Functional Interrogation of the *COL4A1/COL4A2* and *SMAD3* Coronary Artery Disease Loci

By Adam Turner

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of PhD in Biochemistry

Department of Biochemistry, Microbiology and Immunology

Specialization in Human and Molecular Genetics

Faculty of Medicine, University of Ottawa

Ottawa, Ontario, Canada

Abstract

Coronary artery disease (CAD) is the leading cause of death worldwide caused by a complex array of environmental and genetic factors. In recent years, the CARDIoGRAM Consortium has identified dozens of novel CAD-associated loci, including *COL4A1/COL4A2* and *SMAD3*. The goals of this project were two-fold. First, to identify functional CAD-associated SNPs (single nucleotide polymorphisms) at these loci and investigate how these common polymorphisms alter CAD risk. Second, to determine if the *COL4A1/COL4A2* and *SMAD3* loci display both biological interaction and statistical interaction in the context of atherosclerosis. COL4A1 and COL4A2 are critical components of vascular basement membranes and have many additional roles in the vessel wall. SMAD3 is a transcription factor that is a key mediator in the canonical TGFβ signaling pathway. In the first section of this thesis, I show that COL4A1/COL4A2 and SMAD3 display biological interaction in that the TGFβ-mediated upregulation of COL4A1 and COL4A2 is dependent on SMAD3. Furthermore, we were able to identify a *COL4A2-SMAD3* SNP pair that displayed a highly significant statistical interaction for CAD association, highlighting that perturbations of the SMAD3/type IV collagen signaling axis contribute to the pathogenesis of atherosclerosis. The second section details the characterization of a novel functional CAD-associated SNP at the *SMAD3* locus. The rs17293632 SNP, highly linked to the rs56062135 index SNP reported by CARDIoGRAM, disrupts a conserved AP-1 binding site in intron 1 of the *SMAD3* gene. rs17293632 lies within a strong enhancer in many cell types, including arterial smooth muscle cells. The minor, protective allele (T) at rs17293632 disrupts binding of AP-1 proteins, lowers the activity of this enhancer and lowers SMAD3 mRNA levels in humans in both whole blood and carotid plaque tissue. Altogether, AP-1

regulation of *SMAD3* enhancer activity suggests a novel regulatory mechanism relevant to the pathogenesis of CAD. The third section summarizes efforts to characterize the mechanisms whereby the independent rs4773144 and rs9515203 SNPs at the *COL4A1/COL4A2* locus associate highly with CAD. We were unable to identify causal SNPs at the *COL4A1/COL4A2* locus, highlighting the challenges of post-GWAS characterization of some CAD-associated loci. The experiments with SNPs at *COL4A1/COL4A2* demonstrate the challenges in defining mechanisms whereby noncoding DNA variants can lead to common disease.

Acknowledgements

I would first like to thank my supervisor, Dr. Ruth McPherson, for her continued guidance and support along the course of my project, allowing me the opportunity to work on several interesting projects, and providing me the ability to go to many helpful and interesting conferences across North America. I thank the other members of the Atherogenomics Laboratory: Paulina Lau, Dr. Sebastien Soubeyrand, and Amy Martinuk for always discussing various aspects of my project and providing excellent project support. I thank Dr. Majid Nikpay for his bioinformatic and statistical support over the years and Anada Silva for valuable experimental support. I would also like to thank former lab member Dr. Olga Jarinova for her genetics and bioinformatics training early in my project as well as Thet Naing for her help in the lab. I thank my Thesis Advisory Committee members Dr. Patrick Burgon, Dr. Marc Ekker, and Dr. Ilya Ioshikhes for their expertise and suggestions over the past several years. Finally, I greatly thank all of the other students in the Atherogenomics Laboratory (Tara Linseman, Adrianna Douvris, Simran Bhatia, Chris Cole, Kaitlin Kharas, Erica Gotfrit, Michael Seto, Giulia Costa, Laura Fusca, Anh-Thu Dang, and Kaitlyn Beehler) for always being willing to help each other and providing a great and fun work environment.

Table of Contents

List of Abbreviations

3C	chromosome conformation capture
AOS	aneurysms-osteoarthritis syndrome
AoSMCs	primary human aortic smooth muscle cells
AP-1	activator protein 1
apoA-1	apolipoprotein A-1
apoB	apolipoprotein B
apoE	apolipoprotein E
ASE	allele-specific expression
ATAC-seq	assay for transposase accessible chromatin sequencing
bHLH	basic helix loop helix
BMP	bone morphogenic protein
CAD	coronary artery disease
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation sequencing
CVD	cardiovascular disease
DAMPs	damage-associated molecular patterns
DDR	discoidin domain receptor
eQTL	expression quantitative trait locus
ERK	extracellular-signal-regulated kinase
FDR	false discovery rate
GWAS	genome wide association study
HANAC	hereditary angiopathy with nephropathy, aneurysms, and muscle cramps
hASMCs	human arterial smooth muscle cells
hiPSC	human induced pluripotent stem cell
HDAC	histone deacetylase
HDL	high density lipoprotein
ICAM-1	intercellular adhesion molecule 1
JNK	c-Jun N-terminal kinase
LAP	latency associated peptide
LD	linkage disequilibrium
LDL	low density lipoprotein
MAF	minor allele frequency
MAPK	mitogen-activated protein kinase
MI	myocardial infarction
MMPs	matrix metalloproteinases
PAMPs	pathogen-associated molecular patterns
PDGFβ	platelet-derived growth factor β
PMA	phorbol 12-myristate 13-acetate
PPIA	peptidylprolyl isomerase
qRT-PCR	quantitative reverse transcription polymerase chain reaction
SARA	SMAD anchor for activation
SBE	SMAD-binding element
SD	standard deviation

SE	standard error
SMAD3	mothers against decapentaplegic homolog 3
SMC	smooth muscle cell
SNP	single nucleotide polymorphism
TAAD	thoracic aortic aneurysms and dissections
TF	tissue factor
TGFβ	transforming growth factor β
TRE	TPA response element
TPA	12-*O*-tetradecanoylphorbol-13-acetate
VCAM-1	vascular cell adhesion protein 1
VSMC	vascular smooth muscle cell
WT1	Wilm's tumour protein

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

1 INTRODUCTION

1.1 Pathogenesis of Coronary Artery Disease

Cardiovascular disease (CVD) is one of the leading causes of death and disability in Western countries as well as worldwide[1–5]. Coronary artery disease (CAD), the most common form of CVD, is a complex disease resulting from the subendothelial accumulation of atherosclerotic plaques in the coronary arteries. Atherosclerosis is the process underlying CAD and is considered a state of chronic inflammation in response to lipid deposition that ultimately results in the thickening of the arterial wall and often narrowing of the lumen of the coronary artery and restricted blood flow to the myocardium[6]. The processes of atherosclerosis and CAD involve numerous cell types and stages, beginning with plaque development and often ending with plaque rupture and resulting thrombosis and myocardial infarction (MI).

In humans, the first stage of atherosclerosis involves development of early lesions, or "fatty streaks", within the arterial vessel wall (**Figure 1A**). These early lesions consist of cholesterol and lipid-engulfed macrophages, called foam cells, in addition to modified lipoproteins and various types of immune cells, including T cells and to a lesser degree polymorphonuclear leukocytes[7]. Fatty streaks can develop early in life, as studies have shown presence of fatty streaks in the aortas of children, coronary arteries of adolescents, and peripheral vessels of young adults[8–10]. Initiation of fatty streaks occurs through the subendothelial accumulation of lipoproteins, primarily low density lipoprotein (LDL) in focal areas of arteries[11]. Factors determining the location of LDL accumulation in the

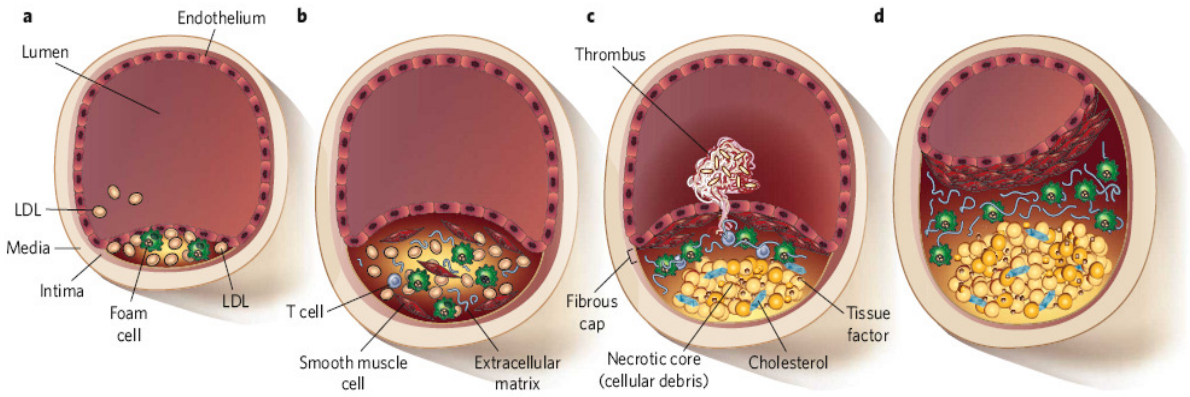

a
Endothelium
Lumen
LDL
Media
Intima
Foam cell
LDL
b
T cell
Smooth muscle cell
Extracellular matrix
c
Thrombus
Fibrous cap
Necrotic core (cellular debris)
Cholesterol
Tissue factor
d

Figure 1. Stages of atherosclerosis progression in a human coronary artery. (A) The initial stage of atherosclerosis begins as a fatty streak. Originally atherogenic lipoproteins such as LDL enter the intima, where they become modified and accumulate in the subendothelial space. These modified lipoproteins become engulfed by macrophages, which often become foam cells that are one of the key features of fatty streaks. (B) In intermediate atherosclerotic lesions, vascular smooth muscle cells from the media migrate to the intima. These vascular smooth muscle cells proliferate and secrete large amounts of extracellular matrix. Intermediate lesions have a heightened inflammatory state, including the presence of other leukocytes and T cells. (C) In advanced lesions prone to plaque rupture and thrombosis, a necrotic core forms from the death of foam cells and release of cell debris and cholesterol. (D) In plaques that do not rupture, the lesion continues to grow and can obstruct the lumen of the coronary artery and lead to angina.

Figure taken from Rader and Daugherty, (2008) Translating molecular discoveries into new therapies for atherosclerosis. *Nature* 451(7181): 904-913, with permission. License agreement number 3800840087532.

arteries include the local vascular endothelium and physical forces such as blood flow and shear stress[7]. Endothelial cells in the tubular regions of arteries, containing uniform and laminar blood flow, are aligned with the direction of blood flow and are less permeable[7]. In contrast, endothelial cells located in regions of arterial curvature or arterial branch points where flow is disturbed show increased permeability to LDL and are much more likely to comprise sites of lesion formation[7,12].

Endothelial activation promotes synthesis of extracellular matrix proteins that result in diffuse intimal thickening and aid in retention of apolipoprotein B (apoB) containing lipoproteins, primarily LDL, in the intima[12]. The presence of negatively charged extracellular matrix proteoglycans on endothelial cells contributes to recruitment/retention of LDL particles within the intima through electrostatic interactions[13,14]. Other apoB containing lipoproteins such as lipoprotein(a) can also accumulate in the intima and promote atherosclerosis at this early stage[7]. Elevated levels of apoB containing particles increase risk of atherosclerosis[13] and in humans the majority of serum cholesterol is carried by LDL particles[6].

Once atherogenic LDLs enter the intima, they undergo modifications such as oxidation, enzymatic activity (by myeloperoxidase or lipoxygenases released from inflammatory cells), and non-enzymatic cleavage[9,12,13]. Oxidative modifications to LDL can occur in both the lipid and apolipoprotein (apoB) components of LDL[6]. The different modifications to the retained lipoproteins are thought to mimic pathogen and/or damage-associated molecular patterns (PAMPs/DAMPs), rendering LDL particles pro-inflammatory[11]. These LDL modifications also lead to LDL aggregation within the extracellular space of the intima[9,12].

The presence of oxidized lipids and oxidized LDL within the intima triggers endothelial cell expression of adhesion molecules as well as endothelial secretion of chemokines. Several cell adhesion molecules expressed by endothelial cells are involved in intimal monocyte recruitment, including VCAM-1 and ICAM-1[6,15]. Endothelial cell expression of adhesion molecules and secretion of cytokines drive infiltration of immune cells into the intima.

Monocytes recruited to the subendothelial space from the inflammatory response differentiate into phagocytic macrophages once they are in the vessel wall and proceed to engulf the modified LDLs[9,12]. Macrophages that take up excess levels of atherogenic lipoproteins become lipid rich foam cells, one of the hallmarks of fatty streaks[9,13,16]. LDL particles that are highly modified are no longer bound by the LDL receptor, but instead are taken up by macrophages by scavenger receptor proteins[6]. Although foam cells harbour large of amounts of cholesterol and cholesteryl esters, they are unable to catabolize cholesterol. In addition to foam cells, infiltration of T cells into the subendothelial space is another key feature of fatty streaks[13,16].

While the fatty streak represents an early stage in atherosclerotic lesion progression, the next stage involves plaque growth and formation of the fibrous cap (**Figure 1B**). As atherosclerosis progresses, the inflammatory situation underway in the early atheroma in combination with the cytokines and growth factors secreted by macrophages and T cells result in smooth muscle cells from the medial layers of blood vessels starting to proliferate and migrate towards the intimal layer[7,17]. Proliferating intimal smooth muscle cells may also ingest modified lipoproteins and further contribute to foam cell formation[6,8]. This smooth muscle cell proliferation and migration involves “activation” or “phenotypic switching” of

smooth muscle cells. Smooth muscle cells in the artery normally have a quiescent and contractile phenotype[11]. The inflammatory context in atherosclerosis results in smooth muscle cells downregulating differentiation markers such as smooth muscle α-actin (ACTA2), smooth muscle myosin heavy chain (MYH11), and SM22α/transgelin (TAGLN)[11]. Increased smooth muscle cell proliferation at this stage of atherosclerosis coincides with increased synthesis of extracellular matrix proteins (primarily collagen), proteoglycans, and other proteins involved in plaque stabilization ("synthetic phenotype")[11]. Eventually a fibrous cap forms that lies over the lipid core and provides structural stability for the atherosclerotic lesion and protects against plaque rupture and thrombosis[7,16]. As the plaque grows, the arterial lumen sometimes narrows that can lead to hampered blood flow to the heart and ischemia induced chest pain or angina[15].

As atherosclerotic plaques develop into more advanced stages, they become more complex and numerous other events occur. One feature of plaques as they continue to progress is apoptosis. Death of lipid laden macrophages and foam cells results in the deposition of tissue factor (TF) in the extracellular space[18,19]. Recent work in a mouse model for CAD has shown that macrophages accumulate in plaques not only through monocyte infiltration but also through local proliferation[20]. Another characteristic of atherosclerotic lesions as they progress is that calcification begins to occur, similar to mechanisms that occur in bone formation[17,21]. Neovascularization also occurs in advanced lesions, where the growth of small vessels from the media could provide a mechanism for entry of inflammatory cells[7].

Rather than being fixed, fibrous caps are dynamic and can undergo substantial remodeling (**Figure 2**)[22]. While fibrous caps are stable, with interstitial collagens (I and III) providing

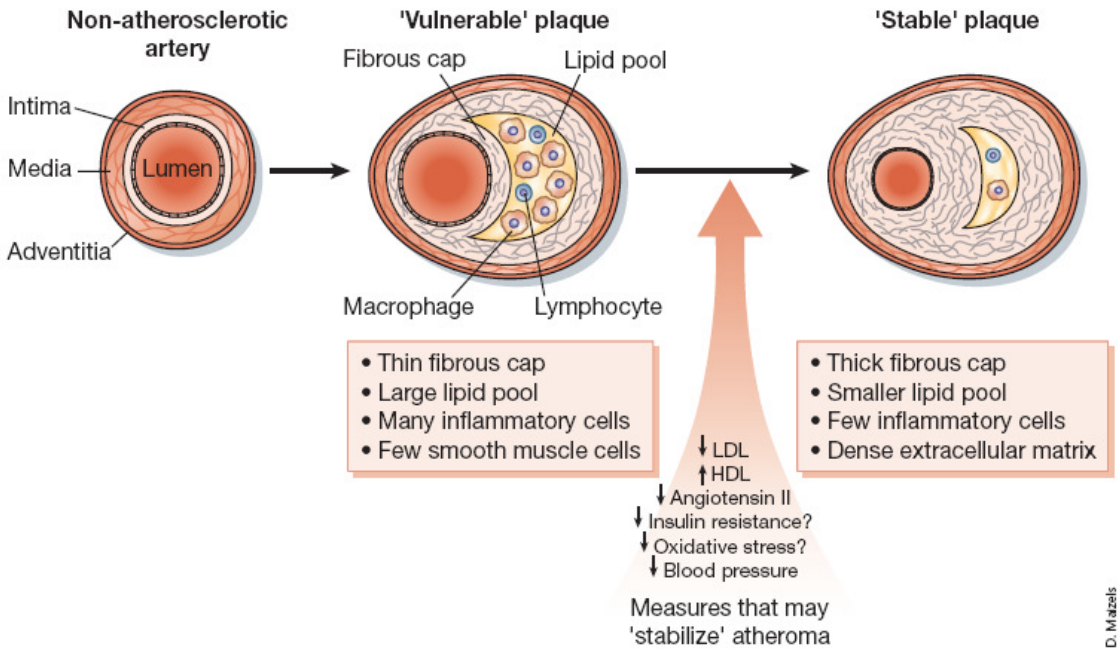

Non-atherosclerotic artery
'Vulnerable' plaque
'Stable' plaque
Intima
Media
Adventitia
Lumen
Fibrous cap
Lipid pool
Macrophage
Lymphocyte
• Thin fibrous cap
• Large lipid pool
• Many inflammatory cells
• Few smooth muscle cells
• Thick fibrous cap
• Smaller lipid pool
• Few inflammatory cells
• Dense extracellular matrix
↓ LDL
↑ HDL
↓ Angiotensin II
↓ Insulin resistance?
↓ Oxidative stress?
↓ Blood pressure
Measures that may 'stabilize' atheroma
D. Mazels

Figure 2. Properties of a healthy non-atherosclerotic artery, an artery containing a stable plaque, and an artery containing a vulnerable plaque that is prone to rupture. During the early stages of atherosclerosis, the intima typically grows outward and does not greatly block the lumen. Plaques that are prone to rupture and lead to myocardial infarction have a prominent lipid pool, a large contingent of inflammatory cells and a thin fibrous cap. Conversely, stable plaques have lower amounts of lipids and inflammatory cells and copious amounts of extracellular matrix that provide a thick fibrous cap that is difficult to rupture.

Figure taken from Libby and Aikawa, (2002) Stabilization of atherosclerotic plaques: New mechanisms and clinical targets. *Nature Medicine* 8(11): 1257-1262, with permission. License agreement number 3800841059061.

mechanical strength, some plaques can progressively become more vulnerable to rupture[22]. These vulnerable plaques have a heightened inflammatory state, thinning of the fibrous cap, abundance of lipid-laden macrophages, and a large necrotic core[6,7,11,22]. Fibrous cap thinning is due to a combination of secretion of matrix metalloproteinases (MMPs) that degrade collagen and extracellular matrix as well as decreased synthesis of collagen by smooth muscle cells[11,22,23]. The necrotic core of vulnerable plaques results from the extracellular accumulation of lipids that accumulate and subsequently coalesce[17]. Plaque necrosis occurs from a combination of lesional macrophage apoptosis and ineffective clearance of these apoptotic cells that results in primary necrosis[11,23].

Although atherosclerotic lesions develop over decades, atherosclerosis is often asymptomatic until plaques rupture, culminating in thrombosis formation, acutely blocking blood flow to the heart and commonly resulting in MI and death. Therefore plaque composition and vulnerability are key components and more important than plaque size in determining the risk of MI in patients[16]. Thrombus formation can be attributed to at least two different events. First, the plaque lesion surface can rupture exposing the thrombogenic subendothelial layer of the blood vessel. Second, the fibrous cap of smooth muscle cells can erode leading to exposure to thrombogenic molecules[9]. Synthesis and release of TF by endothelial cells and macrophages in the plaque contributes heavily to the thrombosis procedure upon plaque rupture/erosion[7].

More recent studies suggest there is more ambiguity than previously believed with respect to the lineal origins of cells within the atherosclerotic plaque. Prior studies investigating functions of macrophages and smooth muscle cells in atherosclerosis have yielded

controversial findings due to ambiguity in properly differentiating between these cell types[24,25]. For example, plaque smooth muscle cells can start to express some macrophage markers and conversely, macrophages can start to express some smooth muscle cell markers[24]. Further complicating detection, this coincides with decreased expression of smooth muscle cell differentiation markers. It is now believed that the smooth muscle cell content of plaques has been greatly underestimated[24].

1.2 Genetics of Coronary Artery Disease

Previous studies have indicated that CAD is a heritable trait that clusters in families. Genetics accounts for approximately 40-50% of CAD risk[26]. Data collected in a cohort of 20,966 Swedish twins collected over 36 years identified the heritability (h^2) of fatal CAD events is 57% in men and 38% in women[27]. Heritability of CAD is greatest for premature CAD. The majority (72%) of early onset CAD cases (men < 55 years, women < 65 years), and 48% of total CAD cases of any age, have a family history of CAD[28].

In general, CAD risk is the result of numerous common genetic polymorphisms[29,30]. However, there have been several molecular genetic studies on rare, Mendelian disorders that predispose to early CAD. Mutations in these genes often drastically affect the levels of either LDL or HDL[31]. Some Mendelian diseases that involve premature CAD include familial hypercholesterolemia (defects in the *LDLR* gene) that results in defective binding of LDL to its receptor, familial defective apoB (defects in the *APOB* gene) that leads to impaired binding of apoB to the LDL receptor, gain of function mutations in *PCSK9*, and apoA1 deficiency (defects in the *APOA1* gene) that leads to low HDL[31].

Many studies that originally sought to dissect the genetic basis of CAD performed linkage analysis in families. While linkage studies are often successful for single-gene disorders, they are more difficult for common and complex diseases and have been criticized for being difficult to reproduce[31]. Other early studies for identifying risk loci for CAD performed candidate gene studies. These studies investigated genes with prior known roles in atherosclerosis pathogenesis, including genes involved in lipoprotein metabolism, inflammation, and thrombosis[31]. Once again, candidate gene studies have been difficult to reproduce with a high occurrence of false positive results[31]. Another drawback to the candidate gene approach is the exclusion of intergenic regions, which constitute a large proportion of the genome that includes DNA elements such as enhancers[32].

1.3 Genome-Wide Association Studies for Coronary Artery Disease

Genome-wide association studies (GWAS) are approaches to examine association of DNA variants, typically single nucleotide polymorphisms (SNPs), with common diseases and complex traits[33]. GWAS are designed to compare DNA markers between a set of cases with a particular disease or phenotype and set of carefully matched controls without the disease/phenotype. GWAS are thus considered to be phenotype-first studies as opposed to genotype-first studies. Most GWAS that have been conducted to date probe for common SNPs and DNA variants, in accordance with the common disease-common variant hypothesis. This hypothesis states that complex diseases such as CAD are the result of the accumulation of the effects of numerous SNPs found commonly in the population[34].

Individually, these disease-associated SNPs would have modest effect size but in combination result in a diseased phenotype.

GWAS are typically performed using chip-based microarray technology that interrogate approximately one million SNPs. The two most common genotyping platforms for GWAS are products from Illumina and Affymetrix, which differ in terms of technology and SNP selection[35]. Typically, the results of a statistical test are denoted as significant and the null hypothesis is rejected if the p value is below 0.05. Using p = 0.05, 5% of the time there will be false positive results where the null hypothesis is rejected when actually true. However, since typically a million SNPs are directly genotyped in a GWAS study, using a p value cut-off of 0.05 will yield a tremendous amount of false-positive disease associations. To eliminate false positive results, published GWAS apply and report Bonferroni correction, which adjusts the p value cut-off for the number of statistical tests performed[35]. If a million SNPs are genotyped in a GWAS study, the p value of 0.05 has to be divided by a million, to give a cut-off of $p < 5 \times 10^{-8}$ (0.05/1,000,000). Arguments are made that Bonferroni correction is overly conservative due to linkage disequilibrium (LD) between GWAS markers, which may result in important disease markers that are nominally significant being missed[35]. An alternative to applying Bonferroni correction for GWAS is to determine the false discovery rate (FDR). Since GWAS directly interrogate only a subset of SNPs in the genome, GWAS datasets undergo a process called imputation to fill in genotypes for the remaining SNPs that were not genotyped[35]. Genetic imputation exploits already known LD patterns and haplotype frequencies from reference panels such as HapMap and the more recent 1000 Genomes Project[35].

Due to the stringent Bonferroni threshold for statistical significance, most GWAS require large sample sizes to generate a sufficient number of disease-associated signals[33]. GWAS typically used a tiered approach, whereby disease-associated SNPs are first identified in a cohort referred to as the discovery set. These SNPs are then often carried forward and genotyped in a cohort referred to as a replication set, which helps in the identification of false-positive associations[33].

The first GWAS for CAD were published in 2007 and all highlighted the 9p21.3 locus as showing the strongest association with CAD. The first study to report 9p21 as highly associated with CAD was published by our laboratory in *Science* that year. A 58 kb interval at this 9p21 locus was consistently associated with CAD in more than 23,000 samples comprising four Caucasian populations, including our own Ottawa Heart Study samples[36]. In the same issue of *Science*, a GWAS containing samples from Iceland and the United States reported variants at the 9p21 locus highly associated with MI[37]. Similar to the results from McPherson *et al*. the SNPs at 9p21 associated with MI are in a large LD block spanning several dozen kilobases. Another notable GWAS for CAD was conducted by Samani *et al*. and published in the *New England Journal of Medicine* later on in 2007[38]. This GWAS used samples from the Welcome Trust Case Control Consortium study (WTCCC), containing 1926 cases with CAD and 2938 controls, as the discovery population and the German MI Family Study, containing 875 cases with CAD and 1644 controls, as the replication cohort. Similar to the studies published in *Science*, the 9p21.3 locus showed the strongest associations with CAD in both the WTCCC and German MI studies[38].

The CAD-associated SNPs at the 9p21 locus are in proximity to the *CDKN2A* and *CDKN2B* genes, which encode the $p16^{INK4A}$, p14ARF, and $p15^{INK4B}$ proteins[36,37]. These proteins are tumour suppressors with important roles in cell proliferation and apoptosis[39]. Furthermore, this 9p21 CAD risk interval overlaps with a recently discovered noncoding RNA termed ANRIL (also known as *CDKN2BAS*) that spans 126.3 kb and partially overlaps with the 5' end of *CDKN2B*[40]. After the identification of 9p21 as the most robust CAD association and the characterization of ANRIL[41], dozens of studies have attempted to functionally characterize CAD variants at this locus. There is currently no consensus as to the mechanisms whereby these risk variants are involved in CAD pathogenesis. This locus does not associate with the traditional CAD risk factors, thus suggesting a novel pathogenic mechanism. In addition to CAD, GWAS have been conducted for individual CAD risk factors, including type II diabetes, blood lipids, smoking behaviour, and blood pressure[30,32].

1.4 CARDIoGRAM and CARDIoGRAMPlusC4D Studies

In order to further elucidate the genetic basis of CAD, the CARDIoGRAM (Coronary ARtery DIsease Genome-wide Replication and Meta-analysis) Consortium was established to generate a meta analysis of CAD GWAS from across North America and Europe. The beginnings of the CARDIoGRAM study consisted of a meta analysis of 14 CAD GWAS comprising 22,233 CAD cases and 64,762 controls all of European descent[42]. This was followed by genotyping the top association signals in 56,682 additional individuals. The original CARDIoGRAM study, published in *Nature Genetics* in 2011, identified 13 novel loci associated with CAD (reaching genome-wide significance threshold of $p<5x10^{-8}$), and

confirmed association of 10 out the 12 loci that had previously been reported to be associated with CAD. For the 13 novel CAD-associated loci, the risk alleles were all common in the population (minor allele frequency greater than 5%), and had odds ratios ranging from 1.06 to 1.17. Most of the novel CAD-associated loci that were identified were not associated with traditional CAD risk factors, suggesting these novel loci are involved with new and unidentified mechanisms in the pathogenesis of CAD. The rs4773144 SNP at the *COL4A1/COL4A2* locus, at chromosome 13q34, one of the two loci investigated in this project, was first identified in this paper. The loci identified in this first CARDIoGRAM paper, along with the lead SNPs of previously published loci, accounted for approximately 10% of the additive genetic variance of CAD[42].

The second CARDIoGRAM paper was published in 2013 in *Nature Genetics*[26] in collaboration with the Coronary Artery Disease (C4D) Genetics Consortium[43] (referred to as CARDIoGRAMplusC4D). The C4D Consortium originally published a meta analysis of four large genome-wide association studies for CAD in a discovery data set consisting of individuals of European and South Asian descent (mostly from India and Pakistan)[43]. This C4D meta analysis identified five novel loci associated with CAD. The original CARDIoGRAM discovery data set (22,233 cases and 74,762 controls) was expanded to include 34 additional CAD sample collections of European and South Asian descent, consisting of 41,513 CAD cases and 65,919 CAD controls. This study identified 15 loci reaching genome-wide significance and increased the number of CAD susceptibility loci to 46. This second CARDIoGRAM paper identified another CAD-associated SNP at the *COL4A1/COL4A2* locus, rs9515203, which is independent from the rs4773144 SNP reported in the first CARDIoGRAM study.

The third CARDIoGRAM/CARDIoGRAMplusC4D paper was published in October 2015 in *Nature Genetics*[44]. The novel feature of this latest CARDIoGRAM paper was use of 1000 Genomes Project data and imputation to expand coverage to investigate many more variants, including insertions/deletions and less common SNPs with a minor allele frequency (MAF) between 0.05 and 1.0 %. The use of 1000 Genomes Project imputation allowed low frequency variants (MAF < 0.05) to be probed which was not possible in previous CARDIoGRAM papers. This CARDIoGRAM paper comprised 48 studies (European, South Asian, and East Asian ancestry) with 60,801 CAD cases and 123,504 controls. After interrogation of 6.7 million common variants (MAF > 0.05) and 2.7 million low frequency variants (0.005<MAF<0.05), ten new CAD-associated loci were identified. 8 of these loci showed significance under an additive model and 2 showed significance under a recessive model. One of the new CAD association signals was rs56062135 in *SMAD3* on chromosome 15, further investigated in this thesis. Joint association analysis conducted in this paper identified 202 FDR variants (q value < 0.05) in 129 loci that were associated with CAD. The majority of these 202 FDR variants were common variants, with half mapping to previously reported GWAS loci. Out of these 202 FDR variants, 7 independent CAD-associated SNPs mapped to or beside the *COL4A1/COL4A2* locus on chromosome 13.

1.5 Coronary Artery Disease Pathway Analysis, Gene-Gene Interactions and Epistasis

Although GWAS have been very successful at identifying dozens of novel CAD-associated loci, including many not associated with traditional risk factors, a number of limitations still

exist[29]. First, the majority of CAD-associated loci reported by CARDIoGRAM have modest effect sizes, with odds ratios typically less than 1.3. In aggregate, these loci when considered independently still only explain a small fraction of CAD heritability (approximately 20%)[44]. The most recent CARDIoGRAMPlusC4D paper also disproved the theory that some of the missing CAD heritability is due to low frequency variants of large effect size[44]. Second, it is difficult to follow up identification of individual GWAS loci with identification of mechanisms and pathways of relevance that contribute to CAD[29] and the causal mechanisms for the majority of these CAD-associated loci remain to be elucidated. In addition, current GWAS for CAD have low power to capture gene-environment interactions.

To delve further into the genetic architecture of CAD and gain more insight into causal CAD pathways, the CARDIoGRAM Consortium performed pathway analysis[45]. A systems genetics approach was conducted by taking large-scale GWAS data for CAD and integrating with pathways (gene sets) from REACTOME. A two stage gene set enrichment analysis was employed using 7 meta analyzed CAD GWAS data sets in the discovery cohort and 9 additional meta analyzed studies in the replication cohort. Overall, 32 REACTOME pathways demonstrated significant association with CAD, including TGFβ/SMAD receptor complex, extracellular matrix integrity, collagen formation, and cell-extracellular matrix interactions[45].

Overall, CAD is a complex disease for which variation between individuals is determined by many variants, all of which have small effects. GWAS search for simple additive effects, whereby all causal variants display independent, additive, and cumulative effects on the trait[46,47]. However, the question arises whether these polymorphisms act independently or

are instead dependent on other polymorphisms across the genome. These interactions between polymorphisms, or interactions between genes, are defined as epistasis. Epistasis can be both functional/biological and statistical. Functional epistasis describes when the biological effect of one variant is dependent on the genotype of another variant. Statistical epistasis refers to interaction variance that explains effects of causal variants that is more than their independent effects[46]. Shedding light on functional epistasis can highlight and further explain important mechanisms in complex diseases. In addition, identifying instances of statistical epistasis may uncover some of the missing heritability of complex diseases such as CAD and be a beneficial tool for prediction of complex traits based on genotype information[46]. The extent in which epistasis influences complex traits is a topic of considerable interest and debate in human genetics and evolution[47].

Although many gene-gene interactions are likely to occur throughout the genome, they have been difficult to detect for a number of reasons. One of these is that since there are millions of SNPs in the human genome, the computational power required to carry out such a high number of pairwise tests is substantial. Second, due to the high number of pairwise tests a high degree of false positives is likely using a cut-off of $p<0.05$. If Bonferroni correction is applied to such a high number of tests, few gene-gene interactions would be strong enough to pass the Bonferroni cut-off value[48]. Large sample sizes are often required for successfully detecting interactions, but this is not always possible[46].

To overcome these hurdles and enable detection of epistasis stemming from human GWAS and improve power, several methods of filtering have been employed. These filtering methods prioritize either candidate genes or SNPs using prior biological knowledge. This

can encompass previously known pathway information, protein-protein interactions, or published GWAS hits[48–51]. Usage of prior biological knowledge reduces the computational power and reduces the number of pairwise tests required for epistasis analysis.

To date only a few studies have reported epistasis in the context of CAD, and these have typically employed candidate gene approaches[52–54]. A recent study using patients from the British Heart Foundation (2,101 CAD cases and 2,426 controls) conducted a primary analysis investigating pairwise interactions among 913 independent common SNPs (MAF>0.1) nominally associated with CAD and a secondary analysis among 11,332 independent common SNPs meeting quality control criteria[55]. For both these analyses several dozen suggestive interactions were detected but did not reach statistical significance after correction for multiple testing.

1.6 *COL4A1/COL4A2* Locus

One of the two loci studied in this project was *COL4A1/COL4A2* that maps to chromosome 13q34[56]. The *COL4A1* and *COL4A2* genes are arranged in a head-to-head conformation and are transcribed on opposite strands. *COL4A2* is transcribed on the forward strand whereas *COL4A1* is transcribed on the reverse strand. COL4A1 and COL4A2 belong to the type IV collagen family, along with other members COL4A3, COL4A4, COL4A5, and COL4A6. *COL4A3-COL4A4* and *COL4A5-COL4A6* have similar pairwise head-to-head orientations on the same chromosome. Unlike genes coding for the fibrillar collagens (ie. types I, II, III, and V collagen), *COL4A1* and *COL4A2* do not have the typical 54 bp exon repeating units[57]. Both *COL4A1* and *COL4A2* are very large genes, being approximately 158 kb and 207 kb in

size, respectively. Furthermore, *COL4A1* contains 52 exons whereas *COL4A2* contains 48 exons[58]. Both genes contain numerous large introns, many of which are several kilobases in size. Despite high degrees of sequence homology of the corresponding proteins, the *COL4A1* and *COL4A2* genes have diverged a substantial amount over time[58].

In relation to their head-to-head orientation, the *COL4A1* and *COL4A2* genes share a common, small, bidirectional promoter region of 127 bp[59,60]. This arrangement of the type IV collagen genes with a bidirectional promoter is also found in mice[61]. By itself, the *COL4A1/COL4A2* bidirectional promoter has no intrinsic transcriptional activity[59]. Instead, transcriptional activity of the *COL4A1* and *COL4A2* genes requires the bidirectional promoter sequence in concert with proximal regulatory sequences beside the promoter at the 5' ends of the *COL4A1* and *COL4A2* genes. These proximal regulatory sequences needed for transcriptional activity consist of both intronic and exonic elements[62]. *In vivo* analysis of the transcription rates of the *COL4A1* and *COL4A2* genes indicates *COL4A1* is transcribed more efficiently than the *COL4A2* gene, at an approximate ratio of 2:1[63]. This subsequently corresponds to the ratio of 2 COL4A1 chains to 1 COL4A2 chain in the protein triple helices[63].

This *COL4A1/COL4A2* bidirectional promoter does not contain a TATA box for transcription initiation, suggesting that there are likely several different sites that could facilitate transcription initiation[58]. The arrangements of the *COL4A1* and *COL4A2* genes on chromosome 13 are not the only cases of genes sharing a common, bidirectional promoter. The *gal1-gal10*[64] and histone H2A-H2B[65] genes in yeast, yp1-yp2[66] and surfeit loci[67] in *Drosophila*, and viral genomes of simian virus-40 (SV40)[68] and adenovirus[69] all have

characterized bidirectional transcriptional units. In humans potentially 10% of genes are part of bidirectional gene pairs (separated by less than 1000 base pairs)[70].

Although the structure of the regulatory elements required for *COL4A1* and *COL4A2* gene expression have been well mapped, little is known about the transcription factors and other elements that bind to the *COL4A1/COL4A2* bidirectional promoter region. A new nuclear factor regulating transcription of both the human *COL4A1* and *COL4A2* genes was discovered that binds the promoter *in vitro* and recognizes a homopyrimidine/purine sequence called a "CTC box." This novel nuclear factor was thus termed CTCBF (CTC-binding factor), and was shown to be necessary for efficient transcription of both genes[71]. In addition to CTCBF, the *COL4A1/COL4A2* bidirectional promoter can bind the general transcription factor Sp1[72], with a Sp1 binding GC box located in the middle of the promoter[59]. There is also a CCAAT box in the promoter ~100 bp upstream of the start site for *COL4A2* transcription that binds a protein called CCAAT-binding factor. Some of these transcription factors have differential effects on transcription of the *COL4A1* and *COL4A2* genes[72]. More recent studies have since clarified the *COL4A1/COL4A2* bidirectional promoter represents an overlapping region of two gene-specific promoters that use the same regulatory elements with differential efficacy[63,72].

Besides regulation at the transcriptional level, the COL4A1 and COL4A2 gene products are regulated by post-transcriptional events such as mRNA stability, alternative splicing, rate of translation, and posttranslational modifications and secretion. While the ratio of steady state of COL4A1 and COL4A2 is often 2:1, these ratios can vary depending on the specific tissue[73,74].

1.7 COL4A1 and COL4A2 Proteins

The COL4A1 and COL4A2 proteins make up the primary structural components of basement membranes, which are specialized extracellular structures/matrices that provide tissue structure and influence cellular behaviour[75]. Basement membranes are amorphous, dense, sheet-like structures that are present throughout the entire body and form compartments within tissues[76]. Typically, basement membranes are 50-100 nm in thickness, are produced by most cell types, and are similar to extracellular matrix[76]. Basement membrane layers separate endothelial and epithelial cell layers from the underlying mesenchyme[75]. The presence of basement membranes is crucial for life since mutations in basement membrane proteins give rise to numerous disorders and phenotypes affecting many different organs[75]. Basement membranes serve as extensions of cellular plasma membranes, protecting tissues from disruptive physical stresses, and are interactive interfaces between cells and the surrounding environment, allowing for transmission of signals between these entities[77]. Basement membrane signaling is traditionally facilitated by integrins, growth factors, and dystroglycan[77].

The most characterized basement membrane components are type IV collagen, laminins, nidogens, and perlecan[75,77]. Laminins are abundant basement membrane heterotrimeric glycoproteins consisting of an α, β, and γ chain that can form numerous different combinations. Laminins are typically cruciform shaped proteins that can bind to other cell membrane and extracellular matrix proteins[75]. Laminins contribute to cell attachment, differentiation, maintenance of cell phenotype, promotion of tissue survival, and cell shape and movement[78]. Nidogens are proteins found in all basement membranes that contain three

globular domains, connected by thin rod-like domains. The roles of nidogens are less clear, however, they help facilitate network formation[75,76]. Perlecan (HSPG2) is a large heparin sulfate proteoglycan that has numerous binding partners, including other basement membrane proteins, integrins, and growth factors. Functions of perlecan include facilitating cell and growth factor signaling as well as basement membrane maintenance[75]. Overall, approximately 50 different proteins make up the basement membrane, including various other forms of collagen[76].

Type IV collagen is only found in basement membranes and consists of six α chains (α1-α6, COL4A1-COL4A6)[79]. These α chain type IV collagen monomers form heterotrimers consisting of various combinations of COL4A1-COL4A6. Most type IV collagen is composed of 2 COL4A1 chains in combination with 1 COL4A2 chain, which are found in all tissues. The distribution of COL4A3, COL4A4, COL4A5, and COL4A6 is restricted to specific tissues[79]. Each type IV collagen chain contains three separate, distinct domains. There is an amino-terminal domain highly enriched in cysteine and lysine amino acids, a long collagenous domain consisting of the three amino acid repeat Gly-X-Y (where X and Y are variable residues) approximately 1400 amino acids in length, and a carboxy terminal non-collagenous domain (NC1)[76,79,80].

The synthesis of COL4A1 and COL4A2 proteins is a complex process involving numerous post-translational modifications. These co-ordinated modifications include removal of a signal peptide in the endoplasmic reticulum, as well as hydroxylation of numerous lysine and proline amino acids[79,81]. Each individual type IV collagen chain is heavily glycosylated, containing ~50 hydroxylysine-linked disaccharide units in the collagenous domain and an

asparagine-linked oligosaccharide unit near the amino terminus[82]. Biosynthesis of COL4A1 and COL4A2 also requires one or more specific molecular chaperone proteins, including heat shock protein 47 (HSP47), which is necessary for proper folding of the individual collagen chains[79]. HSP47 binds to and stabilizes the triple helical region of collagen chains on their way from the endoplasmic reticulum to the Golgi[83,84]. Formation of proper type IV collagen trimers and secretion via the secretory pathway requires several posttranslational modifications mediated by an array of proteins including protein disulfide isomerase (PDI), peptidylprolyl isomerase (PPIA), and various hydroxylases[81]. Secretion of type IV collagen proteins requires trafficking vesicles, of which the protein TANGO1 is a key component, due to the large molecular weight of type IV collagens[85].

Assembly of COL4A1 and COL4A2 into a larger type IV network consists of several tiers at the macromolecular level (**Figure 3**). First, individual COL4A1 and COL4A2 chains initiate assembly into trimers ("protomers") through interaction between three NC1 domains. The NC1 domains are crucial for proper molecular recognition and determine the stoichiometry of the COL4A1 and COL4A2 chains[86]. Next, protomer trimerization occurs like a zipper, going from the carboxy terminus to the amino terminus. The collagenous domain made up of Gly-X-Y amino acid repeats is responsible for forming a helical structure[76,80]. Once type IV collagens form trimeric protomers (2 COL4A1 chains and 1 COL4A2 chain), these protomers then form dimers through binding of NC1 ends to each other, forming a NC1 hexamer. Finally, four type IV collagen protomers can interact via their N-terminal 7S regions to form tetramers[76]. The cysteine and lysine residues at the amino terminus 7S domains are important for interchain cross-linking of four type IV collagen heterotrimers together, via disulfide bonds and lysine-hydroxylysine crosslinks[79,81].

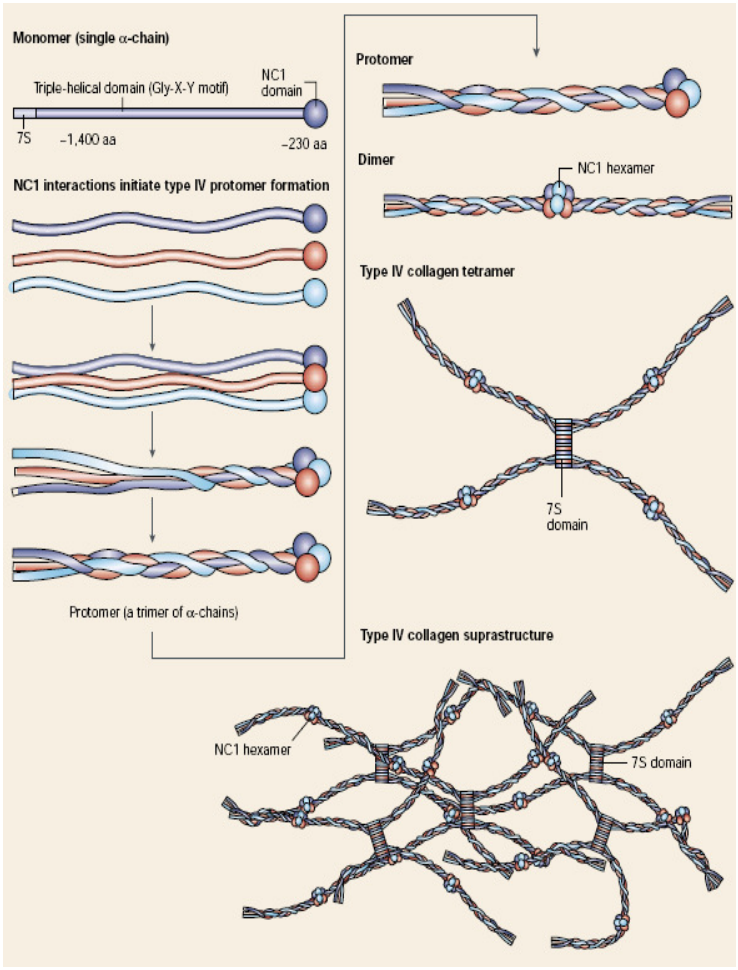

Monomer (single α-chain)
Triple-helical domain (Gly-X-Y motif)
NC1 domain
7S
~1,400 aa
~230 aa
NC1 interactions initiate type IV protomer formation
Protomer (a trimer of α-chains)
Protomer
Dimer
NC1 hexamer
Type IV collagen tetramer
7S domain
Type IV collagen suprastructure
NC1 hexamer
7S domain

Figure 3. Assembly of individual type IV collagen proteins into triple helices (protomers consisting of 3 α chains) and subsequent organization into higher order structures. Each α chain contains an amino terminal 7S domain, a middle triple helical domain, and a carboxy terminal non-collagenous domain (NC1). The NC1 domain of type IV collagens are critical for initiation and assembly of the trimeric structure. Trimerization then occurs similar to a zipper starting at the carboxy terminal end. Higher order type IV collagen structures form from both end-to-end as well as lateral interactions between triple helical protomers.

Figure taken from Kalluri (2003) Basement membranes: Structure, assembly, and role in tumour angiogenesis. *Nature Reviews Cancer* 3(6): 422-433, with permission. License agreement number 3800850110511.

This macromolecular organization of COL4A1 and COL4A2 continues (both end-to-end and lateral interactions between protomers) until a large-scale mesh-like type IV collagen lattice has been assembled at the basement membrane. The triple helical domains of COL4A1 and COL4A2 typically bind to cells through interactions with α1β1 and α2β1 integrins on the cell surface[87]. Other studies have demonstrated that when denatured, type IV collagen can also bind to integrin αvβ3[88,89].

COL4A1 and COL4A2 are found in blood vessels with ubiquitous distribution, where they act to maintain vascular wall integrity during exposure to varying blood pressures[90]. In general, type IV collagen promotes cell adhesion, migration, differentiation, and growth[91]. Due to their important basement membrane functions, COL4A1 and COL4A2 are essential components of tissues and are essential during development. Mice with a null allele of the *Col4a1/Col4a2* locus develop up to E9.5 and display a variable degree of delayed growth. Lethality occurs for these mice between E10.5-E11.5 due to structural deficiencies in basement membranes and integrity failure of Reichert's membrane. These mouse findings suggested COL4A1 and COL4A2 are necessary for maintaining the function and integrity of basement membranes, but that other basement membrane proteins are sufficient for initial membrane deposition and assembly[92].

Besides their roles as structural constituents of the basement membranes, COL4A1 and COL4A2 have other signaling functions. The C-terminal NC1 domains of COL4A1 (arresten) and COL4A2 (canstatin) can both interact with endothelial cells and modulate the process of angiogenesis[93,94]. These NC1 segments are released after proteolytic degradation of the extracellular matrix. Arresten is a 26 kDa protein that functions as an anti-angiogenic

molecule by inhibiting endothelial cell migration, proliferation, and tube formation[93]. Canstatin is a 24 kDa anti-angiogenic protein that inhibits endothelial cell migration, proliferation and can induce apoptosis[94,95]. Arresten, canstatin and other NC1 domains of type IV collagen (also referred to a "matrikines") have been the focus of study in the cancer field as they inhibit *in vivo* tumour growth in several *in vivo* cancer models[94–98]. The tumour suppressor protein p53 increases arresten expression via increased COL4A1 transcription through binding to an enhancer sequence located 26 kb 3' of the *COL4A1* gene. p53 also can stabilize expression of full-length COL4A1 by upregulating α(II) prolyl-hydroxylase and increase release of arresten from COL4A1 through a MMP-dependent mechanism[99]. Arresten binds to cells via integrin α1β1 whereas canstatin binds to cells through either integrin αvβ3 or integrin α3β1[76].

1.8 Roles of Collagen in Coronary Artery Disease

The arterial wall consists of a highly ordered structure of cells and extracellular matrix[100]. Although numerous collagens are expressed in the vessel wall, collagens I and III are the predominant forms where their functions are to impart mechanical/tensile strength[101]. In the intima, endothelial cells lie on a basement membrane rich in type IV collagen and other extracellular matrix proteins[100,101]. The smooth muscle cells in both the intima and media secrete their own basement membranes of which collagen type IV is a large constituent[100,101]. In the normal arterial environment, the smooth muscle cells are embedded in a collagen network containing large amounts of collagens I, III, and V, each layer separated by a well-defined elastic lamina[100,102].

As atherosclerosis progresses, smooth muscle cells switch from a quiescent, contractile phenotype to a proliferative and synthetic phenotype[103]. During this process, smooth muscle cells increase their synthesis of many forms of collagen, which contributes up to 60% of total protein in the plaque[104]. Increased synthesis of collagen and extracellular matrix coincides with increased production of MMPs that in turn can degrade the many forms of collagen and reveal cryptic fragments of collagen. These cryptic collagen fragments influence a number of cellular behaviours in the plaque, including smooth muscle cell migration[100].

With respect to type IV collagen and atherosclerosis, expression is higher in contractile and quiescent smooth muscle cells compared to proliferating and migrating smooth muscle cells[105]. Type IV collagen is proposed to help maintain the differentiated phenotype of smooth muscle cells[106] and the α chain composition may determine the precise contractile function of vascular smooth muscle cells[107]. Collagen IV and other basal laminae proteins can also reduce inflammatory gene expression, reduce LDL uptake in culture, and inhibit extracellular matrix calcification[101,108,109]. Cell migration was reduced when cells were grown in dishes coated with type IV collagen[108]. In contrast, adding type IV collagen to media in culture actually enhances smooth muscle cell migration[110]. Therefore the effect of type IV collagen on smooth muscle cell proliferation and migration might depend on the specific context. Orr *et al*. found that plating smooth muscle cells on collagen IV resulted in elevated expression of smooth muscle cell markers and contractility proteins compared to type I collagen[108]. *In vivo*, type IV collagen protein expression is elevated in advanced atherosclerotic lesions, surrounding smooth muscle cells[111,112].

Typically, type IV collagen binds to smooth muscle cells through integrin receptors or discoidin domain receptors (DDRs). DDRs (DDR1 and DDR2) are widely expressed receptor tyrosine kinases on the cell surface. Type IV collagen can stimulate tyrosine kinase activity for prolonged periods through interaction of triple helical regions with DDR1[79]. Type IV collagen only stimulates phosphorylation of DDR1, not DDR2, and this interaction is independent of β1 integrins[79,113].

Besides roles in smooth muscle function, type IV collagen could potentially have other roles relevant to atherosclerosis. One study demonstrated that macrophages are able to adhere to non-enzymatically glycated collagen IV but not to unmodified collagen IV[114]. These macrophage interactions with glycated type IV collagen were mediated by scavenger receptors, and blocked the ability of macrophages to endocytose acetylated LDL[114].

1.9 COL4A1 and COL4A2 Mutations and Disease

COL4A1 and COL4A2 mutations may alter the structure and function of basement membranes and give rise to numerous vascular abnormalities, many of which pertain to atherosclerosis and MI. Mice with a *Col4a1* mutation display focal detachment of the endothelium from the media and age-dependent defects in vascular function[115]. These defects affect the function of endothelial cells and vascular smooth muscle cells that associate with reduction in basal nitric oxide synthase (NOS) function and increase in smooth muscle nitric oxide sensitivity. These mutant mice have defective deposition of type IV collagen in the basement membrane of descending aortae that leads to activation of the unfolded protein response[115]. A GWAS study identified the rs3742207 coding SNP

(Gln1334His) in COL4A1 as associated with arterial stiffness (measured by pulse wave velocity), which is a predictor of adverse cardiovascular events[116]. In addition, rs3742207 associated with the prevalence of MI in a large scale association study in a Japanese population[117]. No functional studies have been conducted on this SNP concerning the mechanism whereby the Gln to His substitution contributes to these phenotypes. In another GWAS study for coronary artery calcification, the *COL4A1/COL4A2* locus nearly reached genome-wide significance[118].

