) *Dev Biol* **241**(1), 172-182

65. Berger, S. L. (2007) *Nature* **447**(7143), 407-412

66. Jones, P. A., and Baylin, S. B. (2002) *Nat Rev Genet* **3**(6), 415-428

67. Goll, M. G., and Bestor, T. H. (2005) *Annu Rev Biochem* **74**, 481-514

68. Razin, A. (1998) *EMBO J* **17**(17), 4905-4908

69. Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., and Jaenisch, R. (2001) *Nat Genet* **27**(1), 31-39

70. Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G., and Cedar, H. (2007) *J Biol Chem* **282**(16), 12194-12200

71. Razin, A., and Cedar, H. (1991) *Microbiol Rev* **55**(3), 451-458

72. Bird, A. P., and Wolffe, A. P. (1999) *Cell* **99**(5), 451-454

73. Zhang, Y., and Reinberg, D. (2001) *Genes Dev* **15**(18), 2343-2360

74. Clark, S. J., Harrison, J., and Molloy, P. L. (1997) *Gene* **195**(1), 67-71

75. Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (2000) *Nature* **405**(6785), 486-489

76. Duret, L., and Bucher, P. (1997) *Curr Opin Struct Biol* **7**(3), 399-406

77. Hardison, R. C. (2000) *Trends Genet* **16**(9), 369-372

78. Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M., and Frazer, K. A. (2000) *Science* **288**(5463), 136-140

79. Pennacchio, L. A., and Rubin, E. M. (2001) *Nat Rev Genet* **2**(2), 100-109

80. Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., McDowell, J. C., Maskeri, B., Hansen, N. F., Schwartz, M. S., Weber, R. J., Kent, W. J., Karolchik, D., Bruen, T. C., Bevan, R., Cutler, D. J., Schwartz, S., Elnitski, L., Idol, J. R., Prasad, A. B., Lee-Lin, S. Q., Maduro, V. V., Summers, T. J., Portnoy, M. E., Dietrich, N. L., Akhter, N., Ayele, K., Benjamin, B., Cariaga, K., Brinkley, C. P., Brooks, S. Y., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S. L., Huang, M. C., Karlins, E., Laric, P. L., Legaspi, R., Lim, M. J., Maduro, Q.

L., Masiello, C. A., Mastrian, S. D., McCloskey, J. C., Pearson, R., Stantripop, S., Tiongson, E. E., Tran, J. T., Tsurgeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L. S., Young, A. C., Zhang, L. H., Osoegawa, K., Zhu, B., Zhao, B., Shu, C. L., De Jong, P. J., Lawrence, C. E., Smit, A. F., Chakravarti, A., Haussler, D., Green, P., Miller, W., and Green, E. D. (2003) *Nature* **424**(6950), 788-793

81. Dermitzakis, E. T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C., and Antonarakis, S. E. (2003) *Science* **302**(5647), 1033-1035

82. Dermitzakis, E. T., Reymond, A., and Antonarakis, S. E. (2005) *Nat Rev Genet* **6**(2), 151-157

83. Drake, J. A., Bird, C., Nemesh, J., Thomas, D. J., Newton-Cheh, C., Reymond, A., Excoffier, L., Attar, H., Antonarakis, S. E., Dermitzakis, E. T., and Hirschhorn, J. N. (2006) *Nat Genet* **38**(2), 223-227

84. Wardrop, S. L., and Brown, M. A. (2005) *Genomics* **86**(3), 316-328

85. Grice, E. A., Rochelle, E. S., Green, E. D., Chakravarti, A., and McCallion, A. S. (2005) *Hum Mol Genet* **14**(24), 3837-3845

86. Baltimore, D. (2001) *Nature* **409**(6822), 814-816

87. Zhou, J., Leinonen, A., and Tryggvason, K. (1994) *J Biol Chem* **269**(9), 6608-6614

88. Chow, J. C., Yen, Z., Ziesche, S. M., and Brown, C. J. (2005) *Annu Rev Genomics Hum Genet* **6**, 69-92

89. Chadwick, B. P., and Willard, H. F. (2003) *Hum Mol Genet* **12**(17), 2167-2178

90. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007) *Cell* **130**(1), 77-88

91. Brasher, S. V., Smith, B. O., Fogh, R. H., Nietlispach, D., Thiru, A., Nielsen, P. R., Broadhurst, R. W., Ball, L. J., Murzina, N. V., and Laue, E. D. (2000) *EMBO J* **19**(7), 1587-1597

92. Nielsen, A. L., Ortiz, J. A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., and Losson, R. (1999) *EMBO J* **18**(22), 6385-6395

93. Vassallo, M. F., and Tanese, N. (2002) *Proc Natl Acad Sci U S A* **99**(9), 5919-5924

94. Smallwood, A., Esteve, P. O., Pradhan, S., and Carey, M. (2007) *Genes Dev* **21**(10), 1169-1178

95. Lande-Diner, L., and Cedar, H. (2005) *Nat Rev Genet* **6**(8), 648-654

96. Ikeda, K., Iyama, K., Ishikawa, N., Egami, H., Nakao, M., Sado, Y., Ninomiya, Y., and Baba, H. (2006) *Am J Pathol* **168**(3), 856-865

97. Zheng, K., Harvey, S., Sado, Y., Naito, I., Ninomiya, Y., Jacobs, R., and Thorner, P. S. (1999) *Am J Pathol* **154**(6), 1883-1891

98. Heidet, L., Cai, Y., Guicharnaud, L., Antignac, C., and Gubler, M. C. (2000) *Am J Pathol* **156**(6), 1901-1910

99. Sasaki, S., Zhou, B., Fan, W. W., Kim, Y., Barker, D. F., Denison, J. C., Atkin, C. L., Gregory, M. C., Zhou, J., Segal, Y., Sado, Y., Ninomiya, Y., Michael, A. F., and Kashtan, C. E. (1998) *Matrix Biol* **17**(4), 279-291

100. King, K., Flinter, F. A., Nihalani, V., and Green, P. M. (2002) *Hum Genet* **111**(6), 548-554

VITA

Name: Jessica Ashley Moody

Address: Texas A&M University, College of Veterinary Medicine and Biomedical Sciences, College Station, TX 77843-4467

Email Address: jmoody@cvm.tamu.edu

Education:
B.S., Nutritional Science, Texas A&M University, 2003
Ph.D., Veterinary Microbiology, Texas A&M University, 2008

Honors:

2006	Recipient of Fisher Institute Medical Research Award
2003-05	Recipient of Texas A&M University Graduate Merit Fellowship
2003	Designated as the Outstanding Graduating Senior in Nutritional Sciences
2003	Graduated *summa cum laude* from Texas A&M University
2002	Recipient of Northwest Harris County Aggie Moms Club Scholarship
2001	Dean's List
2000	Dean's List

Publications:

Moody, J.A., Boggs, R.M., Porter, W.W., Wellberg, E., and Murphy, K.E. (2008) In preparation

Boggs, R.M., **Moody, J.A.**, Long, C.R., Tsai, K.L., and Murphy, K.E. (2007) *Gene* **404**(1-2), 25-30

Moody, J.A., Famula, T.R., Sampson, R.C., and Murphy, K.E. (2005) *Am J Vet Res* **66**(11), 1900–1902

Moody, J.A., Clark, L.A., and Murphy, K.E. (2005) *The Dog and Its Genome*, Cold Spring Harbor Laboratory Press, Woodbury, NY, 1-18