FUNCTIONAL REGULATION AT THE 9P21.3 GENETIC RISK LOCUS IN CORONARY ARTERY DISEASE (CAD)

By Darlène Antoine

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

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine, University of Ottawa

Ottawa, Ontario, Canada

© Darlène Antoine, Ottawa, Canada, 2015

``To be worthy of something, it is not necessary to have accomplished great things, it is well enough that you have attempted them``

Antoine Dupré

Abstract

The first genetic CAD risk locus to be identified by genome-wide association studies, single nucleotide polymorphisms (SNPs) at 9p21.3 predispose to increased risk of CAD. By bioinformatics scan analysis of the 9p21.3 locus; we interrogated the 59 linked SNPs over the 53,202bp to identify putative transcription factor-binding consensus sequences. We hypothesize that some genetic polymorphisms at the 9p21.3 locus are functional and will disrupt specific regulatory sequences within enhancers. Here, I investigated how polymorphisms affect TEAD-dependent regulation at the 9p21.3 locus, and also how polymorphisms affect GATA factor-dependent regulation at the 9p21.3 locus, using cultured HEK293 and primary human aortic smooth muscle cells (HAoSMCs) to transfect the pGL3-promoter plasmid constructs containing the reference or risk variant sequences (rs10611656, rs4977757, rs10757269, rs9632885). We showed by luciferase reporter assay that the risk allele of the SNPs disrupt activation by various TEAD transcription factors. We also performed electrophoretic mobility shift assay (EMSA) to test for allele-specific transcription factor binding that affect the family of TEAD transcription factors and the GATA factors. EMSA showed binding of TEAD3 and TEAD4, and differential binding for both GATA genotypes, and luciferase reporter assay confirmed that TEAD3 and TEAD4 activate the non-risk but not the risk allele, and for GATA factors no significant activation was shown. Our investigations lead us to conclude that rs10811656 and rs4977757 are functional and disrupt specific TEAD regulatory sequences within enhancers.

Acknowledgments

I remember on December 18th 2012, I was desperately looking for a supervisor, because it was a very difficult situation, in the sense that I was not allowed to postpone anymore my registration at Ottawa University, after several months even a year of chasing a supervisor. It came to my mind on a night snowy day of December 17th to go to the UOHI (University of Ottawa Heart institute), because at that time this was the only affiliated institution to the University of Ottawa where I hadn't contacted yet any supervisor. Well, I went there, with very detailed information about all the supervisors whom I wanted to contact, and Dr. Stewart was the last one who I saw. I arrived at his office asking for him, and a Fellow told me he just left for somewhere else, and if I wanted to see him, I should wait in the conference room, what I indeed did, but it happened that I waited for him for so long that I fell asleep, till someone caught me sleeping and asked me what I was waiting for, and answered to her that I was waiting for Dr. Stewart. She let me know that she was going to look around for me in order to see if she can let him know that someone was waiting for him in the conference room, but after few minutes the Fellow came back with a no answer, and gave me the option of leaving him a memo for which I said yes. I was about to write the memo, and the angel made his appearance. Yes, I've said angel, because Dr. Stewart was the only who has given me the opportunity after several others said no for no specific reasons, to make this dream come true.

And I think that today, I don't only owe him a thank you; otherwise I also owe him my degree. I won't ever thank you enough for this opportunity that you gave me on

pursuing my graduate studies in your lab. You are such a great human being. I pray for you and I will always do.

I also want to thank all my colleagues in the lab for their support and help, specially to Naïf, every time my tedious project got on my nerves, and the experiments didn't work out, you were there to remind me that science is an everyday learning, thinking, and trying over and over again till you get results (which could be positive or negative, but results anyways).

To Tiffany who has helped me catching up on my techniques when I first arrived in the lab, without your help I would be lost for sure. To Ragnar, for helping me out fixing all the details from the examiners reports, thank a lot. To Yanqing, for his technical assistance with all the electronic devices in the lab, and even more. I also want to thank my Thesis Advisory committee members: Dr. McPherson and Dr. Lagace.

To all my sisters and Mom who give me strength and support every time needed to carry on my path. I won't ever love you and thank you enough; you are and will always be my rainbow in my clouds. To my late father, the only man that I know without any doubt believed in me, thanks a lot for teaching me: "with education everything is possible." You are and will always be in my heart.

I dedicate this work to my precious and adorable nephew Matheo that I so much love, you are such a handsome little cutie boy. I won`t stop saying that I know the real beauty through yours eyes!