Due to the wide-ranging presence of basement membranes, mutations in *COL4A1* and *COL4A2* can affect multiple organs and tissues and give rise to numerous cerebral, renal, muscular, and ocular pathologies. Many of these mutations can also affect small vessels and overall result in impaired basement membrane protein levels or assembly. One of the most characterized disorders resulting from *COL4A1/COL4A2* mutations is porencephaly, a rare neurological disease highlighted by the presence of degenerative cavities in the brain[119,120]. Porencephaly typically manifests itself in infants and severe cases can result in significant disability or even death. Mice with a heterozygous mutation that results in an in-frame deletion of exon 40 of *Col4a1* are more susceptible to cerebral haemorrhaging at birth[119]. This mutation affected the secretion of both the mutant and normal type IV collagen[119]. *COL4A1* mutations in patients with porencephaly typically affect either secretion[121] or impair assembly of type IV collagen heterotrimers[122]. Human *COL4A2* mutations have also been implicated in familial porencephaly and small vessel disease[123–125].

A 2007 study published in the *New England Journal of Medicine* proposed a new syndrome called hereditary angiopathy with nephropathy, aneurysms, and muscle cramps (HANAC)[126]. Three *COL4A1* mutations were described in subjects from three separate

families with HANAC, located in either exon 24 or 25. These mutations all affected glycine residues and interrupted the Gly-X-Y amino acid repeat of the collagenous domain. Patients with HANAC develop nephropathy, fragile and damaged blood vessels, and often develop cysts in the kidneys[126]. Several additional *COL4A1* mutations have been subsequently identified in patients with HANAC[121,127].

COL4A1 mutations have also been implicated in small vessel disease in relation to the brain, which underlies approximately 20 to 30% of ischemic strokes and an even greater proportion of intracerebral hemorrhages[121,128–133]. Phenotypic studies in mice suggest *Col4a1* mutations interact with environmental factors to determine the severity of vessel disease[119,128]. *Col4a1* mutations in mice (mostly glycine residues) cause myopathy, numerous brain malformations, and ocular dysgenesis[134]. Similarly, the *COL4A1* gene has also been implicated in Walker-Warburg syndrome, a childhood form of congenital muscular dystrophy involving ocular dysgenesis, congenital myopathy, and neuronal migration defects[135].

1.10 SMAD3 and TGFβ Signaling

The SMAD3 protein is a transcription factor that acts as a key signaling transducer of the TGFβ (transforming growth factor β) signaling pathway. It is an intracellular protein that transduces extracellular signals from TGFβ ligands to the nucleus where it regulates transcription. SMAD3 belongs to the more general SMAD family of transcription factors that can be divided into several groups. There are several receptor regulated SMAD (R-SMAD) proteins: SMAD1, SMAD2, SMAD3, SMAD5, and SMAD8. Upon activation, they

become phosphorylated and form heterotrimeric complexes with the common mediator (co-SMAD) SMAD4, which then shuttle to nucleus and regulate transcription of specific SMAD genes[136].

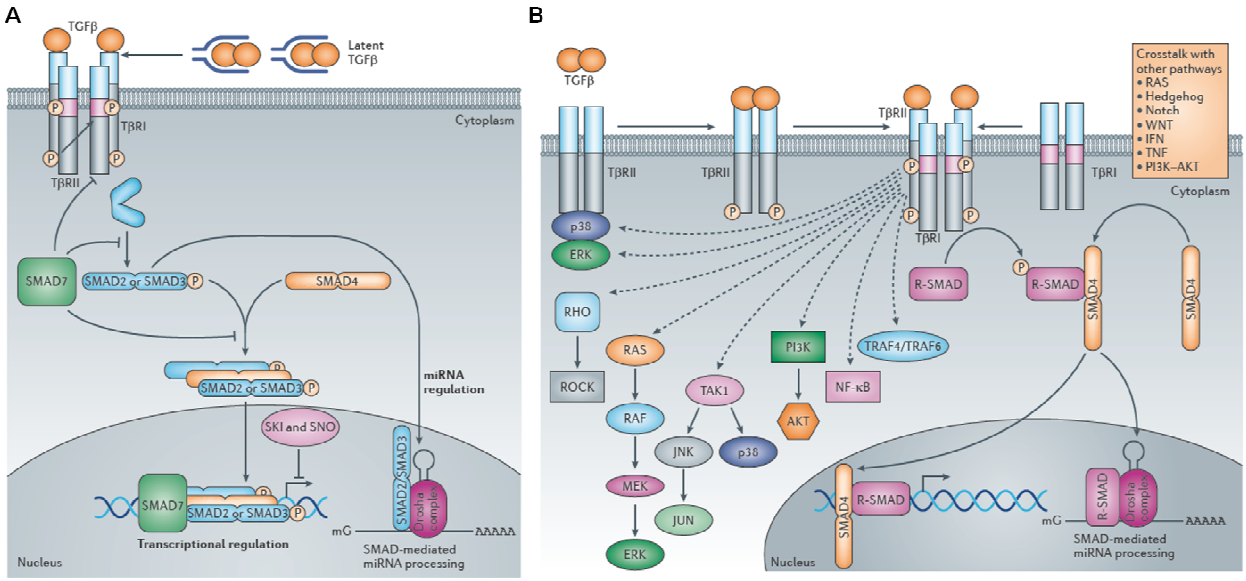
The SMAD signaling pathway is the canonical TGFβ signaling pathway for TGFβ family members. Overall vertebrates contain over 30 different pleiotropic ligands belonging to the broad TGFβ family[137]. The TGFβ superfamily includes the TGFβs, BMPs (bone morphogenic proteins), activins, inhibins, Nodal, myostatin, anti-Müllerian hormone, and growth and differentiation factors[138]. Members of the TGFβ family control a wide variety of cellular functions and their activities regulate various developmental and homeostatic processes[139]. Some notable characteristics of TGFβ family members are growth inhibition of most cell types, stimulation of extracellular matrix protein expression, and induction of apoptosis in epithelial cells[137,138]. TGFβ also has a well described immunomodulary role and TGFβ1-deficient mice die either *in utero* or perinatally due to widespread inflammation[140,141].

Here, the focus will be on the TGFβs as they are the best characterized members of the TGFβ superfamily. In humans, there are three highly homologous isoforms of TGFβ: TGFβ1, TGFβ2, and TGFβ3. These isoforms share the same receptor complexes and signal in similar ways, with the main difference being tissue-specific expression of different isoforms[137]. The R-SMADs activated downstream of the TGFβs are SMAD2 and SMAD3[142]. Every TGFβ ligand is synthesized as a part of a latent precursor complex, consisting of a TGFβ homodimer, a latency associated peptide (LAP), and a latent TGFβ binding protein[137]. TGFβ activation subsequently requires proteolysis of the LAP[137].

In the absence of TGFβ stimulation, both the type I and type II TGFβ receptors exist as inactive homodimers at the cell surface[143]. These TGFβ receptors are serine-threonine kinases, with the type I receptors having a characteristic glycine-serine (GS) sequence upstream of the kinase domain[143]. The exact mechanisms of TGFβ signaling depends on the specific TGFβ ligand, but the canonical TGFβ signaling pathway works through the same general pathway (**Figure 4A**). First, TGFβ binds to the type II TGFβ receptors, which then recruit and phosphorylate type I TGFβ receptors. Interaction of the type II and type I receptors induces phosphorylation of type I receptors at the GS sequence by the type II receptor kinases.

On the kinase domain of the type I receptor, the L45 loop specifies the receptor/SMAD interaction and the phosphorylated GS sequence facilitates the interaction[142]. The catalytic domain of activated type I receptors subsequently phosphorylates the receptor-regulated SMADs (SMAD2 and SMAD3) at C-terminal serine residues[142,143]. Recruitment of SMAD2 and SMAD3 to the TGFβ receptor complex is controlled by a protein termed SARA (SMAD anchor for activation). SARA presents SMAD2 and SMAD3 to the type I TGFβ receptor by binding cooperatively to the non-phosphorylated SMADs and the receptor complex[144].

Once activated, SMAD2/SMAD3 then interact with the common (co) SMAD4, SMAD4. These SMAD complexes then translocate to the nucleus where they regulate transcription of SMAD target genes. Phosphorylated SMAD3 promotes its interaction with the co-activators such as p300 and CBP (cyclic AMP response element-binding protein) that along with the SWI/SNF nucleosome positioning complex help form a transcriptional complex[145]. While phosphorylated SMAD3 along with p300 are synergistically able to augment



C

MH1 Domain
DNA binding
Interaction with transcription factors

MH2 Domain
Activation of transcription
SMAD oligomerization
Interaction with TGFβ receptor

Figure 4. Depictions of the canonical (A) and non-canonical (B) TGFβ signaling pathways and structure of the SMAD3 protein (C). (A) In the canonical TGFβ signaling pathway, TGFβ binds to the type II TGFβ receptor that subsequently recruits and phosphorylates the type I TGFβ receptor. The phosphorylated type I TGFβ receptor phosphorylates the receptor regulated SMAD proteins (such as SMAD3), that then join with the common (co) SMAD, SMAD4. These complexes then translocate to the nucleus to regulate transcription. (B) For non-canonical TGFβ signalling, the activated TGFβ receptor complex transmits signals through various other factors independent of SMAD proteins. TGFβ can cross-talk with several other pathways (eg. ERK, JNK, and p38 MAP Kinase pathways), thus allowing to it to propagates signals through various means. (C) Structural organization and domains of the SMAD3 protein. The conserved MH1 and MH2 domains are separated by a linker sequence. SMAD3 is phosphorylated by the type I TGFβ receptor at the C terminal SSXS sequence, where S represents serine residues. The specific residues phosphorylated by the type I TGFβ receptor (serine 423 and serine 425) are indicated by the orange arrows.

Panels A and B taken from Akhurst and Hata, (2012) Targeting the TGFβ signaling pathway in disease. *Nature Reviews Drug Discovery* 11(10): 790-811, with permission. License agreement number 3800851096691.

transcription[146], SMADs can also interact with transcriptional co-repressors that recruit histone deacetylases (HDACs)[145]. The effects of TGFβ on gene transcription are either positive or negative depending on the target gene and cellular context (ie. availability of SMAD partners, epigenetic context of cell), and few TGFβ target genes are the same across different cell types[145]. In terms of noncanonical TGFβ signaling, TGFβ can exert effects and crosstalk with several other pathways (such as ERK, JNK, and p38 MAP Kinase) independent of SMAD factors[137] (**Figure 4B**).

The *SMAD3* gene encodes a protein 425 amino acids in size consisting of two major domains, a MH1 (mad homology 1) and a MH2 (mad homology 2) domain, which flank a middle linker segment[143] (**Figure 4C**). Overall there are eight SMAD proteins in vertebrates (SMAD1-SMAD8), all of which are structurally similar[143]. The MH1 domain of SMAD3 is capable of binding to DNA and interacting with other transcription factors while the MH2 domain has a wide variety of functions. The MH2 domain, which interacts with and becomes phosphorylated by the TGFβ receptor, is responsible for oligomerization with other SMAD proteins. In addition, the MH2 domain functions in transcription activation as well as interacts with transcriptional coactivators p300 and CBP among other transcription factors[143].

Ligand-induced interaction of SMAD3 with activated type I TGFβ receptors results in phosphorylation of SMAD3 at two serine residues (serine 423 and serine 425) at the C-terminal SSXS motif by the type I receptor[143]. This C-terminal phosphorylation causes SMAD3 to dissociate from the type I TGFβ receptor and interact with either another R-

SMAD protein or SMAD4 via phosphoserine binding pockets in the L3 loop region in the MH2 domain[143].

In the nucleus, SMAD3-SMAD4 complexes are involved in the transcriptional regulation of a wide panel of genes. The MH1 domains of both SMAD3 and SMAD4 can bind DNA with the specific 5'-AGAC-3' sequence, termed SMAD-binding elements (SBE)[144]. A protruding β-hairpin loop in the MH1 domains of SMAD3 and SMAD4 is responsible for binding to SBE DNA elements. Unlike the other SMAD proteins, SMAD2 is unable to bind DNA[144]. However, the affinity of SMAD3 and SMAD4 for binding to DNA is weak and SMAD proteins commonly require complexes with other sequence specific transcription factors for full efficacy[144].

The SMAD family is completed by the presence of two inhibitory SMAD proteins (I-SMADs), SMAD6 and SMAD7. These SMADs lack an MH1 domain and have negative feedback roles in TGFβ signaling through different mechanisms[142]. SMAD6 is cytoplasmic and binds SMAD4, preventing it from interacting the R-SMADs. Conversely, SMAD7 is nuclear and upon TGFβ stimulation shuttles to the cytoplasm and competes with the R-SMADs for TGFβ type I receptor interaction, thereby preventing R-SMAD phosphorylation[142].

In addition to C-terminal phosphorylation, SMAD proteins are regulated at various other levels post-translationally. For instance, the E3-ubiquitin protein ligases SMURF1 and SMURF2 bind to SMADs, leading to transfer of ubiquitin and eventually SMAD degradation via the proteosome. Furthermore, a SMURF2-SMAD7 complex exports to the cytoplasm where it downregulates TGFβ signaling by binding to the TGFβ receptor complex

and promoting its degradation[142]. ERK MAP kinases are able to phosphorylate various serine and threonine residues in the linker region of SMAD proteins[142]. Finally, the proto-oncoproteins Ski and SnoN negatively regulate SMAD signaling through recruitment of transcriptional co-repressors such as HDACs to SMAD complexes and preventing initial formation of SMAD complexes[142].

Besides being a transcriptional mediator in the canonical TGFβ signaling pathway, SMAD3 can also regulate micro RNA (miRNA) biogenesis through facilitating the processing of primary miRNA into precursor miRNA in the nucleus[137,147]. For instance, TGFβ, through R-SMADs including SMAD3, promotes the rapid increase in expression of mature miR-21 by increasing the processing of the primary miR-21 transcripts into precursor miR-21 by the Drosha complex. This induction of miR-21 subsequently promotes the induction of the contractile phenotype for smooth muscle cells. Unlike the canonical SMAD pathway, SMAD4 is not required for this miRNA processing[147,148].

1.11 *SMAD3* Locus and Role in Disease

The human *SMAD3* gene maps to chromosome 15q23 and is ~130 kb in size with 9 exons in total. Most of the exons in *SMAD3* are clustered towards the 3' end of the gene, with a large intron almost 100 kb in size separating exons 1 and 2. SMAD3 is an acronym for "Mothers against decapentaplegic homolog 3", named for SMAD proteins being homologs of the *Drosophila* protein "mothers against decapentaplegic" and the *C. elegans* protein SMA.

To date, over 20 pathogenic variants in the *SMAD3* gene have been characterized, including missense mutations, nonsense mutations, as well as frameshift and splice-site mutations.

One of the common phenotypic consequences of rare *SMAD3* mutations are aortic aneurysms, which can lead to life-threatening conditions such as aortic dissection or rupture[149,150]. Thoracic aortic aneurysms and dissections (TAAD) are a type of aortic aneurysms with a familial clustering and an autosomal dominant pattern of inheritance. Recently *SMAD3* was reported to cause a new syndromic form of TAAD, referred to as aneurysms-osteoarthritis syndrome (AOS) or Loeys-Dietz Syndrome type 3, due to early onset osteoarthritis in the majority of patients. Van de Laar *et al.* in 2011 originally described *SMAD3* mutations in three Dutch families that resulted in aneurysms, dissections, and tortuosity along with early-onset osteoarthritis, mild craniofacial features, and skeletal and cutaneous anormalities[151]. Since then further *SMAD3* mutations in subsequent families presenting with AOS have been characterized, including the phenotypic spectrum of the patients[149,152,153]. Besides arterial aneurysms, dissections and tortuosity, and early osteoarthritis, cardiac abnormalities diagnosed in patients with AOS include congenital heart disease, atrial fibrillation, mitral valve abnormalities, left ventricular hypertrophy, and bicuspid aortic valve[153,154]. These heterozygous *SMAD3* mutations are mostly in the MH2 domain of SMAD3, likely affecting oligomerization of SMAD3 with SMAD4 or SMAD3 transcriptional ability[153]. There are also frame-shift *SMAD3* mutations in the MH1 or linker domains that likely result in truncated transcripts and mRNA decay[153]. While these *SMAD3* AOS mutations are predicted to be loss-of-function, there is paradoxically an increase TGFβ signaling in the aortic wall, include increased levels of TGFβ and total SMAD3[151,153].

Smad3 knockout mice are not embryonic lethal but die between 1 and 8 months, primarily due to defects in immune function, including impaired mucosal immunity and diminished T cell responsiveness to TGFβ[155]. *Smad4* null mice die at approximately embryonic day 6.5

(E6.5), with no signs of gastrulation[156]. Similarly, *Smad2* knockout mice are embryonic lethal[157].

1.12 TGFβ/SMAD3 Signaling in Atherosclerosis

Although TGFβ signaling has wide-ranging and complex effects in the vasculature, the prevailing view in the literature is that TGFβ has an anti-atherogenic role[158]. TGFβ is synthesized both by immune cells in the circulation and atherosclerotic plaque as well as by cells in the vessel wall[159]. There is some argument concerning the association of serum TGFβ levels with CAD prevalence and severity. Several clinical studies report a negative correlation between plasma TGFβ concentrations and the extent of atherosclerotic lesions and prognosis of survival[160,161]. Conversely a recent study of 279 patients with CAD and 268 controls observed higher TGFβ1 and SMAD3 levels in cases[162].

Relevant to atherosclerosis, numerous studies have investigated the effects of TGFβ on vascular smooth muscle cell proliferation in cell culture systems. Although some of the original studies suggested TGFβ promotes smooth muscle cell proliferation[163,164], most studies indicate that TGFβ inhibits the proliferation of vascular smooth muscle cells[165–167], especially under conditions whereby proliferation is stimulated by various growth factors[158]. Besides an anti-proliferative effect, TGFβ also inhibits smooth muscle cell migration[168], another hallmark of plaque development, and increases expression of a number of proteins that help maintain the contractile phenotype of vascular smooth muscle cells[158,165–167].

TGFβ also has an immunomodulary role in the progression of atherosclerosis, especially during the early stages where macrophages and T cells are prominent in the fatty streak[169].

TGFβ can reduce foam cell formation in cultured macrophages[170], reduces accumulation of T cells, and acts a deactivating cytokine for T cells[158,171]. Overexpression of TGFβ1 in macrophages reduces and stabilizes atherosclerotic plaques in apoE deficient mice[172].

TGFβ also has a dynamic role in protection against the development of unstable atherosclerotic plaques[140,158]. First of all, since T cell-rich lesions are key hallmarks of unstable plaques, TGFβ is atheroprotective by reducing T cell accumulation and activation[171]. Secondly, TGFβ increases the stability of atherosclerotic plaques via the enhanced production of numerous extracellular matrix proteins by vascular smooth muscle cells[140,173]. Inhibition of TGFβ by a neutralizing antibody against TGFβ1, 2 and 3 in *Apoe*$^{-/-}$ mice results in an exacerbated atherosclerotic phenotype with a high inflammatory cell content and a decreased collagen content[173]. Similar findings were observed by treating *Apoe*$^{-/-}$ mice with a recombinant soluble type II TGFβ receptor[140]. Further *in vivo* evidence suggesting TGFβ is atheroprotective is based on overexpression of TGFβ in the heart (and subsequently elevated levels in the plasma) of *Apoe*-/- mice reduces plaque growth, prevents aortic dilation, and stabilizes plaque structure[174].

Despite fairly in depth study into TGFβ on a broad scale with respect to atherosclerosis, the roles of SMAD proteins, particularly SMAD3 and its signaling, have not been thoroughly investigated. In nonatherosclerotic healthy artery, SMAD3 expression (as well as SMAD2 and SMAD4) is very low in the intima, which is rich in smooth muscle cells[175]. Using a combination of immunohistochemistry and reverse transcriptase polymerase chain reaction, SMAD proteins were found to be expressed in macrophages of fibrofatty lesions and their expression elevated as monocytes differentiate into macrophages[175]. Finally, smooth muscle

cells abundantly express SMAD3 (in addition to SMAD2 and SMAD4) in aortic fibrous plaques[175].

Original evidence for roles of TGFβ in vascular disorders came from models that investigated balloon-injured rats, where TGFβ levels increased 6 to 24 hours after injury[176]. Furthermore, expression of TGFβ is elevated in human vascular restenosis lesions[177]. Mice that lack *Smad3* display enhanced neointimal hyperplasia and deposit less extracellular matrix in response to vascular injury[178]. Kobayashi *et al.*[179] reported that *Smad3*$^{-/-}$ mice display enhanced neointimal hyperplasia is response to injury, due to increased proliferation of *Smad3*-null smooth muscle cells. Conversely, other *in vivo* studies suggest TGFβ stimulates intimal hyperplasia[180,181]. In rats that underwent left carotid balloon injury, endogenous Smad3 is upregulated[182]. Mice with cardiovascular overexpression of TGFβ1 are embryonic lethal due to abnormal yolk sac vasculogenesis[183]. The same study demonstrated that infecting rat arteries with adenovirus-expressing *Smad3* increases both intimal hyperplasia and vascular smooth muscle cell proliferation[182]. Thus, due to these conflicting findings there is still debate as to the effects of TGFβ signaling in the vasculature.

1.13 AP-1 Transcription Factor

AP-1 is a heterodimeric transcription factor involved in the regulation of a diverse variety of cellular processes, including cell proliferation, transformation, and apoptosis. AP-1 consists of combinations of basic leucine zipper proteins from the Jun (c-Jun, JunB, JunD), Fos (c-Fos, FosB, Fra1, Fra2), ATF, and Maf families[184,185]. The leucine zipper domain of AP-1

proteins is responsible for dimerization, while the basic domain facilitates DNA binding[186,187]. Jun proteins are capable of forming homodimers with each other, whereas Fos proteins cannot homodimerize and require dimerization with Jun proteins for proper DNA binding activity[186,188]. The numerous possible combinations of homodimers and heterodimers in the AP-1 complex thus determines the binding specificity and affinity and subsequently the range of target genes regulated by AP-1[185,186].

AP-1 proteins bind DNA at 12-*O*-tetradecanoylphorbol-13-acetate (TPA) response elements (TRE), with the recognition sequence 5'-TGA[G/C]TCA-3'[184,186]. AP-1 can also bind cAMP response elements (CRE) with the similar recognition sequence 5'-TGACGTCA-3'[184,185,187]. Of all the AP-1 proteins, c-Jun is the most potent transcriptional activator[184,189]. Fos and FosB also have relatively strong transcriptional ability, while JunB, JunD, Fra1, and Fra2 are weak transactivators[185,186]. c-Fos when dimerized with c-Jun increases the transcriptional ability of c-Jun, while JunB antagonizes the transcriptional ability of c-Jun[190]. Overall, regulation of AP-1 is quite complex and can occur at many different levels. These include regulation of AP-1 gene transcription, mRNA turnover, protein turnover, and post-translational modifications that affect protein function and efficiency as a transcription factor. AP-1 can also be regulated via interactions with other proteins and transcription factors that can synergize or inhibit AP-1 transcriptional activity[184].

Besides dimer composition, AP-1 activity is regulated by a number of physiological and environmental factors, including induction by cytokines, growth factors, peptides, cell-matrix interactions, neurotransmitters, bacterial/viral infections, and physical and chemical stresses[184]. All of these stimuli activate various branches of the mitogen activated protein

kinase (MAPK) pathway that elevate AP-1 activity via phosphorylation of the protein subunits. In response to pro-inflammatory cytokines and genotoxic stress, AP-1 is activated typically by the JNK and p38 MAPK cascades. The activated JNK proteins translocate to the nucleus where they phosphorylate c-Jun and increase the transcriptional capability of c-Jun. p38 can induce AP-1 through direct phosphorylation of ATF2, MEF2C, and TCF (ternary complex factor) proteins. Finally, serum and growth factors stimulate AP-1 through the extracellular signal regulated kinase (ERK) pathway[184].

In terms of regulation by the JNK pathway, the major JNK phosphorylation sites on c-Jun are Serine 63 and Serine 73 at the N terminus[184,188]. Activated JNK translocates to the nucleus to phosphorylate c-Jun and this phosphorylation of c-Jun at these sites is required for proper transcriptional function[186,191]. JNK can also phosphorylate JunD that lacks a JNK docking site, but this phosphorylation is inefficient, and the phosphorylation status of JunB is controversial[188].

AP-1 (Jun and Fos) proteins were first identified as viral oncoproteins and many AP-1 component proteins can efficiently transform cells in culture[185]. c-Jun (E12.5) and JunB (E10) knockout mice display embryonic lethality[185] while JunD knockout mice exhibit male sterility and c-Fos knockout mice display osteopetrosis[185]. AP-1 has been reported to have roles in cell proliferation and cell cycle progression, with c-Jun and c-Fos promoting cell proliferation and JunB inhibiting proliferation[184]. AP-1 is also postulated to be a general regulator of several pro-inflammatory signals[192].

As part of their many roles, AP-1 proteins can interact with components of the TGFβ signaling pathway. One study reported c-Fos and c-Jun can co-operate with SMAD3 and

SMAD4 to mediate transcription in response to TGFβ stimulation[193]. This accords with findings that AP-1 regulated gene promoters are often transcriptionally induced by TGFβ[194–198]. In addition, several promoters have overlapping or nearby TREs and SMAD Binding Element sites, though it is possible AP-1 and SMAD factors can bind simultaneously[193,199,200]. These earlier findings were contrasted by reports that AP-1 proteins can antagonistically affect the transcriptional ability of SMAD3[201,202]. For instance, TNFα activated Jun proteins prevent TGFβ-induced SMAD transcriptional function. In addition, off-DNA interactions between c-Jun, JunB, and SMAD3 reduce SMAD3 binding to DNA. Jun family members compete with SMAD3 for the pool of p300 transcriptional co-activator and reduce SMAD3/p300 interactions[201]. A follow up study suggested more dynamic interactions between AP-1 proteins and SMAD3[202]. Though in many cases Jun family members downregulate SMAD3-mediated transactivation, AP-1 dependent promoters can be synergistically activated by SMAD3 and Jun proteins[202].

1.14 General Aims of Study/Hypothesis

Although dozens of CAD-associated loci have been identified from the various CARDIoGRAM *Nature Genetics* papers, identification of functional SNPs that drive disease pathogenesis has only been successfully conducted for a small subset of these loci. There are several mechanisms whereby SNPs associated with common disease can be causal/functional. Functional SNPs can alter the amino acid in the coding sequence of a protein, can disrupt a gene regulatory element such as a promoter or an enhancer, can affect mRNA splicing or translation, can have epigenetic effects, or can affect the function of a noncoding RNA.

This project has three main aims. The first aim is to characterize functional SNPs at the *SMAD3* locus at chromosome 15q22 (**Chapter 3**). The rs56062135 SNP in *SMAD3* was reported in the 2015 CARDIoGRAM *Nature Genetics* study[44]. The second aim is to characterize functional SNPs at the *COL4A1/COL4A2* locus at chromosome 13q34 (**Chapter 4**). The independent rs4773144 and rs9515203 SNPs were reported in the 2013 CARDIoGRAM *Nature Genetics* study[26]. For both of these loci, the CAD-associated SNPs are located in introns and are not linked with SNPs in coding regions or regions that could affect mRNA splicing. We hypothesize that for both the *SMAD3* and *COL4A1/COL4A2* loci, the CAD-associated SNPs (or linked SNPs) disrupt the function of a gene regulatory element that in turn affect transcription of the *SMAD3* and *COL4A1/COL4A2* genes. Altered SMAD3 and COL4A1/COL4A2 expression in cell types relevant to atherosclerosis would alter risk of developing the disease.

The third aim of this project is to investigate whether there is any biological and statistical interaction between COL4A1/COL4A2 and SMAD3 (**Chapter 2**). In addition to these loci associating individually with CAD, TGFβ/SMAD receptor complex, extracellular matrix integrity, collagen formation, and cell-extracellular matrix interactions were all identified as enriched CAD pathways[45]. We hypothesize these loci demonstrate genetic interaction with respect to CAD association and have interrelated roles biologically that contribute to CAD.

2 FUNCTIONAL INTERACTION BETWEEN *COL4A1/COL4A2* AND *SMAD3* RISK LOCI FOR CORONARY ARTERY DISEASE

Adam W. Turner[a], Majid Nikpay[b], Anada Silva[a], Paulina Lau[a], Amy Martinuk[a], Tara A. Linseman[a], Sebastien Soubeyrand[a], Ruth McPherson[a,b,*]

[a]Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada, K1Y 4W7

[b]Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada, K1Y 4W7

Short title: TGFβ/SMAD3 regulation of collagen IV genes

Status: Published in *Atherosclerosis* 242 (2015) 543-552

2.1 Abstract

Objective: The *COL4A1/COL4A2* region on chromosome 13q34 is a highly replicated locus for coronary artery disease (CAD). In the normal arterial wall, type IV collagen acts to inhibit smooth muscle cell proliferation. Its production is in part a function of TGFβ signaling, but the specific regulatory mechanisms, especially in humans, have not been defined. Our aim was to decipher TGFβ signaling components important in the regulation of COL4A1 and COL4A2 and determine whether these components showed genetic interaction with the *COL4A1/COL4A2* locus for CAD association.

Methods and Results: Experiments were performed in primary human aortic smooth muscle cells and HT1080 fibroblasts. Pharmacological inhibition of the TGFβ1 receptor and subsequent SMAD protein phosphorylation by treatment with an ALK5 inhibitor prevented the increase in *COL4A1/COL4A2* mRNA ($p < 0.001$) and protein expression in response to TGFβ1 stimulation. In contrast, inhibition of the non-canonical TGFβ signaling pathways was without effect. siRNA mediated knockdown of SMAD3 and SMAD4 abolished the stimulatory effects of TGFβ1 on COL4A1/COL4A2 ($p < 0.001$) whereas SMAD2 knockdown had no effect. In luciferase reporter assays, neither SMAD3 overexpression nor TGFβ1 treatment altered *COL4A1* or *COL4A2* promoter activity, supportive of more complex regulation of type IV collagen gene expression by the TGFβ/SMAD3 signaling pathway. Epistasis analysis in 5 CAD case/control cohorts revealed that *SMAD3* and *COL4A1/COL4A2* display statistical interaction for CAD association.

Conclusions: These findings demonstrate that SMAD3 is a necessary factor for TGFβ-mediated stimulation of mRNA and protein expression of type IV collagen genes in human

vascular smooth muscle cells. Epistasis analyses further supports the hypothesis that the SMAD3-dependent regulation of COL4A1/COL4A2 may be of functional significance for CAD pathogenesis.

2.2 Introduction

The *COL4A1/COL4A2* gene cluster on chromosome 13 is a highly replicated locus for coronary artery disease (CAD)[26,42]. COL4A1 and COL4A2 are members of the type IV collagen family that comprise the major structural component of basement membranes that line the epithelium of most tissues and the endothelial layer of blood vessels, mediating cell adhesion to the basement membrane via binding to integrin receptors and other cell-surface molecules[75,76,203].

Polymorphisms in *COL4A*1 or *COL4A*2 are associated with several vascular defects, including arterial stiffness and myocardial infarction[116,117]. Genetically engineered mice with *Col4a1* mutations display a complex vascular phenotype, including detachment of the endothelium from the medial layer[115]. The C-terminal non-collagenous (NC1) domains of COL4A1 and COL4A2 have anti-angiogenic activity and inhibit processes including endothelial cell proliferation, migration and tube formation[204]. The role of type IV collagen in atherosclerosis is complex and not completely defined. Thick deposits of type IV collagen are abundant in plaques at advanced stages of atherosclerosis[112]. However, exposure to type IV collagen maintains a contractile and quiescent smooth muscle phenotype[108]. Type IV collagen production is in part regulated by TGFβ1, synthesized within the arterial wall by endothelial cells, monocytes/macrophages, smooth muscle cells, and T cells[159,172]. TGFβ inhibits the proliferation of vascular smooth muscle cells[165] and numerous studies suggest that it has an anti-atherogenic role, including protecting against unstable plaque development[140,158]. Plasma TGFβ1 concentrations are inversely related to the extent of atherosclerotic lesions[160,161] and inhibition of TGFβ by a neutralizing antibody against

TGFβ1, 2 and 3 in *Apoe*$^{-/-}$ mice, resulted in an exacerbated atherosclerotic phenotype with a high inflammatory cell content and a decreased collagen content[173].

The canonical TGFβ signalling pathway involves TGFβ binding to the TGFβ type II receptor which then recruits and subsequently phosphorylates the TGFβ type I receptor[136–138,143,205]. The phosphorylated type I receptor then phosphorylates the receptor-regulated SMAD proteins, SMAD2 and SMAD3. Once phosphorylated, SMAD2 and SMAD3 form complexes with the common (co)-SMAD, SMAD4. These complexes then translocate to the nucleus where they regulate transcription of cognate target genes[136–138,143,205]. Of interest, the *SMAD3* locus has also been linked to CAD[38]. It is well known that TGFβ increases expression of the *COL4A1* and *COL4A2* genes at the transcriptional and protein level. However little is known with respect to specific regulatory mechanisms and most published studies were not in human cells and not performed in cell lines relevant to atherosclerosis[206–210]. Recent large-scale interrogation of genome wide association study (GWAS) data identified extracellular matrix signaling/organization/degradation, collagen formation, and TGFβ/SMAD signaling as enriched CAD-associated pathways[45].

Our goal of this study was to identify TGFβ signaling components that biologically regulate transcription of *COL4A1* and *COL4A2* in humans and determine whether these display genetic interaction with the *COL4A1/COL4A2* locus for CAD association. We sought to determine genetic interaction with respect to *COL4A1/COL4A2* in an effort to uncover relevant biological mechanisms that contribute to CAD. Here, we first clarify the mechanisms whereby TGFβ stimulates expression of the *COL4A1* and *COL4A2* genes using human aortic smooth muscle cells as our model. Secondly, we demonstrate that the

COL4A1/COL4A2 and *SMAD3* loci display epistatic interaction conferring CAD risk. These findings demonstrate the potential importance of SMAD3-dependent regulation of COL4A1/COL4A2 in atherosclerosis and help uncover the genetic architecture of *COL4A1/COL4A2* in relationship to CAD.

2.3 Materials and Methods

Cell culture and TGFβ stimulation

Primary human aortic smooth muscle cells (AoSMCs) from the ascending aorta were purchased from Lonza (catalog #CC-2571). AoSMCs were grown in SmGM-2 Smooth Muscle Cell Basal Medium (Lonza) supplemented with 5% FBS, insulin, hFGF-B, GA-1000, and hEGF. Human TGFβ1 was purchased from R & D Systems (platelet-derived, catalog 100-B) and resuspended in 4 mM HCl containing 0.1% BSA.

Drug treatments

SB431542 and PD98059 were purchased from Sigma; SB203580 and SP600125 were purchased from Cell Signalling Technology. All inhibitors were dissolved in DMSO and further diluted to the appropriate concentrations using sterile water. For inhibitor experiments with AoSMCs, cells were seeded in 6 well plates, grown to confluence, serum starved for 16-24 h, pretreated with inhibitor for 1 h, and subsequently treated with TGFβ1 (10 ng/mL) for 16 h (mRNA analysis) or 24 h (protein analysis).

siRNA knockdown

SMAD3 (s8401, s8402), SMAD4 (s8403), and SMAD2 (s8398) Silencer Select pre-designed siRNAs were purchased from Life Technologies. siRNA transfection was performed using Lipofectamine RNAiMAX (Life Technologies). For each well undergoing

knockdown, the final siRNA concentration was 20 nM; Primary AoSMCs were seeded in 6 well plates and grown to 85% confluence before siRNA transfection.

RNA isolation and qRT-PCR

RNA for all samples was isolated using the HighPure RNA Isolation Kit (Roche). AoSMC RNA (500 ng) and HT1080 cell RNA (1 µg) were reverse-transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche). A combination of oligo(dT) and random hexamer primers were used as per the kit instructions. Quantitative PCR of cDNA samples was performed using LightCycler 480 SYBR Green I Master Mix, and experiments were all run on the LightCycler 480 II machine (Roche). Peptidylprolyl Isomerase A (PPIA) was used as the reference gene for all qPCR experiments. The following primer sequences were used for qPCR analysis: *COL4A1*: forward 5'-TGTGGGCCAGCCAGGCATTG-3' reverse 5'-CAGGGGGTCCGATCGCTCCA-3'; *COL4A2*: forward 5'-AGGGTCGCAGGGAGAGCTGG-3' reverse 5'-TGGGCCTCGTTCCCCTGGAG-3'; *SMAD3*: forward 5'-GCCTTCTGGTGCTCCATCTC-3' reverse 5'-AATAGCGCTGTCACTGAGGCA-3'; *SMAD4*: forward 5'-CTCCTGAGTATTGGTGTTCCAT-3' reverse 5'-CTGACCCAAACATCACCTTCA-3'; *SMAD2:* 5'-GGCCTGATCTTCACAGTCAT-3' reverse 5'-TGATATATCCAGGAGGTGGCG-3'; *PPIA*: forward 5'-ACCGTGTTCTTCGACATTGC-3' reverse 5'-TTCTGTGAAAGCAGGAACCC-3'

Western blotting

Protein samples were harvested at 4°C using RIPA buffer containing fresh Roche cOmplete Protease Inhibitor cocktail and Roche PhosStop Phosphatase Inhibitor cocktail. Protein concentration was determined using the Pierce BCA Protein Assay Kit. 35 µg of each denatured AoSMC sample was loaded onto a 4-15% gradient SDS-PAGE gel (Bio-Rad) and run under reducing conditions. Samples were transferred to nitrocellulose membranes (Bio-Rad) for 2 h at 100V and subsequently blocked for 1 h at room temperature using Odyssey blocking buffer (LI-COR). Blots were incubated with the following primary antibodies: rat anti-human COL4A1 antibody H11 (Chondrex, 1:200 dilution); rat anti-human COL4A2 antibody H22 (Chondrex, 1:200 dilution); rabbit anti-human SMAD3 antibody (Cell Signaling Technology, 1:1000 dilution); rabbit anti-human phospho-SMAD3 (Ser423/425) antibody (C25A9, Cell Signaling Technology, 1:1000 dilution), rabbit anti-human SMAD4 antibody (Cell Signaling Technology, 1:1000 dilution); rabbit anti-human SMAD2 antibody (Cell Signaling Technology, 1:1000 dilution). Mouse anti-human β-actin antibody was used as the loading control throughout these experiments. IRDye secondary antibodies (LI-COR) were used at a concentration of 1:15000 and diluted in Odyssey blocking buffer containing 0.1% Tween 20 and 0.1% SDS. Immunoreactivity was visualized with the LI-COR Odyssey imaging system.

Generation of COL4A1 and COL4A2 reporter plasmids

To assess the transcriptional activity of the *COL4A1* and *COL4A2* genes, we designed reporter constructs mimicking those from previous studies[63,71,72]. However, instead of using

chloramphenicol acetyltransferase (CAT) as the reporter, we generated constructs using pGL3-Basic (Promega) as the backbone, encoding firefly luciferase. These constructs detect transcription from either the *COL4A1* or *COL4A2*-specific direction. All inserts were generated by PCR amplification of genomic DNA and using primers containing restriction enzyme recognition sequences. A description of the cloning steps employed to generate these constructs is given in **Supplementary Materials and Methods**.

Co-expression of type IV collagen reporters and SMAD3

HT1080 cells were plated in 12 well plates and grown to 70% confluency at the time of transfection. Each well was transfected with 500 ng of *COL4A1* reporter, *COL4A2* reporter, or empty pGL3-Basic along with 500 ng of either pCMV5B-Flag-SMAD3 (Addgene) or empty pCMV5. These transfections were performed with FuGENE HD Transfection Reagent (Promega); transfected wells also contained 2% pRL-TK *Renilla* (Promega) to account for transfection efficiency. Cells were harvested for luciferase activity 48 h after transfection. Cells were washed once with 1X PBS and subsequently lysed using 250 µL of 1X Passive Lysis Buffer (Promega) per well and shaking for 15 min on a rocking platform.

Treatment of type IV collagen reporters with TGFβ

AoSMCs were plated in 6 well dishes, grown to 90% confluency and serum starved overnight. Each well was transfected with 2.5 µg of luciferase reporter and subsequently treated with 10 ng/mL TGFβ1 or vehicle control 6 h later. These transfections were

performed with Lipofectamine 3000 Transfection Reagent (Life Technologies); transfected wells also contained 2% pRL-SV40 *Renilla* (Promega) to account for transfection efficiency. Cells were harvested for luciferase activity 48 h after transfection. Cells were washed once with 1X PBS and subsequently lysed using 500 µL of 1X Passive Lysis Buffer (Promega) per well and shaking for 15 min on a rocking platform.

Luciferase assays

Cell lysates in 1X Passive Lysis Buffer were centrifuged for 1 min at 13,000 rpm and 20 µL of supernatant was used per luciferase/*Renilla* assay. Emitted light was quantified using a GloMax luminometer (Promega).

Statistical analysis

All data are expressed as means ± S.D. Significant differences were assessed by one-way ANOVA followed by Bonferroni's post-hoc test.

Epistasis analysis was carried out using PLINK (v1.07) software[211] in 5 independent cohorts, including OHGS_A2, OHGS_C2, DUKE_2, CCGB_2 and ITH_2 where individual genotypes were available. Before the analysis we pruned SNPs and kept only those in linkage equilibrium (LD $r^2 < 0.2$). Furthermore, we only conducted tests for SNPs that had available genotypes in all 5 of the above cohorts. The epistasis test uses the logistic regression and makes a model based on allele dosage for each SNP, A and B, and fits the model in the form of:

$$Y \sim \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \varepsilon$$

Therefore the test for epistasis is based on the β_3 coefficient. The effect size estimates from each cohort were subsequently combined and weighted mean of beta coefficients $(\overline{\beta})$ was calculated based on fixed-effect model[212] as follows:

$$\overline{\beta} = \frac{\sum_{i=1}^{n} w_i \times \beta_i}{\sum_{i=1}^{n} w_i}$$

Where w_i is the inverse variance of the i^{th} study and n is the number of studies

Standard error (SE) of combined effect is:

$$SE(\overline{\beta}) = \sqrt{\frac{1}{\sum_{i=1}^{n} w_i}}$$

Cochran's Q statistic which measures the heterogeneity among studies was calculated as:

$$Q = \sum_{i=1}^{n} w_i (\beta_i - \bar{\beta})^2$$

Q is distributed as a chi-square statistic with $n-1$ degrees of freedom (df).

I^2 index quantifies the degree of heterogeneity and is expressed in percentage of the total variability in a set of effect sizes due to true heterogeneity, that is, to between-studies variability and it was calculated as:

$$I^2 = \frac{(Q - df)}{Q} \times 100$$

Online Supplemental Methods

HT1080 cell culture

The HT1080 cell line was purchased from ATCC and grown in high-glucose DMEM media (Gibco) supplemented with 10% FBS, L-Glutamine, and penicillin/streptomycin.

HT1080 drug treatments

HT1080 cells were seeded in either 6 or 12 well plates, grown to 85% confluence, serum starved for 16-24 h, pretreated with 10 μM inhibitor for 1 h, and subsequently treated with TGFβ1 (2 ng/mL) for 16 h (mRNA analysis) or 24 h (protein analysis).

HT1080 siRNA transfection

HT1080 cells were seeded in 6 or 12 well plates and grown to 60% confluence before siRNA transfection. The total knockdown time for each experiment was 48 h. 24 h after siRNA treatment, cells were treated with or without TGFβ1 (2 ng/mL) for 24 h before harvesting samples. The final siRNA concentration of each well was 20 nM.

Generation of reporter constructs

For the reporter detecting transcription from the *COL4A2* direction, we mimicked the insert from the construct pNA reported by Schmidt *et al*.[72]. This insert contains the promoter, part of exon 1 of *COL4A1* along with exon 1, intron 1, and part of exon 2 of the *COL4A2* gene.

This region was amplified using the primers 5'-GATTGGTACCTGAGCCGGGGCCCCATGGT -3' (KpnI restriction site) and 5'-GATTAAGCTTGGCCCCGGTCAGTCCCACT-3' (HindIII restriction site). After PCR amplification, both the resulting PCR product and pGL3-Basic were digested with KpnI and HindIII before subsequent ligation into the pGL3-Basic multiple cloning site.

To construct a reporter to detect transcription from the *COL4A1* direction, we used the insert from the pBNaSA construct as described previously[72]. This construct consists of the bidirectional promoter flanked by regulatory sequences in the *COL4A1* and *COL4A2* genes. Also included was a splice-acceptor sequence from the *COL4A2* gene[72]. The splice acceptor site was PCR amplified using the primers 5'-GATTAGATCTCGGGCCGCGCACCGCGCTGT-3' (BglII restriction site) and 5'-GATTAAGCTTGGCCCCGGTCAGTCCCACT-3' (HindIII restriction site). Ligation of this sequence was performed after digesting this PCR product and empty pGL3-Basic with BglII and HindIII. The sequence required for *COL4A1* promoter activity was PCR amplified using the primers 5'-GATTGGTACCCGCGAACGCGCGGGCGCGA-3' (KpnI restriction site) and 5'-GATTCTCGAGGGCTCCCAGGCACCCTCAC-'3 (XhoI restriction site). This sequence was inserted into pGL3-Basic after digesting both the PCR product and empty vector with KpnI and XhoI and subsequent ligation. Correct sequences for all constructs were confirmed after the cloning step via sequencing.

HT1080 reporter assays with TGFβ

HT1080 cells were plated in 12 well plates, grown to 90% confluency and serum starved overnight. Each well was transfected with 500 ng of luciferase reporter and subsequently treated with TGFβ1 (2 ng/mL) or vehicle control 6 h later. Transfection was performed with FuGENE HD Transfection Reagent (Promega); transfected wells also contained 2% pRL-TK *Renilla* (Promega) to account for transfection efficiency. Cells were harvested for luciferase activity 48 h after transfection.

2.4 Results

TGFβ1 upregulates COL4A1 and COL4A2 via the SMAD pathway

We first sought to determine whether the effect of TGFβ on COL4A1 and COL4A2 is transmitted via activation of the canonical SMAD pathway or due to a non-SMAD pathway. We elected to investigate primary human AoSMCs given their physiological relevance to CAD. Results were replicated in HT1080 cells (**Supplementary Figures 1-3**), a very well characterized human cell line for *COL4A1* and *COL4A2* transcriptional regulation and promoter studies[59,62,63,71,72]. We utilized the pharmacological inhibitor SB431542 to inhibit activin receptor-like kinase 5 (ALK5), the type I TGFβ receptor kinase, thereby blocking activation and phosphorylation of SMAD2 and SMAD3[213]. In the absence of inhibitor, TGFβ1 stimulation increased COL4A1 and COL4A2 mRNA and protein levels several-fold (**Figures 1A and 1B**). In the presence of SB431542, however, TGFβ1 treatment failed to increase COL4A1 and COL4A2 mRNA and/or protein, indicating that activation by TGFβ requires ALK5 signaling (**Figures 1A and 1B**). Western blotting confirmed that SB431542 treatment blocked SMAD3 phosphorylation at serine residues 423 and 425 (targets of the type I TGFβ receptor) in response to TGFβ1. Taken together, these results suggest that active and phosphorylated SMAD proteins are necessary for the transcriptional effects of TGFβ1 on *COL4A1/COL4A2*.

Next, we sought to assess the contribution of non-SMAD pathways to the upregulation of COL4A1 and COL4A2 by TGFβ by inhibiting three prominent non-SMAD pathways. PD98059 was added to inhibit the ERK branch of the MAP Kinase pathway, specifically by binding to the ERK-specific MAP kinase MEK and preventing phosphorylation of ERK1

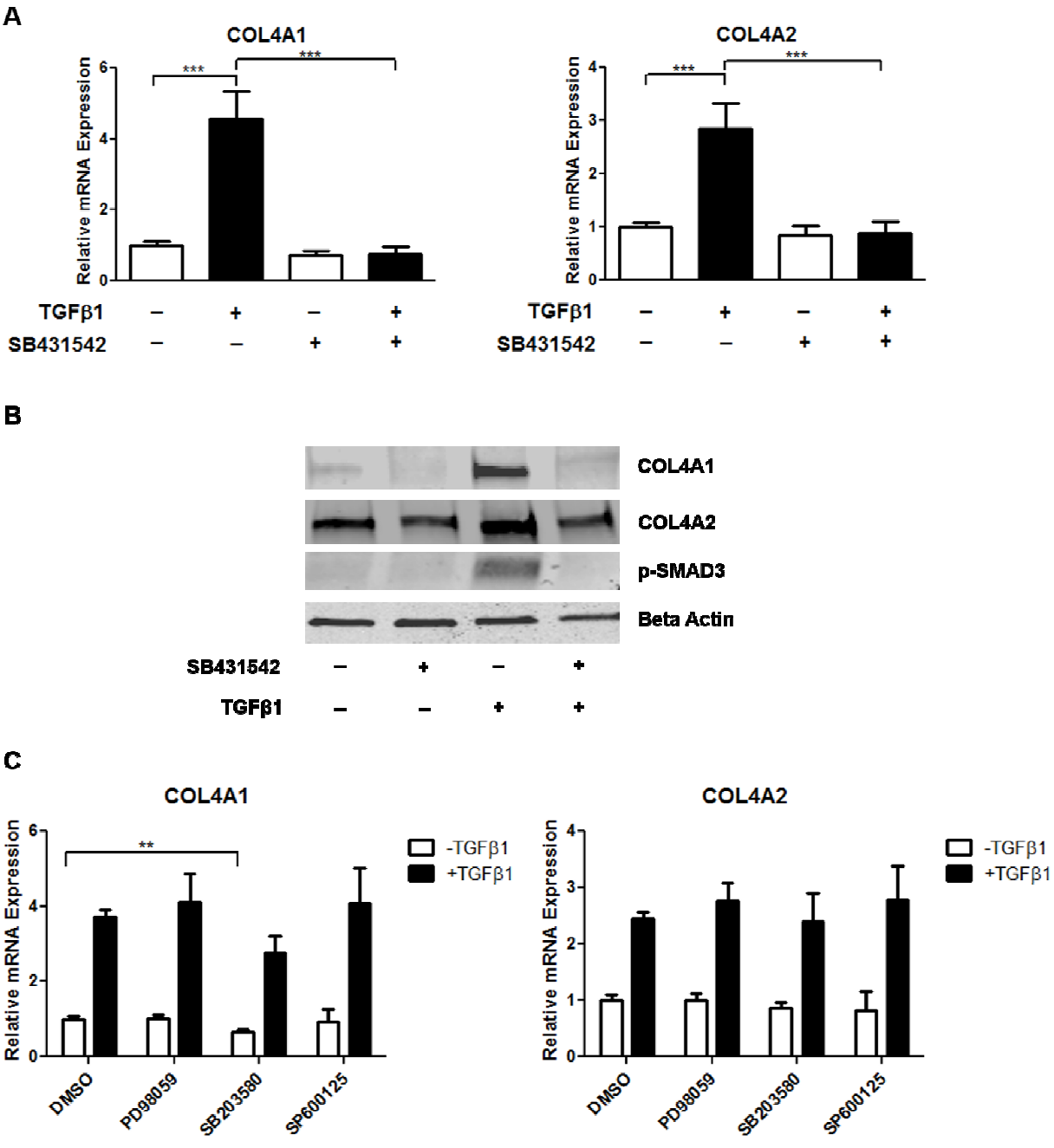
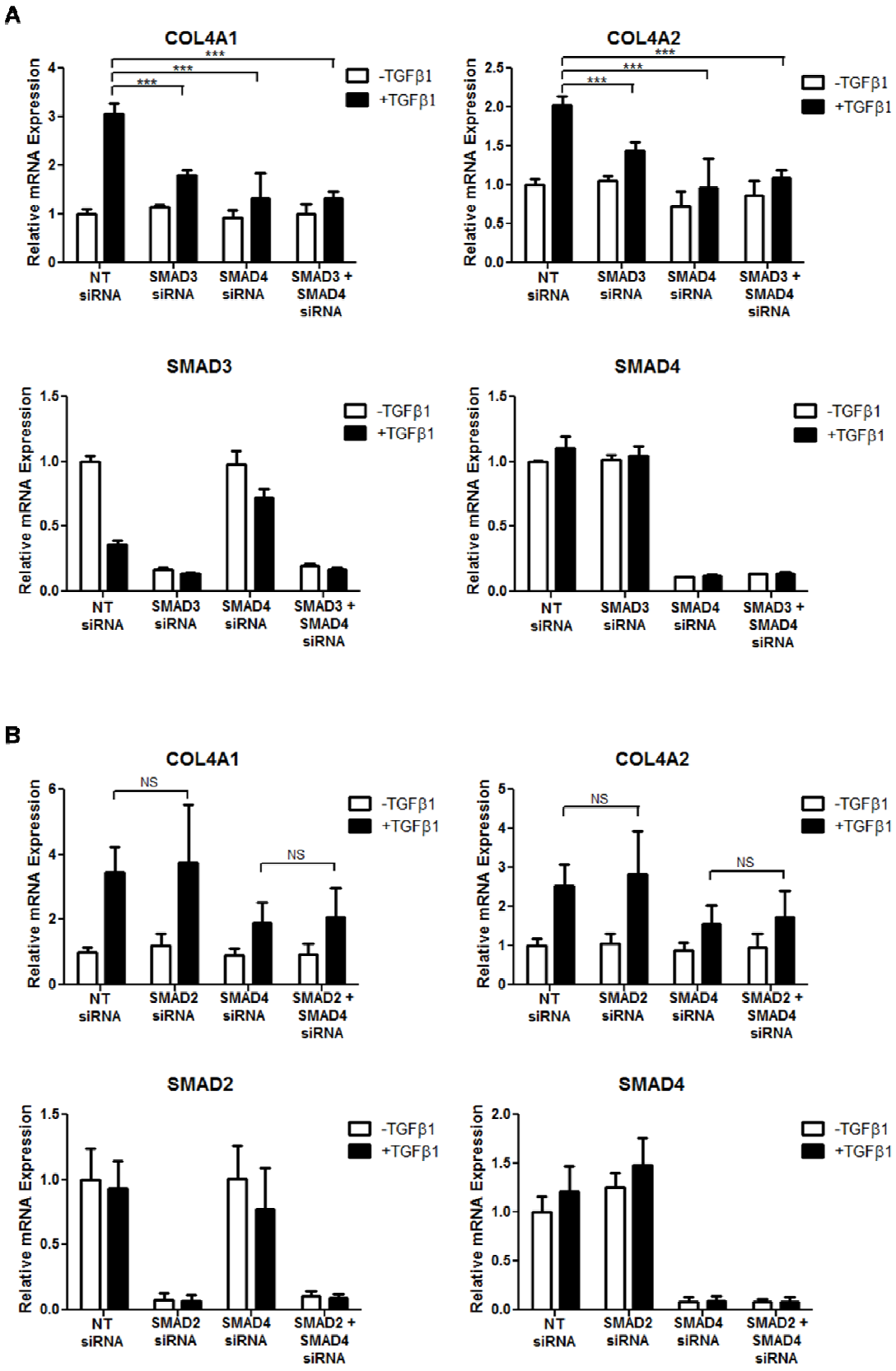

A
COL4A1
Relative mRNA Expression
***
***
TGFβ1 – + – +
SB431542 – – + +
COL4A2
Relative mRNA Expression
***
***
TGFβ1 – + – +
SB431542 – – + +
B
COL4A1
COL4A2
p-SMAD3
Beta Actin
SB431542 – + – +
TGFβ1 – – + +
C
COL4A1
Relative mRNA Expression
**
-TGFβ1
+TGFβ1
DMSO
PD98059
SB203580
SP600125
COL4A2
Relative mRNA Expression
-TGFβ1
+TGFβ1
DMSO
PD98059
SB203580
SP600125

Figure 1. Treating AoSMCs with SB431542 blocks the upregulation of COL4A1 and COL4A2 by TGFβ1 whereas blocking non-SMAD pathways has little effect. AoSMCs were pretreated with 0.1% DMSO or 10 μM of pharmacological inhibitor for 1 h prior to treatment with 10 ng/mL TGFβ1. (A, B) AoSMCs were pretreated with SB431542 and harvested for mRNA 16 h after TGFβ1 treatment (A) and harvested for protein 24 h after TGFβ1 treatment (B). (C) AoSMCs were pretreated with a panel of inhibitors targeting non-SMAD pathways that act through TGFβ and harvested for mRNA 16 h after TGFβ1 treatment. COL4A1 and COL4A2 mRNA levels were measured by qRT-PCR using *PPIA* as the reference gene. Values represent the mean ± S.D. of three independent experiments with three different wells analyzed per condition each experiment. Values are normalized relative to DMSO treatment in the absence of TGFβ1. $^{**}p < 0.01$, $^{***}p < 0.001$, ANOVA with Bonferroni's post-hoc test. Protein levels of COL4A1, COL4A2, phospo-SMAD3 (Ser423/425) and β-actin were measured by Western blot.

and ERK2. SB203580 is a specific inhibitor of the p38 MAPK pathway and SP600125 is an inhibitor of JNK. PD98059 had no effect on COL4A1/COL4A2 levels basally or in the presence of TGFβ stimulation (**Figure 1C**). We observed that p38 MAPK inhibition affected basal levels of COL4A1 mRNA. However, none of the inhibitors investigated measurably affected the fold-change of COL4A1/COL4A2 induction upon TGFβ stimulation (**Figure 1C**). Compared to AoSMCs, p38 MAPK and JNK inhibition had more effect on basal *COL4A1* and *COL4A2* mRNA levels in HT1080 cells (**Supplementary Figure 1C**). These data suggest that the upregulation of *COL4A1/ COL4A2* by TGFβ is due primarily to the activation of the SMAD pathway with minimal contribution from the MAP kinase pathways.

TGFβ1 upregulates COL4A1 and COL4A2 expression in a SMAD3-dependent manner

The relative roles of the canonical SMAD proteins SMAD2, SMAD3, and SMAD4 in TGFβ signaling were then interrogated via siRNA knockdown. We observed that in the absence of TGFβ1, suppression of SMAD2, SMAD3, and SMAD4 expression had no significant effect on COL4A1 and COL4A2 mRNA or protein levels (**Figure 2**). However, suppression of either SMAD3 or SMAD4 blunted the activation of COL4A1 and COL4A2 upon TGFβ1 stimulation and the combined knockdown of SMAD3 and SMAD4 almost completely abolished the stimulatory effects of TGFβ1 in human AoSMCs (**Figures 2A and 2C**). In addition, TGFβ1 stimulation resulted in a decrease in *SMAD3* mRNA levels in AoSMCs as observed previously in numerous cell lines[214–216]. In contrast to SMAD3 and SMAD4, SMAD2 siRNA knockdown had no effect on COL4A1/COL4A2 mRNA or protein levels either basally or in response to TGFβ1 stimulation (**Figures 2B and 2C**). Lastly, targeting


A
COL4A1
Relative mRNA Expression
-TGFβ1
+TGFβ1
***
***
***
NT siRNA
SMAD3 siRNA
SMAD4 siRNA
SMAD3 + SMAD4 siRNA
COL4A2
SMAD3
SMAD4
B
COL4A1
NS
NS
SMAD2 siRNA
SMAD2 + SMAD4 siRNA
COL4A2
SMAD2
SMAD4

C

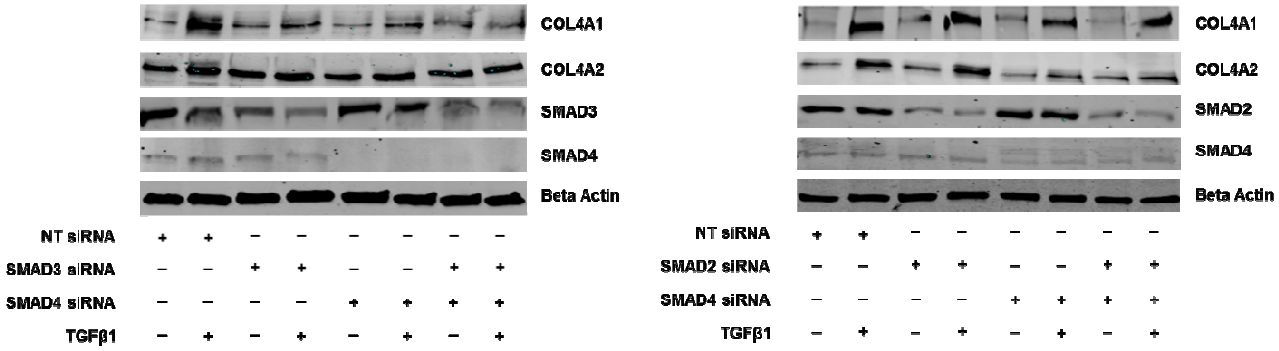
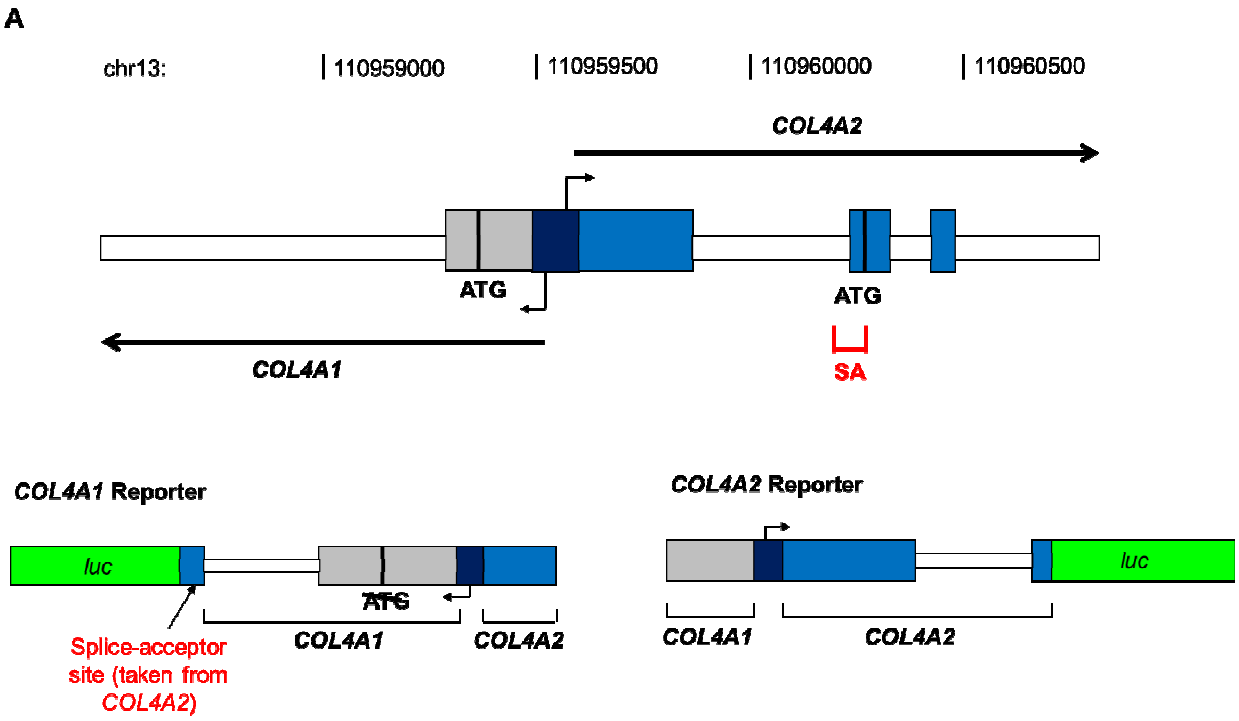

COL4A1
COL4A2
SMAD3
SMAD4
Beta Actin
NT siRNA + + − − − − − −
SMAD3 siRNA − − + + − − + +
SMAD4 siRNA − − − − + + + +
TGFβ1 − + − + − + − +
COL4A1
COL4A2
SMAD2
SMAD4
Beta Actin
NT siRNA + + − − − − − −
SMAD2 siRNA − − + + − − + +
SMAD4 siRNA − − − − + + + +
TGFβ1 − + − + − + − +

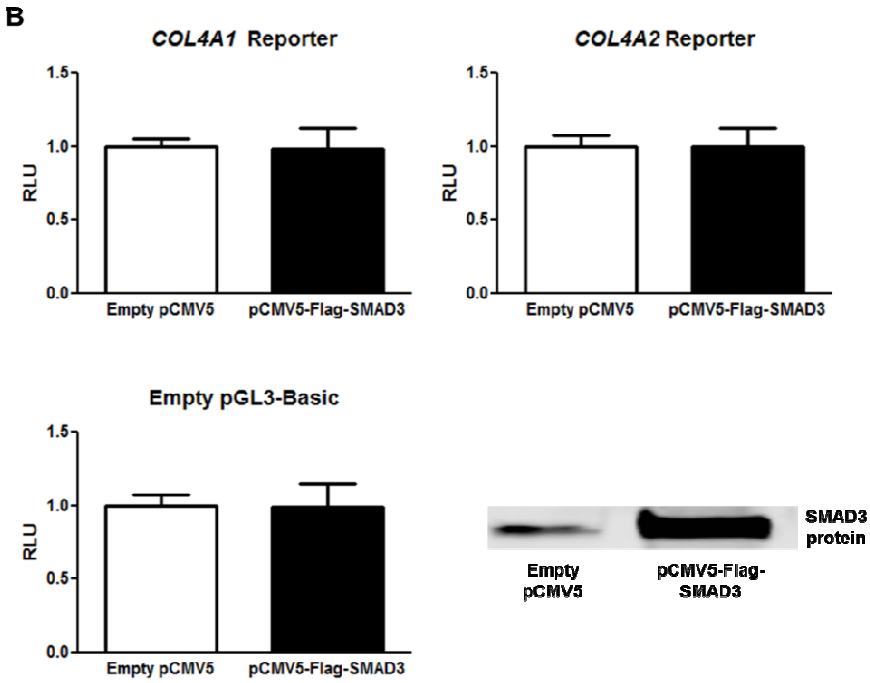
Figure 2. SMAD3, but not SMAD2, knockdown abolishes the upregulation of COL4A1 and COL4A2 by TGFβ1 in AoSMCs. (A) AoSMCs were treated with nontarget siRNA, SMAD3 siRNA, SMAD4 siRNA, or a combination of SMAD3 and SMAD4 siRNA for a total of 48 h. 24 h into the siRNA knockdown, cells were treated with or without 10 ng/mL TGFβ1 for the final 24 h. The total siRNA concentration for each well was 20 nM. (B) AoSMCs were treated with nontarget siRNA, SMAD2 siRNA, SMAD4 siRNA, or a combination of SMAD2 and SMAD4 siRNA for a total of 48 h. 24 h into the siRNA knockdown, cells were treated with or without 10 ng/mL TGFβ1for the final 24 h. The total siRNA concentration for each well was 20 nM. *COL4A1* and *COL4A2* mRNA levels were assessed by qRT-PCR using *PPIA* as the reference gene. Values represent the mean ± S.D. of three independent experiments with three different wells analyzed per condition each experiment. Values are normalized relative to NT siRNA in the absence of TGFβ1. ***$p<0.001$, ANOVA with Bonferroni's post-hoc test. (C) COL4A1, COL4A2, SMAD3, SMAD2 and SMAD4 protein levels were determined by Western blotting, with β-actin as the loading control.

SMAD2 in the context of SMAD4 suppression had the same effect on COL4A1 and COL4A2 levels as SMAD4 knockdown alone. These data indicate that unlike SMAD3 and SMAD4, SMAD2 does not play a role in TGFβ mediated regulation of type IV collagen in the two cell lines we investigated. Of interest, we consistently observed a greater fold increase in COL4A1 mRNA and protein compared to COL4A2 in response to TGFβ1 (**Figures 1 and 2**).

Evidence for an indirect role of SMAD3 in COL4A1/COL4A2 regulation

Despite numerous investigations, a small number of direct SMAD targets have been identified[217,218]. Shared properties of SMAD target genes include rapid induction of mRNA expression upon TGFβ stimulation, presence of a SMAD-binding element (SBE) sequence near the promoter, and promoter activation in response to exogenous TGFβ1 and overexpressed SMAD proteins[217–219]. Having demonstrated that SMAD3 and SMAD4 suppression abolishes the upregulation of COL4A1 and COL4A2 by TGFβ, we next sought to determine whether the effect of SMAD3 was via a direct effect on the bidirectional *COL4A1/COL4A2* promoter. First, we generated luciferase reporter constructs harboring the functional *COL4A1* and *COL4A2* promoter regions (**Figure 3A**). *COL4A1* and *COL4A2* share a common, small, bidirectional promoter that requires the presence of proximal regulatory elements for proper transcription of each respective gene[59,62,63,71,72]. SMAD3 overexpression, confirmed in parallel experiments via Western blot, had no effect on the promoter activities of *COL4A1* and *COL4A2* reporter constructs (**Figure 3B**). Next the responsiveness of the human *COL4A1/A2* promoter region to TGFβ1 treatment was assessed


A
chr13:
110959000
110959500
110960000
110960500
COL4A2
ATG
ATG
COL4A1
SA
COL4A1 Reporter
luc
ATG
Splice-acceptor site (taken from COL4A2)
COL4A1
COL4A2
COL4A2 Reporter
luc
COL4A1
COL4A2


B
COL4A1 Reporter
RLU
1.5
1.0
0.5
0.0
Empty pCMV5
pCMV5-Flag-SMAD3
COL4A2 Reporter
RLU
1.5
1.0
0.5
0.0
Empty pCMV5
pCMV5-Flag-SMAD3
Empty pGL3-Basic
RLU
1.5
1.0
0.5
0.0
Empty pCMV5
pCMV5-Flag-SMAD3
SMAD3 protein
Empty pCMV5
pCMV5-Flag-SMAD3

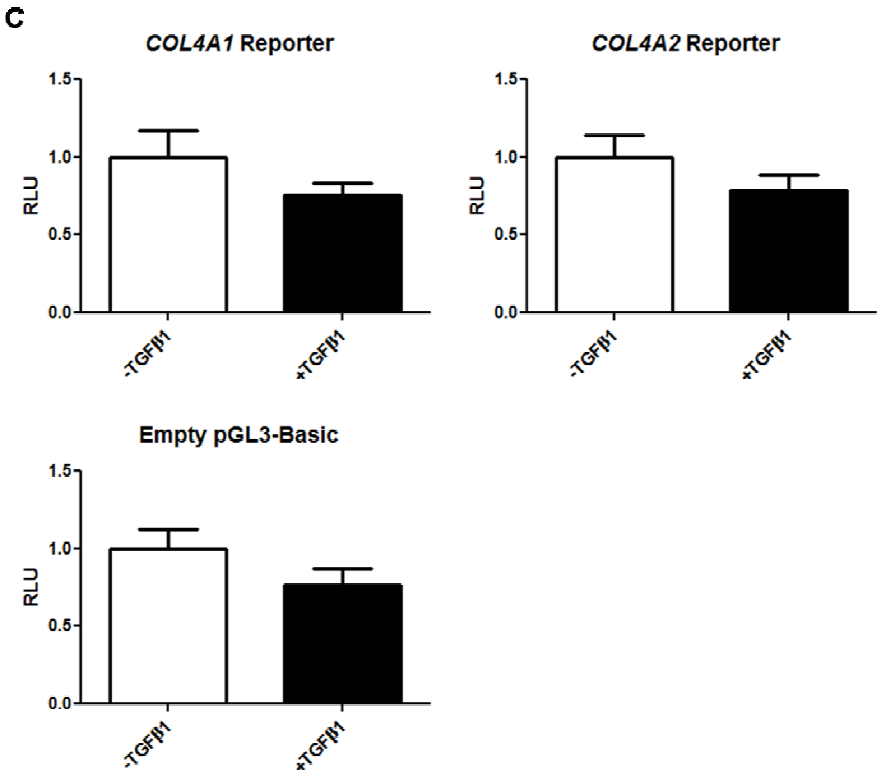

C
COL4A1 Reporter
COL4A2 Reporter
Empty pGL3-Basic
RLU
1.5
1.0
0.5
0.0
-TGFβ1
+TGFβ1

Figure 3. The human *COL4A1/COL4A2* bidirectional promoter alone does not act as a direct SMAD3/TGFβ1 target. (A) Schematics of the *COL4A1* and *COL4A2* promoter constructs used in luciferase assays, adapted from previous studies[62,63,72] (refer to methods). (B) SMAD3 overexpression does not affect luciferase activity of *COL4A1* and *COL4A2* promoter reporter constructs. HT1080 cells were co-transfected with type (IV) collagen reporter constructs along with pCMV5-Flag SMAD3 or empty pCMV5 for 48 h. (C) Exogenous TGFβ treatment does not affect luciferase activity of *COL4A1* or *COL4A2* reporter constructs. AoSMCs were transfected with *COL4A1* and *COL4A2* promoter reporter constructs for 6 h, and then treated with 10 ng/mL TGFβ1 for a subsequent 42 h.

by measuring the activity of the *COL4A1* and *COL4A2* reporter constructs following TGFβ1 treatment. Using a dose of TGFβ1 that resulted in several fold increases in *COL4A1* and *COL4A2* mRNA levels, no upregulation of promoter activity in response to TGFβ1 was observed (**Figure 3C**). We observed treatment with TGFβ lowered the raw firefly luciferase readings of empty pGL3-Basic, *COL4A1* reporter, and *COL4A2* reporter to an equal degree. These findings suggest that SMAD3/TGFβ do not directly act on the human *COL4A1/COL4A2* bidirectional promoter in our system. However, we cannot rule out more indirect stimulatory effects of SMAD3/TGFβ1 on the human *COL4A1/COL4A2* promoter as additional enhancing sequences or proper chromatinization may be required that we could not capture in our luciferase assays. Our data indicate that more besides the previously characterized functional promoter elements alone is required for TGFβ responsiveness. Finally, we directly interrogated the occupancy of the endogenous *COL4A1/COL4A2* promoter by SMAD3 by performing chromatin immunoprecipitation experiments. Using several ChIP-verified SMAD3 antibodies, we were unable to detect enrichment of SMAD3 at the *COL4A1/COL4A2* bidirectional promoter, either basally or in response to TGFβ1 (data not shown).

COL4A1/COL4A2 displays genetic interaction with the SMAD3 locus

The CARDIoGRAM and CARDIoGRAMPlusC4D consortia identified two SNPs at the *COL4A1/COL4A2* locus that were independently associated with CAD. Earlier, Samani *et al*. identified a *SMAD3* SNP that associated with CAD in the WTCCC study[38]. Epistasis is a statistical interaction between two or more genetic loci where the effects are non-additive[220]

and has been hypothesized to play an important role in the genetic determination of complex diseases such as CAD[49,220]. We thus sought to determine whether there is evidence of genetic interaction with respect to CAD association between *COL4A1/COL4A2* SNPs and the *SMAD3* locus.

Epistasis on a genome-wide level has proven challenging due to the computational power required for a large number of pairwise or high-order tests and the need to correct for multiple testing[48,49]. However, filtering to consider only those regions with potential biological interaction reduces the necessity for high computational power and the burden of multiple testing correction[48,49]. We performed gene-gene interaction tests between the *COL4A1/COL4A2* locus and the *SMAD3* locus by performing pairwise tests between SNPs at the two loci. Analysis was conducted in five cohorts from which we had individual genotype data for *COL4A1/COL4A2* and *SMAD3* SNPs; these were subsequently combined in a meta-analysis, comprising a total of over 4500 CAD cases and over 2500 CAD controls. To reduce the number of multiple tests required, we initially performed LD pruning to only consider independent SNPs (LD $r^2 < 0.2$) at the *COL4A1/COL4A2* and *SMAD3* loci. After correcting for multiple testing, we identified a significant interaction between rs72655775 in *COL4A2* and rs12441344 in *SMAD3*, shown in **Table 1** (p-value = 1.6×10^{-6}, Bonferroni-corrected p-value = 6.9×10^{-3}, OR = 2.26). The complete list of tests performed is described in **Supplementary Table 1**; no others reached significance ($p < 0.05$) after Bonferroni correction. Possessing the minor allele C at rs72655775 in combination with the minor allele G at rs12441344 results in a synergistic risk of CAD. This direction of effect was consistent across all five of our cohorts (**Tables 1 and 2**). When considered individually, only rs72655775 was associated with CAD ($p = 0.01$) with a modest odds ratio of 1.25 (**Table 2**),

Table 1. Statistical epistasis between rs72655775(C) and rs12441344(G) in 5 independent cohorts

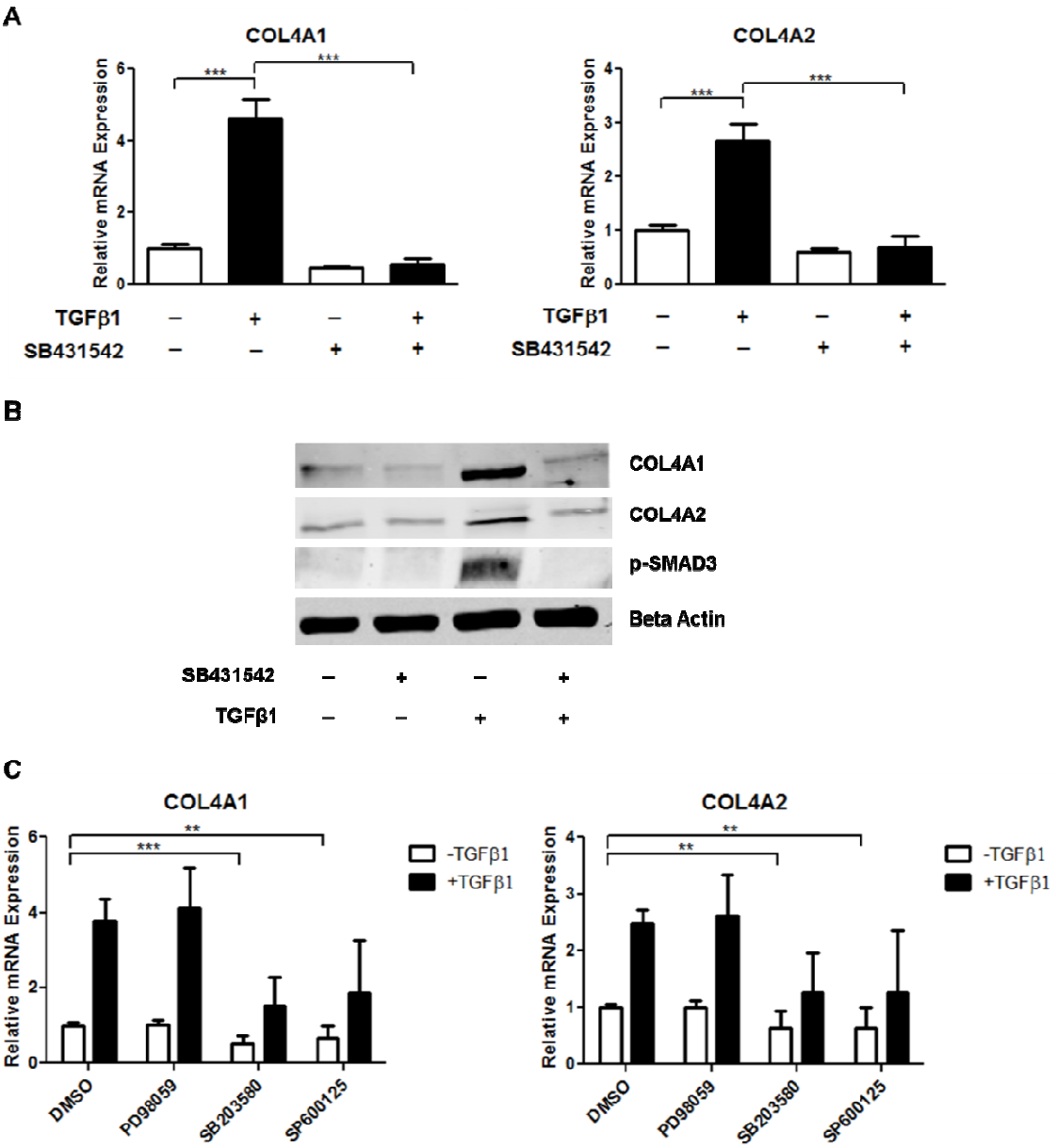
Individual Cohorts				
Study	**Cases/Controls**	**P-value**		**OR (95% CI)**
OHGS_A2	922/1002	0.072		1.92 (0.95-3.89)
OHGS_C2	835/317	0.034		2.56 (1.08-6.06)
DUKE_2	1181/643	0.018		2.18 (1.14-4.16)
CCGB_2	1616/364	0.030		2.44 (1.09-5.45)
ITH_2	402/448	0.021		2.44 (1.14-5.24)
Combined Meta-Analysis				
P-value	**Bonferroni-Adjusted P-value**	**Q**	**I^2**	**OR (95% CI)**
$1.6x10^{-6}$	$6.9x10^{-3}$	0.98	0	2.26 (1.62-3.15)

Table 2. Association of the rs72655775 (*COL4A2*) and rs12441344 (*SMAD3*) SNPs individually with CAD in 5 independent cohorts

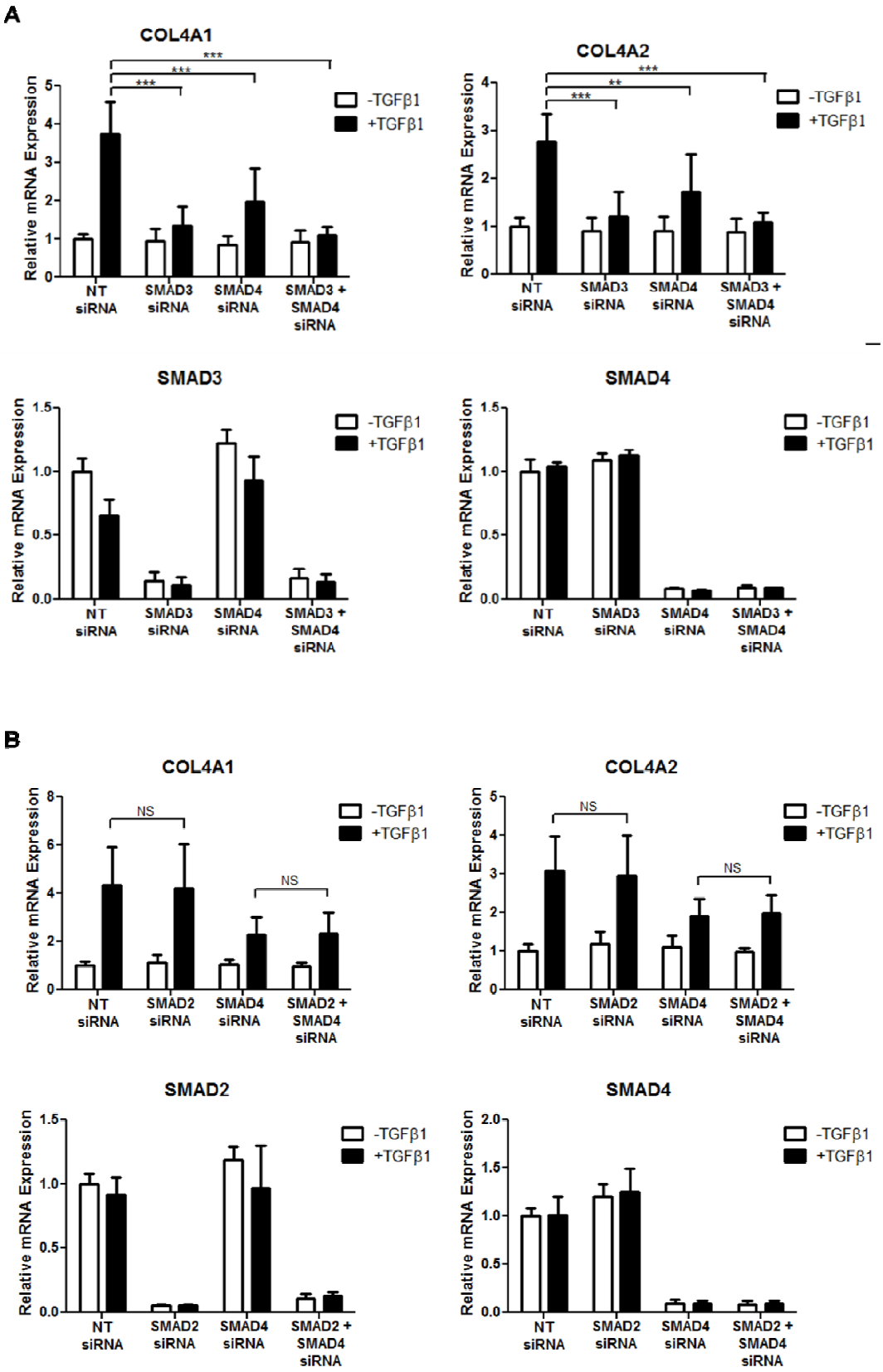
SNP	Minor Allele	Cases/Controls	Minor Allele Frequency (Cases)	Minor Allele Frequency (Controls)	P-value	OR (95% CI)
rs72655775	C	4732/2602	0.059	0.044	9.2×10^{-3}	1.25 (1.06-1.48)
rs12441344	G	4890/2739	0.21	0.22	0.12	0.94 (0.87-1.02)

emphasizing that these SNPs interact together in the context of CAD. Distributions of genotypes for rs72655775 and rs12441344 in cases and controls in the 5 cohorts are shown in **Supplementary Table 2**. Although the rs72655775 SNP has a low minor allele frequency, approximately 6% in CAD cases, 4% in controls (**Table 2**), the C allele for rs72655775 and G allele for rs12441344 co-segregate preferentially in CAD cases compared to controls. Odds ratios for various combinations of genotypes at these two SNPs are given in **Supplementary Table 3**. For each copy of the G allele at rs12441344, the CAD odds ratio progressively increases for carriers of the C allele at rs72655775. These findings demonstrate that epistasis analysis of GWAS data is feasible when filtering based on previous biological knowledge, allowing detection of significant genetic interactions after Bonferroni correction. These results extend the biological data and highlight the importance of SMAD3 in COL4A1/COL4A2 regulation and suggest COL4A1/COL4A2 and SMAD3 have interrelated roles in the pathogenesis of atherosclerosis.

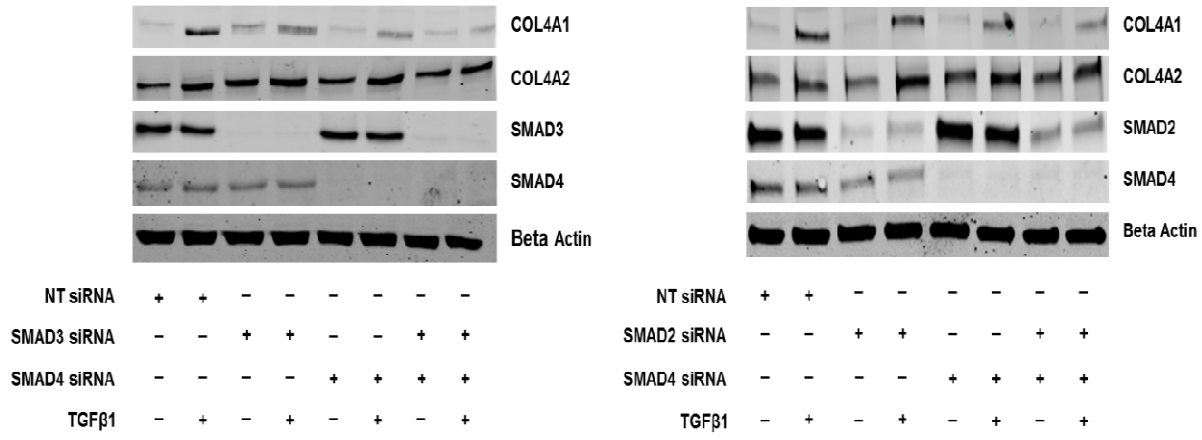
Online Supplementary Figures and Tables



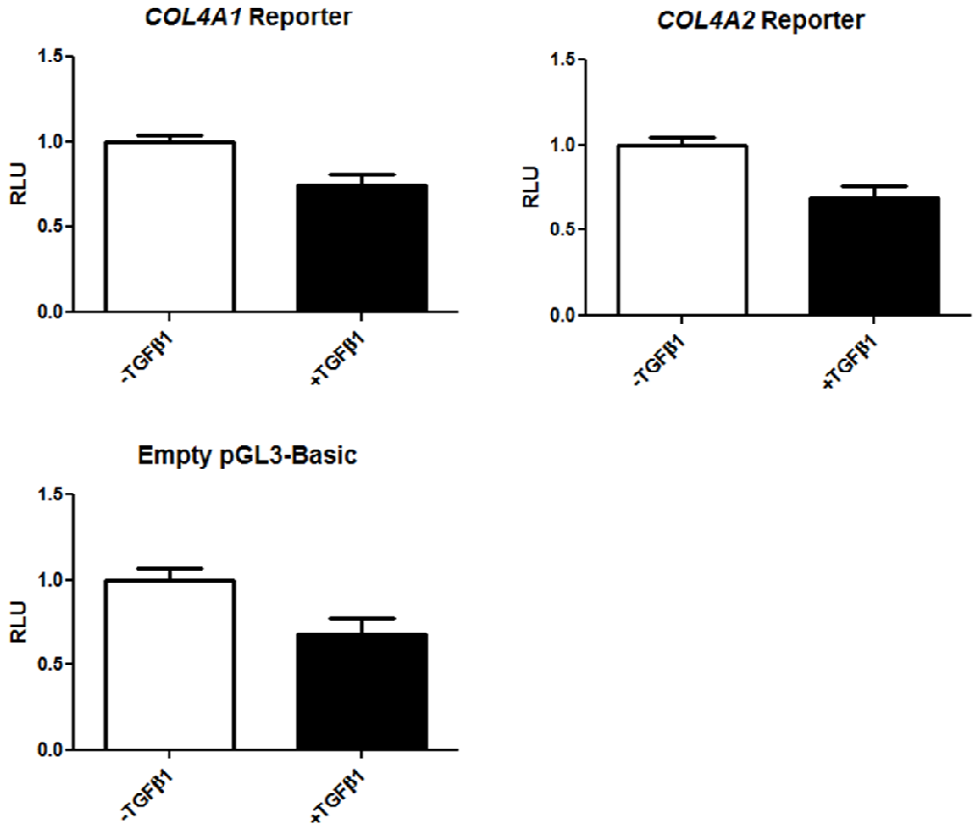
Supplementary Figure 1. Treating HT1080 cells with SB431542 blocks the upregulation of COL4A1 and COL4A2 by TGFβ1 whereas blocking non-SMAD pathways has little effect. HT1080 cells were pretreated with DMSO or 10 μM of pharmacological inhibitor for 1 h prior to treatment with 2 ng/mL TGFβ1. (A, B) HT1080 cells were pretreated with SB431542 and harvested for mRNA 16 h after TGFβ1 treatment (A) and harvested for protein 24 h after TGFβ1 treatment (B). (C) HT1080 cells were pretreated with a panel of inhibitors targeting non-SMAD pathways that act through TGFβ and harvested for mRNA 16 h after TGFβ1 treatment. COL4A1 and COL4A2 mRNA levels were measured by qRT-PCR using *PPIA* as the reference gene. Values represent the mean ± S.D. of three independent experiments with three different wells analyzed per condition each experiment. Values are normalized relative to DMSO treatment in the absence of TGFβ1. **$p < 0.01$, ***$p < 0.001$, ANOVA with Bonferroni's post-hoc test. Protein levels of COL4A1, COL4A2, phospho-SMAD3 (Ser423/425) and β-actin were measured by Western blot.


A
COL4A1
Relative mRNA Expression
-TGFβ1
+TGFβ1
NT siRNA
SMAD3 siRNA
SMAD4 siRNA
SMAD3 + SMAD4 siRNA
COL4A2
SMAD3
SMAD4
B
COL4A1
COL4A2
SMAD2
SMAD4
NS
SMAD2 siRNA
SMAD2 + SMAD4 siRNA

C

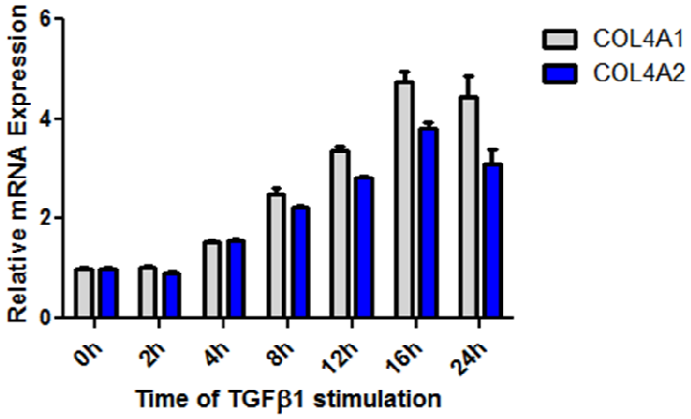

COL4A1
COL4A2
SMAD3
SMAD4
Beta Actin
NT siRNA + + − − − − − −
SMAD3 siRNA − − + + − − + +
SMAD4 siRNA − − − − + + + +
TGFβ1 − + − + − + − +
COL4A1
COL4A2
SMAD2
SMAD4
Beta Actin
NT siRNA + + − − − − − −
SMAD2 siRNA − − + + − − + +
SMAD4 siRNA − − − − + + + +
TGFβ1 − + − + − + − +

Supplementary Figure 2. SMAD3, but not SMAD2, knockdown abolishes the upregulation of COL4A1 and COL4A2 by TGFβ1 in HT1080 cells. (A) HT1080 cells were treated with nontarget siRNA, SMAD3 siRNA, SMAD4 siRNA, or a combination of SMAD3 and SMAD4 siRNA for a total of 48 h. 24 h into the siRNA knockdown, cells were treated with or without 2 ng/mL TGFβ1 for the final 24 h. The total siRNA concentration for each well was 20 nM. (B) HT1080 cells were treated with nontarget siRNA, SMAD2 siRNA, SMAD4 siRNA, or a combination of SMAD2 and SMAD4 siRNA for a total of 48 h. 24 h into the siRNA knockdown, cells were treated with or without 2 ng/mL TGFβ1 for the final 24 h. The total siRNA concentration for each well was 20 nM. *COL4A1* and *COL4A2* mRNA levels were assessed by qRT-PCR using *PPIA* as the reference gene. Values represent the mean ± S.D. of three independent experiments with three different wells analyzed per condition each experiment. Values are normalized relative to NT siRNA in the absence of TGFβ1. $^{**}p < 0.01$, $^{***}p < 0.001$, ANOVA with Bonferroni's post-hoc test. (**C**) COL4A1, COL4A2, SMAD3, SMAD2 and SMAD4 protein levels were determined by Western blotting, with β-actin as the loading control.

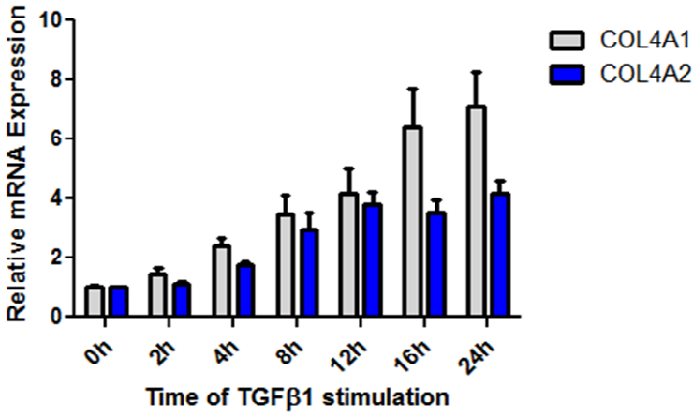

COL4A1 Reporter
RLU
1.5
1.0
0.5
0.0
-TGFβ1
+TGFβ1
COL4A2 Reporter
RLU
1.5
1.0
0.5
0.0
-TGFβ1
+TGFβ1
Empty pGL3-Basic
RLU
1.5
1.0
0.5
0.0
-TGFβ1
+TGFβ1

Supplementary Figure 3. Exogenous TGFβ treatment does not affect luciferase activity of human *COL4A1* or *COL4A2* reporter constructs in HT1080 cells. HT1080 cells were transfected with *COL4A1* and *COL4A2* promoter reporter constructs for 6 h, and then treated with 2 ng/mL TGFβ1 for a subsequent 42 h.

Aortic Smooth Muscle Cells



HT1080 Cells



Supplementary Figure 4. Time course of COL4A1 and COL4A2 mRNA induction by TGFβ1. AoSMCs (top) and HT1080 cells (bottom) were serum-starved for 16 h, treated with TGFβ1 (0 h time point) and harvested for RNA from 2 h to 24 h. Doses of 10 ng/mL and 2 ng/mL TGFβ1 were used for the AoSMCs and HT1080 cells, respectively. mRNA levels were assessed by qRT-PCR using *PPIA* as the reference gene. Values for each gene are normalized relative to 0 h (no TGFβ1 stimulation).

Supplementary Table 1

Too large to print out – available online

Supplementary Table 2. Distributions of genotypes for the rs72655775 (*COL4A2*) and rs12441344 (*SMAD3*) SNPs in the combined analysis of 5 independent cohorts

		CAD Cases (n = 4667)		CAD Controls (n = 2567)	
rs72655775 Genotype	**rs12441344 Genotype**	**n**	**%**	**n**	**%**
CC	GG	1	0.02	0	0.00
	GA	8	0.17	0	0.00
	AA	6	0.13	4	0.16
CT	GG	23	0.49	5	0.19
	GA	185	3.96	50	1.95
	AA	315	6.75	162	6.31
TT	GG	185	3.96	119	4.64
	GA	1353	28.99	841	32.76
	AA	2591	55.51	1386	53.99

Supplementary Table 3. Odds ratios for various combinations of rs12441344 and rs72655775 genotypes

rs12441344 Genotype	CAD Status	rs72655775 Allele counts		OR for rs72655775 (C allele)	95% CI
		C	T		
GG	Cases	25	393	3.1	1.17-8.2
	Controls	5	243		
GA	Cases	201	2891	2.4	1.8-3.3
	Controls	50	1732		
AA	Cases	327	5497	1.0	0.8-1.2
	Controls	170	2934		

2.5 Discussion

Here we have demonstrated in human AoSMCs that the stimulatory effect of TGFβ1 on COL4A1 and COL4A2 mRNA and protein levels is dependent on activation of the SMAD signaling pathway via the type I TGFβ receptor. More specifically, by siRNA knockdown, we show that the upregulation of COL4A1 and COL4A2 by TGFβ1 is dependent on SMAD3 and SMAD4 but not SMAD2. SMAD3/SMAD4 activation via TGFβ stimulation thus represents a critical regulatory hub modulating type IV collagen synthesis. This SMAD3/SMAD4 dependency for stimulation of COL4A1 and COL4A2 expression by TGFβ was also observed in HT1080 cells, the most studied cell line for human type IV collagen promoter regulation (**Supplementary Figures 1 and 2**).

These data do not support a direct transcriptional role for SMAD3 at the *COL4A1/COL4A2* bidirectional promoter region. First, we observed slow increases in *COL4A1* and *COL4A2* mRNA levels upon TGFβ1 stimulation that are more suggestive of an indirect contribution (**Supplementary Figure 4**). Secondly, overexpression of *SMAD3* did not alter the luciferase activity of human *COL4A1* or *COL4A2* promoter constructs. Similarly, TGFβ1 stimulation did not upregulate *COL4A1* or *COL4A2* promoter activity in luciferase assays. This finding contrasts with a study in murine renal tubular epithelial cells that reported transcriptional activity of chimeric collagen (IV) constructs is augmented following TGFβ1 treatment[221]. Another study in murine proximal tubular cells also demonstrated that TGFβ1 increased activity of luciferase reporter constructs consisting of the type IV collagen bidirectional promoter and 5' upstream regulatory sequences[222]. These discrepancies may be due to differences between humans and rodents in promoter sequence, promoter strength,

transcriptional machinery and/or cell type specificity. There is also the possibility that the *COL4A1* and *COL4A2* reporters used in this study containing the promoter sequences either lack appropriate additional enhancing elements or lack the appropriate chromosomal context. Although our human promoter system was not responsive to SMAD3 or TGFβ, it will be worth investigating whether other sequences at this locus could mediate promoter responsiveness to TGFβ. There is correlative data in murine mesangial cells suggesting the transcription factor Sp1 may play a role in mediating TGFβ's effects on type IV collagen[223]. Chromatin immunoprecipitation experiments did not reveal binding of SMAD3 to the *COL4A1/COL4A2* bidirectional promoter, either basally or upon TGFβ1 stimulation *(*data not shown*)*. This is largely consistent with previously published SMAD3 ChIP-seq experiments in other cell types where SMAD3 binding at the *COL4A1/COL4A2* locus was not observed, whether in the presence or absence of TGFβ[224].

Using tetracycline-induced siRNA knockdown of SMAD4 in two immortalized cell lines, Levy *et al*. also suggested that *COL4A1/COL4A2* may be an indirect SMAD target. Specifically, *COL4A1* was described as a late-TGFβ response gene, with significant induction at 6 hours that was not fully dependent on SMAD4[219]. In our experiments we observed that SMAD3 siRNA knockdown or SMAD4 siRNA knockdown alone did not completely abolish the effects of TGFβ1 indicating that there are other unknown mechanisms at play. One possibility is that although we observed a consistent knockdown of at least 80-90% for *SMAD3* and *SMAD4* mRNA (also verified at the protein level), the remaining SMAD3 and SMAD4 could still mediate a small fraction of the canonical TGFβ signaling. Since ALK5 inhibition using SB431542 resulted in a total abrogation of TGFβ's

effects (**Figures 1A and 1B**), it is also possible that blocking ALK5 could impinge on type IV collagen expression via minor, SMAD-independent, effectors.

Since the effect of TGFβ on type IV collagen is believed to be transcriptionally-based[208], further studies are required to investigate specific mechanisms underlying the requirement of SMAD3 for TGFβ1-induced mRNA synthesis. One possibility is the existence of SMAD3 cross-talk with an unknown pathway or regulation of another gene that in turn regulates COL4A1 and COL4A2. SMAD3 may also affect COL4A1 and COL4A2 levels indirectly by regulating microRNAs. For instance, SMAD3 represses the transcription of miR-29 in several cell types by binding to the miR-29 promoter[225,226]. *COL4A1* and *COL4A2* are targets and miR-29 knockdown results in elevated *COL4A1/COL4A2* mRNA levels[227]. There is also evidence from mice and rats suggesting SMAD3 activation correlates with reduced miR-29 levels and increased type IV collagen[228].

By epistasis analysis we identified a significant genetic interaction for CAD risk between rs72655775 at the *COL4A1/COL4A2* locus and rs12441344 at the *SMAD3* locus. This statistical interaction further supports the biological data demonstrating that the regulation of COL4A1/COL4A2 by TGFβ is SMAD3-dependent. Epistasis data also suggest a synergistic rather than additive role between SMAD3 and type IV collagen in CAD pathogenesis. The interaction between rs72655775 in *COL4A2* and rs12441344 in *SMAD3* achieved an odds ratio for CAD of 2.26 (**Table 1**) in contrast with individual odds ratios of 1.25 and 0.94 for rs72655775 and rs12441344, respectively (**Table 2**). The odds ratio for interaction of 2.26 is particularly striking given that index SNPs at most CAD GWAS loci exhibit small effect sizes. Of note, the allele specific odds ratios for lead SNPs at the *COL4A1/COL4A2* locus

(rs4773144, rs9515203) and the *SMAD3* locus (rs17228212) were 1.07, 1.08[26] and 1.21[38] respectively. More studies are required to elucidate the significance of the SNPs highlighted as part of our epistasis analysis but of note rs12441344 (**Table 1**) was associated with sudden cardiac arrest in patients with CAD[229]. Both rs72655775 and rs12441344 are intronic variants, suggesting disruption of noncoding regulatory sequences at these loci could play a synergistic role in altering COL4A1/COL4A2 and SMAD3 expression and signaling.

Perturbations in the TGFβ/SMAD3 signaling axis would be expected to interfere with type IV collagen synthesis thus promoting atherosclerosis. Reduced type IV collagen may destabilize plaques and increase the risk of plaque rupture[26]. During the progression of atherosclerosis, smooth muscle cells switch from a contractile phenotype to a synthetic phenotype[108], accompanied by increased smooth muscle cell proliferation, migration, and fibrosis and decreased synthesis of smooth muscle contractile proteins. Type IV collagen limits the growth of smooth muscle cells, enhances the expression of contractile genes, reduces inflammatory gene expression, reduces matrix calcification, and limits LDL uptake in culture[108]. Type IV collagen expression is higher in quiescent contractile smooth muscle cells as compared to those of a proliferating/migrating phenotype[100,105]. Thus, lower type IV collagen production would be expected to promote smooth muscle cell proliferation and migration. The amount of type IV collagen produced in the basement membrane may also influence a variety of cell-matrix interactions such as binding to integrin receptors or DDR1 (discoidin domain receptor family, member 1) and control the infiltration and accumulation of macrophages, lymphocytes, and lipids in the subendothelial space.

To our knowledge, these are the first experiments in human vascular cells to demonstrate SMAD3-dependency of TGFβ upregulation of *COL4A1* and *COL4A2* expression. We show that the *COL4A1/COL4A2* and *SMAD3* loci, previously reported to individually associate with CAD[26,38,42] interact epistatically to modify CAD risk. Taken together these biological and statistical data support the concept that SMAD3 and COL4A1/COL4A2 play shared roles in modifying atherosclerosis risk.

3 FUNCTIONAL ANALYSIS OF A NOVEL GENOME-WIDE ASSOCIATION STUDY SIGNAL IN *SMAD3* THAT CONFERS PROTECTION FROM CORONARY ARTERY DISEASE

Adam W Turner[1], Amy Martinuk[1], Anada Silva[1], Paulina Lau[1], Majid Nikpay[2], Per Eriksson[3], Lasse Folkersen[3,4], Ljubica Perisic[5], Ulf Hedin[5], Sebastien Soubeyrand[1], Ruth McPherson[1,2]

[1] Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada

[2] Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada

[3] Atherosclerosis Research Unit, Karolinska University Hospital, Stockholm, Sweden

[4] Department of Systems Biology, Technical University of Denmark, Copenhagen, Denmark

[5] Department of Molecular Medicine & Surgery, Karolinska Institute, Stockholm, Sweden

Short title: Functional analysis of the *SMAD3* locus for CAD

Status: Published in *Arteriosclerosis, Thrombosis, and Vascular Biology* 36 (2016) 972-983

3.1 Abstract

Objective – A recent GWAS meta-analysis identified an intronic SNP in *SMAD3*, rs56062135C>T, the minor allele (T) of which associates with protection from CAD. Relevant to atherosclerosis, SMAD3 is a key contributor to TGFβ pathway signaling. Here we seek to identify one or more causal CAD-associated SNPs at the *SMAD3* locus and characterize mechanisms whereby the risk allele(s) contribute to CAD risk.

Approach and Results – By genetic and epigenetic fine mapping, we identified a candidate causal SNP rs17293632C>T (D' 0.97, r^2 0.94 with rs56062135) in intron 1 of *SMAD3* with predicted functional effects. We show that the sequence encompassing rs17293632 acts as a strong enhancer in human arterial smooth muscle cells (hASMC). The common allele (C) preserves an AP-1 site and enhancer function while the protective (T) allele disrupts the AP-1 site and significantly reduces enhancer activity ($p<0.001$). Pharmacological inhibition of AP-1 activity upstream demonstrates that this allele-specific enhancer effect is AP-1 dependent ($p<0.001$). Chromatin immunoprecipitation experiments reveal binding of several AP-1 component proteins with preferential binding to the (C) allele. We show that rs17293632 is an eQTL for *SMAD3* in blood and atherosclerotic plaque with reduced expression of *SMAD3* in carriers of the protective allele. Finally, siRNA knockdown of *SMAD3* in hASMCs increases cell viability, consistent with an anti-proliferative role.