TABLE OF CONTENTS

Abstract..... **II**

Acknowledgements..... **III**

List of Abbreviations..... **VII**

List of Figures..... **VIII**

List of Tables ..... **IX**

Introduction

1.1 Coronary artery disease and Modifiable risk factors..... **1**
1.2 Genetic factors and Loci tied to CAD: Genome wide-association and Linkage studies **2**
1.3 9p21.3 Locus..... **3**
1.4 Long Non-Coding RNA in the 9p21.3 locus..... **5**
1.5 Characteristics of TEAD regulatory enhancer elements at the 9p21.3 region..... **7**
1.6 TEAD and its Co-Activators..... **8**
1.7 TEAD transcription factors and their Role in Gene Regulation..... **9**
1.8 GATA factors in 9p21.3 region..... **10**
1.9 Examining the Functional regulation of the SNPs at the 9p21.3 Locus and TEAD transcription factors..... **11**
1.10 Hypothesis, Aims and Rationale..... **13**

Materials and Methods

2.1 Bioinformatics Scan Analysis..... **14**
2.2 Genotyping..... **14**
2.3 DNA Sequencing..... **15**
2.4 SNPs Constructs..... **15**
2.5 Cell Maintenance..... **16**
2.6 Electrophoretic mobility shift assay..... **16**
2.7 Transient transfection of HEK293 and HAoSMCs..... **17**
2.8 Transient transfection for Luciferase Reporter assay..... **18**
2.9 Luciferase Reporter assay..... **19**
2.10 Bacterial Culture and Plasmid Preparation..... **20**
2.11 Immunoblotting..... **20**
- Preparation of Whole Cell Extracts..... **21**
- SDS-Polyacrylamide Gel Electrophoresis and Protein Transfer to PVDV (Po-

lyvinylidene Difluoride) Membrane…… 21

- Blotting……. 22

2.12 Statistical Analysis……. 22

Results

3.1 Bioinformatics Scan Analysis identified 15 SNPs within the 9p21.3 locus……. 23
3.2 Functional Analysis of the genetic variants at the 9p21.3 locus……... 26
3.3 The Regulatory Functions of TEAD transcription factors for these SNPs…… 30
3.4 Endogenous TEAD in HAoSMCs for the risk and non-risk alleles of the SNPs rs108 11656 and rs4977757……. 38
3.5 GATA transcription factors do not alter the risk and non-risk alleles in 9p21.3 locus.. 41

Discussion

4.1 The Functional Relationship between the Polymorphisms at the 9p21.3 Locus and TEAD transcription factors…… 46
4.2 Bioinformatics analysis of the 9p21.3 locus also identified 2 genetic variants of GA TA factors…… 49
4.3 Study Strengths and Weaknesses…… 50

Conclusions..... 53
References…… 55
List of Tables…….. 63
Appendix A Curriculum Vitae……. 67
Appendix B List of Publications…….. 71
Statement……. 72

List of abbreviations

CAD-coronary artery disease
GWAS-genome wide association studies
SNP-single nucleotide polymorphisms
HEK293-human embryonic kidney 293 cells
HAOSMCS-human aortic smooth muscle cells
TGFβ-transforming growth factor beta
CDKN2A-cyclin dependent kinase inhibitor 2A
CHD- coronary heart disease
AS-atherosclerosis
MI-myocardial infarction
SMAD- mothers against decapentaplegic homolog
DNA-dexoxyribonucleic acid
RNA-ribonucleic acid
PCR-polymerase chain reaction
PBS-phosphate buffer saline
TBST-tris-buffered saline tween
DTT-dithiothreitol
MTAP-methylthioadenosine phosphorylase
VSMCs-sascular smooth muscle cells
TCEP- tris2-carboxyethyl phosphine
RIPA-radioimmunoprecipitation assay buffer
SMC-smooth muscle cells

List of Figures

1 The 9p21.3 is associated with coronary artery disease in GWAS…….. **24**
2 Determining the functionality of SNPs rs10811656 and rs4977757 for TEAD transcription factor bindings in HeLa cells…… **27**
3 Determining enhancer activity in rs10811656 and rs4977757 for TEAD transcription factor binding sites in HEK293 cells…… **33**
4 Over-expressing TEAD4 transcription factor in HAoSMCs…… **36**
5 Determining enhancer activity for endogenous TEAD in rs10811656 and rs4977757 in HAoSMCs…… **39**
6 Binding of GATA transcription factors is affected by 9p21.3 risk alleles ……. **42**
7 The rs9632885 and rs10757269 SNPs that alter GATA binding do not significantly alter the enhancer activity in HAoSMCs..... **44**