Conclusions – The CAD-associated rs17293632C>T SNP represents a novel functional *cis*-acting element at the *SMAD3* locus. The protective (T) allele of rs17293632 disrupts a consensus AP-1 binding site in a *SMAD3* intron 1 enhancer, reduces enhancer activity and *SMAD3* expression, altering hASMC proliferation.

3.2 Introduction

Coronary artery disease (CAD) characterized by the deposition of atherosclerotic plaque in the epicardial arteries remains a leading cause of death and disability[230]. Environmental variables contribute to CAD but genetic factors are believed to be of equal importance. Over the past several years large consortia have conducted meta analyses of genome-wide association studies (GWAS) for CAD[26,42] resulting in the identification of over 45 CAD-associated loci[26]. Together these earlier studies explained approximately 10.6% of the predicted heritability of CAD and for the majority the functional link to atherosclerosis remains unknown.

Recently, using phased haplotypes from the 1000 Genomes Project to impute an enlarged number of SNPs for 48 GWAS, consisting of over 180,000 cases and controls, we identified an additional 10 CAD-associated loci. One of the novel CAD-associated SNPs is rs56062135C>T at the *SMAD3* locus on chromosome 15 and the common allele (C) is associated with increased risk (OR=1.07) relative to the (T) allele[44] (**Table 1**). Of note, this SNP is not in linkage disequilibrium with rs17228212T>C, reported to be associated with CAD in an earlier GWAS[38].

The SMAD3 transcription factor is an important mediator of TGFβ signalling, regulating transcription of genes with SMAD-binding elements. Upon TGFβ stimulation, the type II TGFβ receptor recruits and phosphorylates the type I TGFβ receptor, which subsequently phosphorylates the SMAD3 protein. Phosphorylated SMAD3 then forms a complex with the common (co) SMAD4 that subsequently translocates to the nucleus and regulates transcription[137,143]. Relevant to a role in atherosclerosis, in a systems genetics analysis of

Table 1. Association of the GWAS reported rs56062135[44] SNP and the rs17293632 SNP in intron 1 of *SMAD3* with CAD

									Association model			
									Additive		Recessive	
Lead variant	**Locus**	**Chr**	**A1/A2**	**Effect Allele (A1) freq**	**Imputation quality**	**I^2**	**Heterogeneity P**	**N studies**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
rs56062135	*SMAD3*	15	C/T	0.79	0.98	0.00	0.67	46	1.07 (1.05-1.10)	4.50E-09	1.17 (1.10-1.25)	8.88E-07
rs17293632	*SMAD3*	15	C/T	0.79	0.98	0.00	0.67	46	1.07 (1.05-1.10)	5.72E-09	1.17 (1.11-1.23)	1.22E-06

multiple GWAS studies, we identified TGFβ signalling and SMAD transcriptional activities as enriched pathways for CAD association[45]. However, despite extensive data on the functions of TGFβ with respect to atherosclerosis[158,169], the roles of SMAD proteins, particularly SMAD3 and SMAD3 signaling are less well understood.

SMAD3 is expressed at very low levels in healthy human aorta by immunohistochemistry and qRT-PCR[175]. There is however a marked increase in SMAD3 and other SMAD proteins in fibrofatty lesions, with expression mostly limited to CD68-positive macrophages/macrophage-derived foam cells in these samples. Conversely, in fibrous atherosclerotic plaques, there are high levels of SMAD3 in vascular smooth muscle cells, suggesting that the role of SMAD3 in atherosclerosis depends on cell type and stage of atherosclerosis[175]. Higher SMAD3 expression in smooth muscle cells of the fibrous plaque coincides with TGFβ-mediated synthesis of collagen and other extracellular matrix proteins, which contribute to plaque stability[159]. Rare *SMAD*3 mutations cause aneurysms-osteoarthritis syndrome (AOS), an autosomal dominant form of thoracic aortic aneurysms and dissections (TAAD)[151].

Here we have identified and characterized a functional SNP, rs17293632C>T, in *SMAD3* that resides in a newly identified GWAS locus for CAD. Consistent with functionality, we show that rs17293632C>T is an eQTL in both whole blood and carotid plaque samples with reduced expression of *SMAD3* in carriers of the protective (T) allele. We demonstrate that the protective (T) allele disrupts a consensus AP-1 binding site within a strong intronic enhancer and impairs enhancer activity in hASMCs as well as HeLa and HepG2 cells. Pharmacological blocking of AP-1 signaling and of DNA binding ability markedly

attenuates enhancer activity and ChIP experiments reveal binding of numerous AP-1 components to the region encompassing rs17293632, with impaired binding to the (T) vs. (C) allele. Finally, siRNA knockdown of *SMAD3* in hASMCs increases cell viability, consistent with an anti-proliferative role.

3.3 Materials and Methods

Cell Culture

Primary human arterial smooth muscle cells (hASMCs) were obtained from two sources. Lot HITC6 was provided by the laboratory of Dr. JG Pickering. These cells were derived from primary cultures of smooth muscle cells from the human thoracic artery[231]. They were homozygous for the CC genotype at rs17293632 and all experiments were performed between passages 28-33. The second lot of hASMCs, 0000287836, was purchased from Lonza (catalog #CC-2571) and were heterozygous for rs17293632. These cells were derived from the ascending aorta and experiments were performed between passages 5 and 9. Both lots of hASMCs were grown in SmGM-2 Smooth Muscle Cell Basal Medium (Lonza) supplemented with 5% FBS, insulin, hFGF-B, GA-1000, and hEGF. HeLa cells were purchased from ATCC and grown in high glucose DMEM (Gibco) supplemented with 10% FBS, Penicillin-Streptomycin, and L-glutamine. HepG2 cells were also purchased from ATCC and grown in low glucose DMEM (Gibco) supplemented with 10% FBS, Penicillin-Streptomycin, and L-glutamine.

Generation of *SMAD3* Reporter Constructs

The putative enhancer sequence in intron 1 of *SMAD3* (Chr15:67,441,598-67,443,643) was first PCR amplified with primers: forward 5'-GATTGAGCTCCCTGTTCAGCATTTTGAGTTTC-3' (SacI restriction site); reverse 5'-GATTGCTAGCGAGCTATTGGAGACTGTGAGA-3' (NheI restriction site). PCR was

performed on a previously genotyped genomic DNA template that had a CC genotype at rs17293632 and homozygous for all other SNPs in the amplicon. This PCR product, as well as empty pGL3-Promoter vector (Promega), were both double-digested with SacI and NheI and subsequently PCR or gel purified before ligation. Site-directed mutagenesis was used to change the allele at rs17293632 from C to T and performed using the Q5 site-directed mutagenesis kit (New England BioLabs).

To test for enhancer activity in the context of the human *SMAD3* promoter, we obtained the SMAD3p-Luc plasmid from the laboratory of Dr. Thomas Kelley[232]. This plasmid contains the human *SMAD3* promoter (fragment corresponding to -1879 to +13 of the ATG start site of the *SMAD3* gene) upstream of the firefly luciferase gene in the pGL3-Basic vector. From the pGL3-Promoter-rs17293632-C and pGL3-Promoter-rs17293632-T plasmids we PCR amplified a 347 bp segment using the following primers: forward 5'-GATTGTCGACCTGAGATGGTTGTAAATGTCCC-3' (SalI restriction site); reverse 5'-GATTGTCGACAACTGGCGGCCTTGTTCTAT-3' (SalI restriction site), PCR amplification was conducted using Q5 Hot Start High-Fidelity DNA Polymerase (New England BioLabs). These PCR products plus SMAD3p-Luc, were digested with SalI and subsequently PCR or gel purified before ligation. Correct sequences and insert orientations of all vectors were verified by Sanger sequencing.

Transfection and Luciferase Assays

For transient transfection studies in hASMCs, HITC6 cells were seeded in 6 well dishes and grown to ~85% confluence at the time of transfection. Each well was transfected with 2.5 µg

of firefly luciferase construct and 2% pRL-SV40 *Renilla* (pGL3-Promoter experiments) or 2% pRL-TK *Renilla* (SMAD3p-Luc experiments). hASMCs were harvested 48 hours after transfection by washing once with 1X PBS and lysing with 500 µL of 1X Passive Lysis Buffer (Promega). For transient transfection studies with HeLa and HepG2 cells, cells were seeded in 12 well dishes and grown to 70% confluence at the time of transfection. Each well was transfected with 500 ng of firefly luciferase construct and 2% pRL-TK *Renilla*. The HeLa and HepG2 cells were harvested 24 hours after transfection with 250 µL of 1X Passive Lysis Buffer per well. Transfections for all cell types were conducted using the Lipofectamine 3000 Reagent (Life Technologies). Lysates were centrifuged for 1 minute at 13,000 rpm at 4°C and 20 µL of supernatant were used for the luciferase assay. Luciferase readings for the hASMCs were measured with the Lumat LB 9507 luminometer (Berthold Technologies) and readings for the HeLa and HepG2 cells were measured with the GloMax luminometer (Promega).

Drug Treatments

To determine the effect of SP600125 on *SMAD3* enhancer activity, HITC6 hASMCs were seeded in 6 well dishes and grown to approximately 85% confluence. SP600125 was purchased from Cell Signaling Technology and added to cells for 1 h at a final concentration of 50 µM prior to transfection with the *SMAD3* pGL3-Promoter constructs (2.5 µg reporter per well). The HITC6 cells were harvested 24 h after transfection with 500 µL of 1X Passive Lysis Buffer per well. The final concentration of DMSO in both the control and SP600125 treatments was 0.1%.

PMA Treatments and c-Jun Knockdown

For PMA experiments both HITC6 and lot 287836 cells were seeded in 6 well plates and grown to approximately 85% confluence. At this point, media was changed for 16 hours to SmGM-2 Smooth Muscle Cell Basal Medium (Lonza) supplemented with 0.5% FBS, insulin, hFGF-B, GA-1000, and hEGF. Both lots of cells were treated with a final concentration of 50 nM PMA for the following time points: 0 h, 1 h, 2 h, and 4 h before being harvested for RNA.

For c-Jun siRNA knockdown, HITC6 cells were seeded in 6 well plates and grown to ~80% confluence. Cells were transfected with a final concentration of either 50 nM nontarget siRNA or 50 nM *JUN* siRNA (s7658, Life Technologies). 48 h after transfection, cells were harvested for RNA.

RNA for all samples was isolated using the High Pure RNA Isolation Kit (Roche). 500 ng of RNA was reverse-transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche) using a combination of oligo(dT) and random hexamer primers. Quantitative PCR of cDNA samples was conducted using the LightCycler 480 II machine (Roche). Human *β-actin* was used as the reference gene for qPCR experiments. The following primer sequences were used for qPCR experiments: *SMAD3*: forward 5'-GCCTTCTGGTGCTCCATCTC-3' reverse 5'-AATAGCGCTGTCACTGAGGCA-3'; *JUN*: forward 5'-AGGTTCAGGGTCATGCTCTGTT-3' reverse 5'-ACGTGAAGTGACGGACTGTTCT-3'; *β-actin*: forward 5'-TCCCTGGAGAAGAGCTACGA-3' reverse 5'-ATCTGCTGGAAGGTGGACAG-3'.

Protein Expression

For all protein experiments, either HITC6 or lot 287836 cells were grown in 6 well plates. Protein samples were harvested on ice using RIPA buffer containing fresh Roche cOmplete Protease Inhibitor cocktail and Roche PhosStop Phosphatase Inhibitor cocktail. Protein concentration was determined using the Pierce BCA Protein Assay Kit. 35 µg of each denatured sample was loaded onto a 4-15% gradient SDS-PAGE gel (Bio-Rad) and run under reducing conditions. Samples were transferred to nitrocellulose membranes (Bio-Rad) for 1 h at 100V and subsequently blocked for 1 h at room temperature using Odyssey blocking buffer (SMAD3, c-Jun, and PARP) or 5% BSA (phospho c-Jun). To confirm SP600125 reduced phosphorylation of c-Jun, the membrane was incubated with rabbit Phospho-c-Jun (Ser63) II Antibody #9261 (Cell Signaling Technology, 1:1000 dilution). Total c-Jun was probed by stripping the membrane and re-probing with rabbit c-Jun (H-79) antibody (Santa Cruz Biotechnology, sc-1694). To confirm SMAD3 knockdown in the cell proliferation assays, membranes were incubated with rabbit SMAD3 Antibody #9513 (Cell Signaling Technology, 1:1000 dilution). Apoptosis was assessed by probing with rabbit PARP Antibody #9542 (Cell Signaling Technology, 1:1000 dilution), that detects the normal 116 kDa form of PARP as well as the cleaved 89 kDa protein. For loading controls, all membranes were probed with mouse beta Tubulin Antibody D66 (GeneTex, 1:2000 dilution). IRDye secondary antibodies (LI-COR) were used at a concentration of 1:15000 and membranes were visualized with the LI-COR Odyssey imaging system.

Human Samples for *SMAD3* Whole Blood Analysis

Healthy normal weight individuals of European ancestry[233] were selected for eQTL studies on whole blood. Genomic DNA was genotyped using Affymetrix 6.0 arrays and SNP data imputed using the Phase 1 v3 release from the 1000 Genomes Project. The minor allele frequency for rs17293632 (T) is 0.21 (**Table 1**); 41 subjects in this cohort were homozygous for the (T) allele at rs17293632. Many more subjects were heterozygous or homozygous (C) carriers and 50 of each were selected with careful matching for age and sex to the rs17293632 TT group (**Supplementary Table I**). Fasting blood samples were collected in PAXgene Blood RNA tubes (PreAnalytiX/QIAGEN) and stored at -80°C prior to analysis.

Quantitative Real-Time PCR for SMAD3 Whole Blood Analysis

Whole blood RNA was extracted using the PAXgene Blood RNA kit (PreAnalytiX/QIAGEN). cDNA was synthesized using the Transcriptor First Strand cDNA Synthesis kit (Roche) and a combination of anchored oligo(dT) and random hexamer primers, with 500 ng of patient whole blood RNA as the template. Quantitative PCR for *SMAD3* mRNA was performed with SYBR Green I Master (Roche) and run on the LightCycler 480 II machine (Roche). Peptidylprolyl Isomerase A (*PPIA*) was used as the reference gene for all quantitative PCR experiments. The following primer sequences were used for quantitative PCR analysis: *SMAD3*: forward 5'-GCCTTCTGGTGCTCCATCTC-3' reverse 5'-AATAGCGCTGTCACTGAGGCA-3'; *PPIA*: forward 5'-ACCGTGTTCTTCGACATTGC-3' reverse 5'-TTCTGTGAAAGCAGGAACCC-3'

SMAD3 eQTL Analysis in Primary Tissue

SMAD3 expression in human carotid plaque tissue was investigated using samples from the BiKE (Biobank of Karolinska Endarterectomies) study. The BiKE study at the Karolinska Institute, Stockholm, Sweden comprises a repository of atherosclerotic plaque and plasma samples obtained from patients that underwent carotid endarterectomy, along with clinical parameters, genotype data, RNA expression and proteomics data. Details of the study cohort have been published previously[234–237]. High-density genotyping in BiKE samples was performed using Illumina610w-QuadBead Array and matched with the corresponding plaque samples (n=125). Since rs17293632 was not present on the chip, rs16950687 (D' 0.949, r^2 0.859 with rs17293632; D' 1, r^2 0.953 with rs56062135) was selected as a proxy using the SNaP software from the Broad Institute.

SMAD3 expression was also probed *in vivo* in non-CAD patients from various tissues collected in the ASAP (Advanced Study of Aortic Pathology) study. Patients from the ASAP study had aortic valve and ascending aortic disease and all underwent elective open heart surgery[238]. None of these patients had significant CAD by coronary angiography. Tissues collected included mammary artery intima-media, liver, ascending aorta intima-media, ascending aorta adventitia, and heart. SNPs were genotyped on the Illumina610w-QuadBead Array and rs17293632 imputed using the Mach2.0 algorithm and 1000 Genomes data as reference (imputation quality $r^2 = 0.718$).

Chromatin Immunoprecipitation

Primary hASMCs were seeded in 150 mm dishes and grown to confluence before harvesting. Cells were fixed with 1% formaldehyde before being quenched with glycine. For each round of ChIP, cells from two 150 mm plates (~1 million cells per dish) were combined and lysed with 500 µL of nuclear lysis buffer. Cross-linked chromatin was sheared into fragments of approximately 500 bp using the BioRuptor sonicator (Diagenode) and clarified using centrifugation. PureProteome Protein A/G Magnetic Beads (Millipore) were preblocked with both salmon sperm DNA and BSA for 30 minutes at 4°C. Before the immunoprecipitation, the 500 µL of total chromatin was incubated with 10 µL of resuspended pre-blocked beads to further reduce background. Each chromatin immunoprecipitation used 200 µg of hASMC chromatin, 10 µL of resuspended beads, and 10 µg of either c-Fos (Santa Cruz Biotechnology, K-25: sc-253), c-Jun (Santa Cruz Biotechnology, H-79: sc-1694), JunB (Santa Cruz Biotechnology, 210: sc-73), JunD (Santa Cruz Biotechnology, 329: sc-74), p300 (Santa Cruz Biotechnology, C-20: sc-585), H3K27ac (Abcam, ChIP Grade ab4729), or nonspecific Rabbit IgG. Protein-antibody-bead complexes were incubated overnight at 4°C with rotation. Complexes were washed four times with wash buffer, once with high salt wash buffer, and eluted from the beads with elution buffer containing 1% SDS and 100 mM sodium deoxycholate and heating for 15 minutes at 65°C. Cross-links were reversed by addition of NaCl to a final concentration of 200 mM along with RNase A and shaking at 65°C for 5 hours. Eluted DNA was purified using the GeneJET PCR Purification Kit (Thermo Scientific) and used for quantitative PCR. Quantitative PCR was performed using SYBR Green I Master (Roche) and the LightCycler 480 II machine (Roche). The sequences of primers specific for rs17293632 were FWD: 5'-

CTCCGCGTGAATGTCACTG-3' and REV: 5'- GAGAGGTGAAGAGGGCAAAT-3'. Sequences of negative control primers, amplifying intronic areas of the *CPT1B* gene, are listed in **Supplementary Table II**. Melting curve analysis was conducted for each primer set for every run conducted.

Allele-Specific ChIP-qPCR

A TaqMan SNP Genotyping Assay for the rs17293632 SNP was purchased from Life Technologies (C__33991343_10). Lot 287836 hASMCs, confirmed to be heterozygous at rs17293632, were used for all allele-specific ChIP qPCR experiments. Quantitative PCR was first performed on the immunoprecipitated DNA as described above, and products purified using the GeneJET PCR Purification Kit (Thermo Scientific). TaqMan SNP genotyping assays for rs17293632 were then performed on PCR products to generate the ratio of the (C) to (T) allele for DNA that binds AP-1. For each allele-specific qPCR run, a standard curve was generated consisting of previously genotyped homozygous CC and homozygous TT genomic DNA samples mixed in the following ratios: 9:1, 8:2, 7:3, 6:4, 5:5, 4:6, 3:7, 2:8, and 1:9. The Log2 ratio of FAM/VIC of these standards at cycle 55 was then used to generate the standard curve. The Log2 ratio of FAM/VIC of the immunoprecipitated DNA was then fitted to the standard curve to calculate the ratio of (C) to (T) alleles. Both sonicated input PCR product, as well as unsonicated lot 287836 genomic DNA, which have 1:1 C:T ratios, were included as controls for each run.

Cell Proliferation Assays

Cell proliferation assays were performed using the same protocol for HITC6 and lot 287836 batches of hASMCs. These were first seeded in 24 well plates and grown in SmGM-2 Smooth Muscle Cell Basal Medium (Lonza) supplemented with 5% FBS, insulin, hFGF-B, GA-1000, and hEGF. siRNA transfection was performed when cells reached approximately 45% confluency. Cells were transfected with a final concentration of either 20 nM nontarget siRNA or 20 nM *SMAD3* siRNA. *SMAD3* siRNA oligonucleotides (s8401 and s8402) were purchased from Life Technologies. siRNA transfection was conducted using the Lipofectamine RNAiMAX reagent (Life Technologies) and 6 h after transfection, media was changed and replaced with normal growth medium. After 72 h of knockdown, cell proliferation was measured using the MTT assay (Molecular Probes, Life Technologies). The MTT assay detects living cells, and is based on reduction of the tetrazolium dye MTT to its insoluble form, formazan, that has a purple colour that can be quantified at 595 nm. After MTT incubation at 37°C, DMSO was added as the solubilizing reagent.

Conditional and Joint Association Analysis

To determine if the observed association between *SMAD3* and CAD is mediated by rs56062135 or other SNPs, we used conditional and joint association analysis implemented in GCTA software[239]. This is a two step procedure that involves conditional analysis to identify independent SNPs and secondly, fitting all selected SNPs jointly in a model and excluding SNPs with p values (p-values from joint analysis) that are greater than the cut-off p value, using the software default $p< 5x\ 10^{-8}$ (GWAS significant level) as our threshold.

GCTA method requires summary-level statistics from a meta-analysis of genome-wide association studies (GWAS) and estimated linkage disequilibrium (LD) from a reference sample with individual-level genotype data. We used the SNP summary statistics from the recently published 1000 Genomes–based GWAS meta-analysis of CAD[44] as our meta-analysis file (CARDIoGRAM Meta-analysis). Furthermore, the reference sample must be large enough so that the LD correlations are estimated with little error[239,240], and it can be one of the participating studies of the meta-analysis[239]; therefore, we used the 1000 Genomes post-imputed genotype data from our cohorts that are included in the CARDIoGRAM Meta-analysis and excluded variants with IMPUTE2 info < 0.4, in our final dataset, we have 13,367 subjects with available genotypes at 15M variants that we used as our reference sample for conditional and joint association analysis.

Statistical Analysis

Data displayed represents the mean ± SD for experiments conducted in triplicate, and is representative of 3 independent experiments. Significant differences were assessed by either one-way ANOVA followed by Bonferroni's post-hoc test or unpaired, two-tailed Student's T-test, where appropriate. Figures were generated and analysis performed using the GraphPad Prism software.

3.4 Results

Haplotype analysis of CAD-associated SNPs at the *SMAD3* locus and identification of candidate functional SNPs

The index GWAS SNP rs56062135C>T is intronic to *SMAD3* on chromosome 15 (**Table 1**)[44]. The *SMAD3* gene is approximately 130 kb in size with 9 exons, most clustered towards the 3' end of the gene (**Figure 1A**). The regional association plot of *SMAD3* and chromosome 15 indicated a cluster of CAD-associated SNPs in *SMAD3* in proximity to rs56062135C>T, with weaker CAD-associated signals in nearby genes at this locus (**Figure 1B**). Our conditional and joint association analysis showed that the association between *SMAD3* and CAD is mainly explained by rs56062135, as it was the only SNP that remained significant in the final joint model (**Supplementary Table III**). Further, we investigated the effect of rs56062135 on association of *SMAD3* variants with CAD by doing conditional analysis, as presented in **Supplementary Table V**. None of the variants tagging *SMAD3* reach GWAS significance ($p< 5 \times 10^{-8}$) after conditional analysis and the vicinity SNPs to rs56062135 showed GWAS significant association in the CARDIoGRAM meta-analysis because they are in high linkage disequilibrium with rs56062135 ($r^2 > 0.9$).

We then sought to identify other SNPs in strong LD with this SNP using HaploReg v3 and Haploview[241]. This analysis revealed 6 other intronic SNPs in strong LD ($r^2>0.8$), including rs17293632C>T within intron 1 (**Table 2**). No *SMAD3* exonic SNPs or SNPs predicted to alter splicing were in LD with rs56062135. Haploview software and imported linkage data from the 1000 Genomes Browser reveals 20 separate LD blocks at the *SMAD3* locus; rs56062135C>T and linked SNPs are all found within block 13 (**Supplementary Figure I**).

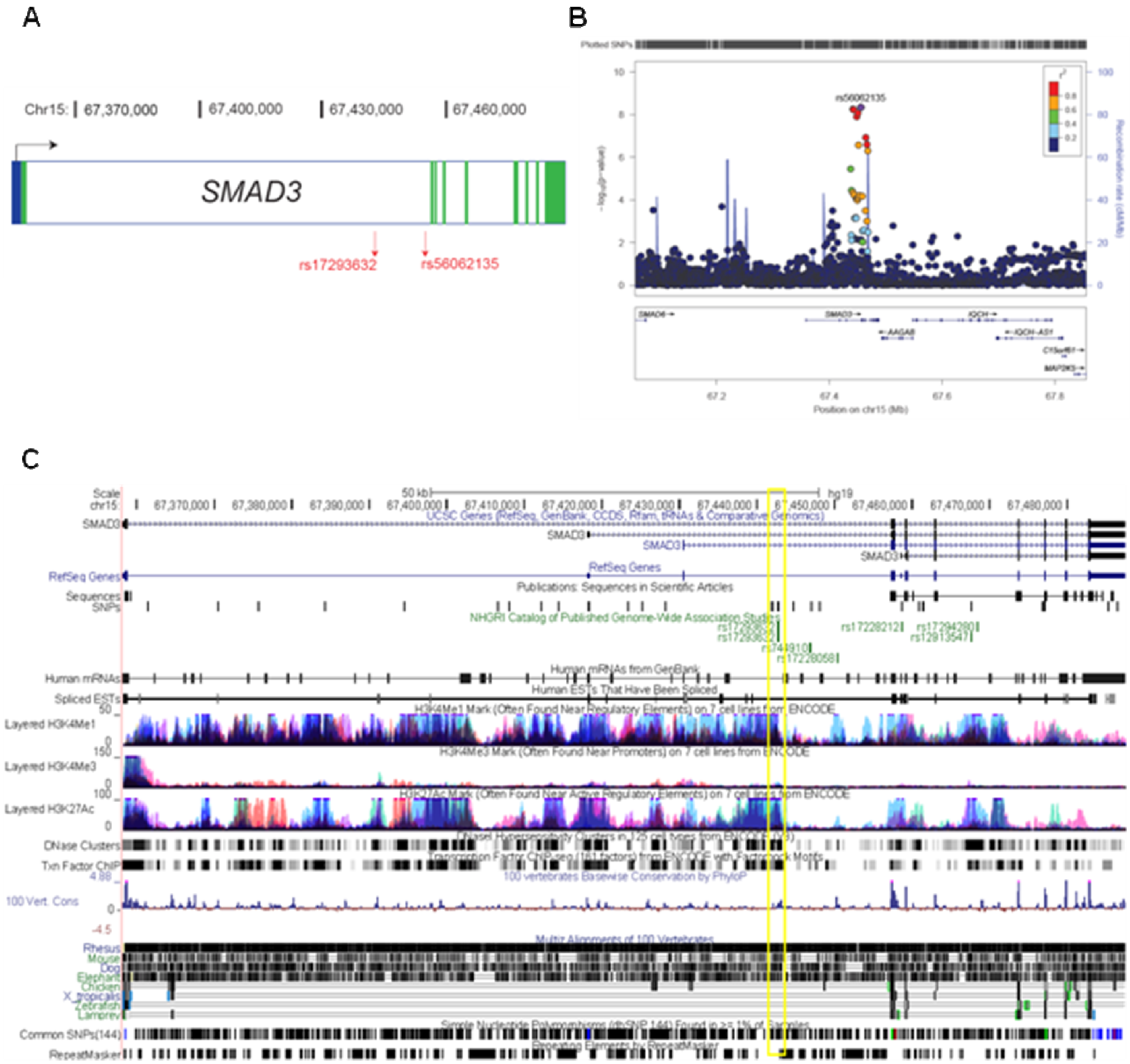

A
Chr15: 67,370,000
67,400,000
67,430,000
67,460,000
SMAD3
rs17293632
rs56062135
B
Plotted SNPs
rs56062135
Position on chr15 (Mb)
Recombination rate (cM/Mb)
C
Scale
chr15:
50 kb
hg19
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
Publications: Sequences in Scientific Articles
Sequences
SNPs
NHGRI Catalog of Published Genome-Wide Association Studies
Human mRNAs
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
Layered H3K4Me1
Layered H3K4Me3
Layered H3K27Ac
DNase Clusters
Txn Factor ChIP
100 Vert. Cons
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(144)
RepeatMasker

Figure 1. Overview and chromatin signatures of the *SMAD3* locus on chromosome 15. (A) *SMAD*3 is found on the long arm of chromosome 15 and contains 9 exons. SNPs denoted in red represent the index rs56062135 GWAS SNP and the functional rs17293632 candidate SNP. (B) Regional association plot for CAD association for *SMAD3* and proximal genes on chromosome 15 generated from 1000 Genomes project imputation data. (C) UCSC genome browser annotation and ENCODE project data for the entire *SMAD3* gene on chromosome 15. SNPs from previously published GWAS are denoted in green. Chromatin regulatory features such as DNase hypersensitivity sites, histone modifications (H3K4me1, H3K4me3, H3K27ac), and predicted transcription factor binding from ChIP-seq data are all included.

Table 2. SNPs in intron 1 of *SMAD3* in strong linkage disequilibrium (r^2>0.8) with rs56062135 and RegulomeDB scores for predicted functionally (1 = very likely to be functional, 6 = not likely to be functional). LD values were obtained from the HaploReg (v3)

SNP	Chromosome	Position (hg19)	D'	r^2	RegulomeDB Score
rs72743461	15	67441750	0.97	0.92	4
rs17293632	15	67442596	0.97	0.94	2a
rs56375023	15	67448363	0.98	0.95	No data
rs17228058	15	67450305	0.99	0.98	5
rs56062135	15	67455630	1.00	1.00	No data
rs72743477	15	67464291	0.98	0.93	5
rs72743482	15	67466599	0.97	0.91	5

This fine-mapping effort led us to conclude that the functional SNP(s) at the *SMAD3* locus are noncoding.

We next attempted to narrow down candidate functional/causal SNPs at this locus. We conducted epigenetic fine-mapping and scanned publically available databases to determine if any of these intronic SNPs lie within a noncoding gene regulatory element such as an enhancer or promoter[242–244]. To help localize truly functional variants amongst a large pool of SNPs, we looked up rs56062135C>T and linked SNPs in the RegulomeDB database. RegulomeDB contains experimental data sets from ENCODE and other sources in addition to computational predictions of regulatory potential[245]. RegulomeDB guides interpretation of regulatory variants and assigns each variant a score, 1 being very likely to be functional and 6 unlikely to be functional. This analysis revealed rs17293632C>T as the only *SMAD3* SNP to represent a suitable candidate functional/causal variant (**Table 2**). ENCODE and regulatory analysis demonstrated that rs17293632C>T is located in a gene regulatory hotspot (**Figure 1C**). HaploReg data indicates rs17293632C>T is located in a region with enhancer histone marks in 8 cell types, DNase hypersensitivity sites in 41 cells types, and alters 25 *in silico* transcription factor binding sites. ChIP-seq data from ENCODE reveal binding of dozens of transcription factors to the DNA region encompassing rs17293632C>T (**Supplementary Figure II**). The rare allele (T) (MAF= 0.21) of rs17293632C>T is protective ($p = 5.72 \times 10^{-9}$) (**Table 1**), comparable to the rs56062135C>T GWAS index SNP. In contrast, rs56062135 and rs56375023 had no available RegulomeDB data (**Table 2**), and ENCODE data did not suggest causal effects.

rs17293632 (T) Disrupts the Function of a Novel Enhancer in Intron 1 of *SMAD3*

We next sought to functionally verify the presence of an enhancer by investigating the region containing rs17293632C>T and 1 kb 5' and 3' flanking sequences (**Supplementary Figure III**). This 2 kb sequence contains several SNPs in addition to rs17293632C>T, as shown in **Figure 2A**. Pairwise linkage disequilibrium values between SNPs in this region are shown in **Supplementary Figure IV**. We cloned this 2 kb sequence for dual luciferase assays by PCR amplification from genomic DNA homozygous CC at rs17293632 and homozygous for all other SNPs. This sequence was inserted into the pGL3-Promoter vector (Promega) that contains an SV40 promoter upstream of the firefly luciferase gene and is used to evaluate putative enhancer sequences (**Figure 2B**). Since enhancer effects are commonly cell-type specific[246,247], experiments were performed in primary hASMCs of relevance to CAD. When the 2 kb sequence from *SMAD3* intron 1 was cloned into pGL3-Promoter with the C allele present for rs17293632, we observed a 10 fold increase in luciferase expression relative to no insert ($p<0.001$) (**Figure 2C**). When the rs17293632 allele (C) was changed to (T), via site-directed mutagenesis, the activity of this *SMAD3* enhancer was reduced by half ($p<0.001$). This indicated that in hASMCs, rs17293632(T) disrupts binding of one or more transcription factors or chromatin modifying enzymes. In HeLa and HepG2 cells, we found this 2 kb sequence to also act as a strong to moderate enhancer with an approximately four fold increase in luciferase activity compared to empty vector ($p<0.001$) (**Figure 2C**). Interestingly, the effects of this enhancer in HeLa and HepG2 cells are almost completely dependent on rs17293632C>T genotype, as the presence of the (T) allele nearly fully disrupts enhancer activity. To further examine the contribution of the

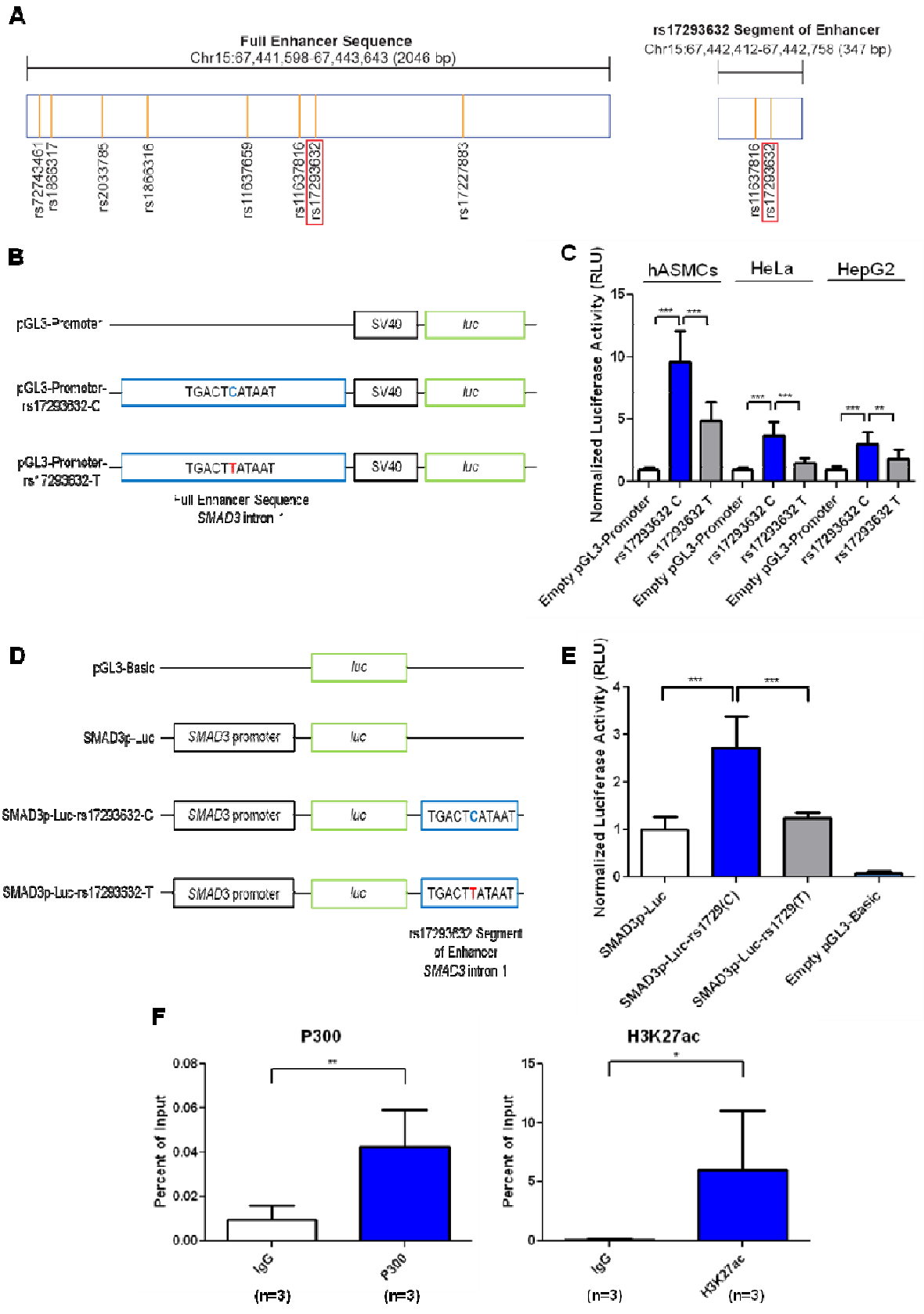

A
Full Enhancer Sequence
Chr15:67,441,598-67,443,643 (2046 bp)
rs17293632 Segment of Enhancer
Chr15:67,442,412-67,442,758 (347 bp)
rs72743461
rs1868317
rs2033785
rs1863316
rs11637659
rs11637816
rs17293632
rs17227883
B
pGL3-Promoter
SV40
luc
pGL3-Promoter-rs17293632-C
TGACTCATAAT
pGL3-Promoter-rs17293632-T
TGACTTATAAT
Full Enhancer Sequence
SMAD3 Intron 1
C
hASMCs
HeLa
HepG2
Normalized Luciferase Activity (RLU)
Empty pGL3-Promoter
rs17293632 C
rs17293632 T
D
pGL3-Basic
SMAD3p-Luc
SMAD3 promoter
SMAD3p-Luc-rs17293632-C
SMAD3p-Luc-rs17293632-T
rs17293632 Segment of Enhancer
SMAD3 Intron 1
E
Normalized Luciferase Activity (RLU)
SMAD3p-Luc
SMAD3p-Luc-rs1729(C)
SMAD3p-Luc-rs1729(T)
Empty pGL3-Basic
F
P300
H3K27ac
Percent of Input
IgG
(n=3)

Figure 2. rs17293632 resides within a strong enhancer in hASMCs and other cell types and is responsible for transcriptional activity of this enhancer. (A) Schematic of the different enhancer sequences containing rs17293632 investigated in dual luciferase assays. SNPs located within these enhancer sequences are denoted by their relative position. (B) Overview of the *SMAD3* enhancer reporter constructs with the SV40 promoter used for dual luciferase assays. These constructs contain the putative ~2 kb full *SMAD3* enhancer sequence (containing either the (C) or (T) allele at rs17293632) upstream of the SV40 promoter and firefly luciferase gene of the pGL3-Promoter vector that allows for identification of functional enhancer sequences. Empty pGL3-Promoter vector was employed as a negative control. (C) Firefly luciferase values for the SV40 promoter *SMAD3* enhancer reporter constructs transfected in HITC6 hASMCs, HeLa cells, and HepG2 cells. Firefly luciferase values were normalized to co-transfected *Renilla* luciferase. The hASMCs were harvested for dual luciferase assays 48 hours after transfection and HeLa and HepG2 cells harvested 24 hours after transfection. Plotted values represent mean ± SD of three independent experiments (triplicates each experiment). ***$p<0.001$, **$p<0.01$, ANOVA with Bonferroni's post-hoc test. (D) Overview of the *SMAD3* enhancer reporter constructs with the *SMAD3* promoter used for dual luciferase assays. 347 bp enhancer segments (rs17293632 allele either (C) or (T)) were inserted downstream of the *SMAD3* promoter and firefly luciferase gene in the SMAD3p-Luc plasmid. (E) Firefly luciferase values for the *SMAD3* promoter *SMAD3* enhancer reporter constructs that were transfected in HITC6 hASMCs and harvested 48 hours after transfection, with raw firefly luciferase values normalized using co-transfected *Renilla* luciferase. Plotted values represent mean ± SD of three independent experiments (triplicates each experiment). ***$p<0.001$, ANOVA with Bonferroni's post-hoc test. (F) Chromatin immunoprecipitation data from HITC6 hASMCs homozygous CC at rs17293632 shows enrichment of both p300 protein as well as the H3K27ac chromatin mark at the rs17293632 intronic sequence compared to non-specific IgG. Values are from three independent experiments and displayed as percentage of input DNA. Error bars denote SD. **$p<0.01$, *$p<0.05$, two-tailed unpaired Student's T-test.

enhancer in the *SMAD3* context, we cloned a 347 bp subset of this enhancer (**Figure 2A**) with both alleles for rs17293632 into the SMAD3p-Luc plasmid containing the human *SMAD3* promoter in the pGL3-Basic backbone[232] (**Figure 2D**). The enhancer was inserted 3' of both the *SMAD3* promoter sequence and firefly luciferase gene to mimic the orientation in the genome. As seen with the longer enhancer above, this 347 bp enhancer acts as an enhancer in the context of the *SMAD3* promoter in hASMCs and the rs17293632 (T) allele almost completely disrupts enhancer function ($p<0.001$) (**Figure 2E**).

Next, we sought to verify that the *SMAD3* intron 1 sequence containing rs17293632C>T acts as an active enhancer in hASMCs by performing ChIP with an H3K27ac antibody. Relative to non-specific IgG, H3K27ac was highly enriched at rs17293632 ($p<0.05$) and suggests this enhancer is functionally active in these cells (**Figure 2F**). The histone mark H3K27ac is typically deposited by the protein p300, a transcriptional co-activator commonly found at enhancers[248,249] that ENCODE predicted to strongly bind rs17293632 in other cell types. Indeed, binding of p300 ($p<0.01$) to the rs17293632 sequence was enriched in the hASMC system (**Figure 2F**), further supporting the presence of a functional enhancer.

SMAD3 rs17293632C>T is an eQTL in Whole Blood and Human Atherosclerotic Plaque Tissue

An expression quantitative trait locus (eQTL) effect provides strong evidence for a functional effect of a given SNP[250,251]. We first searched common publicly available online eQTL databases to determine if rs17293632C>T or LD SNPs are associated with transcript levels of *SMAD3* or another gene. We originally queried both the eQTL Browser from the

University of Chicago and the NCBI eQTL Browser, yet neither rs17293632C>T nor its linked SNPs appeared as eQTLs. Furthermore, we could not identify any reported eQTL SNP within the *SMAD3* gene in an electronic search. We then sought to determine whether rs17293632C>T acts as an eQTL for *SMAD3* expression using PAXgene® derived mRNA from whole blood of healthy individuals. Imputed rs17293632C>T genotypes from the Affymetrix 6.0 array were verified experimentally using a TaqMan SNP Genotyping Assay specific for rs17293632. Since a previous study reported sex specific effects on *SMAD3* expression[252], we selected an even number of male and female subjects across the three genotypes for rs17293632C>T and all were matched for age and body mass index. We detected a significant association between rs17293632 genotype and *SMAD3* mRNA levels in whole blood ($p<0.05$, one-way ANOVA) (**Figure 3A**).

We next queried whether rs17293632C>T associates with *SMAD3* expression in human atherosclerotic carotid plaque samples[234–237]. Microarray data from plaque tissue in the BiKE cohort, available for n=125 patients, was analyzed for association with rs17293632C>T. Since neither this SNP nor rs56062135C>T were present on the chip, rs16950687A>G (D' 0.949, r^2 0.859 with rs17293632; D' 1, r^2 0.953 with rs56062135) was used as a proxy SNP. As shown, rs16950687A>G was significantly associated with *SMAD3* mRNA levels in carotid lesions ($p<0.05$) (**Figure 3B**). We also examined association of rs17293632C>T with *SMAD3* expression in several different tissues from the Advanced Study of Aortic Pathology (ASAP)[238]. eQTL data from the ascending aorta adventitia was directionally consistent with the effects in blood and plaque tissue, but did not reach significance. Taken together these data provide support for rs17293632C>T as an eQTL for

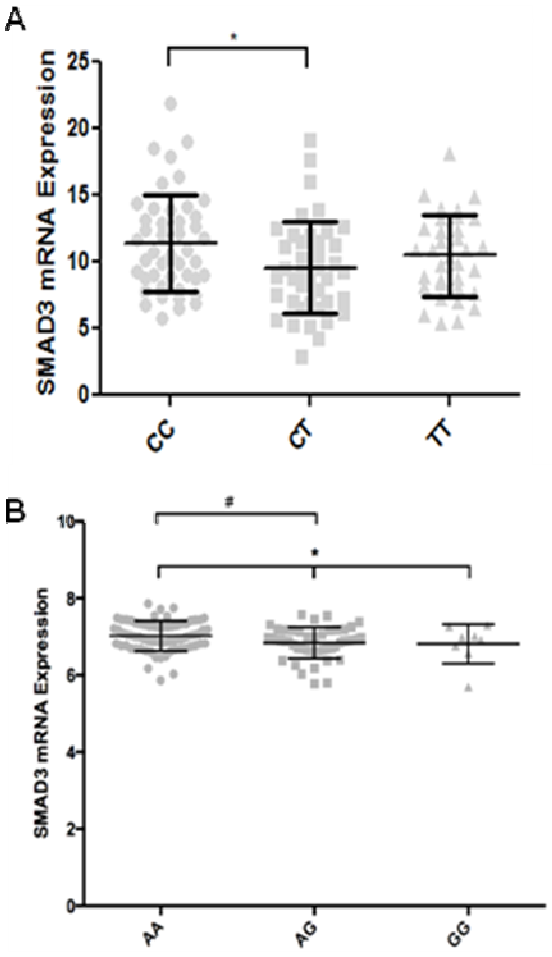

A
SMAD3 mRNA Expression
25
20
15
10
5
0
*
CC
CT
TT
B
SMAD3 mRNA Expression
10
8
6
4
2
0
#
*
AA
AG
GG

Figure 3. rs17293632 represents a *cis*-eQTL for *SMAD3* expression in human whole blood and carotid plaque tissue. (A) *SMAD3* mRNA levels in whole blood was measured across the three genotypes for rs17293632 by quantitative RT-PCR and normalized to expression levels of *PPIA*. The number of patients for the CC, CT, and TT genotypes were 50, 50, and 41, respectively. Whole blood RNA was extracted from PAXgene Blood RNA tubes taken from fasting healthy individuals matched for age and gender. Lines on the graph represent mean ± SD. *p<0.05, ANOVA with Bonferroni's post-hoc test. (B) Association of rs16950687, proxy SNP for rs17293632, with *SMAD3* mRNA expression in n=125 genotyped carotid plaque samples from the BiKE study. *p<0.05, one-way ANOVA, #p<0.05, T-test between the AA and AG genotypes.

SMAD3 mRNA expression in human whole blood and plaque tissue, consistent with an effect on gene regulation.

Enhancer Activity at Intron 1 of *SMAD3* is AP-1 Dependent

Although numerous transcription factors were predicted to bind the *SMAD3* intron 1 enhancer, the protective allele (T) of rs17293632 is predicted to disrupt the consensus binding sequence, 5'-TGA[G/C]TCA-3', for the transcription factor AP-1 (**Figure 4A**). To investigate the role of AP-1 at the *SMAD3* intron 1 enhancer, we first sought to determine the effects of inhibiting AP-1 signalling pharmacologically using SP600125, an inhibitor of the JNK pathway that blocks both the phosphorylation and transcriptional activity of c-Jun, a key player in AP-1 mediated transcription[253]. In hASMCs, blocking AP-1 activation via the JNK pathway was shown to strongly diminish enhancer activity of the full *SMAD3* intron 1 reporter containing the protective (C) allele at rs17293632 (**Figure 4B**). As predicted, SP600125 did not alter the enhancer activity of the construct containing rs17293632 (T), where the AP-1 site is already disrupted. Reduced c-Jun phosphorylation in response to SP600125 treatment was confirmed via Western blot (**Supplementary Figure V**). These data demonstrate that the AP-1 transcription factor is crucial for proper enhancer function at this specific location.

Next, we sought to determine if activating AP-1 increases endogenous SMAD3 mRNA levels by using the commonly used phorbol ester PMA (phorbol 12-myristate 13-acetate, also referred to as 12-*O*-Tetradecanoylphorbol-13-acetate/TPA). Over the course of 1 to 4 hours PMA stimulation, SMAD3 mRNA levels increase in two separate lots of hASMCs

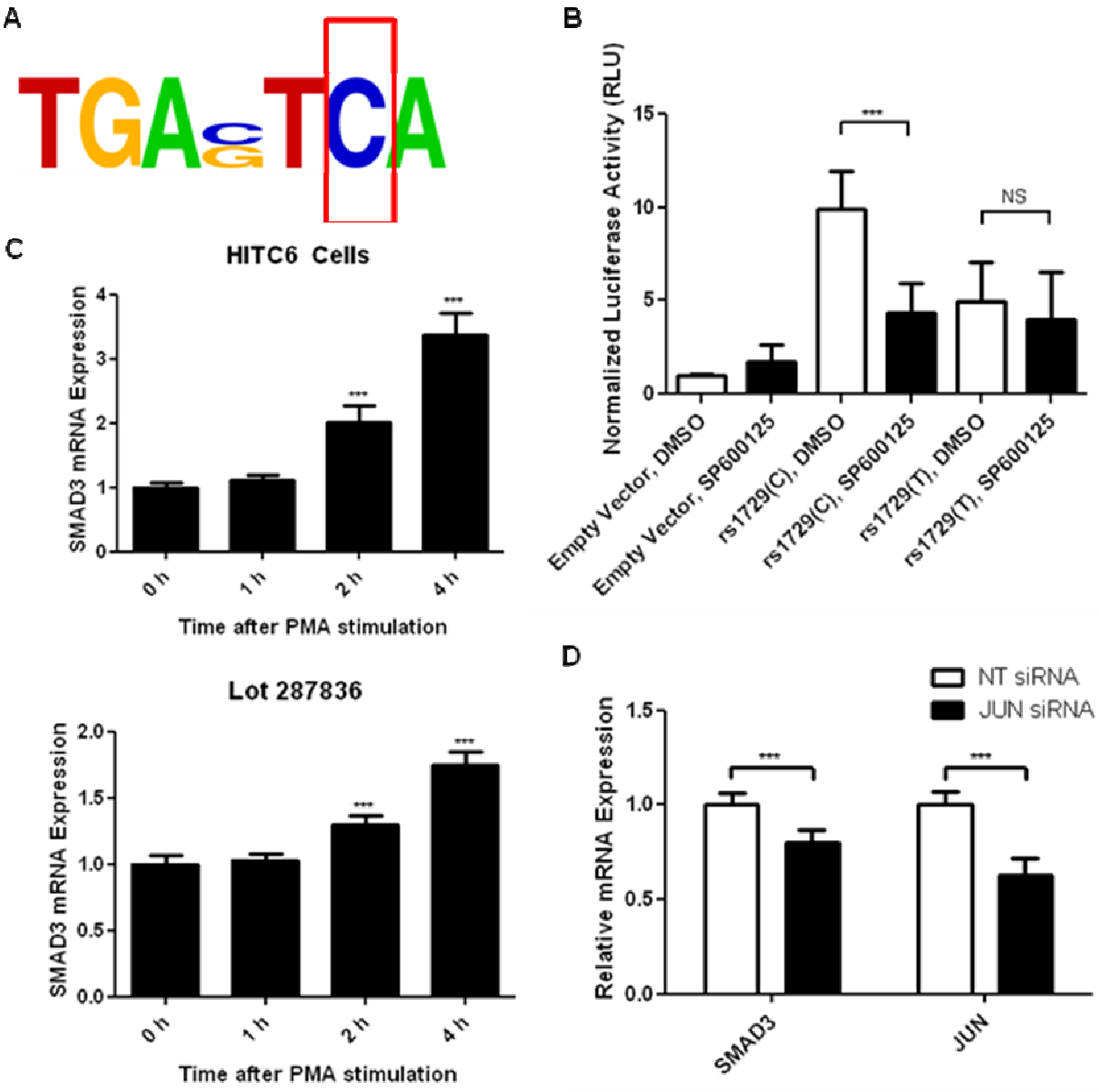

A
TGACGTCA
B
Normalized Luciferase Activity (RLU)
15
10
5
0
***
NS
Empty Vector, DMSO
Empty Vector, SP600125
rs1729(C), DMSO
rs1729(C), SP600125
rs1729(T), DMSO
rs1729(T), SP600125
C
HITC6 Cells
SMAD3 mRNA Expression
4
3
2
1
0
***
***
0 h
1 h
2 h
4 h
Time after PMA stimulation
Lot 287836
SMAD3 mRNA Expression
2.0
1.5
1.0
0.5
0.0
***
***
0 h
1 h
2 h
4 h
Time after PMA stimulation
D
NT siRNA
JUN siRNA
Relative mRNA Expression
1.5
1.0
0.5
0.0
***
***
SMAD3
JUN

Figure 4. AP-1 regulates enhancer activity at *SMAD3* intron 1 and regulates expression of SMAD3 mRNA. (A) Depiction of the consensus AP-1 binding site, 5'-TGA[G/C]TCA-3', with the location of the rs17293632 SNP highlighted in the red box. (B) HITC6 hASMCs were pretreated with 50 μM SP600125, which blocks phosphorylation of c-Jun, for 1 hour, prior to transfection with the full-length *SMAD3* enhancer reporter vectors or empty pGL3-Promoter vector for 24 hours. Values are plotted as mean ± SD of three independent experiments (triplicates each experiment). ***p<0.001, NS = not significant, ANOVA with Bonferroni's post-hoc test. (C) HITC6 and lot 287836 hASMCs were treated with a final concentration of 50 nM PMA and harvested for RNA at 0, 1, 2, and 4 hours. *SMAD3* mRNA levels were assessed by quantitative RT-PCR and normalized to expression levels of *β-actin*. Values represent mean ± SD of three independent experiments (triplicates each experiment). ***p<0.001 compared to 0 h, ANOVA with Bonferroni's post-hoc test. (D) HITC6 hASMCs were transfected with either 50 nM nontarget siRNA or 50 nM *JUN* siRNA for 48 hours. *SMAD3* and *JUN* mRNA levels were assessed by quantitative RT-PCR and normalized to expression levels of *β-actin.* Values represent mean ± SD of three independent experiments (triplicates each experiment). ***p<0.001 compared to NT siRNA, two- tailed unpaired Student's T-test.

(**Figure 4C**). Finally, knocking down c-Jun expression via siRNA decreases SMAD3 mRNA over the course of 48 hours ($p<0.001$), suggesting c-Jun and AP-1 normally increase SMAD3 transcription (**Figure 4D**). By contrast, knocking down other AP-1 components such as c-Fos, JunB, and JunD did not produce significant changes in SMAD3 mRNA (data not shown).

Allele-Specific Binding of AP-1 Components for rs17293632 in Primary Arterial Smooth Muscle Cells

ENCODE ChIP-seq data (Supplementary Figure II) indicate that AP-1 proteins c-Fos, FosL1, c-Jun, JunB, and JunD are able to bind the rs17293632 site in the *SMAD3* intron but with little data on vascular cell types. We thus performed ChIP experiments in HITC6 hASMCs homozygous for the C allele. Protein-DNA complexes were pulled down using antibodies specific for c-Fos, c-Jun, JunB, and JunD as well as a nonspecific IgG control. We demonstrate very strong binding of c-Fos and JunD to rs17293632(C) as well as moderate binding of JunB and c-Jun (**Figure 5A**). AP-1 ChIP products were also amplified for negative control intronic genomic regions (**Supplementary Figure VI**). Relative to the input DNA, we observed much weaker binding of AP-1 proteins to the negative control genomic areas (intronic sequences selected that are not predicted to bind AP-1).

We next sought to determine if changing rs17293632 (C) to (T) affects binding of AP-1 proteins by performing allele-specific ChIP-qPCR, using lot 287836 primary hASMCs that were heterozygous for rs17293632. To assess the allele-specific binding of AP-1 we first performed ChIP and followed with a nested PCR-TaqMan genotyping method for

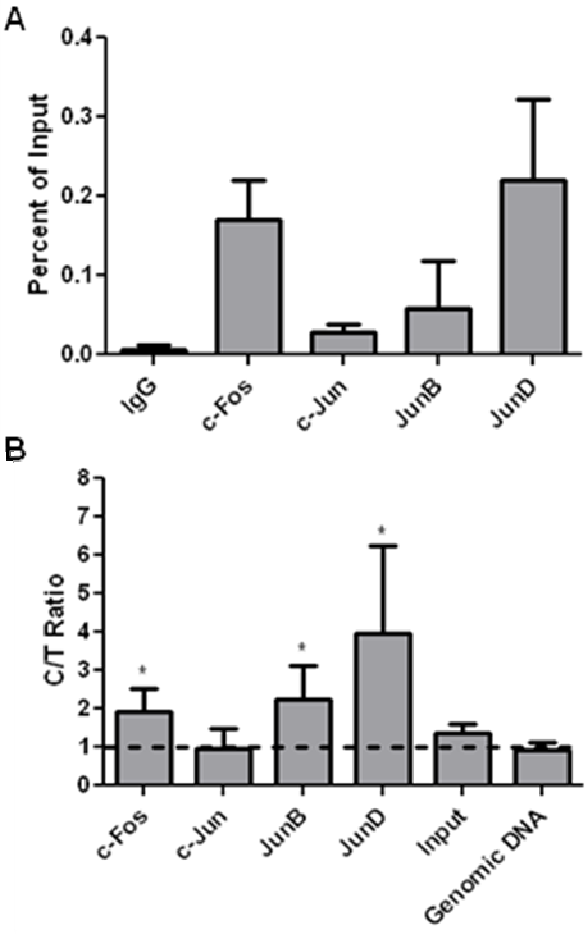

A
Percent of Input
0.4
0.3
0.2
0.1
0.0
IgG
c-Fos
c-Jun
JunB
JunD
B
C/T Ratio
8
7
6
5
4
3
2
1
0
c-Fos
c-Jun
JunB
JunD
Input
Genomic DNA

Figure 5. AP-1 proteins bind to the rs17293632 sequence and preferentially bind the common (C) allele. (A) Chromatin immunoprecipitation data from HITC6 hASMCs shows enrichment of AP-1 component proteins c-Fos, c-Jun, JunB, and JunD at the intronic enhancer comprising rs17293632. Values on the y-axis are expressed as percentage of the input chromatin DNA and are representative of three independent experiments. Error bars denote SD. (B) Allele-specific ChIP was conducted in lot 287836 hASMCs, heterozygous at rs17293632. Eluted ChIP DNA values are expressed as the ratio of the rs17293632 (C) to (T) allele and are representative of three independent ChIP/PCR/TaqMan experiments. *$p<0.05$, T-test compared to genomic DNA TaqMan control. Error bars denote SD.

quantification. Protein/DNA complexes were immunoprecipitated with c-Fos, c-Jun, JunB, and JunD antibodies. After PCR amplification of the eluted ChIP DNA, TaqMan genotyping was performed to determine the ratios of the (C) to (T) alleles for rs17293632 in the amplified ChIP products. These TaqMan values were then applied to a standard curve consisting of different ratios of the two alleles (generated from homozygous genotypes). We find strong enrichment of (C) allele binding for both c-Fos and JunD ($p<0.05$) (**Figure 5B**), both AP-1 members that bound the strongest to this region as shown in **Figure 5A**. JunB also demonstrated enrichment for the C allele ($p<0.05$), whereas c-Jun showed, surprisingly, no enrichment (see Discussion). As expected hASMC lot 287836 input and genomic DNA control samples had C:T ratios around 1.0. Overall, these allele-specific ChIP data demonstrate enriched binding of AP-1 to the common (C) vs. protective (T) allele, consistent with the above luciferase and eQTL data. Altered binding of AP-1 would lower the efficacy of the *SMAD3* intron 1 enhancer and thus reduce *SMAD3* transcription and mRNA levels.

SMAD3 Knockdown Increases Human Arterial Smooth Muscle Cell Proliferation

SMAD3 knockdown in normal smooth muscle cell medium resulted in a 15-25% increase in viable cells ($p<0.01$) (**Figure 6**) with very similar results for both lots of hASMCs. These findings indicate that SMAD3 has a negative effect on viability in hASMCs, which likely reflects an effect on cell proliferation. Since previous studies have shown that TGFβ induces apoptosis via a SMAD-dependent mechanism[254,255], we determined protein levels of PARP (Poly (ADP-ribose) polymerase) a DNA repair protein that is a key caspase target.

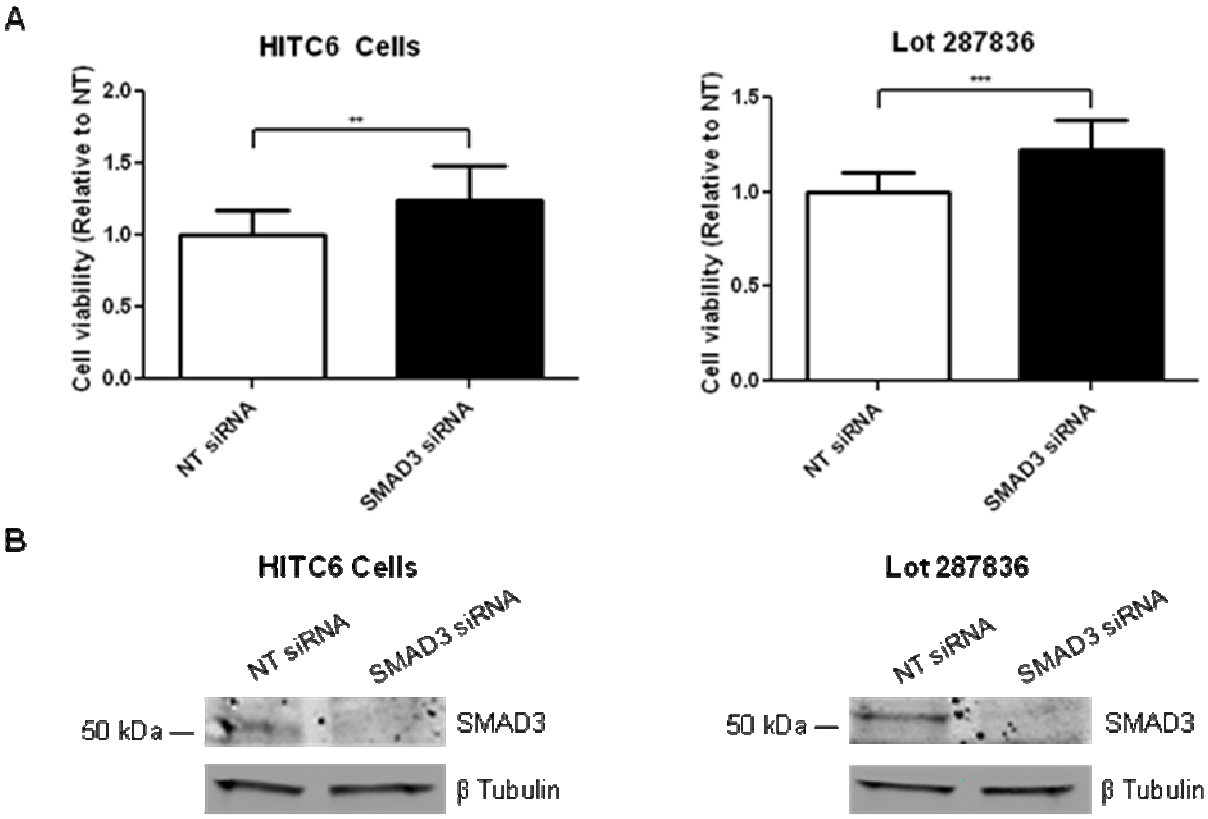
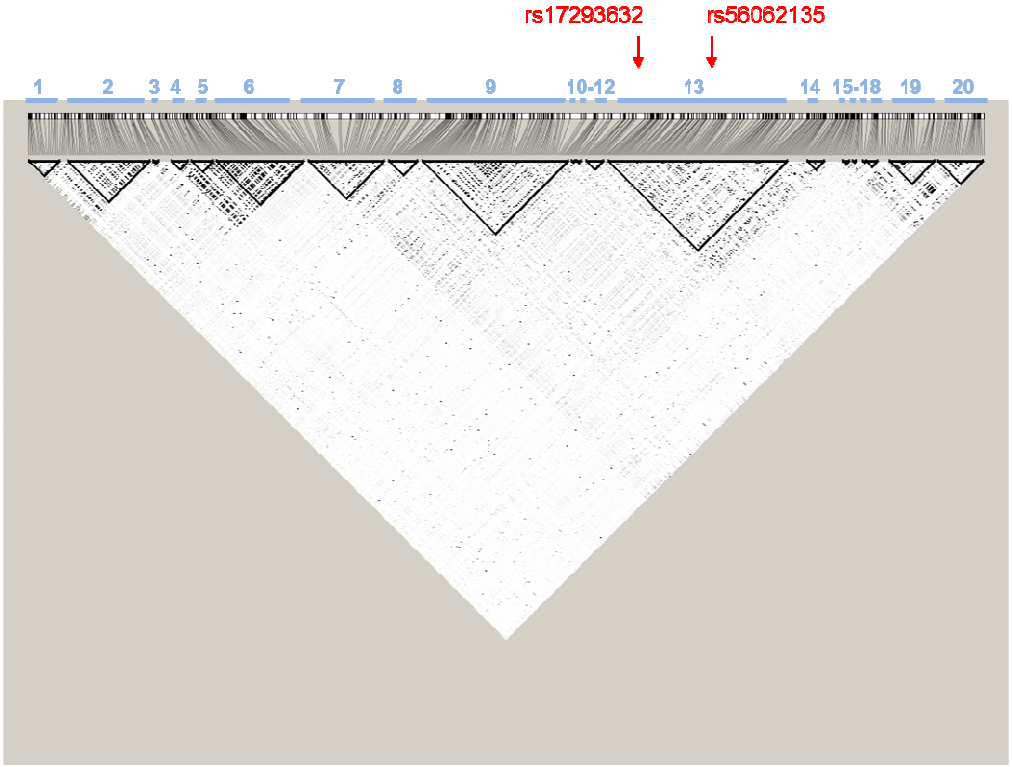

A
HITC6 Cells
Cell viability (Relative to NT)
2.0
1.5
1.0
0.5
0.0
**
NT siRNA
SMAD3 siRNA
Lot 287836
***
B
HITC6 Cells
NT siRNA
SMAD3 siRNA
50 kDa —
SMAD3
β Tubulin
Lot 287836

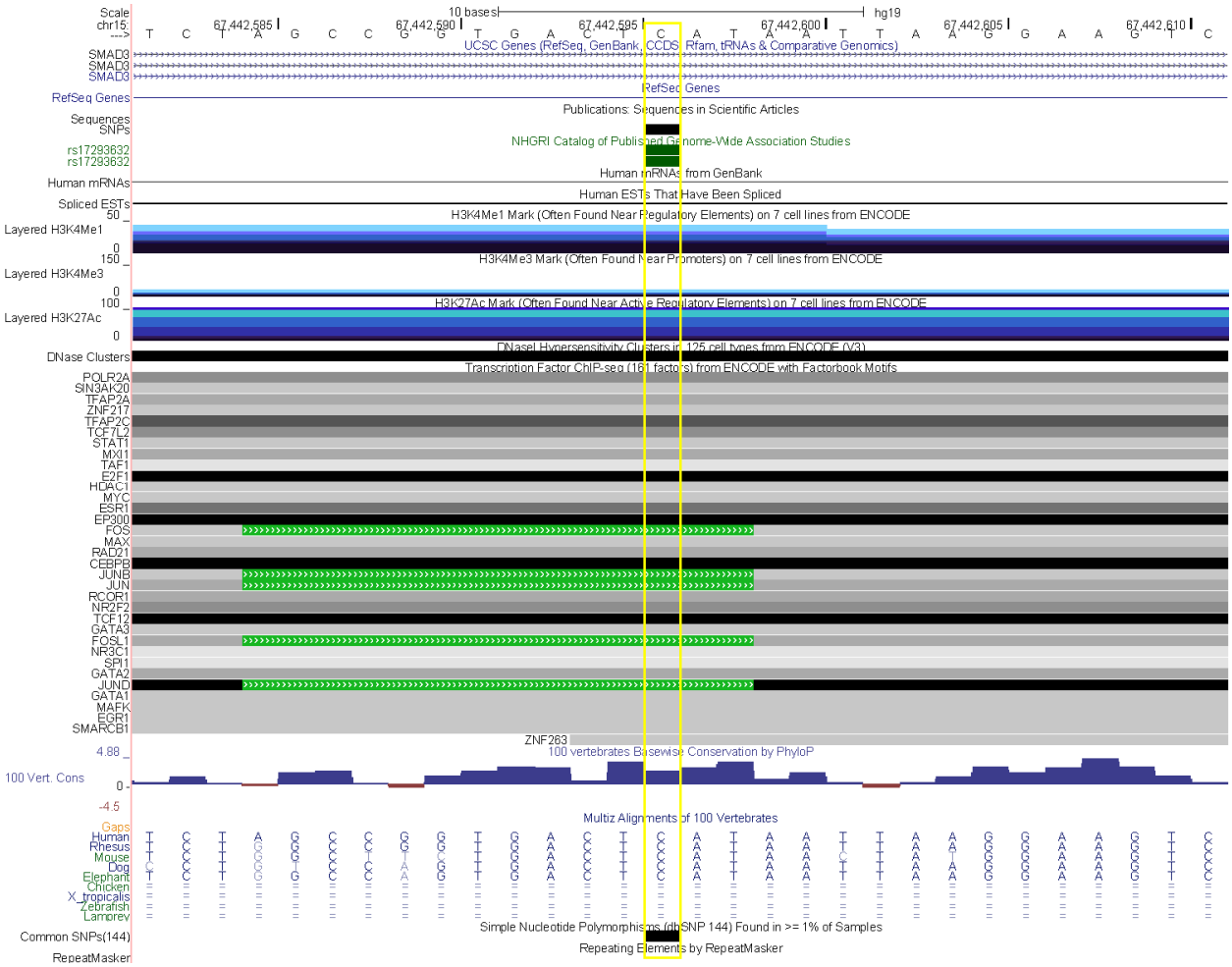
Figure 6. *SMAD3* knockdown in two separate lots of primary human arterial SMCs increases cell proliferation/viability. (A) HITC6 and lot 287836 cells were transfected with either 20 nM nontarget siRNA or 20 nM *SMAD3* siRNA and harvested for the MTT assay 72 h later to evaluate cell proliferation/cell viability. Values displayed represent the means ± SD of three independent experiments. ***$p<0.001$, **$p<0.01$, ANOVA with Bonferroni's post-hoc test. (B) *SMAD3* knockdown at the protein level was confirmed via Western blot, with human β-tubulin used as the loading control.

Caspase-mediated cleavage of PARP is a common marker of apoptosis, detectable by Western blotting as the conversion of the 116 kDa full-length protein to its 89 kDa form[256]. We could not detect the characteristic 89 kDa form of PARP in nontarget siRNA treated hASMCs, and observed no difference in this marker of apoptosis between nontarget and *SMAD3* knockdown samples (**Supplementary Figure VII**). Furthermore, the microscopic appearance of hASMCs did not differ between nontarget siRNA and *SMAD3* siRNA treatment, supportive of an increase in human ASMC proliferation in response to *SMAD3* siRNA knockdown (**Figure 6**). Thus these data indicate that SMAD3 knockdown is not repressing a putative underlying apoptosis but rather suggest that SMAD3 regulates cell proliferation via other means.

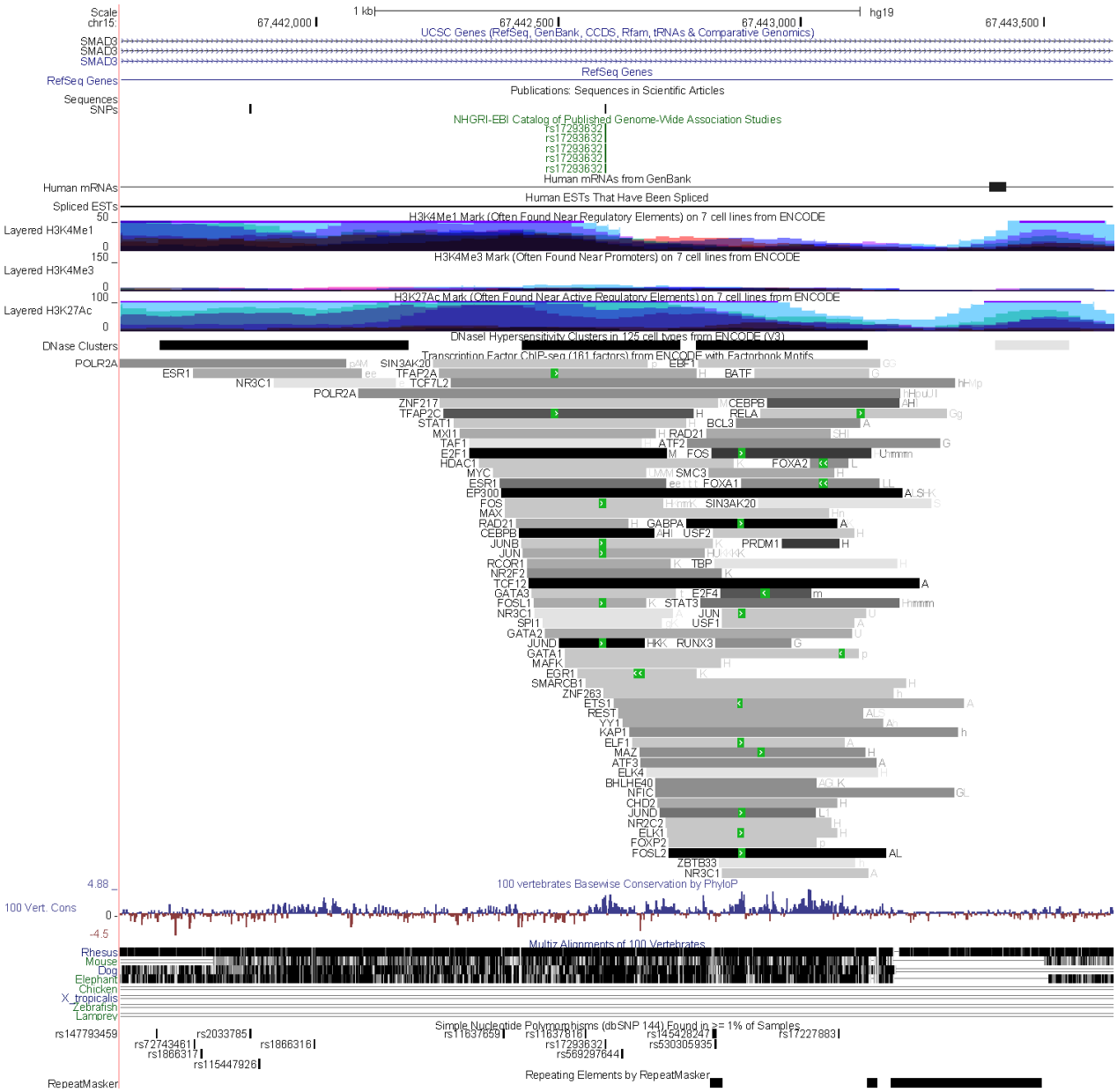
Online Only Supplemental Figures and Tables



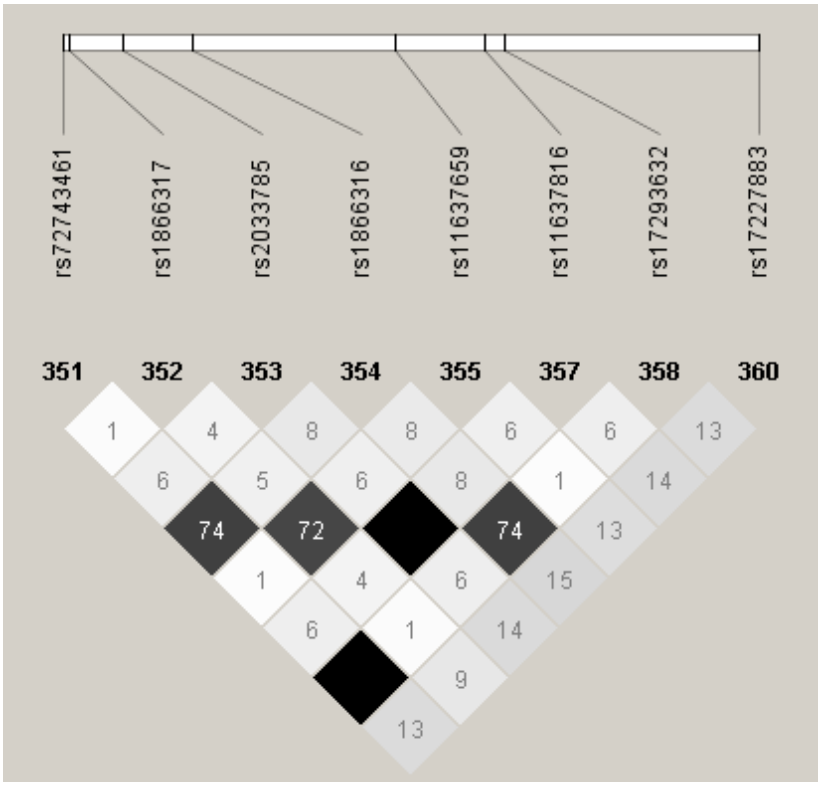
Supplementary Figure I. Visualization of the various LD blocks in the *SMAD3* gene using the Haploview software and linkage data taken from the 1000 Genomes Browser.


Scale
chr15
10 bases
hg19
67,442,585
67,442,590
67,442,595
67,442,600
67,442,605
67,442,610
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SMAD3
RefSeq Genes
Sequences
SNPs
Publications: Sequences in Scientific Articles
NHGRI Catalog of Published Genome-Wide Association Studies
rs17293632
Human mRNAs
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me1
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
Layered H3K4Me3
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
POLR2A
SIN3AK20
TFAP2A
ZNF217
TFAP2C
TCF7L2
STAT1
MXI1
TAF1
E2F1
HDAC1
MYC
ESR1
EP300
FOS
MAX
RAD21
CEBPB
JUNB
JUN
RCOR1
NR2F2
TCF12
GATA3
FOSL1
NR3C1
SPI1
GATA2
JUND
GATA1
MAFK
EGR1
SMARCB1
ZNF263
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
Multiz Alignments of 100 Vertebrates
Gaps
Human
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Simple Nucleotide Polymorphisms (dbSNP 144) Found in >= 1% of Samples
Common SNPs(144)
Repeating Elements by RepeatMasker
RepeatMasker

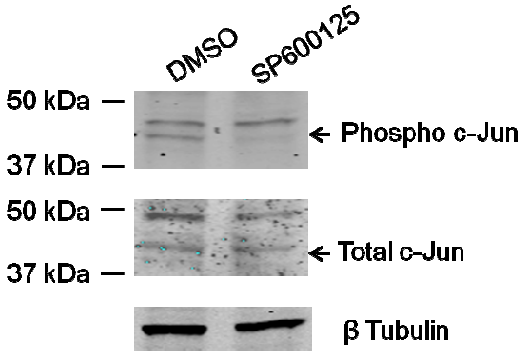
Supplementary Figure II. UCSC genome browser annotation and ENCODE project data for the regional view of the rs17293632 SNP.


Scale
chr15:
1 kb
hg19
67,442,000
67,442,500
67,443,000
67,443,500
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SMAD3
RefSeq Genes
Publications: Sequences in Scientific Articles
Sequences
SNPs
NHGRI-EBI Catalog of Published Genome-Wide Association Studies
rs17293632
Human mRNAs
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
Layered H3K4Me1
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me3
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
Layered H3K27Ac
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
POLR2A
ESR1
NR3C1
TCF7L2
ZNF217
TFAP2C
STAT1
MXI1
TAF1
E2F1
HDAC1
MYC
EP300
FOS
MAX
RAD21
CEBPB
JUNB
JUN
RCOR1
NR2F2
TCF12
GATA3
FOSL1
SPI1
GATA2
JUND
GATA1
MAFK
SMARCB1
ZNF263
ETS1
REST
YY1
KAP1
ELF1
MAZ
ATF3
ELK4
BHLHE40
NFIC
CHD2
NR2C2
FOXP2
ZBTB33
SIN3AK20
EBF1
BATF
CEBPB
RELA
BCL3
RAD21
ATF2
FOS
FOXA2
SMC3
FOXA1
SIN3AK20
GABPA
USF2
PRDM1
TBP
E2F4
STAT3
JUN
USF1
RUNX3
FOSL2
NR3C1
4.88
100 Vert. Cons
0
-4.5
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Simple Nucleotide Polymorphisms (dbSNP 144) Found in >= 1% of Samples
rs147793459
rs2033785
rs727434461
rs1866316
rs1866317
rs115447926
rs11637659
rs11637816
rs17293632
rs5692976441
rs145428247
rs5303059351
rs17227883
RepeatMasker
Repeating Elements by RepeatMasker

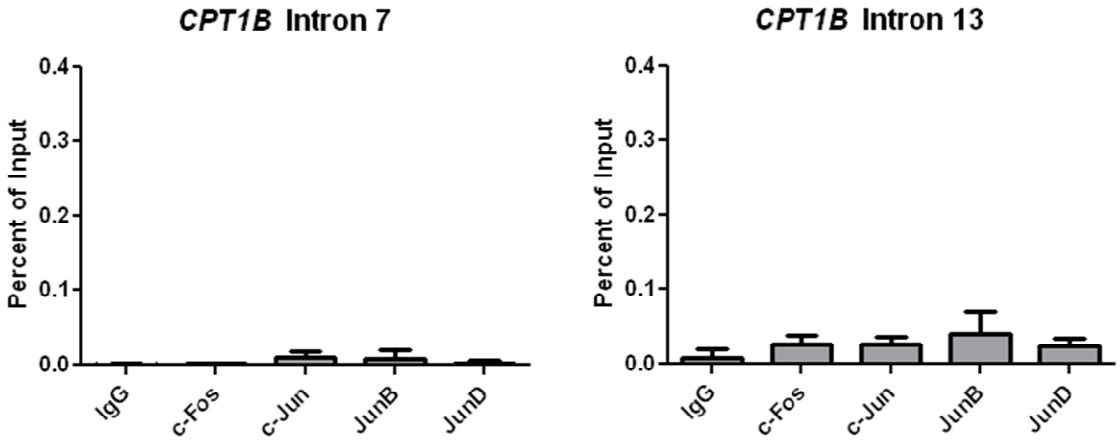
Supplementary Figure III. UCSC genome browser annotation and ENCODE project data for the entire full *SMAD3* intron 1 enhancer sequence characterized.


rs72743461
rs1866317
rs2033785
rs1866316
rs11637659
rs11637816
rs17293632
rs17227883
351
352
353
354
355
357
358
360
1
4
8
8
6
6
13
6
5
6
8
1
14
74
72
74
13
1
4
6
15
6
1
14
9
13

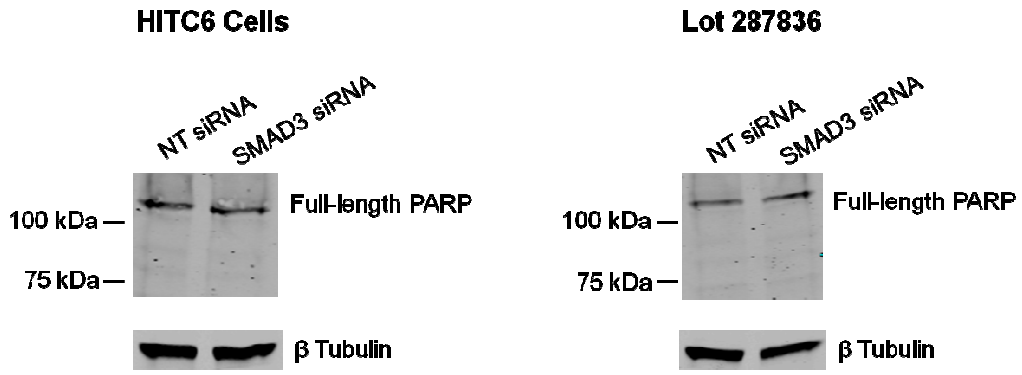
Supplementary Figure IV. LD heatmap generated using the Haploview software and linkage data taken from the 1000 Genomes Browser displaying pairwise linkage disequilibrium for SNPs located within the *SMAD3* intron 1 enhancer. Numbers within the blocks represent pairwise r^2 values multiplied by 100. White to grey to black blocks represent increasing r^2 values.


DMSO
SP600125
50 kDa
37 kDa
Phospho c-Jun
50 kDa
37 kDa
Total c-Jun
β Tubulin

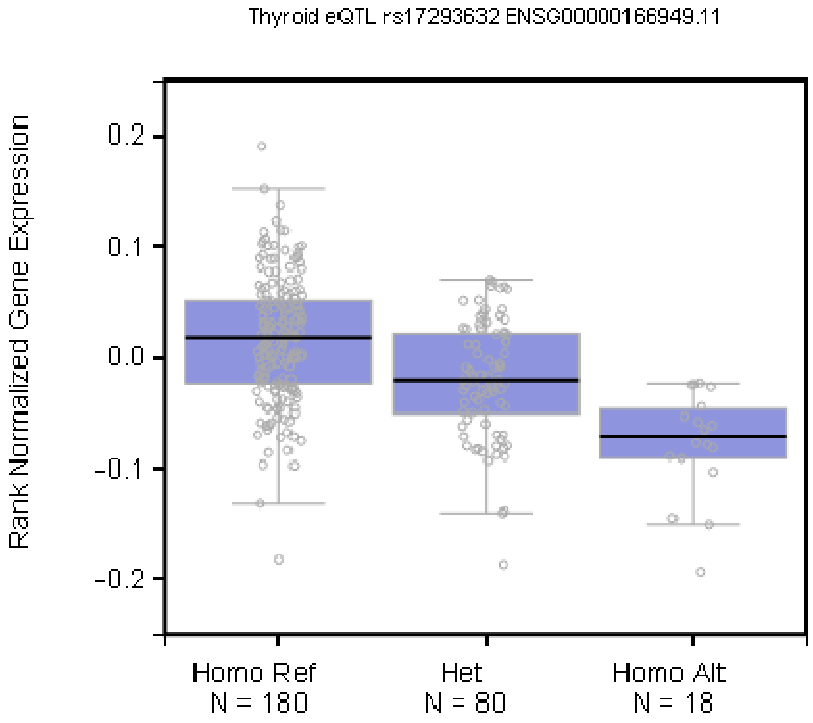
Supplementary Figure V. Reduced levels of phospho c-Jun in response to SP600125 were confirmed via Western blotting, with human β-tubulin used as the loading control. HITC6 hASMCs were treated with either 50 μM SP600125 or DMSO control for 24 hours prior to harvesting.


CPT1B Intron 7
Percent of Input
0.4
0.3
0.2
0.1
0.0
IgG
c-Fos
c-Jun
JunB
JunD
CPT1B Intron 13
Percent of Input
0.4
0.3
0.2
0.1
0.0
IgG
c-Fos
c-Jun
JunB
JunD

Supplementary Figure VI. Negative control DNA regions amplified for AP-1 chromatin immunoprecipitation. Crosslinked DNA-protein complexes from HITC6 cells were immunoprecipitated with the antibodies for AP-1 components c-Fos, c-Jun, JunB, and JunD. The ChIP DNA was then amplified with primers specific for two regions of the *CPT1B* gene (intron 7 and intron 13). Values are expressed as the percentage of input chromatin DNA. Values denote the mean of three independent ChIP experiments ± SD.


HITC6 Cells
NT siRNA
SMAD3 siRNA
100 kDa
75 kDa
Full-length PARP
β Tubulin
Lot 287836
NT siRNA
SMAD3 siRNA
100 kDa
75 kDa
Full-length PARP
β Tubulin

Supplementary Figure VII. *SMAD3* siRNA knockdown does not affect cleavage of PARP in both lots of hASMCs. HITC6 and lot 287836 cells were transfected with either 20 nM nontarget siRNA or 20 nM *SMAD3* siRNA and harvested for protein 72 hours later. Lysates were probed with PARP antibody that recognizes both the full length, 116 kDa form of the protein, as well as the 89 kDa form that serves as a marker for cells undergoing apoptosis. Human β-tubulin was used as the loading control for both lots of cells.


Thyroid eQTL rs17293632 ENSG00000166949.11
Rank Normalized Gene Expression
0.2
0.1
0.0
-0.1
-0.2
Homo Ref
N = 180
Het
N = 80
Homo Alt
N = 18

Supplementary Figure VIII. eQTL box plot for rs17293632 and SMAD3 mRNA expression in thyroid tissue from the GTEx eQTL Browser. ENSG00000166949.11 refers to SMAD3 levels.

Supplementary Table I. Characteristics of the three groups of healthy individuals selected for rs17293632 eQTL analysis

Cohort	N Male	N Female	Mean Age (± SD)	Mean BMI (± SD)
CC	25	25	45.68 (15.78)	21.00 (1.70)
CT	25	25	43.68 (14.30)	20.20 (1.98)
TT	19	22	44.65 (15.82)	20.60 (1.80)

Supplementary Table II. Sequences of the negative control intronic DNA regions used for AP-1 chromatin immunoprecipitation

Region	Amplicon Size	Forward primer	Reverse primer
CPT1B Intron 7	123	GGAAGAGGGAGACAGAACACC	GCACAAGGTCCTACAGAGGAA
CPT1B Intron 13	101	TCAGCACTTCCCTCTTACCC	GGCCAAGACATCTTCAGGAG

Supplementary Table III. Conditional and joint association analysis results for *SMAD3* locus

		CARDIoGRAM Meta-analysis			Conditional and joint analysis		
Variant	**Effect allele**	**Beta**	**SE**	**p-value**	**Beta**	**SE**	**p-value**
rs56062135	C	0.07	0.01	4.52E-09	0.07	0.01	4.54E-09

Supplementary Table IV. SNPs within the 2 kb enhancer in intron 1 (chr15:67,441,598-67,443,643) of the *SMAD3* gene that are in linkage disequilibrium with published GWAS SNPs

Reported GWAS SNP (MAF)	Disease or Trait	Reference	Enhancer SNP (MAF)	D'	r^2
rs56062135 (0.21)	Coronary artery disease	Nikpay et al. *Nat Genet* 2015	rs72743461 (0.22)	0.97	0.92
			rs1866316 (0.28)	0.97	0.66
			rs17293632 (0.21)	0.97	0.94
rs17228212 (0.29)	Coronary artery disease	Samani et al. *N Engl J Med* 2007	rs17227883 (0.29)	0.98	0.96
rs12441344 (0.22)	Risk of sudden cardiac arrest in patients with coronary artery disease	Tseng et al. *Heart Rhythm* 2009	rs2033785 (0.23)	0.98	0.90
			rs11637816 (0.23)	0.98	0.90
rs17293632 (0.21)	Inflammatory bowel disease	Jostins et al. *Nature* 2012	rs72743461 (0.22)	1	1
			rs1866316 (0.28)	1	0.70
			rs17293632 (0.21)	1	1
rs17293632 (0.21)	Crohn's disease	Franke et al. *Nat Genet* 2010	rs72743461 (0.22)	1	1
			rs1866316 (0.28)	1	0.70
			rs17293632 (0.21)	1	1
rs17228058 (0.21)	Allergy susceptibility	Hinds et al. *Nat Genet 2013*	rs72743461 (0.22)	0.96	0.92
			rs1866316 (0.28)	0.96	0.65
			rs17293632 (0.21)	0.97	0.93
rs744910 (0.52)	Asthma	Moffat *N Engl J Med* 2010	rs72743461 (0.22)	-0.98	0.29
			rs2033785 (0.23)	0.99	0.27
			rs1866316 (0.28)	-0.97	0.40
			rs11637659 (0.80)	1	0.27
			rs11637816 (0.23)	0.99	0.27
			rs17293632 (0.21)	-0.98	0.28
			rs17227883 (0.29)	0.96	0.35
rs12913547 (0.24)	Central corneal thickness and keratoconus	Lu et al. *Nat Genet* 2013	rs2033785 (0.23)	0.84	0.70
			rs11637816 (0.23)	0.84	0.70

rs17294280 (0.24)	Asthma and hay fever	Ferreira et al. *J Allergy Clin Immunol* 2014	rs72743461 (0.22)	0.88	0.66
			rs1866316 (0.28)	0.74	0.46
			rs17293732 (0.21)	0.68	0.89

Supplementary Table V.

Too large to include in thesis document - available online.

3.5 Discussion

Despite the success of recent genome-wide association studies, there has been limited progress in understanding the function of the multiple risk loci identified. Multiple hurdles exist including the fact that most are located in intergenic regions or introns rather than coding sequences. Secondly, the identified polymorphism is most often not causal but rather in linkage disequilibrium with a neighboring or even distal causal polymorphism. Thirdly, while the association is at the level of genomic DNA, the relevance may be restricted to a particular tissue or organ. Lastly, the direct contribution of any given locus could be temporally restricted; for instance, early transient activation followed by long-standing enduring epigenetic repercussions.

Here we have investigated a recently discovered GWAS signal near rs56062135C>T in *SMAD3*[44] that is not linked to a previously reported GWAS signal at rs17228212C>T[38]. By genetic and epigenetic fine mapping, we identified a SNP, rs17293632C>T, in almost perfect LD with rs56062135C>T, in a region with chromatin histone marks suggestive of enhancer activity. Functional studies confirmed that rs17293632C>T is located within a strong enhancer sequence in primary hASMCs, HeLa and HepG2 cells. The common (C) allele maintains the conserved AP-1 binding site and results in greater AP-1 enhancer binding and maximal enhancer activity, whereas the protective (T) allele (MAF 0.21) disrupts the AP-1 recognition sequence, reduces AP-1 binding and impairs enhancer activity. We propose that this enhancer is active in numerous cell types and that rs17293632(T) disrupts enhancer function via a universal mechanism. Stimulation with the AP-1 activator PMA increases endogenous SMAD3 mRNA levels in hASMCs while c-Jun

siRNA knockdown decreases SMAD3 mRNA, suggesting AP-1 positively modulates SMAD3 transcription.

While AP-1 regulation is complex and its transcriptional activity can be stimulated by numerous factors, we observed that the presence of 5% serum was sufficient to produce an enhancer effect at *SMAD3* intron 1 likely due to the presence of various growth factors and cytokines. ChIP data (**Figure 5A**) demonstrate highly enriched binding in HITC6 cells of the AP-1 components c-Fos and JunD, with lesser but still enriched binding of c-Jun and JunB compared to nonspecific IgG. This suggests that in the presence of 5% FBS, AP-1 is able to sufficiently bind the rs17293632(C) sequence. We were surprised not to find more enriched binding of c-Jun, a central component of all AP-1 complexes[184,189], to the *SMAD3* intron 1 enhancer, especially since c-Jun siRNA knockdown lowers SMAD3 mRNA.

Although c-Jun is important in the transcriptional activity of AP-1, it is possible that our system did not have the proper stimulatory context to maximize c-Jun binding to DNA. Furthermore, c-Jun expression is low in many cell types and its expression requires elevation by appropriate stimuli[189]. The c-Jun antibody used may also not have been optimal for ChIP experiments. In addition, the epitope recognized by the c-Jun antibody may not be accessible in our ChIP material due the potential of the formaldehyde cross-linking of protein to DNA masking effect. By contrast, c-Fos and JunD bound strongly to the rs17293632 sequence in an allele-dependent fashion, yet their siRNA knockdowns did not affect SMAD3 expression. Potential reasons for this are that the c-Fos and JunD proteins are less potent transcription activators compared to c-Jun or that there is functional redundancy between some AP-1 family members[186,257].

Here we also demonstrate ChIP enriched binding of p300 at the rs17293632 enhancer. p300 is a transcriptional co-activator by acting as a histone acetyltransferase that is able to relax chromatin structure and facilitate binding of other transcription factors and can also act as an adaptor protein, suggesting that p300 may help the rs17293632 *SMAD3* intron 1 enhancer loop around and interact with/activate the *SMAD3* promoter area that is located in the 5' direction. Mechanisms of p300 recruitment to this region, roles of other transcription factors from Supplementary Figure II ChIP-seq data, and effect of rs17293632 on chromatin modifications warrant future investigation.

Luciferase assays consistently confirmed an effect of rs17293632C>T on enhancer activity. However, our data with the 2 kb enhancer sequence indicated changing the allele from (C) to (T) reduced reporter activity significantly but not completely relative to the empty pGL3-Promoter. Similarly, treatment with SP600125 reduced activity of the *SMAD3* reporter with the rs17293632 C allele, to the same levels as the *SMAD3* reporter with the T allele but both had greater activity in comparison to empty pGL3-Promoter, and treatment with SP600125 had no effect when the AP-1 site was disrupted. Thus the 2 kb *SMAD3* intron 1 sequence investigated still has some enhancer activity in hASMCs independent of the rs17293632 SNP studied here. As indicated in **Supplementary Figure III**, other transcription factors are predicted to bind a few hundred base pairs away from the cluster of transcription factors that bind to the immediate rs17293632 SNP. ENCODE data also demonstrate H3K4me1 and H3K27ac enhancer histone marks across the entire 2 kb intronic sequence studied. Nevertheless, our results indicate that rs17293632 contributes to a large fraction, perhaps the majority, of this intronic enhancer regulation.

During the course of these experiments, another report highlighted rs17293632C>T as a strong candidate functional SNP. Farh *et al*.[258] developed a fine-mapping algorithm to identify candidate causal variants for numerous autoimmune diseases and provide ChIP-seq data for AP-1 in HeLa cells heterozygous at rs17293632. They observed robust binding of AP-1 to the rs17293632 (C) allele but not to the sequence with the (T) allele (31 AP-1 reads for the C allele vs. 1 read for the T allele). Their findings are directionally similar to our allele-specific TaqMan experiments with AP-1 (**Figure 5B**). However, their study did not provide further experimental evidence with respect to rs17293632 function. Although rs17293632 was used as an example of a likely causal SNP, the specific AP-1 protein(s) investigated in their ChIP-seq experiment was not reported.

Further evidence for rs17293632C>T as a causal SNP at this locus is provided by eQTL data for *SMAD3* expression in both whole blood and vascular tissue. We elected to focus solely on *SMAD3* for eQTL analysis because rs17293632 is located within *SMAD3* and no other genes in the region had plausible links to atherosclerosis. Although we demonstrate that rs17293632 is a *cis*-eQTL for *SMAD3* expression, we cannot rule out *trans*-eQTL effects. There are several genes in the vicinity of *SMAD3*/rs17293632 that may also be regulated by rs17293632 and future studies in this regard will be of interest. Although p-values in eQTL analysis were nominally significant at $p<0.05$, they would not pass the multiple correction threshold if we probed for many genes in addition to *SMAD3*. Recent studies have shown some eQTLs are shared amongst tissues[259,260], while others are tissue specific[260,261]. While we show eQTL data in whole blood and vascular tissue, it will be important to determine whether rs17293632 is an eQTL in various other tissues since this SNP has pleiotropic effects on other diseases (**Supplementary Table IV**). In the midst of our experiments,

rs17293632 was deposited in the GTEx eQTL database as an eQTL for SMAD3 in thyroid with the same direction of effect ($p=1.9 \times 10^{-13}$, **Supplementary Figure VIII**)

While these fine-mapping, bioinformatic, and experimental findings demonstrate that rs17293632C>T is a functional SNP at this locus, others may exist. Table 2 shows that this SNP is in almost perfect linkage disequilibrium with six other SNPs at the *SMAD3* locus. While the latter do not appear to significantly disrupt transcription factor binding sites or have strong regulatory chromatin modifications compared to rs17293632C>T, it is plausible that they could still affect gene regulation. As indicated in **Figure 1C**, the *SMAD3* locus as a whole appears to contain many gene regulatory hotspots, with several regions containing strong H3K27ac, H3K4me1, and clusters of transcription factor binding in ENCODE ChIP-seq data.

Other reports indicate the enhancer sequence in *SMAD3* encompassing rs17293632C>T has relevance to phenotypes other than atherosclerosis. Of note, a cluster of SNPs in vicinity of rs17293632 and towards the 3' end of the *SMAD3* gene (**Figure 1C**) have previously been associated with various diseases (**Supplementary Table IV**), most with immune components in their pathophysiology. These include Crohn's disease, inflammatory bowel disease, asthma and atopy[262–266]. As we have demonstrated, the *SMAD3* intron 1 enhancer is functional across numerous cell types and thus effects of rs17293632C>T are likely to apply across a wide variety of cell types expressing AP-1 family member proteins. Affecting enhancer function and subsequently *SMAD3* expression would be expected to alter TGFβ signaling with pleiotropic effects giving rise to various conditions and diseases. Similar to Crohn's disease and inflammatory bowel disease, atherosclerosis has an

immune/inflammatory component and macrophages and T cells play important roles in pathogenesis. It could thus be of interest to determine if immune and inflammatory cytokines are able to activate AP-1 signaling upstream of binding to rs17292632. Since the rs17293632 SNP and 5'-TGA[G/C]TCA-3' AP-1 recognition sequence is conserved between humans and mice, generation of mouse models with different alleles for rs17293632 or deletion of the enhancer sequence entirely may further unravel the contribution of this AP-1 regulatory mechanism. Alternatively, genome editing of the AP-1 binding site using CRISPR/Cas9 in human cell lines could be used to further clarify its role.

It is well established that SMAD3 is an important mediator of TGFβ signalling and previous studies suggest TGFβ signaling has a protective effect in atherosclerosis[158,159,169]. However, the role of SMAD3 in the vasculature is somewhat less clear. As reported by Kobayashi *et al*.[179], *Smad3* null mice showed enhanced neointimal hyperplasia with decreased matrix deposition in response to femoral artery injury. *Smad3*$^{-/-}$ intima were also shown to contain more proliferating vascular SMCs and less collagen as compared to wild type intima. In agreement with these findings, here we find that in human arterial SMCs (**Figure 6**) *SMAD3* knockdown increases cell viability consistent with an anti-proliferative role of SMAD3. In contrast, Tsai *et al*.[182] reported that over-expression of *Smad3* via adenoviral delivery to injured rat carotid artery or in cultured vascular SMCs increased SMC proliferation in response to TGFβ.

Irrespective of these discrepant findings, it is not immediately evident how decreased *SMAD3* expression, and increased smooth muscle cell proliferation in carriers of rs17293632 (T) would be associated with protection from CAD. This finding may suggest

that decreased *SMAD3* expression has a beneficial effect at a specific stage of atherosclerosis or in a specific cellular/tissue context. It is known that *SMAD3* is expressed at low levels in normal artery but increases dramatically in response to injury and is elevated in atherosclerotic plaque[175,182]. Attenuation of this response in carriers of the rs17293632 (T) allele may somehow be protective. Secondly, while effects on vascular smooth muscle cells are plausible in terms of linking this SNP to atherosclerosis, it is possible that reduced *SMAD3* expression in another cell type in the vessel wall (endothelial, macrophage, T cell) would have a anti-atherogenic effect. Our findings that expression of the *SMAD3* gene is under positive regulation by AP-1 is notable in that there are reports Jun proteins can antagonistically affect function of the SMAD3 protein[202], suggesting dynamic interplay between these factors.

In summary, of the dozens of CAD-associated loci identified by the GWAS approach, few have been functionally characterized. Here we fine mapped a newly reported GWAS locus to identify a causal SNP that represents a novel functional *cis*-acting element at the *SMAD3* locus on chromosome 15. The protective rs17293632 (T) allele disrupts a consensus AP-1 binding site, resulting in impaired AP-1 binding to the *SMAD3* intron 1 enhancer, and reduced enhancer activity that in turn correlates with lower *SMAD3* expression in blood and human plaque.

URLs

HaploReg v3: http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php

Haploview: https://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview

eQTL Browser, University of Chicago: http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/

NCBI eQTL Browser: http://www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi

GTEx eQTL Browser:http://www.gtexportal.org

RegulomeDB: http://www.regulomedb.org/index

4 FUNCTIONAL RELATIONSHIP OF THE *COL4A1/COL4A2* LOCUS AT 13q34 TO CORONARY ARTERY DISEASE

Status: not yet submitted

4.1 Introduction

Coronary artery disease (CAD) is the leading cause of death worldwide and is characterized by the buildup of atherosclerotic plaque in the epicardial arteries. Although genetics is believed to account for ~50% of inter-individual risk for developing CAD[267], the genetic factors that contribute to disease are poorly defined. The CARDIoGRAM Consortium has conducted type 2 meta-analyses of CAD genome-wide association studies to identify novel genetic loci associated with CAD, and have published three recent studies in *Nature Genetics*[26,42,44]. The most recent CARDIoGRAMPlusC4D study comprised over 185,000 combined CAD cases and controls and interrogated 6.7 million common variants (SNPs and insertions/deletions) and 2.7 million low frequency variants. These papers have now identified 58 CAD-associated loci reaching genome-wide significance ($p<5 \times 10^{-8}$) and 202 FDR (false discovery rate) variants with $q<0.05$. Many of the novel CAD-associated loci have unknown roles in CAD pathogenesis and are not involved with traditional CAD risk factors. One of these is the *COL4A1/COL4A2* locus at chromosome 13q34, which has not been widely characterized with respect to atherosclerosis. These genes encode the large COL4A1 and COL4A2 proteins that form triple helices and are key constituents of basement membranes[79].

Type IV collagen triple helices consist primarily of two COL4A1 molecules combined with one COL4A2 molecule that assemble via their C-terminal globular domains. These collagen IV triple helices end up forming a meshwork that comprises the major structural component

of basement membranes[79]. The basement membrane is a thin sheet of fibers lining the epithelium of most tissues as well as lining the endothelial layer of blood vessels[75]. The COL4A1 and COL4A2 proteins also have functional roles in angiogenesis, cell proliferation, and cell migration[94,95,268,269]. Type IV collagen normally functions to maintain the quiescent, contractile phenotype of vascular smooth muscle cells[100,105]. Basement membrane proteins including type IV collagen also reduce inflammatory gene expression, reduce LDL uptake in culture, and lower calcification of extracellular matrix[101,108,109].

The *COL4A1* and *COL4A2* genes are located adjacent to each other on chromosome 13q34 and are encoded on opposite strands (**Figure 1**). The *COL4A1* gene is ~158 kb and consists of 52 exons while the *COL4A2* gene is ~207 kb and has 48 exons. They share a common, bidirectional promoter of only 127 bp[59]. However, transcription of each gene requires the presence of both upstream and downstream activating sequences. Exon 1 and intron 1 of both *COL4A1* and *COL4A2* contain elements required for proper transcription. The bidirectional promoter lacks a traditional TATA box but binds transcription factors such as Sp1, CCAAT-binding protein, and CTC binding factor (CTCBF)[63,270]. Here, we attempt to characterize the functional roles of the first two CAD-associated SNPs at the *COL4A1/COL4A2* locus reported by the CARDIoGRAM Consortium. The rs4773144 index SNP at this locus was reported in the first CARDIoGRAM paper[42], while the independent rs9515203 index SNP was reported in the follow up CARDIoGRAMPlusC4D paper[26] (**Table 1**). Interestingly the rs4773144 and rs9515203 SNPs are not linked (r^2=0.01)[26], meaning they represent independent CAD association signals. This work highlights the important and emerging roles of SNPs in noncoding regions of the genome in the pathogenesis of complex diseases such as CAD.