List of Tables

1 Primers of TEAD1-4 DNA Sequencing…….. **63**
2 Primers for PCR and Sequencing of the 2 GATA SNPs……. **63**
3 Oligonucleotides used for the SNP constructs rs10811656 and rs4977757 for the risk and non risk-alleles……. **64**
4 Sequences for the binding elements for the risk and non-risk alleles with the respective transcription factors involved in the disruptive binding…….. **65**
5 Sequences of the oligonucleotides used for EMSA to test transcription factor binding to SNPs……. **66**

1 Introduction

1.1 Coronary Artery Disease and Modifiable Risk Factors

Coronary artery disease (CAD) is the major cause of morbidity and mortality in most developed countries [1]. According to experts, coronary heart disease (CHD) is expected to become the leading cause of mortality and morbidity in the world by the year of 2020 [2]. CAD is defined as a narrowing of the coronary arteries that supply blood and oxygen to the heart muscle, when there is accumulation of substances in these arteries, which is termed as plaques, the condition is called atherosclerosis. Atherosclerosis is a form of chronic inflammation associated with the secretion of many cytokines, which are produced by T cells, macrophages, endothelial cells, and smooth muscle cells [3], this form of inflammation is also known to be progressive and multifactorial disease that mostly affects elastic and muscular arteries [4].

Although considered a disease of advancing age, approximately 15% of cases are diagnosed before age 65 [5]. In general, coronary atherosclerosis is a complex polygenic disorder, caused by multiple genetic factors, environmental factors, and interactions among these factors [6]. Studies have shown that acute coronary syndrome such as unstable angina, myocardial infarction (MI), and sudden death occur when unstable plaques rupture or ulcerate leading to platelet accumulation and activation, fibrin deposition, thrombus formation and possible vessel occlusion [7,8]. However, rare single-gene forms of atherosclerosis also exist [9].

The so-called ``traditional`` risk factors, such as dyslipidemia, diabetes mellitus, and hypertension, have significant genetic and environmental components [10]. Modification of conventional risk factors has shown a 30% to 40% reduction in mortality and

morbidity [11, 12]. The cumulative risk for CAD in males by age 70 is 35% and by age 90 is 49%. Women typically develop CAD about 10 years later than men with a cumulative risk of 24% and 32% by ages 70 and 90, respectively [13].

Moreover, risk factors that increased mortality by 20% to 30% include high blood pressure, cigarette smoking, total cholesterol (TC), LDL-C, HDL-C, and diabetes [14,15]. Also, factors such as diabetes, left ventricular hypertrophy, family history of premature CHD, and ERT have also been considered in defining conventional modifiable risk factors for CHD [15].

1.2 Genetic Factors and Loci tied to CAD: Genome-wide Association and Linkage Studies.

Studies utilizing genome scan approaches have found novel genetic loci associated with CAD, which might provide additional insight to genetic factors contributing to atherosclerosis. Investigators usually consider the manifestation of CAD as the interaction of several genetic and nongenetic factors. Genetic and nongenetic factors that can be classified into 2 classes as: environmental and behavorial, which predispose individuals to the most forms of CAD, are known to be prevalent in developed countries, including smoking, inactivity, excess calories, and high fat intake. Thus, it has been shown that when CAD occurs at a young age or if the phenotype is severe, the presence of multiple and highly penetrant genetic factors, is likely [16].

As of 2013, 45 loci had been identified by Genome-Wide Association Studies (GWAS), including 15 novel loci discovered in a large meta-analysis called CARDIoGRAM [17]. Notably, only 12 of these loci were used to show that cardiovascular

risk prediction goes beyond the traditional risk factors [18]. Considerable evidence shows that the genetic variants for CAD are very common, occurring on average in 50% of the population with a frequency varying from 2% to 91%, and the majority of these genetic variants for CAD are located in DNA sequences that do not code for protein, which means that the risk variant mediates its increased risk for CAD directly or indirectly through regulation of DNA sequences that code for protein [19].

The first genetic risk variant for CAD was the 9p21.3 that was simultaneously discovered in 2007 by different independent GWAS worldwide [20,21,22,23]. This variant on the p arm of chromosome 9 at the position 21-3 (9p21.3) is the most well-known and replicated in case of CHD [24]. Furthermore, the 9p21.3 locus contains multiple CAD-associated single nucleotide polymorphisms (SNPs) in strong linkage disequilibrium, spanning a genomic region of over 50kb [25, 26, 27, 28].