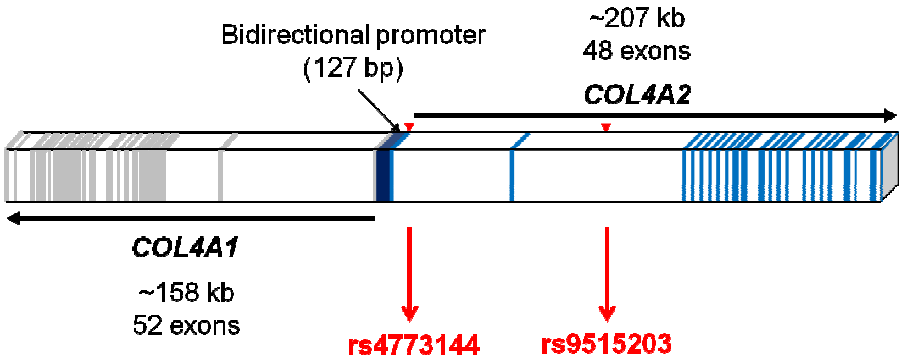

Bidirectional promoter
(127 bp)
~207 kb
48 exons
COL4A2
COL4A1
~158 kb
52 exons
rs4773144
rs9515203

Figure 1. Schematic of the *COL4A1* (52 exons) and *COL4A2* (48 exons) genes on chromosome 13q34. These genes are arranged in a head-to-head fashion and share a common, bidirectional promoter of only 127 bp. *COL4A2* is transcribed on the forward strand of genomic DNA, whereas *COL4A1* is transcribed on the reverse strand. The red arrows represent positions of the CARDIoGRAM risk SNPs. *COL4A1* exons are annotated with grey boxes, *COL4A2* exons are annotated with blue boxes, and introns are annotated by the white boxes. The rs4773144 SNP is located in proximity to the bidirectional promoter, whereas the rs9515203 SNP is in the middle of a 70 kb intron.

Table 1. Properties of the rs4773144 and rs9515203 index SNPs at the *COL4A1/COL4A2* locus associated with CAD. Values reported are from the CARDIoGRAMPlusC4D paper published in *Nature Genetics* in January 2013[26].

SNP ID	Position (build 37)	Role	Risk allele frequency	Combined p-Value	Odds Ratio (95% CI)
rs4773144	13:110,960,712	Promoter, Intron	0.44 (G)	1.43×10^{-11}	1.07
rs9515203	13:111,049,623	Intron	0.74 (T)	5.85×10^{-12}	1.08

4.2 Materials and Methods

Human Subjects

All CAD cases and controls were selected from the Ottawa Heart Study cohort as described previously[36]. CAD cases had severe, premature CAD with onset before the age of 55. Control subjects with no history or symptoms of CAD consisted of men aged >65 years and women aged >70 years. Details of CARDIoGRAM cohorts and conditional and joint analysis are explained in the most recent CARDIoGRAM Nature Genetics paper[44].

Cell Culture

HT1080 fibrosarcoma cells were purchased from ATCC and grown in high glucose DMEM media (Gibco) supplemented with 10% FBS, L-Glutamine, and penicillin-streptomycin. Primary human arterial smooth muscle cells (hASMCs), lot HITC6, were obtained from the laboratory of Dr. JG Pickering. These cells were derived from primary cultures of smooth muscle cells from the human thoracic artery[231] and experiments performed between passages 28 and 33. HITC6 cells were grown in SmGM-2 Smooth Muscle Cell Basal Medium (Lonza) supplemented with 5% FBS, insulin, hFGF-B, GA-1000, and hEGF.

Generation of *COL4A1/COL4A2* Promoter Constructs

For the reporter detecting transcription from the *COL4A2* direction, we mimicked the insert from the construct pNA reported by Schmidt *et al*.[72]. This insert contains the promoter, part

of exon 1 of *COL4A1* along with exon 1, intron 1, and part of exon 2 of the *COL4A2* gene. This region was amplified using the primers 5'-GATTGGTACCTGAGCCGGGGCCCCATGGT -3' (KpnI restriction site) and 5'-GATTAAGCTTGGCCCCGGTCAGTCCCACT-3' (HindIII restriction site). After PCR amplification, both the resulting PCR product and pGL3-Basic were digested with KpnI and HindIII before subsequent ligation into the pGL3-Basic multiple cloning site.

To construct a reporter to detect transcription from the *COL4A1* direction, we used the insert from the pBNaSA construct as described previously[72]. This construct consists of the bidirectional promoter flanked by regulatory sequences in the *COL4A1* and *COL4A2* genes. Also included was a splice-acceptor sequence from the *COL4A2* gene[72]. The splice acceptor site was PCR amplified using the primers 5'-GATTAGATCTCGGGCCGCGCACCGCGCTGT-3' (BglII restriction site) and 5'-GATTAAGCTTGGCCCCGGTCAGTCCCACT-3' (HindIII restriction site). Ligation of this sequence was performed after digesting this PCR product and empty pGL3-Basic with BglII and HindIII. The sequence required for *COL4A1* promoter activity was PCR amplified using the primers 5'-GATTGGTACCCGCGAACGCGCGGGCGCGA-3' (KpnI restriction site) and 5'-GATTCTCGAGGGCTCCCAGGCACCCTCAC-'3 (XhoI restriction site). This sequence was inserted into pGL3-Basic after digesting both the PCR product and empty vector with KpnI and XhoI and subsequent ligation.

Once the regions necessary for *COL4A1* and *COL4A2* transcription were successfully inserted into the pGL3-Basic, site directed mutagenesis was conducted on the original *COL4A1* construct to change the rs12429420 allele. For the *COL4A2* reporters, site-directed

mutagenesis was used to generate the proper combinations of alleles for rs35466678, rs7327528, and rs76536922. All of these mutagenesis reactions were conducted using PfuUltra II fusion HS DNA Polymerase (Agilent Technologies). Correct sequences for all constructs were confirmed after the cloning step via sequencing.

Generation of rs9515203 Enhancer Constructs

To test for enhancer activity near the rs9515203 SNP, we cloned a full-length, ~2 kb sequence (R1+R2+R3) in addition to the R1, R2, and R3 subsequences into the pGL3-Promoter vector (Promega) that contains an SV40 promoter. For each insert, we PCR amplified genomic DNA using primers with BamHI restriction sites (**Table 2**). Next, these PCR products as well as empty pGL3-Promoter were digested with BamHI prior to subsequent ligation. Each insert with BamHI sites was cloned into pGL3-Promoter in both forward and reverse orientations. To change the allele at rs9515203, site-directed mutagenesis was performed on the full and R3 constructs (in the forward orientation) using the Q5 Site Directed Mutagenesis Kit (New England BioLabs).

Transfection and Dual Luciferase Assays

For transient transfection studies with the *COL4A1/COL4A2* promoter reporter constructs, HT1080 cells were seeded in 12 well dishes and grown to ~70% confluence at the time of transfection. Each well was transfected with 1 µg of *COL4A1/COL4A2* promoter reporter and 2% pRL-SV40 *Renilla*. HT1080 cells were harvested 24 hours after transfection by

Table 2. Primer sequences to amplify regulatory regions near the rs9515203 SNP in intron 4 of the *COL4A2* gene.

Construct	Insert Size (bp)	Primer	Sequence (5'-3')
Full Enhancer (R1+R2+R3)	1966	FWD	GATTGGATCCTCTTTAAAGCAGTTCTGGAT
		REV	GATTGGATCCAGCCCTTCAAAATGTAATGG
R1	802	FWD	GATTGGATCCTCTTTAAAGCAGTTCTGGAT
		REV	GATTGGATCCGCATACCATCTGACACGTT
R2	281	FWD	GATTGGATCCATAGAGAGATTTCAGAGGTC
		REV	GATTGGATCCAACAGTCGGCCAGA
R3	883	FWD	GATTGGATCCTATCCGCCCCCATT
		REV	GATTGGATCCAGCCCTTCAAAATGTAATGG

washing once with 1X PBS and lysing with 250 µL of Passive Lysis Buffer (Promega). HT1080 cell transfections were performed with X-tremeGENE 9 transfection reagent (Roche).

For transient transfection studies in hASMCs, HITC6 cells were seeded in 6 well dishes and grown to ~90% confluence at the time of transfection. Each well was transfected with 2.5 µg of firefly luciferase construct and 2% pRL-SV40 *Renilla*. hASMCs were harvested 24 hours after transfection by washing once with 1X PBS and lysing with 500 µL of 1X Passive Lysis Buffer (Promega). hASMC transfections were performed with the Lipofectamine 3000 reagent (Life Technologies). All harvested lysates were centrifuged for 1 minute at 13,000 rpm at 4°C. Dual luciferase assays for both cell types were performed using the Promega GloMAX 96 Microplate luminometer.

COL4A1/COL4A2 Expression Analysis

COL4A1, COL4A2, and proximal gene expression at chromosome 13q34 were probed *in vivo* in non-CAD patients from various tissues collected in the ASAP (Advanced Study of Aortic Pathology) study. Patients from the ASAP study had aortic valve and ascending aortic disease and all underwent elective open heart surgery[238]. None of these patients had significant CAD by coronary angiography. Tissues collected included mammary artery intima-media, liver, ascending aorta intima-media, ascending aorta adventitia, and heart. SNPs were genotyped on the Illumina610w-QuadBead Array and rs9515203 imputed using the Mach2.0 algorithm and 1000 Genomes data as reference (imputation quality $r^2 = 0.33$) in

138 patients. rs9515203 was also directly genotyped on a custom cardiometabochip in 106 patients.

4.3 Results

Functional characterization of the rs4773144 SNP

One of the original reported CAD-associated SNPs at the *COL4A1/COL4A2* locus was rs4773144, located in intron 3-4 of *COL4A2* and very close (~1 kb) to the bidirectional *COL4A1/COL4A2* promoter (**Figure 1, Figure 2**). Our original hypothesis was that the rs4773144 SNP from CARDIoGRAM is linked to SNPs in the nearby promoter region that in turn functionally alter promoter activity and *COL4A1* and/or *COL4A2* transcription. The Ensembl and UCSC genome browsers had deposited several SNPs located in regions previously reported to be necessary for *COL4A1* and/or *COL4A2* transcription (**Figure 2**). Regulatory features of the *COL4A1/COL4A2* bidirectional promoter and surrounding area, such as histone modifications and ENCODE ChIP-seq data, are shown in **Figure 3**. While the *COL4A1/COL4A2* bidirectional promoter has been well characterized, UCSC genome browser and ENCODE annotation reveal high levels of H3K4me3 (commonly found near promoters), and moderate to high levels of H3K27ac (found in proximity to active regulatory elements) (**Figure 3**)[242,243].

At the start of this project, imputation (such as from the 1000 Genomes Project) was not widely implemented. Therefore we began by manually sequencing the *COL4A1/COL4A2* promoter region (blue box) (**Figure 2**) and genotyped the SNPs in this region in approximately 500 CAD cases and 500 CAD controls from our Ottawa Heart Study cohort to identify CAD-associated SNPs linked with rs4773144. Out of the manually sequenced SNPs at the promoter area, four showed the strongest association with CAD: rs12429420, rs35466678, rs7327528, and rs76536922. Currently, with recent advances in imputation

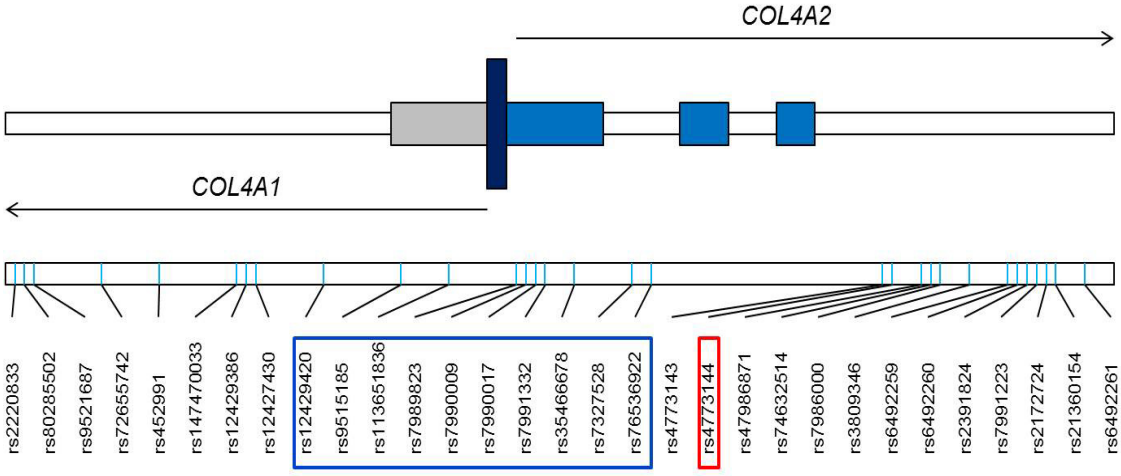

COL4A2
COL4A1
rs2220833
rs80285502
rs9521687
rs72655742
rs452991
rs147470033
rs12429386
rs12427430
rs12429420
rs951518
rs113651836
rs7989823
rs7990009
rs7990017
rs7991332
rs35466678
rs7327528
rs76536922
rs4773143
rs4773144
rs47986871
rs74632514
rs7986000
rs3809346
rs6492259
rs6492260
rs2391824
rs7991223
rs2172724
rs21360154
rs6492261

Figure 2. Schematic of the *COL4A1/COL4A2* bidirectional promoter region. Exons are denoted by thick boxes, whereas introns are denoted by thin lines/boxes. The SNPs in the blue box were the ones originally sequenced and genotyped and subsequently imputed using CARDIoGRAM and 1000 Genomes Project data. The rs4773144 SNP in the red box is the index SNP from CARDIoGRAM.

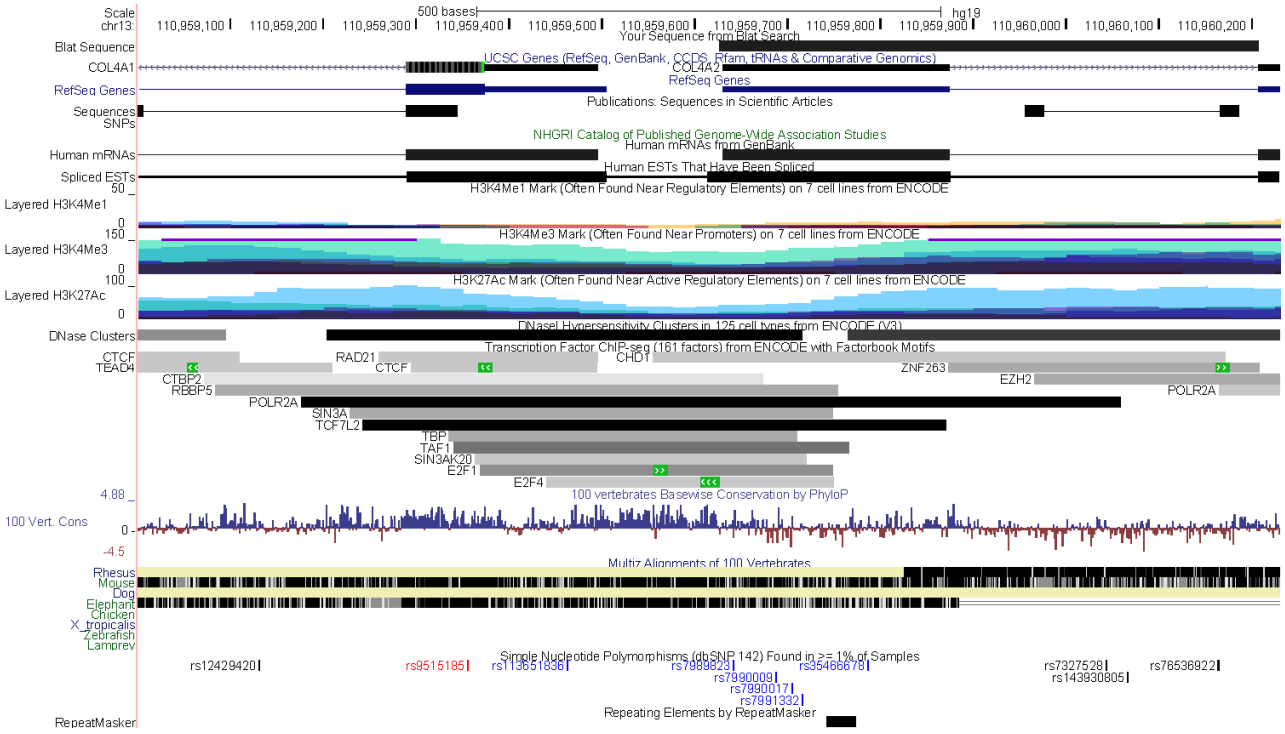

Scale
chr13:
500 bases
hg19
110,959,100
110,959,200
110,959,300
110,959,400
110,959,500
110,959,600
110,959,700
110,959,800
110,959,900
110,960,000
110,960,100
110,960,200
Blat Sequence
Your Sequence from Blat Search
COL4A1
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
COL4A2
RefSeq Genes
RefSeq Genes
Sequences
Publications: Sequences in Scientific Articles
SNPs
NHGRI Catalog of Published Genome-Wide Association Studies
Human mRNAs
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
50
Layered H3K4Me1
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
0
150
Layered H3K4Me3
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
0
100
Layered H3K27Ac
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
0
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
CTCF
RAD21
CHD1
TEAD4
CTCF
ZNF263
CTBP2
EZH2
RBBP5
POLR2A
POLR2A
SIN3A
TCF7L2
TBP
TAF1
SIN3AK20
E2F1
E2F4
4.88
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
0
-4.5
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples
rs12429420
rs9515185
rs113651836
rs7989825
rs35466678
rs7990009
rs7990017
rs7991332
rs7327528
rs76536922
rs143930805
RepeatMasker
Repeating Elements by RepeatMasker

Figure 3. UCSC genome browser and ENCODE project annotation for the *COL4A1/COL4A2* bidirectional promoter and proximal regulatory elements.

using 1000 Genomes Project data, we now have access to association data for these promoter SNPs in all of the CARDIoGRAM cohorts (**Table 3**). In the much larger CARDIoGRAM cohort, rs12429420, rs35466678, rs7327528, and rs76536922 remained the promoter SNPs showing the strongest CAD association. These four promoter SNPs have lower minor allele frequencies (~5-20%) than rs4773144 (~40%), but have high D' values with rs4773144 (**Table 4**). From our *in silico* analysis neither rs12429420, rs35466678, rs7327528, or rs76536922 are predicted to disrupt any consensus transcription factor binding sequences (**Figure 3**).

We performed luciferase assays to determine if these four CAD-associated SNPs in the *COL4A1/COL4A2* bidirectional promoter region affected reporter activity of *COL4A1* and *COL4A2* promoter constructs. We cloned the regions necessary for promoter activity into pGL3-Basic and performed site-directed mutagenesis to create various allelic combinations (**Figure 4A**). These constructs were transfected into HT1080 fibrosarcoma cells and luminescence measured 24 hours after transfection. The majority of papers studying *COL4A1/COL4A2* promoter regulation performed reporter assays in HT1080 cells, which ubiquitously express the COL4A1 and COL4A2 proteins[63,71,72]. Our dual luciferase assays in HT1080 cells revealed the *COL4A1* and *COL4A2* regions cloned into pGL3-Basic acted as strong promoters, but we observed no SNP effects on promoter activity (**Figure 4B**).

Thus far it has been difficult to pinpoint functional SNPs near rs4773144 and the *COL4A1/COL4A2* bidirectional promoter. While SNPs at the bidirectional promoter appeared to be the best functional candidates, it is also possible that the intronic rs4773144 SNP itself, or SNPs in linkage disequilibrium could be functional. As **Table 5** shows,

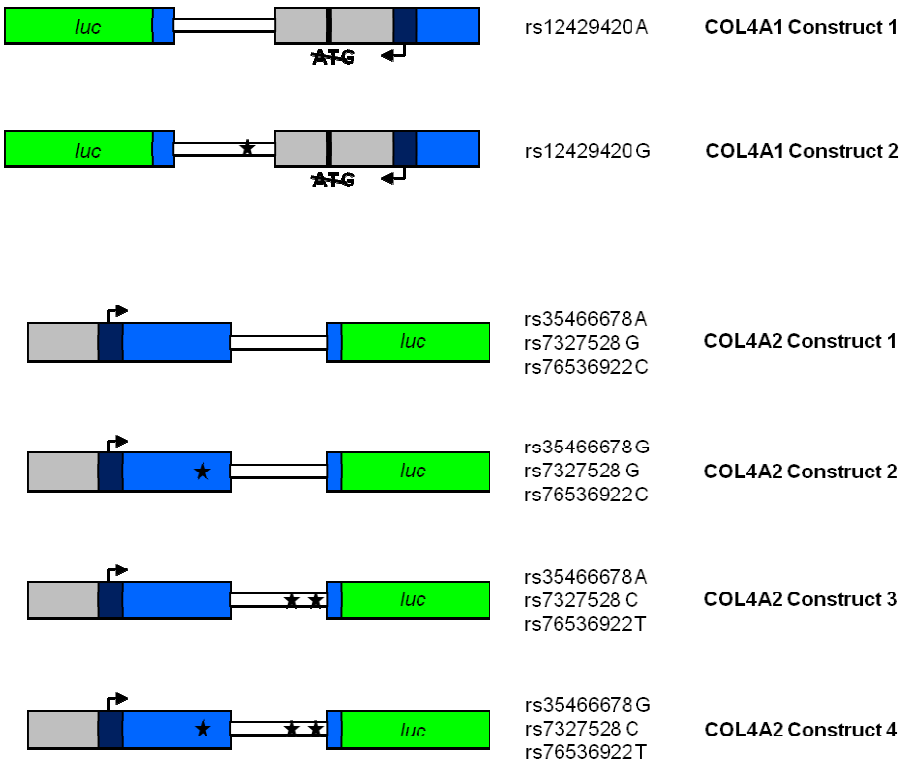
Table 3. Association of SNPs at the *COL4A1/COL4A2* bidirectional promoter region with CAD in conditional and joint analysis of CARDIoGRAM data using 1000 Genomes imputation

SNP	chr	Bp_hg19	Effect allele	Other allele	Effect allele freq	beta	SE	p-value	q-value	N studies
rs12429420	13	110,959,131	A	G	0.8338	-0.0545	0.0150	0.0003	0.1387	43
rs113651836	13	110,959,464	C	A	0.7134	0.0292	0.0118	0.0135	0.4014	42
rs7989823	13	110,959,643	C	A	0.5728	0.0320	0.0124	0.0097	0.6952	40
rs7990009	13	110,959,688	C	G	0.7012	0.0290	0.0116	0.0125	0.4543	42
rs7990017	13	110,959,705	T	C	0.5022	0.0026	0.0108	0.8076	0.3978	46
rs7991332	13	110,959,717	T	C	0.6975	0.0299	0.0116	0.0101	0.2933	42
rs35466678	13	110,959,787	A	G	0.8270	-0.0530	0.0150	0.0004	0.2245	43
rs7327528	13	110,960,044	C	G	0.0595	0.1047	0.0233	6.76E-06	0.8923	38
rs76536922	13	110,960,164	T	C	0.0550	0.0960	0.0241	6.86E-05	0.7730	38
rs4773144	13	110,960,712	A	G	0.5726	-0.0524	0.0103	3.87E-07	0.0384	43

Table 4. Linkage disequilibrium (D' and r^2) of *COL4A1/COL4A2* promoter SNPs with the rs4773144 index SNP from CARDIoGRAM in Ottawa Heart Study samples

SNP	chr	Bp_hg19	D'	r^2
rs12429420	13	110,959,131	1.000	0.206
rs113651836	13	110,959,464	1.000	0.322
rs7989823	13	110,959,643	0.118	0.007
rs7990009	13	110,959,688	1.000	0.346
rs7990017	13	110,959,705	0.183	0.020
rs7991332	13	110,959,717	1.000	0.352
rs35466678	13	110,959,787	1.000	0.215
rs7327528	13	110,960,164	0.859	0.063
rs76536922	13	110,960,712	0.971	0.074

A


luc
ATG
rs12429420A
COL4A1 Construct 1
luc
ATG
rs12429420G
COL4A1 Construct 2
luc
rs35466678A
rs7327528G
rs76536922C
COL4A2 Construct 1
luc
rs35466678G
rs7327528G
rs76536922C
COL4A2 Construct 2
luc
rs35466678A
rs7327528C
rs76536922T
COL4A2 Construct 3
luc
rs35466678G
rs7327528C
rs76536922T
COL4A2 Construct 4

B

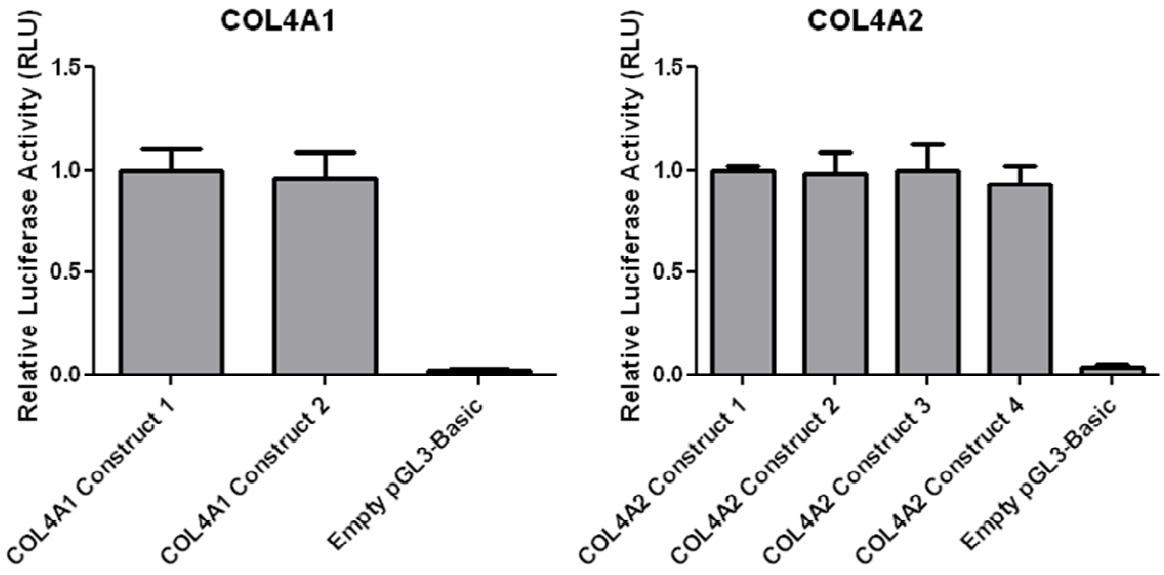

COL4A1
Relative Luciferase Activity (RLU)
1.5
1.0
0.5
0.0
COL4A1 Construct 1
COL4A1 Construct 2
Empty pGL3-Basic
COL4A2
Relative Luciferase Activity (RLU)
1.5
1.0
0.5
0.0
COL4A2 Construct 1
COL4A2 Construct 2
COL4A2 Construct 3
COL4A2 Construct 4
Empty pGL3-Basic

Figure 4. CAD-associated SNPs at the *COL4A1/COL4A2* bidirectional promoter do not affect luciferase activity *in vitro* in either the *COL4A1* or *COL4A2* direction. (A) Schematic of the *COL4A1* and *COL4A2* reporters investigated, containing the DNA elements necessary for transcription of *COL4A1* and *COL4A2*. * denote presence/locations of CAD risk alleles. (B) HT1080 cells were transfected with *COL4A1* and *COL4A2* promoter reporter constructs for 24 hours before being harvested for dual luciferase assays. Each well was co-transfected with 2% pRL-SV40 *Renilla* to control for transfection efficiency. Values represent the average of three independent experiments (± SD), each experiment having three replicates for each condition. RLU, relative light units.

Table 5. SNPs in intron 3-4 of *COL4A2* in strong linkage disequilibrium with the CARDIoGRAM rs4773144 SNP

SNP ID	Position (hg19)	D' with rs4773144	r^2 with rs4773144
rs4773143	13:110,960,685	1	1
rs4773144	13:110,960,712	1	1
rs7986871	13:110,960,789	1	1
rs3809346	13:110,960,943	1	1

rs4773144 is in perfect LD with three nearby SNPs (rs4773143, rs7986871, and rs3809346) in the same intron. This intronic region possesses chromatin marks suggestive of either an active enhancer or promoter (**Figure 5**). Furthermore, ENCODE ChIP-seq data suggests RNA Polymerase II as well as EZH2 bind to the region containing rs4773144, indicating this intronic sequence could potentially transcribe a noncoding RNA or be a target of chromatin modifying complexes.

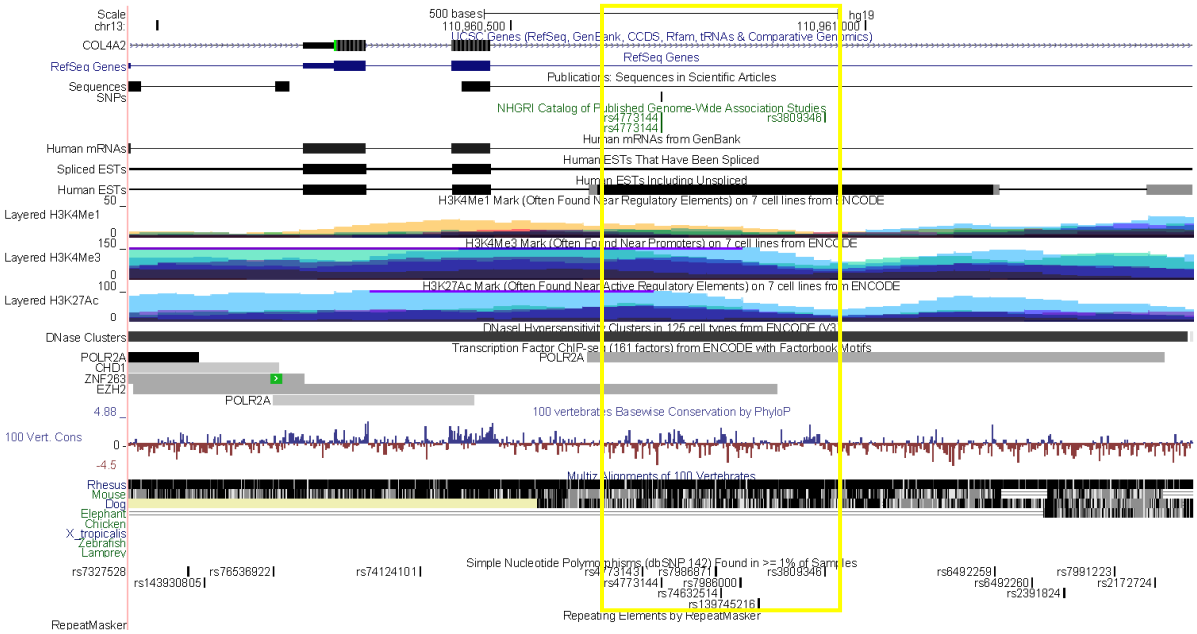

Scale
chr13:
500 bases
110,960,500
110,961,000
hg19
COL4A2
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
RefSeq Genes
Sequences
Publications: Sequences in Scientific Articles
SNPs
NHGRI Catalog of Published Genome-Wide Association Studies
rs4773144
rs38093481
rs4773144
Human mRNAs
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
Human ESTs
Human ESTs Including Unspliced
50
Layered H3K4Me1
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
0
150
Layered H3K4Me3
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
0
100
Layered H3K27Ac
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
0
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
POLR2A
POLR2A
CHD1
ZNF263
EZH2
POLR2A
4.88
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
0
-4.5
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples
rs7327528
rs143930805
rs76536922
rs74124101
rs4773143
rs7986871
rs38093481
rs4773144
rs7986000
rs74632514
rs139745216
rs6492259
rs6492260
rs7991223
rs2391824
rs2172724
Repeating Elements by RepeatMasker
RepeatMasker

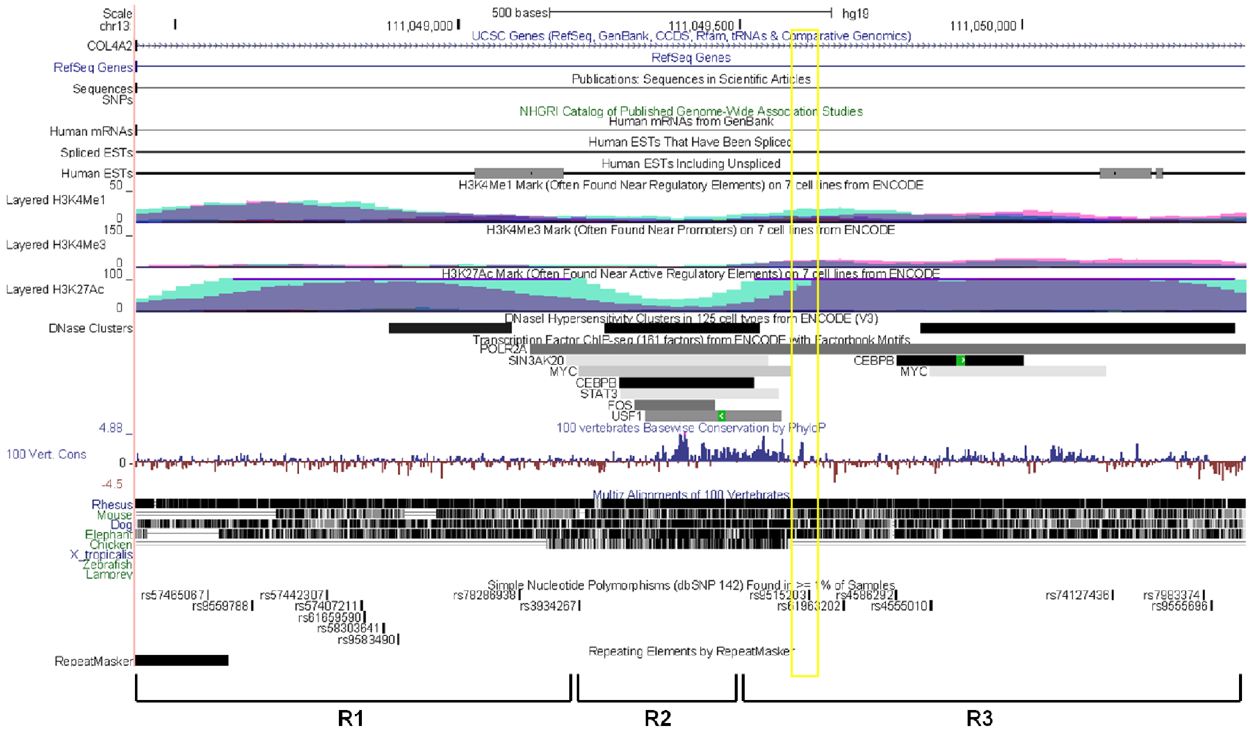
Figure 5. UCSC genome browser and ENCODE project annotation of intron 3-4 of *COL4A*2 containing the CARDIoGRAM rs4773144 SNP, approximately 1 kb from the *COL4A1/COL4A2* bidirectional promoter. The yellow box represents the location of rs4773144 and its tightly linked SNPs.

Functional characterization of the rs9515203 SNP

The rs9515203 SNP at the *COL4A1/COL4A2* locus reported by the CARDIoGRAMPlusC4D Consortium[26] is located in the middle of a 70 kb intron in *COL4A2* (**Figure 1**). The common allele T at rs9515203 is associated with increased risk of CAD, while the minor allele C (MAF 0.26) is protective (**Table 1**). Bioinformatics analysis reveals this SNP not strongly linked with any other SNPs, suggesting an independent CAD signal that contributes to CAD pathogenesis by itself. According to the UCSC genome browser, the sequence around this SNP has several features characteristic of enhancers (**Figure 6A**). Typically chromatin marks signature of enhancers are H3K4 monomethylation (H3K4me1), the absence of H3K4 trimethylation (H3K4me3), the presence of DNase I hypersensitivity sites, and acetylation of H3K27 (H3K27ac)[242,243]. In the 2 kb region around the rs9515203 SNP, ENCODE data showed two large peaks of H3K27 acetylation, a DNase hypersensitivity site, moderate levels of H3K4me3, and a fair degree of H3K4me1. Furthermore, several transcription factors including C/EBP Beta, RNA Polymerase II, and USF1 are predicted to bind to the sequence near rs9515203 (**Figure 6A**).

We arbitrarily chose this 2 kb sequence and divided it into three subsequences (referred to as R1, R2, and R3) based on the chromatin signatures. We then cloned the full 2 kb sequence (R1+R2+R3) into the pGL3-Promoter vector (Promega) as well as the R1, R2, and R3 subsequences individually (both forward and reverse orientations). The pGL3-Promoter vector has basal promoter activity and insertion of functional enhancer sequences will result in increased luciferase activity. As **Figure 6B** shows, luciferase assays in primary hASMCs shows the full 2 kb sequence (R1+R2+R3) increases promoter activity by roughly tenfold.

A


Scale
chr13
500 bases
hg19
111,049,000
111,049,500
111,050,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
COL4A2
RefSeq Genes
Publications: Sequences in Scientific Articles
NHGRI Catalog of Published Genome-Wide Association Studies
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
Human ESTs Including Unspliced
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me1
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
Layered H3K4Me3
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
POLR2A
SIN3AK20
MYC
CEBPB
STAT3
FOS
USF1
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples
Repeating Elements by RepeatMasker
RepeatMasker
R1
R2
R3

B

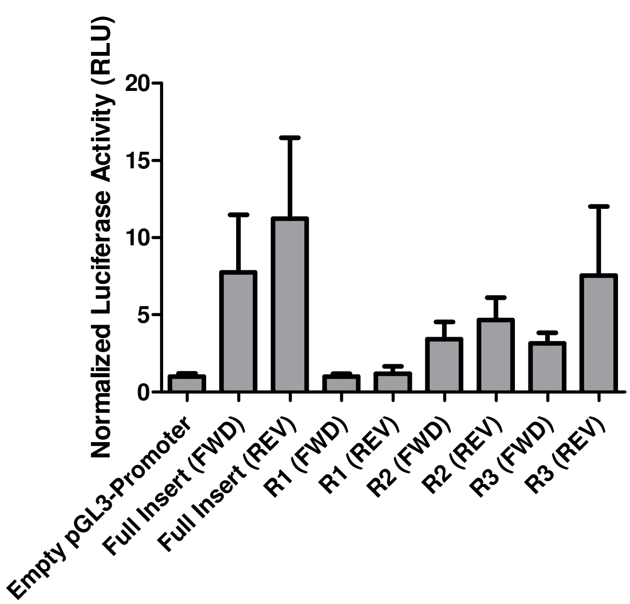

Normalized Luciferase Activity (RLU)
0
5
10
15
20
Empty pGL3-Promoter
Full Insert (FWD)
Full Insert (REV)
R1 (FWD)
R1 (REV)
R2 (FWD)
R2 (REV)
R3 (FWD)
R3 (REV)

Figure 6. Enhancer properties of DNA in proximity to the rs9515203 SNP. (A) UCSC genome browser and ENCODE annotation of the DNA sequence near the rs9515203 CARDIoGRAMPlusC4D SNP (yellow box). This sequence consists of chromatin histone marks suggesting the existence of an enhancer. This 2 kb sequence was divided into subsequences R1, R2, and R3. (B) The R2 and R3 sequences have enhancer activity in primary hASMCs, whereas the R1 sequence does not. Each sequence was cloned into the pGL3-Promoter vector and 2.5 µg of each construct was transfected into hASMCs. Cells were harvested 24 h after transfection and assayed for luciferase activity. Each well was co-transfected with 2% pRL-SV40 *Renilla* to control for transfection efficiency. Values represent the average of three independent experiments (± SD), each experiment having three separate wells for each condition. RLU, relative light units.

The R1 sequence had no enhancer effect, while the R2 and R3 (containing rs9515203) sequences produce approximate three to seven fold increases in enhancer activity. The R2 and R3 sequences appear to have additive effects for enhancer activity at this *COL4A2* intron since their combined effects match the full length insert.

Next, since the rs9515203 SNP lies within the R3 region (**Figure 6A**), we sought to determine if changing the allele at rs9515203 has an effect on enhancer activity in luciferase assays for both the full length enhancer sequence and the R3 sequence alone. Site-directed mutagenesis was performed to generate constructs containing both the T (risk) and C (protective) alleles. However, the rs9515203 SNP had no effect on enhancer activity in experiments in hASMCs (**Figure 7**). The activity of the *Renilla* control vector was the same for both rs9515203 genotypes.

We next determined whether rs9515203 is an expression quantitative trait locus (eQTL) for *COL4A1* and/or *COL4A2* expression in humans *in vivo*. eQTLs are DNA sequence variants leading to altered expression levels of mRNAs, and provide very strong evidence suggesting a functional impact on gene regulation. Our lab has access to hundreds of PAXgene Blood RNA tubes (PreAnalytiX/QIAGEN). Despite access to these samples, we were never able to detect *COL4A1* or *COL4A2* mRNA in whole blood using qRT-PCR. qRT-PCR experiments gave poor Cp values, poor melting curves, and the presence of multiple products. Type IV collagen is low in blood since blood cells do not express high amounts of basement membrane proteins. In addition, the numerous forms of collagen in humans and the high frequency of repeating elements in collagen likely resulted in the poor melting curves and nonspecific products.

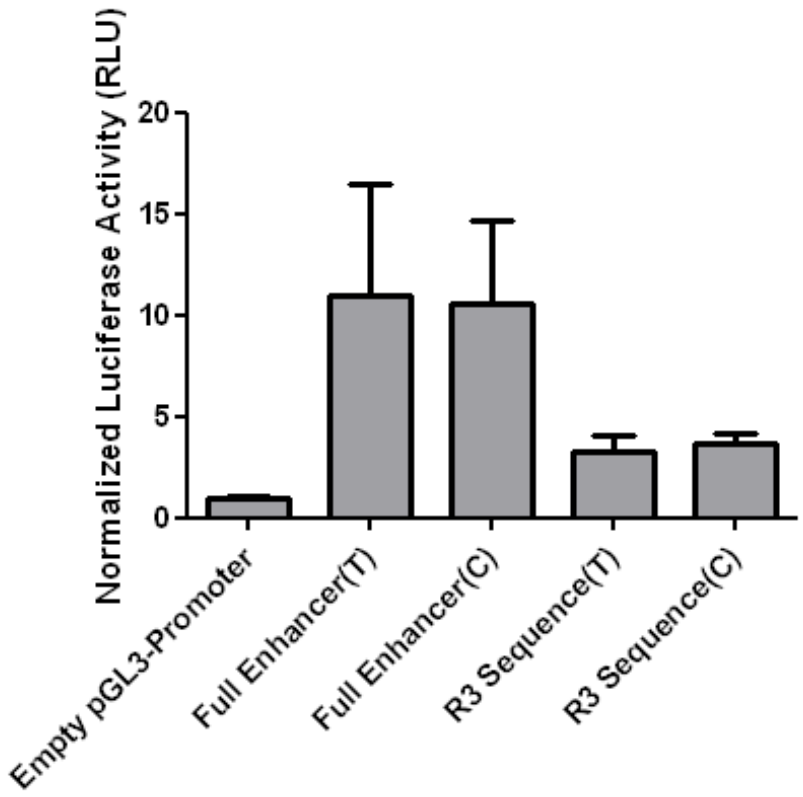

Normalized Luciferase Activity (RLU)
20
15
10
5
0
Empty pGL3-Promoter
Full Enhancer(T)
Full Enhancer(C)
R3 Sequence(T)
R3 Sequence(C)

Figure 7. The rs9515203 SNP has no effect on enhancer activity in *in vitro* luciferase assays. 2.5 µg of each construct was transfected into hASMCs and co-transfected with 2% pRL-SV40 *Renilla* to control for transfection efficiency. Cells were harvested 24 h after transfection and assayed for luciferase activity. Values represent the average of three independent experiments (± SD), each experiment having three separate wells for each condition. RLU, relative light units.

In order to determine if rs9515203 was an eQTL for *COL4A1* and/or *COL4A2* in humans *in vivo*, we collaborated with the Advanced Study of Aortic Pathology (ASAP) group in Sweden, which has collected several tissues from patients undergoing either aortic valve surgery or surgery for aortic aneurysm. These patients did not have significant CAD as assessed by coronary angiography. Tissues collected in the ASAP study included mammary artery intima-media, liver, ascending aorta intima-media, ascending aorta adventitia, and heart. The rs9515203 SNP was not present on the Illumina610w-QuadBead Array typically used by the ASAP group so had to be imputed for n=138 patient samples (**Figure 8A**). In ascending aortic intima-media, the rs9515203 SNP has an eQTL effect on both *COL4A2* (stronger) and *COL4A1* (weaker).

The box plot for the imputed rs9515203 eQTL effect in ascending aorta intima-media is shown in **Figure 8B**. rs9515203 was also directly genotyped on a custom cardiometabochip in a subset of ASAP samples (n=106), where we again observed a significant effect in ascending aorta intima-media (**Figure 8C**). As both **Figures 8B and 8C** show, the minor, protective rs9515203 (C) allele associates with lower *COL4A2* mRNA expression. Since we observe an eQTL effect for rs9515203 in the ascending aorta intima-media of patients without noticeable CAD, it will be of interest to assess the effect of this SNP in diseased plaque tissue.

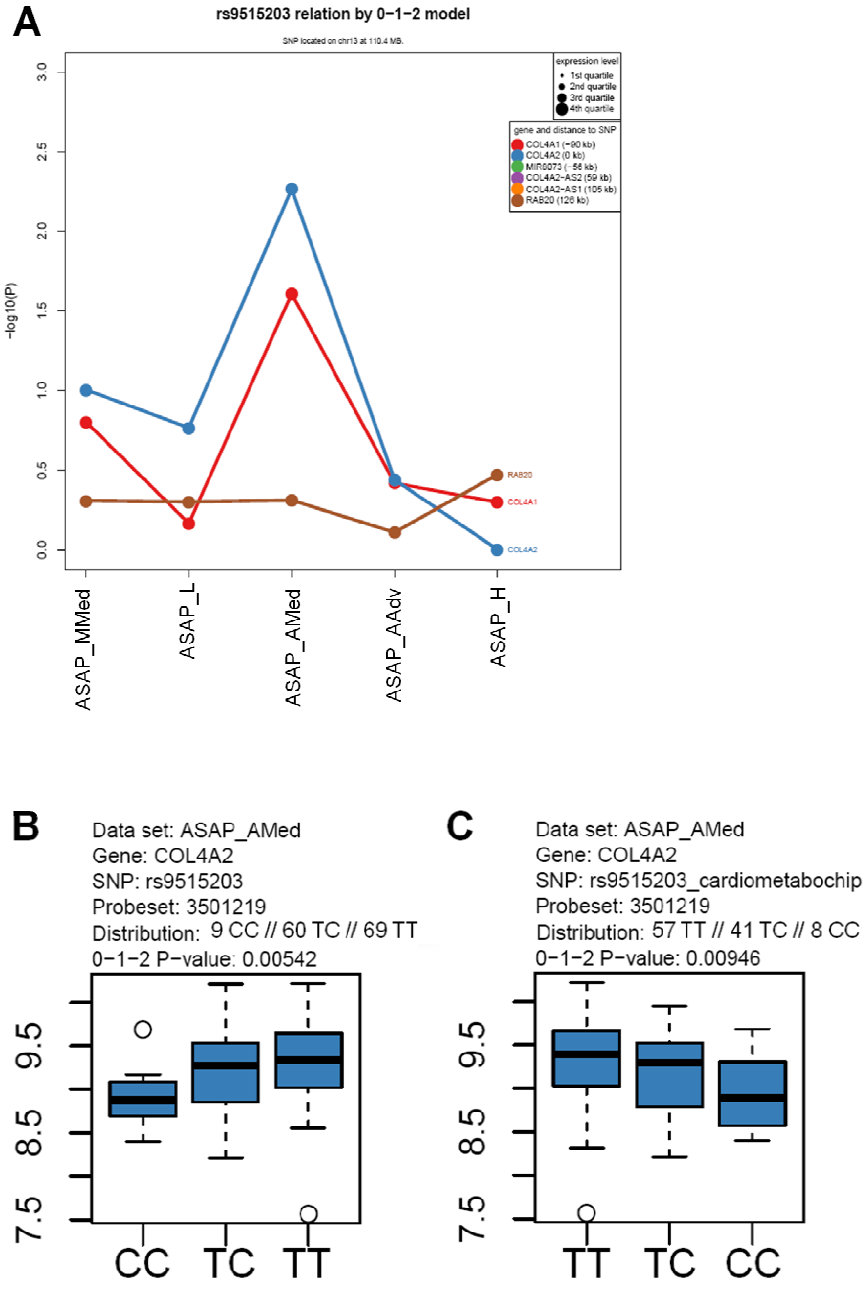

A
rs9515203 relation by 0−1−2 model
SNP located on chr13 at 110.4 MB.
expression level
1st quartile
2nd quartile
3rd quartile
4th quartile
gene and distance to SNP
COL4A1 (−90 kb)
COL4A2 (0 kb)
MIR6073 (−56 kb)
COL4A2-AS2 (59 kb)
COL4A2-AS1 (105 kb)
RAB20 (126 kb)
−log10(P)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
RAB20
COL4A1
COL4A2
ASAP_MMed
ASAP_L
ASAP_AMed
ASAP_AAdv
ASAP_H
B
Data set: ASAP_AMed
Gene: COL4A2
SNP: rs9515203
Probeset: 3501219
Distribution: 9 CC // 60 TC // 69 TT
0−1−2 P−value: 0.00542
7.5
8.5
9.5
CC
TC
TT
C
Data set: ASAP_AMed
Gene: COL4A2
SNP: rs9515203_cardiometabochip
Probeset: 3501219
Distribution: 57 TT // 41 TC // 8 CC
0−1−2 P−value: 0.00946
7.5
8.5
9.5
TT
TC
CC

Figure 8. rs9515203 is an eQTL for *COL4A1* (weaker, red line) and *COL4A2* (stronger, blue line) mRNA expression *in vivo* in humans in the ascending aorta intima-media. (A) Genotyped samples (n=138) from the Advanced Study of Aortic Pathology (ASAP) study were assessed for mRNA expression of the *COL4A1*, *COL4A2*, and proximal genes. Tissues analyzed from this study are mammary artery intima-media (MMed), liver (L), ascending aorta intima-media (AMed), ascending aorta adventitia (AAdv), and heart (H). (B) Box plot showing *COL4A2* mRNA expression in ascending aorta intima-media (AMed) as a function of rs9515203 genotype imputed from the Illumina610w-QuadBead Array. (C) Box plot showing COL4A2 mRNA expression in ascending aorta intima-media (AMed) as a function of rs9515203 genotype directly typed on the custom cardiometabochip in n=106 patients from the ASAP study.

Functional characterization of further CAD-associated SNPs at the *COL4A1/COL4A2* locus

Recently published conditional and joint analysis of CARDIoGRAMPlusC4D data using 1000 Genomes Project imputation (Nikpay et al. *Nat Genet* 2015) has now revealed 202 FDR variants (*q* value < 0.05), seven of which are at or near the *COL4A1/COL4A2* locus (**Table 6**). The relative locations of these seven FDR SNPs are detailed in **Figure 9**. Of these seven CAD-associated SNPs at the *COL4A1/COL4A2* locus, rs4773141 is in moderate linkage disequilibrium with rs4773144 (D' 0.90, r^2 0.62), and rs9515203 has already been characterized. The p values for some of these novel *COL4A1/COL4A2* SNPs are more robust in comparison to prior index SNPs in the other CARDIoGRAM papers. Epistasis analysis of the seven SNPs from **Table 6** revealed these SNPs showed no pairwise interaction for CAD association (data not shown).

Table 6. List of the seven FDR variants (out of 202 total) at the *COL4A1/COL4A2* locus ($q < 0.05$) from the most recent CARDIoGRAMPlusC4D study[44] employing conditional and joint analysis

SNP	chr	Bp_hg19	cM	Effect allele	freq	beta	SE	p-value	q-value	h2_05
rs11617955	13	110,818,102	121.0122	T	0.8936	0.0985	0.0163	1.39E-09	0.0000	0.0004
rs4773141	13	110,954,353	121.5649	C	0.6408	-0.0750	0.0118	1.78E-10	0.0000	0.0006
rs11838776	13	111,040,681	122.1888	G	0.7367	-0.0743	0.0115	8.86E-11	0.0000	0.0005
rs9515203	13	111,049,623	122.1999	T	0.7606	0.0671	0.0117	1.08E-08	0.0002	0.0004
rs34905765	13	111,100,780	122.3948	C	0.8982	-0.0807	0.0162	6.41E-07	0.0069	0.0003
rs56003851	13	111,114,176	122.4161	C	0.8035	0.0849	0.0133	1.97E-10	0.0000	0.0005
rs61969072	13	111,380,701	123.0967	T	0.8267	-0.0584	0.0125	2.87E-06	0.0233	0.0002

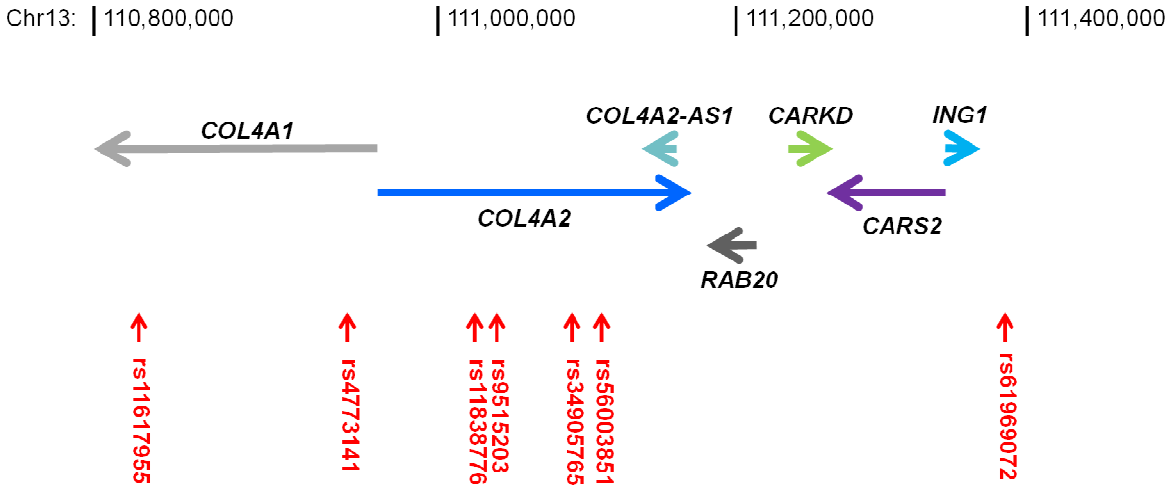

Chr13:
110,800,000
111,000,000
111,200,000
111,400,000
COL4A1
COL4A2-AS1
CARKD
ING1
COL4A2
CARS2
RAB20
rs11617955
rs4773141
rs9515203
rs11838776
rs34905765
rs56003851
rs61969072

Figure 9. Schematic of the *COL4A1/COL4A2* locus and surrounding area at chromosome 13q34. Red arrows represent the locations of the seven independent CAD-associated FDR SNPs recently reported by CARDIoGRAM using 1000 Genomes Project imputation.

4.4 Discussion

From conditional and joint analysis of CARDIoGRAM data using 1000 Genomes Project imputation, 7 of the 202 FDR variants map to the *COL4A1/COL4A2* locus at 13q34[44]. The high frequency of independent, robust CAD-associated signals at this locus highlights COL4A1 and COL4A2 as emerging targets in atherosclerosis. What is particularly interesting about this locus is when analyzing all 202 reported FDR variants from the latest CARDIoGRAM study, the other CAD loci have fewer independent signals. This high frequency of independent signals at *COL4A1/COL4A2*, in combination with the intronic/intergenic positions of these SNPs, suggests to us that this locus harbours numerous gene regulatory hotspots.

We postulate the function of these SNPs (or linked SNPs) is disruption of gene regulatory elements that in turn regulate *COL4A1* and/or *COL4A2* transcription. *COL4A1* and *COL4A2* represent very plausible target genes for effects of these CAD-associated SNPs due to their critical roles in basement membranes found in all blood vessels. That being said, we cannot rule out *trans*-eQTL effects of these SNPs on unknown distal genes since chromatin can loop and facilitate long-range interactions. Assuming the primary mechanisms whereby the *COL4A1/COL4A2* SNPs associate with CAD is through local effects on *COL4A1* and *COL4A2* transcription, this also demonstrates remarkable pleiotropy at this locus. That is, numerous different independent mechanisms exist at this locus whereby SNPs alter *COL4A1/COL4A2* expression and increase risk of CAD. It will be helpful to acquire eQTL data for these seven FDR SNPs in cells/tissues relevant to CAD, as it would provide strong functional evidence for effects on gene expression.

With respect to the rs4773144 and rs9515203 SNPs originally reported in the first two CARDIoGRAM papers[26,42], we have yet to clarify the functional mechanisms behind their robust associations with CAD. For rs4773144, we elected to focus on four SNPs in the *COL4A1/COL4A2* bidirectional promoter region, with lower minor allele frequencies than rs4773144. These four promoter SNPs are in modest LD with rs4773144 (**Table 4**), and associate very highly with CAD themselves (**Table 3**). However, in HT1080 cells these SNPs had no effect on *COL4A1/COL4A2* promoter activity when the *COL4A1/COL4A2* bidirectional promoter region was cloned into the pGL3-Basic vector. Although HT1080 cells represent a fibrosarcoma cell line, we conducted preliminary experiments with them because they are the most comprehensively used cell type for experiments dealing with *COL4A1/COL4A2* promoter regulation[63,71,72]. When we first studied these collagen promoter SNPs we were unable to successfully transfect the more physiologically relevant hASMCs in our laboratory. With the development of newer transfection reagents such as Lipofectamine 3000 (Life Technologies), it may be of interest to test the collagen promoter SNPs in these primary cells. Furthermore, while we ruled out signal saturation in our dual luciferase assays, the *COL4A1* and *COL4A2* constructs still gave a very strong firefly luciferase signal. Therefore, *COL4A1/COL4A2* SNP effects might be difficult to identify in the context of such a strong promoter in the HT1080 cell system.

rs4773144 and its tightly linked SNPs (**Table 5**) all lie within a putative enhancer/promoter sequence in intron 3 of the *COL4A2* gene (**Figure 5**). Functional annotation of *COL4A2* intron 3 with ENCODE ChIP-seq data suggests binding of RNA Polymerase II and EZH2 (**Figure 5**), an enzyme that participates in histone methylation and transcriptional repression[271]. To our knowledge there are several groups investigating the functional

significance of the rs4773144 SNP at 13q34. One of these groups looked at the intronic four SNP haplotype (encompassing rs4773144) through in depth *in silico* analysis and follow-up functional studies[272]. From *in silico* analysis, all four SNPs showed some evidence of functionality, with rs3809346 representing the strongest functional candidate. Gel shift experiments suggested these four intronic SNPs do not alter protein binding and luciferase reporter assays interrogating these SNPs and the corresponding DNA region were inconclusive. Finally, eQTL studies showed carriers of the rs4773144 risk allele (G) showed a trend towards a decrease in type IV collagen mRNA expression in kidney[272].

For the rs9515203 SNP, located in intron 4 of the *COL4A2* gene, we demonstrated this SNP lies within an enhancer sequence in hASMCs. However, according to ENCODE ChIP-seq data rs9515203 is not predicted to disrupt binding of any transcription factors. In our luciferase assays in hASMCs rs9515203 does not affect luciferase activity of the enhancer constructs (**Figure 7**). While luciferase assays did not reveal a function for rs9515203, it is possible this SNP could affect chromatin accessibility in the genomic DNA that could not be captured *in vitro*. Finally, while rs9515203 lies within a functional enhancer, **Figure 6** shows there is substantial H3K4me3, indicative of promoter histone modifications. Future experiments interrogating whether the rs9515203 genomic region can act as a promoter may be warranted by cloning this sequence into pGL3-Basic. Interestingly, RNA Polymerase II is predicted to bind the rs9515203 sequence from ENCODE ChIP-seq data. There could potentially be a promoter for a noncoding RNA that rs9515203 could affect through unknown mechanisms.

We identified rs9515203 as a *cis*-eQTL for *COL4A2* (and *COL4A1* to a lesser degree) expression in ascending aortic intima-media (**Figure 8**), suggesting rs9515203 somehow regulates gene expression at this locus. Since smooth muscle cells are a major constituent of the intimal and medial layers of the coronary artery, this rs9515203 effect on type IV collagen (especially *COL4A2*) expression has a plausible functional impact on atherosclerosis progression. We were surprised that the minor, protective (C) allele at rs9515203 associates with lower *COL4A2* expression since the literature suggests type IV collagen is normally atheroprotective. Type IV collagen levels correlate with reduced proliferation and migration of vascular smooth muscle cells[100,105]. Lower type IV collagen could also be predicted to result in a less dense basement membrane and potential increased infiltration of monocytes, lipoproteins and T cells in the arterial wall. Of note, this ASAP eQTL effect is in healthy tissue, which may not accord with diseased plaque.

For rs9515203 and the rest of the seven FDR *COL4A1/COL4A2* SNPs, future experiments are required to determine how alterations in COL4A1/COL4A2 mRNA and protein levels affect risk of atherosclerosis. While most studies point toward an atheroprotective role for COL4A1 and COL4A2, a few studies have reported increased type IV collagen expression in advanced lesions as atherosclerosis progresses[111,112]. One of these studies of human atherosclerotic lesions found thick deposits of type IV collagen around elongated smooth muscle cells in advanced lesions[112]. None of the type IV collagen deposits in advanced lesions were located in the center of the atheroma. Occasional deposits of type IV collagen were observed in calcified tissues, along with vascularisation of lesions, where small vessels were delineated by type IV collagen[112]. While basement membranes were often thicker in regions of calcification in this study, we could not find any further links between type IV

collagen and plaque calcification in the literature. Though higher type IV collagen in advanced lesions could suggest type IV collagen promotes atherosclerosis, it could also be a by-product of the buildup of smooth muscle cells and phenotypic change back to the contractile state at advanced stages. Multi-layered basement membrane formation around smooth muscle cells has been proposed to be a defense mechanism against vascular hemodynamic forces and an indicator of cell senescence[112,273].

Another important aspect to note is that for both the above immunohistochemical analyses as well as many prior functional studies of type IV collagen and atherosclerosis, there is no way to delineate COL4A1 and COL4A2 from each other or from the other four alpha chains of type IV collagen. While COL4A1 and COL4A2 are the predominant forms of type IV collagen in basement membranes, we cannot assume COL4A1 and COL4A2 have the exact same role in CAD pathogenesis or that other forms of type IV collagen do not have an effect.

The rs4773144 and rs9515203 SNPs at the *COL4A1/COL4A2* locus demonstrate the difficulty in identifying causal CAD-associated SNPs. First, these polymorphisms are common and have modest effect sizes (odds ratios <1.10). According to the common disease-common variant hypothesis, where CAD is the result of the combination of many common variants of modest effect, these *COL4A1/COL4A2* SNPs would have subtle biological effects that are difficult to capture in functional assays. For example, in luciferase reporter assays, 5-10% effects of SNPs on promoter/enhancer activity are hard to capture with statistical significance. If CAD-associated SNPs had drastic deleterious effects on biological function, they would be selected against over the course of evolution and not be

common in the population. Second, SNP effects on gene expression relevant to disease are often cell-type specific[274,275] and specific cell culture conditions may be required to capture an effect. Third, there are many mechanisms (several potentially unknown) by which polymorphisms in noncoding regions can have functional effects. While effects on transcription factor binding and activity of gene regulatory elements are the most obvious candidates, intronic SNPs can affect chromatin looping independent of transactivation[276], can have epigenetic effects (ie. affect methylation)[277], or affect long noncoding RNA function[278]. Ultimately, ~50% of CAD SNPs lie outside ENCODE regulatory elements or coding regions in genes, suggesting the presence of unknown mechanisms[279].

In terms of future experiments of worth for the *COL4A1/COL4A2* locus, it would be beneficial to conduct a chromosome-conformation capture (3C) based technique. This would allow us to investigate the three-dimensional organization of the 13q34 locus and determine if the intronic *COL4A1/COL4A2* regions harboring CAD-associated SNPs interact with other genes *in trans*. Another beneficial future experiment will be to take advantage of CRISPR/Cas9 technology to delete the *COL4A1/COL4A2* intronic regions that contain the CAD-associated SNPs. To do this we would first generate human induced pluripotent stem cells (hiPSCs), which offer the advantages of cell plasticity, numerous passages, and genetic homogeneity. After using CRISPR/Cas9 to delete the intronic sequences, we could then differentiate the hiPSCs into either VSMCs or endothelial cells, the cell types relevant to atherosclerosis that produce basement membrane proteins. Depending on the findings of the 3C-based technique, we could then assess effects of deleting these regions on COL4A1 expression, COL4A2 expression, or expression of a distal gene. In VSMCs we could assess

effects of these CAD-associated intronic regions on parameters such as proliferation, migration, apoptosis, inflammation, and adhesion.

5 GENERAL DISCUSSION

Gene-gene interactions and pathway analysis are relevant in atherosclerosis

To date, CARDIoGRAMPlusC4D has reported 58 CAD-associated loci that have reached genome-wide significance and has further identified 202 FDR variants (q<0.05)[44]. These CAD-associated variants are a supplement to the hundreds of other common variants identified from GWAS associated with various other common and complex diseases, ranging from diabetes to psychiatric disorders to autoimmune disorders. For most complex traits and diseases, the proportion of disease heritability explained by SNPs reaching GWAS significance is approximately 10%. However, for some diseases such as Crohn's Disease and multiple sclerosis, or quantitative traits such as height or blood lipids, GWAS SNPs can explain approximately 10-20% of heritability[280]. Thus, for CAD and other complex human diseases, the question remains as to wherein all of the missing heritability lies. The recently published 2015 CARDIoGRAM study employing 1000 Genomes Project imputation helped to fill in some of the missing gaps of CAD heritability, explaining approximately 20% of the heritability of CAD[44]. The more GWAS can be applied to understand the heritability of CAD, the better the ability to clinically predict the occurrence of CAD and MI[281].

An important point to note is that the number of discovered variants reaching significance in a GWAS study is strongly correlated with the experimental sample size[280]. Therefore, increasing sample sizes should certainly increase the number of variants that reach statistical significance after accounting for multiple testing. Due to this multiple testing threshold, many important SNPs with functional roles in CAD pathogenesis have been missed due to not achieving the $p<5x10^{-8}$ threshold[280].

We are currently in the midst of the post-GWAS era, trying to understand the functional and biological implications of the hundreds of GWAS SNPs, including dozens of GWAS SNPs for CAD. As previously mentioned, some of these post-GWAS strategies include pathway analysis, gene-gene interactions, and gene-environment interactions. These strategies not only may uncover some of the missing statistical heritability, but also can shed light on the molecular mechanisms whereby CAD-associated loci are involved in disease. This is especially applicable for many of the CAD-associated loci not associated with traditional CAD risk factors.

Here, not only did we search for functional CAD-associated SNPs at the *COL4A1/COL4A2* and *SMAD3* loci independently, we sought to determine whether SNPs at these two loci display interaction in terms of CAD association. In five independent CAD/control cohorts, we identified epistasis between the (C) allele at rs72655775 in *COL4A2* and the (G) allele at rs12441344 in *SMAD3* (**Chapter 2, Table 1**) that synergistically increase CAD risk. This interaction had an odds ratio of 2.26 and was significant after correcting for multiple testing. Both rs72655775 and rs12441344 are intronic SNPs and in strong to moderate linkage disequilibrium with numerous other SNPs. The functional roles of these SNPs/linked SNPs need to be uncovered, including mechanisms that explain their synergistic effects on CAD association. While epistasis analysis was conducted in all of the cohorts for which our lab has genotype information for, replication in other CARDIoGRAM data sets would add even further strength to this genetic interaction between *COL4A1/COL4A2* and *SMAD3*.

From a literature search, we would predict lower levels of both COL4A1/COL4A2 and SMAD3 proteins would increase risk of atherosclerosis and CAD. In terms of biological

interaction between COL4A1/COL4A2 and SMAD3, we believe that hASMCs represent the cell line of most relevance. Type IV collagen expression is higher in smooth muscle cells with a quiescent, contractile phenotype compared to activated smooth muscle cells that are proliferating and migrating[100,105]. Second, most prior studies suggest TGFβ is anti-atherogenic, including inhibiting smooth muscle cell proliferation and migration[158]. Lowering both type IV collagen and SMAD3 (thus disrupting TGFβ signaling) would be predicted to increase smooth muscle cell proliferation and migration and be pro-atherogenic according to the traditional view of smooth muscle cell proliferation in CAD.

This traditional view that aberrant vascular smooth muscle cell (VSMC) proliferation promotes formation of plaques, whereas VSMCs in advanced plaques are beneficial by improving plaque stability has recently been challenged[282]. The traditional view was based on the assumption of a homogenous VSMC population in the vessel wall that are distinct from other cell types such as macrophages based on expressed cell-specific markers. More recent studies have demonstrated cells in the plaque expressing VSMC markers can arise from various origins, including macrophages. Furthermore, recent studies have also suggested VSMC proliferation can have reparative effects in the atherosclerotic plaque and is not a primary driver of plaque formation[282]. In contrast, VSMC cell death and senescence and now believed to strongly contribute to atherogenesis and plaque instability[282].

Further clarification as to whether smooth muscle cell proliferation is beneficial or detrimental in CAD pathogenesis is still required and functional experiments to decipher the mechanisms whereby rs72655775 (C) and rs12441344 (G) (or linked alleles) would synergistically increase CAD risk. Interrogating common pathways shared by CAD-

associated loci and epistasis is relevant in the sense that one of the ultimate goals of GWAS is to translate findings into clinical applications that can lead to the development of new therapeutic strategies[283]. The type IV collagen/SMAD3 signaling axis could represent one of these pathways. Our eQTL data from Chapters 3 and 4 is opposite of these predictions, where protective alleles for rs17293632/*SMAD3* (**Chapter 3, Figure 3**) and rs9515203/*COL4A2* (**Chapter 4, Figure 8**) associate with lower levels of the respective gene.

Post-GWAS characterization of CAD loci

In addition to interrogating pathways and epistasis, other post-GWAS strategies include fine-mapping disease-associated SNPs, using eQTL data to help interpret GWAS findings, allele-specific expression (ASE) experiments, and functional experiments both *in vitro* and *in vivo*[283,284]. With recent advances in imputation using 1000 Genomes Project data, it is currently much easier to fine-map SNPs and interrogate the full extent of genetic variation at a specific locus, including information of lower frequency variants[283]. One aspect of fine-mapping we did not investigate was the haplotype structure across populations, since it has been suggested studying different ethnic groups will help in this approach[283].

Identifying disease-related eQTLs are also a valuable post-GWAS resource in that correlation between genotype and gene expression is very useful in identify genes involved in complex diseases[251]. Recently the eQTL databases have greatly expanded to include many more tissues, including the availability of data from the GTEx project[260]. Since the effects of many SNPs on gene expression are cell and tissue specific[260,261], SNPs can now be

interrogated in tissues of relevance to disease pathogenesis. Furthermore, trait-associated SNPs are more likely to be eQTLs[250], which will aid in prioritizing functional SNPs at a locus and unravelling disease mechanisms. The evolution of systems genetics will aid tremendously in integrating genetic data with expression data, gene-gene interaction data, and functional data[220,283]. This systems genetics approach takes into account intermediate phenotypes alongside clinical phenotypes and considers gene-environment interactions[29].

One of the most notable findings from GWAS is that the majority of disease-associated (including CAD) SNPs are located in noncoding regions of the genome (intronic or intergenic)[44,278], which are becoming recognized as increasingly relevant[285]. There are few common CAD-associated nonsynonymous SNPs implicated in atherogenesis. One example is the *ADAMTS7* locus, where a nonsynonymous variant (rs3825807, Ser214Pro) may impair the function of the ADAMTS7 protein in VSMCs and associates with reduced CAD risk[284,286]. eQTL studies reveal ADAMTS7 protective alleles associate with lower ADAMTS7 expression and *Adamts7* deletion is atheroprotective in *Ldlr*$^{-/-}$ and *ApoE*$^{-/-}$ mice[284,287].

Functional elements located in noncoding regions of the genome could include promoter, enhancer, silencer, and insulator sequences that affect gene regulation. CAD-associated SNPs may also be in proximity or disrupt function of a noncoding RNA[284,288]. The CAD-associated SNPs reaching genome-wide significance at the *SMAD3* and *COL4A1/COL4A2* loci (and linked SNPs) all map to noncoding regions. We believe that the identification of rs17293632 in *SMAD3* as a novel CAD-associated SNP represents a very good example of the road from association to function. However, in the case of the rs4773144 and rs9515203

SNPs at the *COL4A1/COL4A2* locus, the mechanisms behind causality remain elusive, highlighting the likely complex molecular nature of SNP/trait associations.

Elsewhere, the identification and characterization of causal noncoding CAD-associated SNPs has only been reported for a small number of CARDIoGRAM loci. For example, the chromosome 1p13 locus has been robustly associated with CAD, in addition to LDL cholesterol and myocardial infarction. Musunuru *et al*. characterized a common noncoding SNP, rs12740374, that creates a new C/EBP binding site and alters the expression of the *SORT1* gene in the liver[251]. SORT1 subsequently regulates VLDL secretion by the liver via altering levels of plasma LDL cholesterol and VLDL particle levels[251]. Miller *et al*. characterized the rs12190287 SNP at the *TCF21* locus that disrupts an AP-1 binding site that regulates signaling through both platelet-derived growth factor β (PDGFβ) and Wilms tumor 1 (WT1) pathways[289]. rs12190287 alters transcription factor binding, chromatin organization, and TCF21 expression. Recently the rs93459379 SNP at 6p24 has been characterized as a causal variant that affects the levels of the *PHACTR1* gene[290]. This study demonstrated rs93459379 is an eQTL for PHACTR1 in human coronary arteries and affects binding of the MEF2 transcription factor in endothelial cell extracts. Deletion of the *PHACTR1* MEF2 binding site using CRISPR/Cas9 in heterozygous endothelial cells results in 35% less PHACTR1 expression[290]. Another recent study by Almontashiri *et al*. demonstrated that CAD-associated SNPs at 9p21.3 disrupt TEAD transcription factor binding and the TEAD3-dependent induction of p16 by TGFβ, partially accounting for the CAD risk at this locus[291].

GWAS SNPs that map to noncoding DNA regions could directly disrupt the binding sequence of a transcription factor or have an epigenetic effect and alter chromatin accessibility at these elements, though these are not mutually exclusive. In addition, many reported intronic variants can affect splicing, either by occurring within exon-intron splice junctions, affecting splice enhancer sequences, or activating cryptic splice sites[292]. Although rs17293632 directly disrupted a conserved, consensus AP-1 binding site, many intronic SNPs predicted to be causal do not lie within a transcription factor binding site. For example, in the study by Farh *et al*. assessing causal variants in autoimmune disease, only a fraction altered recognizable transcription factor sequence motifs[258]. However, noncoding disease variants that are not in a transcription factor-binding motif may still be able to have effects on gene regulation. Biochemical and genetic manipulations have revealed motifs adjacent to transcription factor binding sites have the potential to influence transcription factor activity[293]. Support for functional roles of non-canonical sequences is from the high sequence conservation of enhancers outside of known transcription factor motifs and the complex structural interactions and looping of enhancers[258]. Thus the rs4773144 and rs9515203 *COL4A1/COL4A2* SNPs may affect gene regulation via one of these mechanisms or a further, unknown mechanism.

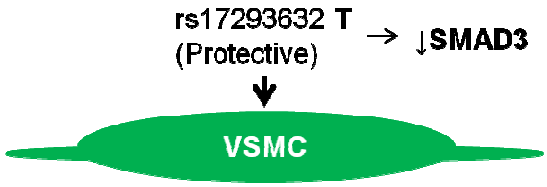
Recently, many new strategies are being developed to prioritize causal variants that are noncoding. One example of an experimental approach is ATAC-seq (Assay for Transposase Accessible Chromatin), that relies on direct *in vitro* transposition of sequencing adapters into regions of open chromatin using hyperactive Tn5 transposase[294]. Therefore, amplifiable DNA fragments suitable for downstream high throughout DNA sequencing are preferentially generated at open chromatin positions. In contrast, Tn5 transposase has

difficulty accessing areas of less open/closed chromatin due to steric hindrance[294]. The ATAC-seq procedure requires fewer cells and less starting material compared to other methods that gauge chromatin accessibility, making in suitable for application in primary cells. With respect to functional annotations, several statistical algorithms and approaches have recently been developed that aim to better prioritize causal variants[295].

Further insight into the rs17293632 SNP at the *SMAD3* locus

As mentioned, we chose hASMCs for the majority of our experiments investigating the rs17293632 SNP due to coronary artery smooth muscle cells composing up to 80% of the contents of the atherosclerotic plaque[24] and expressing most of the reported CAD-associated genes. The phenotypic modulation of arterial SMCs in the transition from a contractile state to a proliferative and synthetic state is also key in CAD pathogenesis[11,296]. As shown in **Chapter 3, Figure 6**, knocking down SMAD3 increases hASMC proliferation in two separate lots of cells, suggestive of anti-proliferative role of SMAD3. Thus, based on the traditional view of VSMC proliferation it is not apparent how the rs17293632 (T) allele, that decreases SMAD3 levels and would be expected to increase smooth muscle cell proliferation, would confer protection from CAD. However, as has been recently postulated, smooth muscle cell proliferation may have a reparative effect in the plaque[282]. The (T) allele for rs17293632 that associates with lower SMAD3 levels may actually be beneficial according to this more recent view by promoting repairs of ruptured plaques (**Figure 1**).

Since SMAD3 expression in plaque VSMCs increases as atherosclerosis progresses[175], perhaps lower SMAD3 confers protection through alternative mechanisms. In human



Senescence

↓**SMAD3** = ↓ Senescence
↓ Inflammation
↓ Adhesion Molecule Expression (eg. ICAM-1)
↓ Chemokine Production (eg. MCP-1)
↓ Innate Immune Receptor Production (eg. Toll-Like Receptor 4)

Apoptosis

↓**SMAD3** = ↓ Apoptosis
↑ VSMC Content
↑ Plaque Stability

Proliferation

↓**SMAD3** = ↑ VSMC Proliferation
↑ Ability to Repair Ruptured Plaques

Figure 1. Potential mechanisms whereby the rs17293632 (T) allele could have a protective effect against CAD in VSMCs. At late stages of atherosclerosis, VSMC proliferation is postulated to be beneficial by aiding in the repair of ruptured plaques. In contrast, VSMC apoptosis and senescence are believed to be detrimental by reducing plaque stability and promoting an inflammatory environment. Based off the Review by Bennett *et al.*[282]

plaques VSMCs display both increased susceptibility to apoptosis as well as early and increased senescence[282]. Although we did not observe any changes in apoptosis from **Chapter 3**, **Supplementary Figure VII**, there are indications SMAD3 can promote both apoptosis[254] and senescence[297]. In these cases, the protective (T) allele for rs17293632 that reduces SMAD3 levels would reduce both VSMC apoptosis and senescence. This would result in a reduced inflammatory state and a more stable plaque (**Figure 1**).

Together, VSMC apoptosis and senescence promote inflammation, recruitment of monocytes, and secretion of mitogens by VSMCs[282]. With respect to apoptosis, VSMC death results in a less stable plaque that is prone to rupture and symptomatic plaques exhibit increased levels of VSMC apoptosis compared to stable lesions. With respect to senescence, plaque VSMCs have shortened telomeres (a hallmark of senescence) compared to the healthy vessel wall[282]. Furthermore, aged VSMCs have increased IL-6 levels, increased MCP-1 levels, and upregulation of adhesion molecules such as ICAM-1, generating a pro-inflammatory environment[282].

Although rs17293632 disrupts AP-1 binding and *SMAD3* enhancer activity in hASMCs, it is very possible that this SNP is functional in other cell types in the vessel wall or plaque relevant to CAD. It is surprising the protective allele (T) at rs17293632 lowers *SMAD3* mRNA levels both *in vitro* and *in vivo*. Since most evidence points towards TGFβ signaling being anti-atherogenic, and SMAD3 is a critical mediator of TGFβ signaling, we would have expected the protective allele to increase SMAD3 levels across various cell types. In endothelial cells TGFβ inhibits expression of adhesion molecules that are involved in leukocyte recruitment[298]. In cultured macrophages TGFβ reduces foam cell formation[170],

and has a role in inhibiting the activation and proliferation of T cells[171]. Finally, TGFβ inhibits VSMC proliferation and migration and increases expression of extracellular matrix proteins that can stabilize plaques. We believe the direction of effect of rs17293632 on enhancer activity is consistent across many cell lines in tissues from both of luciferase experiments and our *in vivo* data in whole blood and carotid plaque.

Therefore, the role of SMAD3 in atherosclerosis remains to be elucidated. Searching other GWAS, we did not identify an association of the *SMAD3* locus with coronary artery calcification or traditional CAD risk factors. While not mentioned in earlier sections, TGFβ/SMAD3 may play a role in recruiting monocytes to the sites of early atherosclerotic lesions through increasing expression of monocyte chemoattractant protein-1 (MCP-1). TGFβ stimulated production of MCP-1 by vascular smooth muscle cells and in a rat carotid injury model, *Smad3* overexpression significantly increased MCP-1 expression after vascular injury[299]. A similar effect is observed in human endothelial cells, where TGFβ via SMAD3 also increases MCP-1 levels[300]. However, another group reported TGFβ1 through Smad3 in mice inhibits MCP-1 expression in macrophages[301], suggesting a possible cell-type specific effect that warrants further investigation. TGFβ also stimulates leukocyte chemotaxis and production of proteoglycans by smooth muscle cells that could also contribute to early monocyte recruitment[159].

As shown in **Chapter 3, Table 2**, rs17293632 is in very strong linkage disequilibrium with six other SNPs, all located towards the 3' end of the *SMAD3* gene. We chose to focus on the rs17293632 SNP due to bioinformatics data and RegulomeDB indicating it was by the far the strongest functional candidate of the seven SNPs. Again, although we demonstrate

rs172923632 as causal SNP, we cannot discount any potential functionality for the six other SNPs. While no obvious putative functional roles for these other *SMAD3* SNPs was apparent, it remains possible that they affect gene regulation through an unknown mechanism. Although unlikely, one or more of these linked SNPs could have additive effects alongside rs17293632. Ideally it would have been beneficial to test the genomic DNA regions of these linked SNPs for gene regulatory activity using reporter assays, similar to those conducted for rs17293632. However, this would have required exhaustive, numerous cloning steps alongside many site-directed mutagenesis reactions. Therefore, for practicality we did not pursue investigating all of the less promising CAD-associated *SMAD3* SNPs.

With respect to future directions with the rs17293632 SNP, although we demonstrated an effect on SMAD3, experiments such as a 3C-based method would be beneficial to determine if the rs17293632 enhancer sequence could interact with and regulate gene expression of another locus *in trans*. These 3C-based methods would be able to capture potential chromatin looping effects and capture three-dimensional chromatin architecture.

Since we demonstrated rs17293632 is a functional CAD-associated SNP and highlighted that SMAD3 is likely to have important functions in CAD pathogenesis, the next steps will be to further clarify SMAD3's targets in cells/tissues of relevance to CAD. To our knowledge no SMAD3 ChIP-seq has been conducted with respect to the binding of SMAD3 in vascular cell lines or tissues relevant to CAD. It will be noteworthy to determine whether SMAD3 binds to promoters or regulatory elements in proximity to other CAD-associated genes or genes with important roles in the vasculature. Furthermore, ChIP-seq in cell

lines/tissues relevant to CAD may help clarify the indirect regulation of the *COL4A1* and *COL4A2* genes by SMAD3.

Other future experiments to more directly interrogate the function of the rs17293632 SNP/*SMAD3* intron 1 enhancer would be to use hiPSC technology along with CRISPR/Cas9 technology. The advantages of hiPSC technology are that they come from an isogenic background and can be differentiated into numerous cell types. Once the hiPSCs have been obtained, CRISPR/Cas9 could be employed to either delete the entire *SMAD3* intron 1 enhancer sequence or to modify the rs17293632 SNP alone. The hiPSCs could then be differentiated into various cell types relevant to atherosclerosis, such as VSMCs, endothelial cells, or myeloid cells. Once the hiPSCs have been differentiated, we could then design experiments (depending on the cell type) and assess effects of rs17293632/*SMAD3* intron 1 on parameters such as gene expression, proliferation, inflammation, and adhesion molecule expression.

Conclusions

Although nearly 60 CAD-associated loci have now been uncovered from published GWAS, identification of causal genes and variants has been limited. The high frequency of common CAD-associated polymorphisms in noncoding DNA sequences such as intergenic regions or introns highlights the importance of noncoding regulatory elements in complex disease. These GWAS findings have revealed novel genes and pathways in the pathogenesis of atherosclerosis, several of which may lead to new CAD therapies in the future.

The studies conducted in this thesis attempted to uncover the molecular mechanisms behind association of the following two novel loci with CAD: *COL4A1/COL4A2* and *SMAD3*. For the *SMAD3* locus at chromosome 15q23, I show that the rs17293632 SNP, tightly linked with the rs56062135 index SNP, disrupts an AP-1 binding sequence in an intronic enhancer. The rs17293632 protective allele (T) lowers AP-1 transcription factor binding, enhancer activity, and associates with lower SMAD3 mRNA *in vivo*. For the *COL4A1/COL4A2* locus at chromosome 13q34, the effects of rs4773144, rs9515203, and five other independent CAD-associated SNPs still need to be elucidated. Finally, COL4A1/COL4A2 and SMAD3 demonstrate both biological interaction and statistical interaction for CAD association. Future directions from this project include clarifying the roles of the COL4A1, COL4A2, and SMAD3 proteins in atherosclerosis progression and the cell types where they have effects. Genome editing techniques such as CRISPR/Cas9 would also be beneficial to directly interrogate the roles of rs17293632 and the CAD-associated *COL4A1/COL4A2* SNPs in a cell type relevant to CAD.

6 REFERENCES

1. World Health Organization. The 10 leading causes of death in the world, 2000 and 2012. 2013. http://www.who.int/mediacentre/factsheets/fs310/en/.
2. Heidenreich PA, Trogdon JG, Khavjou OA, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation*. 2011;123(8):933-944. doi:10.1161/CIR.0b013e31820a55f5.
3. Mozaffarian D, Benjamin EJ, Go AS, et al. *Heart Disease and Stroke Statistics--2015 Update: A Report from the American Heart Association*. Vol 131. 2015. doi:10.1161/CIR.0000000000000152.
4. Kessler T, Erdmann J, Schunkert H. Genetics of coronary artery disease and myocardial infarction - 2013. *Curr Cardiol Rep*. 2013;15(6). doi:10.1007/s11886-013-0368-0.
5. O'Donnell CJ, Nabel EG. Genomics of cardiovascular disease. *N Engl J Med*. 2011;365:2098-2109.
6. Glass CK, Witztum JL. Atherosclerosis: the road ahead. *Cell*. 2001;104(4):503-516. doi:10.1016/S0092-8674(01)00238-0.
7. Lusis AJ. Atherosclerosis. *Nature*. 2000;407(6801):233-241. doi:10.1038/35025203.
8. Ross R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature*. 1993;362(6423):801-809. doi:10.1038/362801a0.
9. Rader DJ, Daugherty A. Translating molecular discoveries into new therapies for atherosclerosis. *Nature*. 2008;451(7181):904-913. doi:10.1038/nature06796.
10. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *New Engl J*. 2005;352:1685-1695. doi:10.1056/NEJM200507283530425.
11. Tabas I, García-Cardeña G, Owens GK. Recent insights into the cellular biology of atherosclerosis. *J Cell Biol*. 2015;209(1):13-22. doi:10.1083/jcb.201412052.
12. Moore KJ, Sheedy FJ, Fisher EA. Macrophages in atherosclerosis: a dynamic balance. *Nat Rev Immunol*. 2013;13(10):709-721. doi:10.1038/nri3520.
13. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. *Nat Med*. 2011;17(11):1410-1422. doi:10.1038/nm.2538.
14. Kwon GP, Schroeder JL, Amar MJ, Remaley AT, Balaban RS. Contribution of macromolecular structure to the retention of low-density lipoprotein at arterial branch points. *Circulation*. 2008;117(22):2919-2927. doi:10.1161/CIRCULATIONAHA.107.754614.
15. Libby P. Inflammation in atherosclerosis. *Nature*. 2002;420(6917):868-874. doi:10.1038/nature01323.
16. Weissberg PL. Coronary disease: Atherogenesis: current understanding of the causes of atheroma. *Heart*. 2000;83(2):247-252. doi:10.1136/heart.83.2.247.
17. Libby P. Pathophysiology of coronary artery disease. *Circulation*. 2005;111(25):3481-3488. doi:10.1161/CIRCULATIONAHA.105.537878.
18. Geng Y-J. Progression of atheroma: a struggle between death and procreation. *Arterioscler Thromb Vasc Biol*. 2002;22(9):1370-1380. doi:10.1161/01.ATV.0000031341.84618.A4.

19. Bogdanov VY, Balasubramanian V, Hathcock J, Vele O, Lieb M, Nemerson Y. Alternatively spliced human tissue factor: a circulating, soluble, thrombogenic protein. *Nat Med*. 2003;9(4):458-462. doi:10.1038/nm841.
20. Robbins CS, Hilgendorf I, Weber GF, et al. Local proliferation dominates lesional macrophage accumulation in atherosclerosis. *Nat Med*. 2013;19(9):1166-1172. doi:10.1038/nm.3258.
21. Demer LL. Vascular calcification and osteoporosis: inflammatory responses to oxidized lipids. *Int J Epidemiol*. 2002;31(4):737-741. doi:10.1093/ije/31.4.737.
22. Libby P, Aikawa M. Stabilization of atherosclerotic plaques: new mechanisms and clinical targets. *Nat Med*. 2002;8(11):1257-1262. doi:10.1038/nm1102-1257.
23. Moore KJ, Tabas I. Macrophages in the pathogenesis of atherosclerosis. *Cell*. 2011;145(3):341-355. doi:10.1016/j.cell.2011.04.005.
24. Shankman LS, Gomez D, Cherepanova OA, et al. KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat Med*. 2015;21(6):628-637. doi:10.1038/nm.3866.
25. Gomez D, Owens GK. Smooth muscle cell phenotypic switching in atherosclerosis. *Cardiovasc Res*. 2012;95(2):156-164. doi:10.1093/cvr/cvs115.
26. Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25-33. doi:10.1038/ng.2480.
27. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: A 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002;252(3):247-254. doi:10.1046/j.1365-2796.2002.01029.x.
28. Williams RR, Hunt SC, Heiss G, et al. Usefulness of cardiovascular family history data for population-based preventive medicine and medical research (The Health Family Tree Study and the NHLBI Family Heart Study). *Am J Cardiol*. 2001;87(2):129-135. doi:10.1016/S0002-9149(00)01303-5.
29. Lusis AJ. Genetics of atherosclerosis. *Trends Genet*. 2012;28(6):267-275. doi:10.1016/j.tig.2012.03.001.
30. Musunuru K, Kathiresan S. Genetics of coronary artery disease. *Annu Rev Genomics Hum Genet*. 2010;11:91-108. doi:10.1146/annurev-genom-082509-141637.
31. Watkins H, Farrall M. Genetic susceptibility to coronary artery disease: from promise to progress. *Nat Rev Genet*. 2006;7(3):163-173. doi:10.1038/nrg1805.
32. Patel RS, Ye S. Genetic determinants of coronary heart disease: new discoveries and insights from genome-wide association studies. *Heart*. 2011;97(18):1463-1473. doi:10.1136/hrt.2010.219675.
33. Manolio TA. Genomewide association studies and assessment of risk of disease. *N Engl J Med*. 2010;363(21):166-176. doi:10.1056/NEJMc1010310#SA1.
34. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-753. doi:10.1038/nature08494.
35. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol*. 2012;8(12):e1002822. doi:10.1371/journal.pcbi.1002822.
36. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316(5830):1488-1491. doi:10.1126/science.1142447.

37. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316(5830):1491-1493. doi:10.1126/science.1142842.
38. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443-453.
39. Kim WY, Sharpless NE. The regulation of INK4/ARF in cancer and aging. *Cell*. 2006;127(2):265-275. doi:10.1016/j.cell.2006.10.003.
40. Pasmant E, Laurendeau I, Héron D, Vidaud M, Vidaud D, Bièche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res*. 2007;67(8):3963-3969. doi:10.1158/0008-5472.CAN-06-2004.
41. Jarinova O, Stewart AFR, Roberts R, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol*. 2009;29(10):1671-1677. doi:10.1161/ATVBAHA.109.189522.
42. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43(4):333-338. doi:10.1038/ng.784.
43. Peden JF, Hopewell JC, Saleheen D, et al. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43(4):339-344. doi:10.1038/ng.782.
44. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130. doi:10.1038/ng.3396.
45. Ghosh S, Vivar J, Nelson CP, et al. Systems genetics analysis of genome-wide association study reveals novel associations between key biological processes and coronary artery disease. *Arterioscler Thromb Vasc Biol*. 2015;35(7):1712-1722. doi:10.1161/ATVBAHA.115.305513.
46. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. *Nat Rev Genet*. 2014;15(11):722-733. doi:10.1038/nrg3747.
47. Hemani G, Shakhbazov K, Westra H-J, et al. Detection and replication of epistasis influencing transcription in humans. *Nature*. 2014;508(7495):249-253. doi:10.1038/nature13005.
48. Sun X, Lu Q, Mukheerjee S, Crane PK, Elston R, Ritchie MD. Analysis pipeline for the epistasis search - statistical versus biological filtering. *Front Genet*. 2014;5(106):1-7. doi:10.3389/fgene.2014.00106.
49. Ma L, Clark AG, Keinan A. Gene-based testing of interactions in association studies of quantitative traits. *PLoS Genet*. 2013;9(2):e1003321. doi:10.1371/journal.pgen.1003321.
50. Sun Y V. Integration of biological networks and pathways with genetic association studies. *Hum Genet*. 2012;131(10):1677-1686. doi:10.1007/s00439-012-1198-7.
51. Sun Y V, Kardia SLR. Identification of epistatic effects using a protein-protein interaction database. *Hum Mol Genet*. 2010;19(22):4345-4352. doi:10.1093/hmg/ddq356.

52. Araújo MA, Goulart LR, Cordeiro ER, et al. Genotypic interactions of renin-angiotensin system genes in myocardial infarction. *Int J Cardiol*. 2005;103(1):27-32. doi:10.1016/j.ijcard.2004.07.009.
53. Tsai C, Hwang J, Ritchie MD, et al. Renin-angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: Detection of high order gene-gene interaction. *Atherosclerosis*. 2007;195(1):172-180. doi:10.1016/j.atherosclerosis.2006.09.014.
54. Lucas G, Lluís-Ganella C, Subirana I, et al. Hypothesis-based analysis of gene-gene interactions and risk of myocardial infarction. *PLoS One*. 2012;7(8):e41730. doi:10.1371/journal.pone.0041730.
55. Musameh MD, Wang WYS, Nelson CP, et al. Analysis of gene-gene interactions among common variants in candidate cardiovascular genes in coronary artery disease. *PLoS One*. 2015;10(2):e0117684. doi:10.1371/journal.pone.0117684.
56. Griffin CA, Emanuel BS, Hansen JR, Caunee WK, Myers JC. Human collagen genes encoding basement membrane α1(IV) and α2(IV) chains map to the distal long arm of chromosome 13. *Proc Natl Acad Sci USA*. 1987;84(2):512-516.
57. Soininen R, Huotari M, Ganguly A, Prockop DJ, Tryggvason K. Structural organization of the gene for the alpha 1 chain of human type IV collagen. *J Biol Chem*. 1989;264(23):13565-13571. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2701944.
58. Tryggvason K, Soininen R, Hostikka SL, Ganguly A, Huotari M, Prockop DJ. Structure of the human type IV collagen genes. *Ann N Y Acad Sci*. 1990;580:97-111. http://www.ncbi.nlm.nih.gov/pubmed/11068203.
59. Pöschl E, Pollner R, Kühn K. The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure. *EMBO J*. 1988;7(9):2687-2695. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=457057&tool=pmcentrez&rendertype=abstract.
60. Soininen R, Huotari M, Tryggvason K. The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. *J Biol Chem*. 1988;2(14):17217-17220.
61. Burbelo PD, Martin GR, Yamada Y. Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. *Proc Natl Acad Sci U S A*. 1988;85(24):9679-9682. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=282835&tool=pmcentrez&rendertype=abstract.
62. Pollner R, Fischer G, Pöschl E, Kühn K. Regulation of divergent transcription of the genes coding for basement membrane type IV collagen. *Ann N Y Acad Sci*. 1990;580:44-54. http://www.ncbi.nlm.nih.gov/pubmed/2337305.
63. Pollner R, Schmidt C, Fischer G, Kühn K, Pöschl E. Cooperative and competitive interactions of regulatory elements are involved in the control of divergent transcription of human Col4A1 and Col4A2 genes. *FEBS Lett*. 1997;405(1):31-36. doi:10.1016/S0014-5793(97)00152-X.

64. Johnston M, Davis RW. Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae. *Mol Cell Biol*. 1984;4(8):1440-1448. doi:10.1128/MCB.4.8.1440.
65. Osley MA, Gould J, Kim S, Kane MY, Hereford L. Identification of sequences in a yeast histone promoter involved in periodic transcription. *Cell*. 1986;45(4):537-544. doi:3518945.
66. Garabedian MJ, Hung MC, Wensink PC. Independent control elements that determine yolk protein gene expression in alternative Drosophila tissues. *Proc Natl Acad Sci U S A*. 1985;82(5):1396-1400. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=397268&tool=pmcentrez&rendertype=abstract.
67. Williams TJ, Fried M. The MES-1 murine enhancer element is closely associated with the heterogeneous 5' ends of two divergent transcription units. *Mol Cell Biol*. 1986;6(12):4558-4569. doi:10.1128/MCB.6.12.4558.Updated.
68. Gidoni D, Kadonaga JT, Barrera-Saldaña H, Takahashi K, Chambon P, Tjian R. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. *Science*. 1985;230(4725):511-517. doi:10.1126/science.2996137.
69. Natarajan V, Madden MJ, Salzman NP. Proximal and distal domains that control in vitro transcription of the adenovirus IVα2 gene. *Proc Natl Acad Sci U S A*. 1984;81(20):6290-6294. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6593702.
70. Trinklein NT, Force Aldred S, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. An abundance of bidirectional promoters in the human genome. *Genome Res*. 2004;14(1):62-66. doi:10.1101/gr.1982804.
71. Fischer G, Schmidt C, Opitz J, Cully Z, Kühn K, Pöschl E. Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription. *Biochem J*. 1993;292:687-695. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1134168&tool=pmcentrez&rendertype=abstract.
72. Schmidt C, Fischer G, Kadner H, Genersch E, Kühn K, Pöschl E. Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. *Biochim Biophys Acta*. 1993;1174(1):1-10. http://www.ncbi.nlm.nih.gov/pubmed/8334157.
73. Boot-Handford RP, Kurkinen M, Prockop DJ. Steady-state levels of mRNAs coding for the type IV collagen and laminin polypeptide chains of basement membranes exhibit marked tissue-specific stoichiometric variations in the rat. *J Biol Chem*. 1987;262(26):12475-12478. http://www.ncbi.nlm.nih.gov/pubmed/3624269.
74. Kleinman HK, Ebihara I, Killen PD, et al. Genes for basement membrane proteins are coordinately expressed in differentiating F9 cells but not in normal adult murine tissues. *Dev Biol*. 1987;122(2):373-378. doi:0012-1606(87)90302-2 [pii].
75. Van Agtmael T, Bruckner-Tuderman L. Basement membranes and human disease. *Cell Tissue Res*. 2010;339(1):167-188. doi:10.1007/s00441-009-0866-y.
76. Kalluri R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat Rev Cancer*. 2003;3(6):422-433. doi:10.1038/nrc1094.

77. Yurchenco PD. Basement membranes: Cell scaffoldings and signaling platforms. *Cold Spring Harb Perspect Biol*. 2011;3(2):1-27. doi:10.1101/cshperspect.a004911.
78. Colognato H, Yurchenco PD. Form and function: The laminin family of heterotrimers. *Dev Dyn*. 2000;218(2):213-234. doi:10.1002/(SICI)1097-0177(200006)218:2<213::AID-DVDY1>3.0.CO;2-R.
79. Khoshnoodi J, Pedchenko V, Hudson BG. Mammalian collagen IV. *Microsc Res Tech*. 2008;71(5):357-370. doi:10.1002/jemt.20564.
80. Prockop DJ, Kivirikko KI. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem*. 1995;64:403-434. doi:10.1146/annurev.bi.64.070195.002155.
81. Mao M, Alavi M V., Labelle-Dumais C, Gould DB. *Type IV Collagens and Basement Membrane Diseases: Cell Biology and Pathogenic Mechanisms*. Vol 76. Elsevier Ltd; 2015. doi:10.1016/bs.ctm.2015.09.002.
82. Hudson BG, Stephen T, Tryggvasonll K. Type IV collagen: structure, gene organization, and role in human diseases. *J Biol Chem*. 1993;3(38).
83. Ono T, Miyazaki T, Ishida Y, Uehata M, Nagata K. Direct in vitro and in vivo evidence for interaction between Hsp47 protein and collagen triple helix. *J Biol Chem*. 2012;287(9):6810-6818. doi:10.1074/jbc.M111.280248.
84. Koide T, Aso A, Yorihuzi T, Nagata K. Conformational requirments of collagenous peptides for recognition by the chaperone protein HSP47. *J Biol Chem*. 2000;275(36):27957-27963.
85. Malhotra V, Erlmann P, Nogueira C. Procollagen export from the endoplasmic reticulum. *Biochem Soc Trans*. 2015;43(1):104-107. doi:10.1042/BST20140286.
86. Khoshnoodi J, Sigmundsson K, Cartailler J-P, Bondar O, Sundaramoorthy M, Hudson BG. Mechanism of chain selection in the assembly of collagen IV: a prominent role for the α2 chain. *J Biol Chem*. 2006;281(9):6058-6069. doi:10.1074/jbc.M506555200.
87. Vandenberg P, Ries A, Luckenbill-edds L. Characterization of a type IV collagen major cell binding site with affinity to the α1β1 and the α2β1 integrins. *J Cell Biol*. 1991;113(6):1475-1483.
88. Xu J, Rodriguez D, Petitclerc E, et al. Proteolytic exposure of a cryptic site within collagen type IV is required for angiogenesis and tumor growth in vivo. *J Cell Biol*. 2001;154(5):1069-1079. doi:10.1083/jcb.200103111.
89. Xu J, Rodriguez D, Kim JJ, Brooks PC. Generation of monoclonal antibodies to cryptic collagen sites by using subtractive immunization. *Hybridoma*. 2000;19(5):375-385. doi:10.1089/02724570050198893.
90. Sado Y, Kagawa M, Naito I, et al. Organization and expression of basement membrane collagen IV genes and their roles in human disorders. *J Biochem*. 1998;123(5):767-776. http://www.ncbi.nlm.nih.gov/pubmed/9562604.
91. Madri JA. Extracellular matrix modulation of vascular cell behaviour. *Transpl Immunol*. 1997;5(3):179-183. doi:10.1016/S0966-3274(97)80035-4.
92. Pöschl E, Schlötzer-Schrehardt U, Brachvogel B, Saito K, Ninomiya Y, Mayer U. Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development*. 2004;131(7):1619-1628. doi:10.1242/dev.01037.

93. Colorado PC, Torre A, Kamphaus G, et al. Anti-angiogenic cues from vascular basement membrane collagen. *Cancer Res*. 2000;60(9):2520-2526. http://www.ncbi.nlm.nih.gov/pubmed/10811134.
94. Kamphaus GD, Colorado PC, Panka DJ, et al. Canstatin, a novel matrix-derived inhibitor of angiogenesis and tumor growth. *J Biol Chem*. 2000;275(2):1209-1215. http://www.ncbi.nlm.nih.gov/pubmed/10625665.
95. Petitclerc E, Boutaud A, Prestayko A, et al. New functions for non-collagenous domains of human collagen type IV. *J Biol Chem*. 2000;275(11):8051-8061.
96. Sudhakar A, Boosani CS. Signaling mechanisms of endogenous angiogenesis inhibitors derived from type IV collagen. *Gene Regul Syst Bio*. 2007;1:217-226. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2759143&tool=pmcentrez&rendertype=abstract.
97. Mundel TM, Kalluri R. Type IV collagen-derived angiogenesis inhibitors. *Microvasc Res*. 2007;74(2-3):85-89. doi:10.1016/j.mvr.2007.05.005.
98. Monboisse JC, Oudart JB, Ramont L, Brassart-Pasco S, Maquart FX. Matrikines from basement membrane collagens: A new anti-cancer strategy. *Biochim Biophys Acta - Gen Subj*. 2014;1840(8):2589-2598. doi:10.1016/j.bbagen.2013.12.029.
99. Assadian S, El-Assaad W, Wang XQD, et al. p53 inhibits angiogenesis by inducing the production of arresten. *Cancer Res*. 2012;72(5):1270-1279. doi:10.1158/0008-5472.CAN-11-2348.
100. Adiguzel E, Ahmad PJ, Franco C, Bendeck MP. Collagens in the progression and complications of atherosclerosis. *Vasc Med*. 2009;14(1):73-89. doi:10.1177/1358863X08094801.
101. Barnes MJ, Farndale RW. Collagens and atherosclerosis. *Exp Gerontol*. 1999;34(4):513-525. http://www.ncbi.nlm.nih.gov/pubmed/10817807.
102. Chistiakov DA, Sobenin IA, Orekhov AN. Vascular extracellular matrix in atherosclerosis. *Cardiol Rev*. 2013;21(6):270-288. doi:10.1097/CRD.0b013e31828c5ced.
103. Campbell JH, Campbell GR. The role of smooth muscle cells in atherosclerosis. *Curr Opin Lipidol*. 1994;5(5):323-330.
104. Smith E. The influence of age and atherosclerosis on the chemistry of aortic intima. 2. Collagen and mucopolysaccharides. *J Atheroscler Res*. 1965;5(2):241-248.
105. Okada Y, Katsuda S. Collagen synthesis by cultured arterial smooth muscle cells during spontaneous phenotypic modulation. *Acta Pathol Jpn*. 1990;40:157-164.
106. Hultgårdh-Nilsson A, Lövdahl C, Blomgren K, Kallin B, Thyberg J. Expression of phenotype- and proliferation-related genes in rat aortic smooth muscle cells in primary culture. *Cardiovasc Res*. 1997;34(2):418-430. doi:10.1016/S0008-6363(97)00030-8.
107. Seki T, Naito I, Oohashi T, Sado Y, Ninomiya Y. Differential expression of type IV collagen isoforms, alpha5(IV) and alpha6(IV) chains, in basement membranes surrounding smooth muscle cells. *Histochem Cell Biol*. 1998;110(4):359-366. http://www.ncbi.nlm.nih.gov/pubmed/9792414.
108. Orr WA, Lee MY, Lemmon JA, et al. Molecular mechanisms of collagen isotype-specific modulation of smooth muscle cell phenotype. *Arterioscler Thromb Vasc Biol*. 2009;29(2):225-231. doi:10.1161/ATVBAHA.108.178749.

109. Moiseeva EP. Adhesion receptors of vascular smooth muscle cells and their functions. *Cardiovasc Res*. 2001;52(3):372-386. http://www.ncbi.nlm.nih.gov/pubmed/11738054.
110. Nelson PR, Yamamura S, Kent KC. Extracellular matrix proteins are potent agonists of human smooth muscle cell migration. *J Vasc Surg*. 1996;24(1):25-33. doi:10.1016/S0741-5214(96)70141-6.
111. Murata K, Motayama T, Kotake C. Collagen types in various layers of the human aorta and their changes with the atherosclerotic process. *Atherosclerosis*. 1986;60(3):251-262. doi:10.1016/0021-9150(86)90172-3.
112. Katsuda S, Okada L, Minamoto T, Oda Y, Matsui Y, Nakanishi I. Collagens in human atherosclerosis. Immunohistochemical analysis using collagen type-specific antibodies. *Arter Thromb Vasc Biol*. 1992;12:494-502.
113. Vogel W, Brakebusch C, Fässler R, Alves F, Ruggiero F, Pawson T. Discoidin domain receptor 1 is activated independently of beta(1) integrin. *J Biol Chem*. 2000;275(8):5779-5784. http://www.ncbi.nlm.nih.gov/pubmed/10681566.
114. El Khoury J, Thomas CA, Loike JD, Hickman SE, Cao L, Silverstein SC. Macrophages adhere to glucose-modified basement membrane collagen IV via their scavenger receptors. *J Biol Chem*. 1994;269(14):10197-10200.
115. Van Agtmael T, Bailey MA, Schlötzer-Schrehardt U, et al. Col4a1 mutation in mice causes defects in vascular function and low blood pressure associated with reduced red blood cell volume. *Hum Mol Genet*. 2010;19(6):1119-1128. doi:10.1093/hmg/ddp584.
116. Tarasov K V, Sanna S, Scuteri A, et al. COL4A1 is associated with arterial stiffness by genome-wide association scan. *Circ Cardiovasc Genet*. 2009;2(2):151-158. doi:10.1161/CIRCGENETICS.108.823245.
117. Yamada Y, Kato K, Oguri M, et al. Genetic risk for myocardial infarction determined by polymorphisms of candidate genes in a Japanese population. *J Med Genet*. 2008;45(4):216-221. doi:10.1136/jmg.2007.054387.
118. O'Donnell CJ, Kavousi M, Smith A V, et al. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation*. 2011;124(25):2855-2864.
119. Gould DB, Phalan FC, Breedveld GJ, et al. Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science*. 2005;308(5725):1167-1171. doi:10.1126/science.1109418.
120. Breedveld G, de Coo IF, Lequin MH, et al. Novel mutations in three families confirm a major role of COL4A1 in hereditary porencephaly. *J Med Genet*. 2006;43(6):490-495. doi:10.1136/jmg.2005.035584.
121. Weng YC, Sonni A, Labelle-Dumais C, et al. COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. *Ann Neurol*. 2012;71(4):470-477. doi:10.1002/ana.22682.
122. Yoneda Y, Haginoya K, Kato M, et al. Phenotypic spectrum of COL4A1 mutations: Porencephaly to schizencephaly. *Ann Neurol*. 2013;73(1):48-57. doi:10.1002/ana.23736.
123. Verbeek E, Meuwissen MEC, Verheijen FW, et al. COL4A2 mutation associated with familial porencephaly and small-vessel disease. *Eur J Hum Genet*. 2012;20(8):844-851. doi:10.1038/ejhg.2012.20.

124. Yoneda Y, Haginoya K, Arai H, et al. De novo and inherited mutations in COL4A2, encoding the type IV collagen α2 chain cause porencephaly. *Am J Hum Genet*. 2012;90(1):86-90. doi:10.1016/j.ajhg.2011.11.016.
125. Ha TT, Sadleir LG, Mandelstam SA, et al. A mutation in COL4A2 causes autosomal dominant porencephaly with cataracts. *Am J Med Genet Part A*. 2016;170:1059-1063. doi:10.1002/ajmg.a.37527.
126. Plaisier E, Gribouval O, Alamowitch S, et al. COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps. *N Engl J Med*. 2007;357(26):2687-2695. doi:10.1056/NEJMoa071906.
127. Plaisier E, Chen Z, Gekeler F, et al. Novel COL4A1 mutations associated with HANAC syndrome: a role for the triple helical CB3[IV] domain. *Am J Med Genet A*. 2010;152A(10):2550-2555. doi:10.1002/ajmg.a.33659.
128. Gould DB, Phalan FC, van Mil SE, et al. Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med*. 2006;354(14):1489-1496. doi:10.1056/NEJMoa053727.
129. Vahedi K, Kubis N, Boukobza M, et al. COL4A1 mutation in a patient with sporadic, recurrent intracerebral hemorrhage. *Stroke*. 2007;38(5):1461-1464. doi:10.1161/STROKEAHA.106.475194.
130. Bilguvar K, DiLuna ML, Bizzarro MJ, et al. COL4A1 mutation in preterm intraventricular hemorrhage. *J Pediatr*. 2009;155(5):743-745. doi:10.1016/j.jpeds.2009.04.014.
131. Shah S, Ellard S, Kneen R, et al. Childhood presentation of COL4A1 mutations. *Dev Med Child Neurol*. 2012;54(6):569-574. doi:10.1111/j.1469-8749.2011.04198.x.
132. Lemmens R, Maugeri A, Niessen HWM, et al. Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. *Hum Mol Genet*. October 2012:1-7. doi:10.1093/hmg/dds436.
133. Vahedi K, Alamowitch S. Clinical spectrum of type IV collagen (COL4A1) mutations: a novel genetic multisystem disease. *Curr Opin Neurol*. 2011;24(1):63-68. doi:10.1097/WCO.0b013e32834232c6.
134. Kuo DS, Labelle-Dumais C, Gould DB. COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. *Hum Mol Genet*. 2012;21(R1):R97-R110. doi:10.1093/hmg/dds346.
135. Labelle-Dumais C, Dilworth DJ, Harrington EP, et al. COL4A1 mutations cause ocular dysgenesis, neuronal localization defects, and myopathy in mice and walker-warburg syndrome in humans. *PLoS Genet*. 2011;7(5). doi:10.1371/journal.pgen.1002062.
136. Moustakas A, Souchelnytskyi S, Heldin CH. Smad regulation in TGF-β signal transduction. *J Cell Sci*. 2001;114(Pt 24):4359-4369. http://www.ncbi.nlm.nih.gov/pubmed/11792802.
137. Akhurst RJ, Hata A. Targeting the TGFβ signalling pathway in disease. *Nat Rev Drug Discov*. 2012;11(10):790-811. doi:10.1038/nrd3810.
138. Miyazwa K, Shinozaki M, Hara T, Furuya T, Miyazono K. Two major Smad pathways in TGF-β superfamily signalling. *Genes to cells*. 2002;7:1191-1204.
139. Attisano L, Wrana JL. Signal transduction by the TGF-β superfamily. *Science*. 2002;296(5573):1646-1647. doi:10.1126/science.1071809.

140. Lutgens E, Gijbels M, Smook M, et al. Transforming growth factor-β mediates balance between inflammation and fibrosis during plaque progression. *Arterioscler Thromb Vasc Biol*. 2002;22(6):975-982. doi:10.1161/01.ATV.0000019729.39500.2F.
141. Shull MM, Ormsby I, Kier AB, et al. Targeted disruption of the mouse transforming growth factor-β1 gene results in multifocal inflammatory disease. *Nature*. 1992;359(6397):693-699. doi:10.1038/355242a0.
142. Massagué J, Seoane J, Wotton D. Smad transcription factors. *Genes Dev*. 2005;19(23):2783-2810. doi:10.1101/gad.1350705.
143. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-β family signaling. *Nature*. 2003;425(6958):577-584.
144. Ten Dijke P, Miyazono K, Heldin CH. Signaling inputs converge on nuclear effectors in TGF-B signaling. *Trends Biochem Sci*. 2000;25(2):64-70. doi:10.1016/S0968-0004(99)01519-4.
145. Massagué J. TGFβ signalling in context. *Nat Rev Mol Cell Biol*. 2012;13(10):616-630. doi:10.1038/nrm3434.
146. Janknecht R, Wells NJ, Hunter T. TGF-β-stimulated cooperation of Smad proteins with the coactivators CBP/p300. *Genes Dev*. 1998;12:2114-2119. doi:10.1101/gad.12.14.2114.
147. Davis-Dusenbery BN, Hata A. Smad-mediated miRNA processing. *RNA Biol*. 2011;8(1):71-76. doi:10.4161/rna.8.1.14299.
148. Davis BN, Hilyard AC, Lagna G, Hata A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature*. 2008;454(7200):56-61. doi:10.1038/nature07086.
149. Wischmeijer A, Van Laer L, Tortora G, et al. Thoracic aortic aneurysm in infancy in aneurysms-osteoarthritis syndrome due to a novel SMAD3 mutation: further delineation of the phenotype. *Am J Med Genet Part A*. 2013;161(5):1028-1035. doi:10.1002/ajmg.a.35852.
150. Lindsay ME, Dietz HC. Lessons on the pathogenesis of aneurysm from heritable conditions. *Nature*. 2011;473(7347):308-316. doi:10.1038/nature10145.
151. Van de Laar IMBH, Oldenburg RA, Pals G, et al. Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. *Nat Genet*. 2011;43(2):121-126. doi:10.1038/ng.744.
152. Regalado ES, Guo D-C, Villamizar C, et al. Exome sequencing identifies SMAD3 mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. *Circ Res*. 2011;109(6):680-686. doi:10.1161/CIRCRESAHA.111.248161.
153. Van de Laar IMBH, van der Linde D, Oei EHG, et al. Phenotypic spectrum of the SMAD3-related aneurysms-osteoarthritis syndrome. *J Med Genet*. 2012;49(1):47-57. doi:10.1136/jmedgenet-2011-100382.
154. Van der Linde D, van de Laar IMBH, Bertoli-Avella AM, et al. Aggressive cardiovascular phenotype of aneurysms-osteoarthritis syndrome caused by pathogenic SMAD3 variants. *J Am Coll Cardiol*. 2012;60(5):397-403. doi:10.1016/j.jacc.2011.12.052.
155. Yang X, Letterio JJ, Lechleider RJ, et al. Targeted disruption of SMAD3 results in impaired mucosal immunity and diminished T cell responsiveness to TGF-β. *EMBO J*. 1999;18(5):1280-1291. doi:10.1093/emboj/18.5.1280.

156. Sirard C, De L, Elia A, et al. The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes Dev*. 1998;12:107-119.
157. Nomura M, Li E. Smad2 role in mesoderm formation, left-right patterning and craniofacial development. *Nature*. 1998;393(6687):786-790. doi:10.1038/31693.
158. Grainger DJ. Transforming growth factor β and atherosclerosis: so far, so good for the protective cytokine hypothesis. *Arterioscler Thromb Vasc Biol*. 2004;24(3):399-404. doi:10.1161/01.ATV.0000114567.76772.33.
159. Bobik A. Transforming growth factor-βs and vascular disorders. *Arterioscler Thromb Vasc Biol*. 2006;26(8):1712-1720. doi:10.1161/01.ATV.0000225287.20034.2c.
160. Tashiro H, Shimokawa H, Sadamatu K, Yamamoto K. Prognostic significance of plasma concentrations of transforming growth factor-β in patients with coronary artery disease. *Coron Artery Dis*. 2002;13(3):139-143. http://www.ncbi.nlm.nih.gov/pubmed/12131016.
161. Grainger DJ, Kemp PR, Metcalfe JC, Liu AC, Lawn RM. The serum concentration of active transforming growth factor-β is severly depressed in advanced atherosclerosis. *Nat Med*. 1995;1:74-79.
162. Chen C, Lei W, Chen W, et al. Serum TGF-β1 and SMAD3 levels are closely associated with coronary artery disease. *BMC Cardiovasc Disord*. 2014;14:18. doi:10.1186/1471-2261-14-18.
163. Battegay EJ, Raines EW, Seifert RA, Bowen-Pope DF, Ross R. TGF-β induces bimodal proliferation of connective tissue cells via complex control of an autocrine PDGF loop. *Cell*. 1990;63(3):515-524. doi:10.1016/0092-8674(90)90448-N.
164. Stouffer GA, Owens GK. Angiotensin II-induced mitogenesis of spontaneously hypertensive rat-derived cultured smooth muscle cells is dependent on autocrine production of transforming growth factor-β. *Circ Res*. 1992;70(4):820-828. doi:10.1161/01.RES.70.4.820.
165. Owens GK, Geisterfer AAT, Yang YW, Komoriya A. Transforming growth factor-β-induced growth inhibition and cellular hypertrophy in cultured vascular smooth muscle cells. *J Cell Biol*. 1988;107:771-780.
166. Bjorkerud S. Effects of transforming growth factor-beta 1 on human arterial smooth muscle cells in vitro. *Arterioscler Thromb Vasc Biol*. 1991;11(4):892-902. doi:10.1161/01.ATV.11.4.892.
167. Grainger DJ, Kemp PR, Witchell CM, Weissberg PL, Metcalfe JC. Transforming growth factor β decreases the rate of proliferation of rat vascular smooth muscle cells by extending the G2 phase of the cell cycle and delays the rise in cyclic AMP before entry into M phase. *Biochem J*. 1994;299 (Pt 1):227-235. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1138046&tool=pmcentrez&rendertype=abstract.
168. Kojima S, Harpel PC, Rifkin DB. Lipoprotein (a) inhibits the generation of transforming growth factor beta: an endogenous inhibitor of smooth muscle cell migration. *J Cell Biol*. 1991;113(6):1439-1445.
169. Mallat Z, Tedgui A. The role of transforming growth factor beta in atherosclerosis: novel insights and future perspectives. *Curr Opin Lipidol*. 2002;13(5):523-529. doi:10.1097/00041433-200210000-00008.

170. Argmann CA, Diepstraten CH Van Den, Sawyez CG, et al. Transforming growth factor-beta1 inhibits macrophage cholesteryl ester accumulation induced by native and oxidized VLDL remnants. *Arterioscler Thromb Vasc Biol*. 2011;21:2011-2018.
171. Gojova A, Esposito B, Gourdy P, Ardouin P, Tedgui A, Mallat Z. Specific abrogation of transforming growth factor-β signaling in T cells alters atherosclerotic lesion size and composition in mice. *Blood*. 2003;102(12):4052-4059. doi:10.1182/blood-2003-05-1729.Z.M.
172. Reifenberg K, Cheng F, Orning C, et al. Overexpression of TGF-β1 in macrophages reduces and stabilizes atherosclerotic plaques in ApoE-deficient mice. *PLoS One*. 2012;7(7):e40990. doi:10.1371/journal.pone.0040990.
173. Mallat Z, Gojova A, Marchiol-Fournigault C, et al. Inhibition of transforming growth factor-β signaling accelerates atherosclerosis and induces an unstable plaque phenotype in mice. *Circ Res*. 2001;89(10):930-934. doi:10.1161/hh2201.099415.
174. Frutkin AD, Otsuka G, Stempien-Otero A, et al. TGF-β1 limits plaque growth, stabilizes plaque structure, and prevents aortic dilation in apolipoprotein E-null mice. *Arterioscler Thromb Vasc Biol*. 2009;29(9):1251-1257. doi:10.1161/ATVBAHA.109.186593.
175. Kalinina N, Agrotis A, Antropova Y, et al. Smad expression in human atherosclerotic lesions: evidence for impaired TGF-β/Smad signaling in smooth muscle cells of fibrofatty lesions. *Arterioscler Thromb Vasc Biol*. 2004;24(8):1391-1396. doi:10.1161/01.ATV.0000133605.89421.79.
176. Majesky MW, Lindner V, Twardzik DR, Schwartz SM, Reidy MA. Production of transforming growth factor β during repair of arterial injury. *J Clin Invest*. 1991;88(September):904-910.
177. Nikol S, Isner JM, Pickering JG, Kearney M, Leclerc G, Weir L. Expression of transforming growth factor-β is increased in human vascular restenosis lesions. *J Clin Invest*. 1992;90(4):1582-1592. doi:10.1172/JCI116027.
178. Ashcroft GS, Yang X, Glick a B, et al. Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. *Nat Cell Biol*. 1999;1(5):260-266. doi:10.1038/12971.
179. Kobayashi K, Yokote K, Fujimoto M, et al. Targeted disruption of TGF-β-Smad3 signaling leads to enhanced neointimal hyperplasia with diminished matrix deposition in response to vascular injury. *Circ Res*. 2005;96(8):904-912. doi:10.1161/01.RES.0000163980.55495.44.
180. Kanzaki T, Tamura K, Takahashi K, et al. In vivo effect of TGF-β1. Enhanced intimal thickening by administration of TGF-β1 in rabbit arteries with a balloon catheter. *Arterioscler Thromb Vasc Biol*. 1995;15:1951-1957.
181. Wolf YG, Rasmussen LM, Ruoslahti E. Antibodies against transforming growth factor-β1 suppress intimal hyperplasia in a rat model. *J Clin Invest*. 1994;93:1172-1178.
182. Tsai S, Hollenbeck ST, Ryer EJ, et al. TGF-β through Smad3 signaling stimulates vascular smooth muscle cell proliferation and neointimal formation. *Am J Physiol Heart Circ Physiol*. 2009;297(2):H540-H549. doi:10.1152/ajpheart.91478.2007.
183. Agah R, Prasad KSS, Linnemann R, Firpo MT, Quertermous T, Dichek DA. Cardiovascular overexpression of transforming growth factor-β1 causes abnormal

yolk sac vasculogenesis and early embryonic death. *Circ Res*. 2000;86(10):1024-1030. doi:10.1161/01.RES.86.10.1024.
184. Shaulian E, Karin M. AP-1 as a regulator of cell life and death. *Nat Cell Biol*. 2002;4(5):E131-E136. doi:10.1038/ncb0502-e131.
185. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. *Nat Rev Cancer*. 2003;3(11):859-868. doi:10.1038/nrc1209.
186. Hess J, Angel P, Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci*. 2004;117(25):5965-5973. doi:10.1242/jcs.01589.
187. Chinenov Y, Kerppola TK. Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene*. 2001;20:2438-2452.
188. Shaulian E. AP-1 - The Jun proteins: Oncogenes or tumor suppressors in disguise? *Cell Signal*. 2010;22(6):894-899. doi:10.1016/j.cellsig.2009.12.008.
189. Karin M, Liu Z, Zandi E. AP-1 function and regulation. *Curr Opin Cell Biol*. 1997;9(2):240-246. doi:10.1016/S0955-0674(97)80068-3.
190. Shaulian E, Karin M. AP-1 in cell proliferation and survival. *Oncogene*. 2001;20:2390-2400. doi:10.1038/sj.onc.1204383.
191. Karin M. The regulation of AP-1 activity by mitogen-activated protein kinases. *J Biol Chem*. 1995;270(28):16483-16486. doi:10.1098/rstb.1996.0008.
192. Schonthaler HB, Guinea-Viniegra J, Wagner EF. Targeting inflammation by modulating the Jun/AP-1 pathway. *Ann Rheum Dis*. 2011;70 Suppl 1(Suppl 1):i109-i112. doi:10.1136/ard.2010.140533.
193. Zhang Y, Feng XH, Derynck R. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature*. 1998;394(6696):909-913. doi:10.1038/29814.
194. De Groot RP, Kruijer W. Transcriptional activation by TGFβ1 mediated by the DYAD symmetry element (DSE) and the TPA responsive element (TRE). *Biochem Biophys Res Commun*. 1990;168(3):1074-1081. doi:10.1152/ajpendo.00521.2007.
195. Felts SJ, Stoflet ES, Eggers CT, Getz MJ. Tissue factor gene transcription in serum-stimulated fibroblasts is mediated by recruitment of c-Fos into specific AP-1 DNA-binding complexes. *Biochemistry*. 1995;34(38):12355-12362.
196. Jin G, Howe PH. Regulation of clusterin gene expression by transforming growth factor beta. *J Biol Chem*. 1997;272(42):26620-26626. http://www.ncbi.nlm.nih.gov/pubmed/9334243.
197. Keeton MR, Curriden SA, van Zonneveld AJ, Loskutoff DJ. Identification of regulatory sequences in the type 1 plasminogen activator inhibitor gene responsive to transforming growth factor beta. *J Biol Chem*. 1991;266(34):23048-23052. http://www.ncbi.nlm.nih.gov/pubmed/1744101.
198. Wrana JL, Attisano L, Cárcamo J, et al. TGFβ signals through a heteromeric protein kinase receptor complex. *Cell*. 1992;71(6):1003-1014. doi:0092-8674(92)90395-S [pii].
199. Yingling JM, Datto MB, Wong C, Frederick JP, Liberati NT, Wang XF. Tumor suppressor Smad4 is a transforming growth factor beta-inducible DNA binding protein. *Mol Cell Bio*. 1997;17(12):7019-7028.
200. Shi Y, Wang Y-F, Jayaraman L, Yang H, Massagué J, Pavletich NP. Crystal structure of a Smad MH1 domain bound to DNA. *Cell*. 1998;94(5):585-594. doi:10.1016/S0092-8674(00)81600-1.

201. Verrecchia F, Pessah M, Atfi A, Mauviel A. Tumor necrosis factor-α inhibits transforming growth factor-beta/Smad signaling in human dermal fibroblasts via AP-1 activation. *J Biol Chem*. 2000;275(39):30226-30231. doi:10.1074/jbc.M005310200.
202. Verrecchia F, Vindevoghel L, Lechleider RJ, Uitto J, Roberts AB, Mauviel A. Smad3/AP-1 interactions control transcriptional responses to TGF-β in a promoter-specific manner. *Oncogene*. 2001;20(26):3332-3340. doi:10.1038/sj.onc.1204448.
203. Kühn K. Basement membrane (type IV) collagen. *Matrix Biol*. 1995;14(6):439-445. http://www.ncbi.nlm.nih.gov/pubmed/7795882.
204. Sund M, Xie L, Kalluri R. The contribution of vascular basement membranes and extracellular matrix to the mechanics of tumor angiogenesis. *APMIS*. 2004;112(7-8):450-462. doi:10.1111/j.1600-0463.2004.t01-1-apm11207-0806.x.
205. Derynck R, Akhurst RJ, Balmain A. TGF-β signaling in tumor suppression and cancer progression. *Nat Genet*. 2001;29(2):117-129. doi:10.1038/ng1001-117.
206. Madri JA, Pratt BM, Tucker AM. Phenotypic modulation of endothelial cells by transforming growth factor depends upon the composition and organization of the extracellular matrix. *J Cell Biol*. 1988;106:1375-1384.
207. Janat MF, Liau G. Transforming growth factor β1 is a powerful modulator of platelet-derived growth factor action in vascular smooth muscle cells. *J Cell Physiol*. 1992;150(2):232-242. doi:10.1002/jcp.1041500203.
208. Grande JP, Melder D, Zinsmeister A, Killen P. Transforming growth factor-beta 1 induces collagen IV gene expression in NIH-3T3 cells. *Lab Investig*. 1993;69(4):387-395.
209. Grande JP, Melder DC, Zinsmeister AR. Modulation of collagen gene expression cytokines: Stimulatory effect of transforming growth factor-β1, with divergent effects of epidermal growth factor and tumour necrosis factor-α on collagen type I and collagen type IV. *J Lab Clin Med*. 1997;130(5):476-486.
210. Li J, Qu X, Ricardo SD, Bertram JF, Nikolic-Paterson DJ. Resveratrol inhibits renal fibrosis in the obstructed kidney: potential role in deacetylation of Smad3. *Am J Pathol*. 2010;177(3):1065-1071. doi:10.2353/ajpath.2010.090923.
211. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795.
212. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet*. 2013;14(6):379-389. doi:10.1038/nrg3472.
213. Inman GJ, Francisco NJ, Callahan JF, et al. SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like (ALK) receptors ALK4, ALK5, and ALK7. *Mol Pharmacol*. 2002;62(1):65-74.
214. Baugé C, Cauvard O, Leclercq S, Galéra P, Boumédiene K. Modulation of transforming growth factor beta signalling pathway genes by transforming growth factor beta in human osteoarthritic chondrocytes: involvement of Sp1 in both early and late response cells to transforming growth factor beta. *Arthritis Res Ther*. 2011;13(1):R23. doi:10.1186/ar3247.
215. Poncelet A-C, de Caestecker MP, Schnaper HW. The transforming growth factor-β/SMAD signaling pathway is present and functional in human mesangial cells. *Kidney Int*. 1999;56(4):1354-1365. doi:10.1046/j.1523-1755.1999.00680.x.

216. De Caestecker MP, Piek E, Roberts AB. Role of transforming growth factor-β signaling in cancer. *J Natl Cancer Inst*. 2000;92(17):1388-1402. http://www.ncbi.nlm.nih.gov/pubmed/10974075.
217. Verrecchia F, Mauviel A. Transforming growth factor-β signaling through the Smad pathway: role in extracellular matrix gene expression and regulation. *J Invest Dermatol*. 2002;118(2):211-215. doi:10.1046/j.1523-1747.2002.01641.x.
218. Verrecchia F, Chu ML, Mauviel A. Identification of novel TGF-β/Smad gene targets in dermal fibroblasts using a combined cDNA microarray/promoter transactivation approach. *J Biol Chem*. 2001;276(20):17058-17062. doi:10.1074/jbc.M100754200.
219. Levy L, Hill CS. Smad4 dependency defines two classes of transforming growth factor β (TGF-β) target genes and distinguishes TGF-β-induced epithelial-mesenchymal transition from its antiproliferative and migratory responses. *Mol Cell Biol*. 2005;25:8108-8125. doi:10.1128/MCB.25.18.8108.
220. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. 2014;15(1):34-48. doi:10.1038/nrg3575.
221. Grande JP, Warner GM, Walker HJ, et al. TGF-β1 is an autocrine mediator of renal tubular epithelial cell growth and collagen IV production. *Exp Biol Med*. 2002;227:171-181.
222. Kunico GS, Alvarez R, Killen PD, Neilson EG, Paul D. Transforming growth factor-β modulation of the α1(IV) collagen gene in murine proximal tubular cells. *Am J Physiol Renal Physiol*. 1996;271:F120-F125.
223. Silbiger S, Lei J, Ziyadeh FN, et al. Estradiol reverses TGF-β1-stimulated type IV collagen gene transcription in murine mesangial cells mechanisms. *Am J Physiol Ren Physiol*. 1998;274:F1113-F1118.
224. Zhang Y, Handley D, Kaplan T, et al. High throughput determination of TGFβ1/SMAD3 targets in A549 lung epithelial cells. *PLoS One*. 2011;6(5):e20319. doi:10.1371/journal.pone.0020319.
225. Qin W, Chung ACK, Huang XR, et al. TGF-β/Smad3 signaling promotes renal fibrosis by inhibiting miR-29. *J Am Soc Nephrol*. 2011;22(8):1462-1474. doi:10.1681/ASN.2010121308.
226. Zhou L, Wang L, Lu L, Jiang P, Sun H, Wang H. Inhibition of miR-29 by TGF-beta-Smad3 signaling through dual mechanisms promotes transdifferentiation of mouse myoblasts into myofibroblasts. *PLoS One*. 2012;7(3):e33766. doi:10.1371/journal.pone.0033766.
227. Cushing L, Kuang PP, Qian J, et al. miR-29 is a major regulator of genes associated with pulmonary fibrosis. *Am J Respir Cell Mol Biol*. 2011;45(2):287-294. doi:10.1165/rcmb.2010-0323OC.
228. Ekman M, Bhattachariya A, Dahan D, Uvelius B, Albinsson S, Swärd K. Mir-29 repression in bladder outlet obstruction contributes to matrix remodeling and altered stiffness. *PLoS One*. 2013;8(12):e82308. doi:10.1371/journal.pone.0082308.
229. Tseng ZH, Vittinghoff E, Musone SL, et al. Association of TGFBR2 polymorphism with risk of sudden cardiac arrest in patients with coronary artery disease. *Heart Rhythm*. 2009;6(12):1745-1750. doi:10.1016/j.hrthm.2009.08.031.
230. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006;3(11):e442. doi:10.1371/journal.pmed.0030442.

231. Li S, Fan Y-S, Chow LH, et al. Innate diversity of adult human arterial smooth muscle cells: cloning of distinct subtypes from the internal thoracic artery. *Circ Res*. 2001;89(6):517-525. doi:10.1161/hh1801.097165.
232. Lee JY, Elmer HL, Ross KR, Kelley TJ. Isoprenoid-mediated control of SMAD3 expression in a cultured model of cystic fibrosis epithelial cells. *Am J Respir Cell Mol Biol*. 2004;31:234-240. doi:10.1165/rcmb.2003-0447OC.
233. Cole CB, Nikpay M, Lau P, et al. Adiposity significantly modifies genetic risk for dyslipidemia. *J Lipid Res*. 2014;55(11):2416-2422. doi:10.1194/jlr.P052522.
234. Perisic L, Aldi S, Sun Y, et al. Gene expression signatures, pathways and networks in carotid atherosclerosis. *J Intern Med*. 2016;279(3):293-308.
235. Perisic L, Hedin E, Razuvaev A, et al. Profiling of atherosclerotic lesions by gene and tissue microarrays reveals PCSK6 as a novel protease in unstable carotid atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2013;33(10):2432-2443. doi:10.1161/ATVBAHA.113.301743.
236. Folkersen L, Persson J, Ekstrand J, et al. Predicted of ischemic events on the basis of transcriptomic and genomic profiling in patients undergoing carotid endarterectomy. *Mol Med*. 2012;18(4):669-775. doi:10.2119/molmed.2011.00479.
237. Razuvaev A, Ekstrand J, Folkersen L, et al. Correlations between clinical variables and gene-expression profiles in carotid plaque instability. *Eur J Vasc Endovasc Surg*. 2011;42(6):722-730. doi:10.1016/j.ejvs.2011.05.023.
238. Jackson V, Petrini J, Caidahl K, et al. Corrigendum to "Bicuspid aortic valve leaflet morphology in relation to aortic root morphology: A study of 300 patients undergoing open-heart surgery." *Eur J Cardio-thoracic Surg*. 2012;41(2):471. doi:10.1093/ejcts/ezr168.
239. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369-375. doi:10.1038/ng.2213.
240. Pardo L, Bochdanovits Z, de Geus E, et al. Global similarity with local differences in linkage disequilibrium between the Dutch and HapMap-CEU populations. *Eur J Hum Genet*. 2009;17(6):802-810. doi:10.1038/ejhg.2008.248.
241. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263-265. doi:10.1093/bioinformatics/bth457.
242. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15(4):272-286. doi:10.1038/nrg3682.
243. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39(3):311-318. doi:10.1038/ng1966.
244. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-112. doi:10.1038/nature07829.
245. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797. doi:10.1101/gr.137323.112.

246. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43-49. doi:10.1038/nature09906.
247. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247.
248. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*. 2011;12(1):7-18. doi:10.1038/nrg2905.
249. Visel A, Blow MJ, Li Z, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009;457(7231):854-859. doi:10.1038/nature07730.
250. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6(4):e1000888. doi:10.1371/journal.pgen.1000888.
251. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010;466(7307):714-719. doi:10.1038/nature09266.
252. Busque L, Belisle C, Provost S, Giroux M, Perreault C. Differential expression of SMAD3 transcripts is not regulated by cis-acting genetic elements but has a gender specificity. *Genes Immun*. 2009;10(2):192-196. doi:10.1038/gene.2008.101.
253. Bennett BL, Sasaki DT, Murray BW, et al. SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase. *Proc Natl Acad Sci U S A*. 2001;98(24):13681-13686. doi:10.1073/pnas.251194298.
254. Jang C-W, Chen C-H, Chen C-C, Chen J, Su Y-H, Chen R-H. TGF-β induces apoptosis through Smad-mediated expression of DAP-kinase. *Nat Cell Biol*. 2002;4(1):51-58. doi:10.1038/ncb731.
255. Wildey GM, Patil S, Howe PH. Smad3 potentiates transforming growth factor β (TGFβ)-induced apoptosis and expression of the BH3-only protein Bim in WEHI 231 B lymphocytes. *J Biol Chem*. 2003;278(20):18069-18077. doi:10.1074/jbc.M211958200.
256. Oliver FJ, de la Rubia G, Rolli V, Ruiz-Ruiz MC, de Murcia G, Menissier-de Murica J. Importance of poly(ADP-ribose) polymerase and its cleavage in apoptosis. *J Biol Chem*. 1998;273(50):33533-33539.
257. Mechta-Grigoriou F, Gerald D, Yaniv M. The mammalian Jun proteins: redundancy and specificity. *Oncogene*. 2001;20(19):2378-2389.
258. Farh KK-H, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337-343. doi:10.1038/nature13835.
259. Bullaughey K, Chavarria CI, Coop G, Gilad Y. Expression quantitative trait loci detected in cell lines are often present in primary tissues. *Hum Mol Genet*. 2009;18(22):4296-4303. doi:10.1093/hmg/ddp382.
260. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80-)*. 2015;348(6235):648-660. doi:10.1126/science.1262110.

261. Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type dependent manner. *Science (80-).* 2010;325(5945):1246-1250. doi:10.1126/science.1174148.Common.
262. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119-124. doi:10.1038/nature11582.
263. Franke A, McGovern DPB, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010;42(12):1118-1125. doi:10.1038/ng.717.
264. Hinds DA, McMahon G, Kiefer AK, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet*. 2013;45(8):907-911. doi:10.1038/ng.2686.
265. Moffat MF, Gut IG, Demenaid F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010;363:1211-1221.
266. Ferreira MAR, Matheson MC, Tang CS, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J Allergy Clin Immunol*. 2014;133(6):1564-1571. doi:10.1016/j.jaci.2013.10.030.
267. Peden JF, Farrall M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum Mol Genet*. 2011;20(R2):R198-R205. doi:10.1093/hmg/ddr384.
268. Sudhakar A, Nyberg P, Keshamouni VG, Mannam AP. Human α1 type IV collagen NC1 domain exhibits distinct antiangiogenic activity mediated by α1β1 integrin. 2005;115(10). doi:10.1172/JCI24813.include.
269. Nyberg P, Xie L, Sugimoto H, et al. Characterization of the anti-angiogenic properties of arresten, an alpha1beta1 integrin-dependent collagen-derived tumor suppressor. *Exp Cell Res*. 2008;314(18):3292-3305. doi:10.1016/j.yexcr.2008.08.011.
270. Schmidt C, Pollner R, Pöschl E, Kühn K. Expression of human collagen type IV genes is regulated by transcriptional and post-transcriptional mechanisms. *FEBS Lett*. 1992;312(2-3):174-178. http://www.ncbi.nlm.nih.gov/pubmed/1426248.
271. Schwartz YB, Pirrotta V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet*. 2013;14(12):853-864. doi:10.1038/nrg3603.
272. Braund PS. Functional analysis of novel genetic markers of coronary artery disease identified by genome-wide association studies. *Thesis*. 2015.
273. Ross R, Wight T, Strandness E, Thiele B. Human atherosclerosis I . Cell constitution and characteristics of advanced lesions of the superficial femoral artery. *Am J Pathol*. 1984;114(1):79-93.
274. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet*. 2013;9(8). doi:10.1371/journal.pgen.1003649.
275. Fu J, Wolfs MGM, Deelen P, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet*. 2012;8(1):e1002431. doi:10.1371/journal.pgen.1002431.
276. Qiu X, Vu TH, Lu Q, et al. A complex deoxyribonucleic acid looping configuration associated with the silencing of the maternal Igf2 allele. *Mol Endocrinol*. 2008;22(6):1476-1488. doi:10.1210/me.2007-0474.

277. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*. 2010;465(7299):721-727. doi:10.1038/nature09230.
278. Edwards SL, Beesley J, French JD, Dunning M. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779-797. doi:10.1016/j.ajhg.2013.10.012.
279. Miller CL, Pjanic M, Quertermous T. From locus association to mechanism of gene causality: the devil is in the details. *Arter Thromb Vasc Biol*. 2015;35(10):2079-2081. doi:10.1038/nature14590.5.
280. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90(1):7-24. doi:10.1016/j.ajhg.2011.11.029.
281. Davies RW, Dandona S, Stewart AFR, et al. Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. *Circ Cardiovasc Genet*. 2010;3(5):468-474. doi:10.1161/CIRCGENETICS.110.946269.
282. Bennett MR, Sinha S, Owens GK. Vascular smooth muscle cells in atherosclerosis. *Circ Res*. 2016;118(4):692-702. doi:10.1161/CIRCRESAHA.115.306361.
283. Maouche S, Schunkert H. Strategies beyond genome-wide association studies for atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2012;32(2):170-181. doi:10.1161/ATVBAHA.111.232652.
284. Nurnberg ST, Zhang H, Hand NJ, et al. From loci to biology: functional genomics of genome-wide association for coronary disease. *Circ Res*. 2016;118(4):586-606. doi:10.1161/CIRCRESAHA.115.306464.
285. Lander ES. Initial impact of the sequencing of the human genome. *Nature*. 2011;470(7333):187-197. doi:10.1038/nature09792.
286. Pu X, Xiao Q, Kiechl S, et al. ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *Am J Hum Genet*. 2013;92(3):366-374. doi:10.1016/j.ajhg.2013.01.012.
287. Bauer RC, Tohyama J, Cui J, et al. Knockout of Adamts7, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. *Circulation*. 2015;131(13):1202-1213. doi:10.1161/CIRCULATIONAHA.114.012669.
288. Douvris A, Soubeyrand S, Naing T, et al. Functional analysis of the TRIB1 associated locus linked to plasma triglycerides and coronary artery disease. *J Am Heart Assoc*. 2014;3(3):e000884. doi:10.1161/JAHA.114.000884.
289. Miller CL, Anderson DR, Kundu RK, et al. Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus. *PLoS Genet*. 2013;9(7):e1003652. doi:10.1371/journal.pgen.1003652.
290. Beaudoin M, Gupta RM, Won HH, et al. Myocardial infarction-associated SNP at 6p24 interferes with MEF2 binding and associates with PHACTR1 expression levels in human coronary arteries. *Arterioscler Thromb Vasc Biol*. 2015;35(6):1472-1479. doi:10.1161/ATVBAHA.115.305534.
291. Almontashiri NAM, Antoine D, Zhou X, et al. 9p21.3 coronary artery disease risk variants disrupt TEAD transcription factor-dependent TGFβ regulation of p16 expression in human aortic smooth muscle cells. *Circulation*. 2015;132(21):1969-1978. doi:10.1161/CIRCULATIONAHA.114.015023.

292. Cooper DN. Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Hum Genomics*. 2010;4(5):284-288. doi:10.1186/1479-7364-4-5-284.
293. Gordan R, Shen N, Dror I, et al. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep*. 2013;3(4):1093-1104. doi:10.1016/j.celrep.2013.03.014.
294. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213-1218. doi:10.1038/nmeth.2688.
295. Trynka G, Westra H-J, Slowikowski K, et al. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *Am J Hum Genet*. 2015;97(1):139-152. doi:10.1016/j.ajhg.2015.05.016.
296. Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol Rep*. 2004;84(3):767-801.
297. Vijayachandra K, Lee J, Glick AB. Smad3 regulates senescence and malignant conversion in a mouse multistage skin carcinogenesis model. *Cancer Res*. 2003;63(13):3447-3452.
298. Gamble JR, Khew-Goodall Y, Vadas MA. Transforming growth factor-beta inhibits E-selectin expression on human endothelial cells. *J Immunol*. 1993;150(10):4494-4503.
299. Zhang F, Tsai S, Kato K, et al. Transforming growth factor-beta promotes recruitment of bone marrow cells and bone marrow-derived mesenchymal stem cells through stimulation of MCP-1 production in vascular smooth muscle cells. *J Biol Chem*. 2009;284(26):17564-17574. doi:10.1074/jbc.M109.013987.
300. Ma J, Wang Q, Fei T, Han JJ, Chen Y. MCP-1 mediates TGF-β–induced angiogenesis by stimulating vascular smooth muscle cell migration. *Blood*. 2007;109(3):987-995. doi:10.1182/blood-2006-07-036400.The.
301. Feinberg MW, Shimizu K, Lebedeva M, et al. Essential role for Smad3 in regulating MCP-1 expression and vascular inflammation. *Circ Res*. 2004;94(5):601-608. doi:10.1161/01.RES.0000119170.70818.4F.

7 CONTRIBUTIONS

Dr. Majid Nikpay used the PLINK (v1.07) software to perform epistasis analysis between the *COL4A1/COL4A2* and *SMAD3* loci in five independent cohorts. Majid Nikpay also conducted conditional and joint analysis of the *SMAD3* and *COL4A1/COL4A2* loci as part of the 2015 *Nature Genetics* CARDIoGRAM paper. Anada Silva and Amy Martinuk aided with tissue culture, chromatin immunoprecipitation, transfections, RNA isolation, and cDNA synthesis. The BiKE Study Group (Biobank of Karolinska Endarterectomies, Karolinska Institute, Sweden) provided us with eQTL data from carotid plaque tissue and the ASAP Group (Advanced Study of Aortic Pathology, Karolinska Institute, Sweden) provided us with eQTL data from five human tissues.

8 APPENDIX I *PHACTR1*: FUNCTIONAL CLUES LINKING A GWAS LOCUS TO CORONARY ARTERY DISEASE

Adam W. Turner, Ruth McPherson

Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada, K1Y 4W7

Status: Published in *Arteriosclerosis, Thrombosis, and Vascular Biology* 35 (2015) 1293-1295

Recent genome-wide association studies for coronary artery disease (CAD) have identified more than 45 novel loci, the majority of which are not associated with traditional CAD risk factors[1,2]. The lead SNPs for each of these loci are generally common, displaying allele frequencies from 0.13 to 0.91 and allele specific effect sizes (odds ratios) for CAD of between 1.06 and 1.3.

Notably, the majority of CAD signals identified by the GWAS approach are in non-coding regions of the genome, not unexpected given that only 1% of the genome is protein coding. An important part of the non-coding genome is under purifying selection[3], implying important regulatory functions supported by the fact that an excess of GWAS signals are close to genic regions. However, for most of these variants including the *PHACTR1* gene region on chromosome 6p24, the causal biological mechanisms have remained unclear.

In this issue of ***ATVB***, Lettre and colleagues have sought to: confirm previous reports that rs9349379A>G is the lead CAD associated SNP in this region, determine by eQTL analysis if *PHACTR1* is likely to be the causal gene and using a combination of bioinformatic and laboratory analysis to explore possible causal mechanisms relating the risk allele (G) (MAF 0.41) to decreased *PHACTR1* expression.

By resequencing and imputation from the 1000 Genomes data set, in accord with previous findings, they identify rs9349379 as the top SNP associated with CAD in Montreal Heart Institute Biobank samples consisting of 1176 MI cases and 1996 French Canadian controls (G allele OR: 1.37; $p=8.4X10^{-6}$). No other SNP in this region remained significant following conditional analysis on rs9349379.

Expression quantitative trait loci (eQTL) are genetic variants that associate with the RNA transcript level of a given gene. The majority of those identified by the GWAS approach appear to have *cis*-effects, thus acting on the expression of nearby genes. Evidence that a given SNP associates with expression of a candidate gene (eQTL) further supports a potentially causal association. Given that these effects are often cell specific, eQTL analysis in a relevant cell or tissue is important. Since available eQTL data sets for this SNP did not include coronary artery cell types, analysis was performed in a sample of 25 genotyped coronary artery specimens. Despite the small sample size, the risk allele was found to associate with reduced expression of *PHACTR1* but not with expression of all other coding genes in a 1 Mb region on either side of rs9349379.

As a complementary approach to eQTL analysis, public databases, generated by the ENCODE[4] and Roadmap Epigenomics projects[5] can be used to identify predicted functional regulatory elements[6]. An important caveat is that these analyses require genome-wide chromatin data from a relevant cell type.[7] These computational approaches to the identification of regulatory sequences need to be followed up by more specific methodologies[8, 9]. Thus, it can first be determined whether a risk locus is likely to harbor functional *cis*-acting regulatory modules whose activity is altered by a particular risk variant prior to standard molecular biology approaches. Here, bioinformatic analysis using Roadmap data and *in silico* searches predicted that rs9349379(G) interrupts a MEF2 binding site. This was confirmed *in vitro* by electrophoretic mobility shift assay (EMSA) with HUVEC nuclear extracts and by supershift assay, MEF2 was shown to be at least one of the nuclear proteins interacting with this region.

Finally to support the premise that the MEF2 binding site is functionally important, they employed CRISPR/Cas9 genome editing in human embryonic stem cells. They identified a clone carrying a heterozygous deletion of a 34 bp segment encompassing the MEF2 binding site that in differentiated endothelial cells attenuated *PHACTR1* expression.

In summary, this important functional analysis of *PHACTR1* illustrates the multi-faceted approach required to unravel mechanisms of a GWAS locus in common disease. In addition, rs9349379 does not associate with other traditional CAD risk factors, highlighting altered binding of MEF2 and/or other factors at this region could representing a very interesting novel CAD mechanism.

Nonetheless, important questions remain. Although the risk allele (G) associates with decreased *PHACTR1* expression in human coronary arteries, it is not entirely clear which arterial wall cell type is susceptible to altered *PHACTR1* expression. Since functional studies were carried out in HUVEC, the authors suggest that arterial endothelial cells are key. The authors have also not firmly established that disruption of MEF2 binding *per se* is causal since other transcription factors may potentially bind to this site, and partial knockdown of two isoforms of *MEF2* did not alter *PHACTR1* expression. Of note, a putative association of a structural variant in *MEF2A* with CAD risk was refuted by later studies. This study cannot rule out other isoforms of *MEF2* potentially having roles in *PHACTR1* regulation.

Perhaps more importantly, little information is yet available on how the *PHACTR1* gene product may relate to atherosclerosis. It is known to modulate protein phosphatase 1 (PP1)

activity *in vitro* and to interact with actin. The locus associates with chronic CAD and coronary artery calcium rather than acute MI supportive of effects on atherosclerosis *per se*. However, the rs9349379(G) allele for CAD has opposite effects on risk for cervical artery dissection[10] suggesting complex effects on arterial wall integrity. The MEF2 binding sequence and rs9349379 are conserved between humans and mice, so future studies incorporating a murine model of atherosclerosis may shed further light on this fascinating but unfinished story.

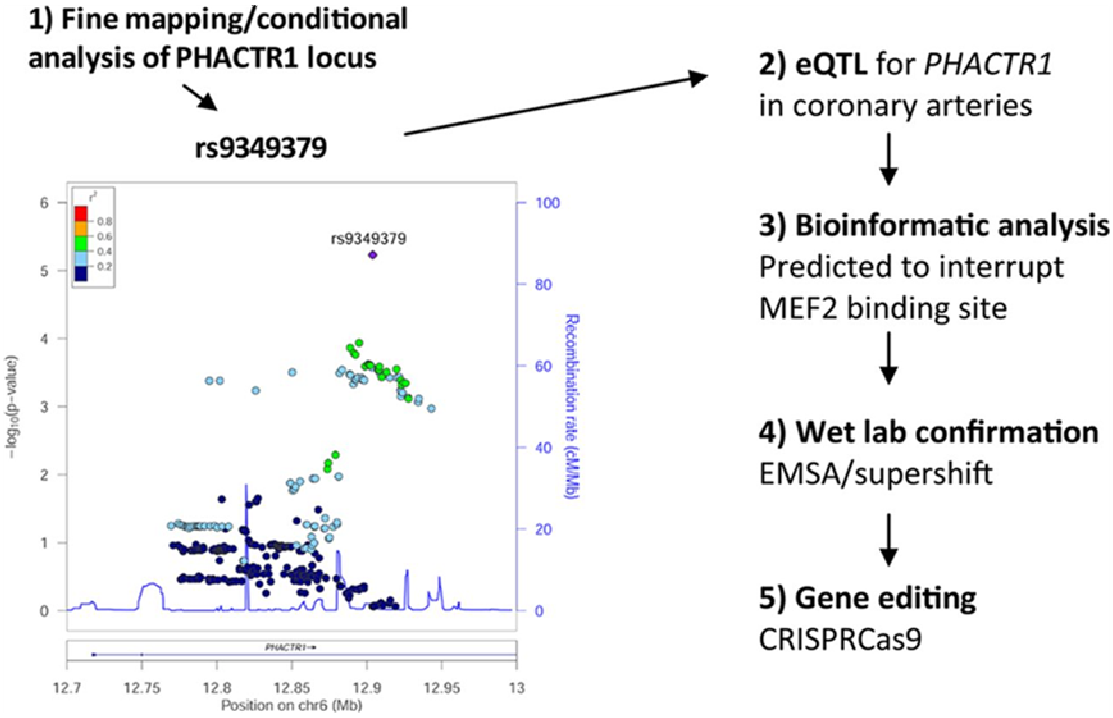

1) Fine mapping/conditional analysis of PHACTR1 locus
rs9349379
2) eQTL for PHACTR1 in coronary arteries
3) Bioinformatic analysis
Predicted to interrupt MEF2 binding site
4) Wet lab confirmation
EMSA/supershift
5) Gene editing
CRISPRCas9
rs9349379
Recombination rate (cM/Mb)
Position on chr6 (Mb)

Figure. (1) By resequencing, imputation from the 1000-genome data set and conditional analysis, rs9349379A>G is identified as the top single-nucleotide polymorphism (SNP) associated with coronary artery disease in Montreal Heart Institute Biobank. (2) Expression quantitative trait locus (eQTL) analysis in 25 genotyped coronary artery specimens shows the risk allele (G) to be associated with reduced expression of *PHACTR1* but not with expression of all other coding genes in a 1-Mb region on either side of rs9349379. (3) Bioinformatic analysis using Roadmap data and in silico searches predicts that rs9349379(G) interrupts a myocyte enhancer factor (MEF)2–binding site. (4) This is confirmed in vitro by electrophoretic mobility shift assay (EMSA) with human umbilical vein endothelial cell nuclear extracts, and by supershift assay, MEF2A and MEF2C are shown to be among the nuclear proteins interacting with this region. (5) CRISPR/Cas9 genome editing is performed in human embryonic stem cells, and a clone carrying a heterozygous deletion of a 34-bp segment encompassing the MEF2-binding site is shown in differentiated endothelial cells to attenuate *PHACTR1* expression.

References

(1) Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011;43(4):333-8.

(2) Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 2012 December 2;45(1):25-33.

(3) Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 2010;6(12):e1001025.

(4) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 2011 April;9(4):e1001046.

(5) Chadwick LH. The NIH Roadmap Epigenomics Program data resource. *Epigenomics* 2012 June;4(3):317-24.

(6) Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. *Cell* 2011 September 30;147(1):57-69.

(7) Boyle AP, Song L, Lee BK, London D, Keefe D, Birney E, Iyer VR, Crawford GE, Furey TS. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res* 2011 March;21(3):456-64.

(8) Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009 October;10(10):669-80.

(9) Song L, Zhang Z, Grasfeder LL, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* 2011 October;21(10):1757-67.

(10) Debette S, Kamatani Y, Metso TM et al. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat Genet* 2015 January;47(1):78-83.

Adam Turner
University of Ottawa Heart Institute
H4227 – 40 Ruskin St. – Ottawa, Ontario, K1Y 4W7

Education

- PhD in Biochemistry – Specialization in Human and Molecular Genetics (2011-present)
  University of Ottawa, Ottawa, ON, Canada

- Masters of Science in Biochemistry – Specialization in Human and Molecular Genetics (2009-2011) – Transferred to PhD
  University of Ottawa, Ottawa, ON, Canada

- Bachelor of Science (Honours) in Biochemistry with Co-op (2003-2009)
  University of Ottawa, Ottawa, ON, Canada

- High School Diploma (1999-2003)
  John McCrae Secondary School, Ottawa, ON, Canada

Conferences

- Canadian Lipoprotein Conference (October 2015)
  Toronto, Ontario, Canada
  Oral Presentation
- American Society of Human Genetics Annual Meeting (October 2015)
  Baltimore, Maryland, USA
  Poster Presentation
- Arteriosclerosis, Thrombosis, and Vascular Biology Annual Meeting (May 2014)
  Toronto, Ontario, Canada
  Poster Presentation
- American Heart Association Scientific Sessions (November 2013)
  Dallas, Texas, USA
  Poster Presentation
- Canadian Lipoprotein Conference (September 2013)
  Mont Tremblant, Quebec, Canada
  Poster Presentation
- American Society of Human Genetics Annual Meeting (October 2012)
  San Francisco, California, USA
  Poster Presentation
- American Society of Human Genetics Annual Meeting (October 2011)
  Montreal, Quebec, Canada
  Poster Presentation

- Arteriosclerosis, Thrombosis, and Vascular Biology Annual Meeting (April 2011)
  Chicago, Illinois, USA
  Poster Presentation
- Canadian Lipoprotein Conference (October 2010)
  Niagara on the Lake, Ontario, Canada
  Oral Presentation

Publications

Turner, AW. *et al*. Functional analysis of a novel genome-wide association study signal in SMAD3 that confers protection from coronary artery disease. *Arterioscler Thromb Vasc Biol*. **36**, 972-983 (2016)

Turner, AW. *et al.* Functional interaction between COL4A1/COL4A2 and SMAD3 risk loci for coronary artery disease. *Atherosclerosis.* **242**, 543-552 (2015)

Turner, AW and McPherson, R. PHACTR1: Functional clues linking a genome-wide association study locus to coronary artery disease. *Arterioscler Thromb Vasc Biol*. **35**, 1293-1295 (2015)

Scholarships and Awards

- Best Poster Award – 2013 Canadian Lipoprotein Conference
- University of Ottawa Admission Scholarship – Doctorate (2011)
- University of Ottawa Excellence Scholarship (2011)
- Ontario Graduate Scholarship (2010)
- University of Ottawa Admission Scholarship (2009)
- Dean's List, University of Ottawa
- University of Ottawa Merit Scholarships (2006-2009)
- University of Ottawa Admission Scholarship (2003)

Research Internships

- University of Ottawa Heart Institute, Ottawa, ON (September 2009- present)
  Supervisor: Dr. Ruth McPherson
  Research Project: Functional analysis of noncoding SNPs at the *COL4A1/COL4A2* and *SMAD3* loci

- University of Ottawa Heart Institute, Ottawa, ON (May 2007-August 2007, May 2008- April 2009)
  Supervisor: Dr. Patrick Burgon
  Research Project: Characterization of a novel lamin A/C binding protein

- Canada Border Services Agency (January 2007-April 2007)
  Supervisor: Dr. Rachel Ng
  Project: Human forensics and tobacco genetics/genomics

Other Job Placements

- Canadian Biotechnology Secretariat, Industry Canada (September 2005-December 2006)
  Supervisor: Joshua Belinko
  Role: Development and maintenance of a new government of Canada biotechnology website