Since the discovery of this locus, many investigators have established and replicated the association of the 9p21.3 locus with CAD and myocardial infarction. Some studies have theorized that 9p21.3 promotes MI rather than atherosclerosis, suggesting that MI occurs on a substrate of coronary atherosclerosis [22, 23]. While others based their studies on showing that the 9p21.3 locus is associated with CAD by revealing that targeted deletion in the 9p21.3 non-coding interval can activate the proliferation of vascular smooth muscle cells as well as their diminished senescence [29].

1.3 9p21.3 Locus

The 9p21.3 was the first genetic risk locus to be identified by the GWAS, which genetic variants predispose to increased risk of CAD. It is a chromosomal region relatively

replete of open reading frames, and the closest genes are a cluster consisting of CDKN2A-CDKN2B [30]. For instance, the CDKN2A/2B loci are well recognized as tumor-suppressor genes that are involved in the regulation of cell cycle, aging, senescence and apoptosis [31].

The 9p21.3 region, termed as the *`9p21.3 CAD risk interval`*, encompasses multiple genetic variants in tight linkage disequilibrium, spanning a 53-kb region, and considered to be the most significant locus for CAD [32]. The CDKN2A encodes both p16 (INK4A), a negative regulator of cyclin dependent kinases, and p14 alternative reading frame (ARF), an activator of p53. However, the exact function of CDKN2BAS is still unknown, but it has been shown to regulate gene expression of CDKN2A/2B, and SNPs in this locus have been associated with cardiovascular disease, cancer and other diseases in GWAS [33, 34]. The 9p21.3 region also includes a cluster of type I interferon (IFN) genes, which encode pleiotropic cytokines that exhibit strong antiviral, antiproliferative and immunomodulatory effects [35].

In addition, the 9p21.3 locus was reported to give a significant genomic signal for other diseases, like type 2 diabetes, aortic or intracranial aneurysms, peripheral disease or cancers [36, 37]. It is estimated that the world population is heterozygote for the 9p21.3 locus with a 25% increased risk and homozygote with a 50% increased risk, which classes the 9p21.3 risk allele in the same range of magnitude as the population-attributable risk of hypertension (28% in men and 29% in women) [38]. Further, as was recently highlighted the 9p21.3 risk alleles are carried by 75% of the world population (excluding black Africans) and confer risk for coronary atherosclerosis independently of known risk

factors [39]. It also has been highlighted that the 9p21 risk variant for CAD is perhaps the most robust genetic variant and the most studied of those risk variants with unknown function. This risk variant was introduced into the genome after the arrival of higher primates. Also, the 9p21 risk variant is highly conserved in the human genome [40].

This risk locus is of particular interest because the genetic variants fall within an intergenic region that is more considered to be a gene desert: a region of 53,202 [41]. The genetic risk at the 9p21.3 locus consists of a cluster of 59 linked single nucleotide polymorphisms (SNPs) over a 53,202 bp region. A study has revealed a growing consensus sequence of the 59 linked SNPs in this region that alter DNA sequence and disrupt or create transcription factor binding sites that may alter gene expression [42].

1.4 Long Non-Coding RNA and the 9p21.3 Locus

ANRIL encodes a long antisense noncoding RNA in the INK4A locus that overlaps with the 9p21.3 risk interval and may have distinct regulatory roles in the expression of adjacent protein coding genes, which includes methylthioadenosine phosphorylase (MTAP) and the 2 cyclin kinases inhibitors (CDKN2A/B) [43]. For instance, MTAP was identified by GWAS as a naevus and melanoma associated gene that encodes an enzyme that has a role in polyamine metabolism. Further, loss of this gene expression can exert a tumor-promoting effect [44, 45].

The 9p21.3 risk locus contains no protein coding genes or known microRNAs. However, the 9p21.3 locus does contain a sequence for an antisense RNA (ANRIL, CDKN2BAS) [41]. It has been shown that the risk locus overlaps exons 13-20 of a recently identified large, non-coding, antisense RNA in the INK4 locus, also known as

CDKN2BAS [46].

Generally, long non-coding RNAs are not transcribed in large quantities in the genome, and may have a known functional significance, because as many other protein coding genes, they are not highly conserved between species, and contain short open reading frames (ORFs) from 50-100 amino acids, and can span large genomic distances [47,48]. A recent study has revealed that the ANRIL may be involved in the atherosclerotic process such as in thrombogenesis, vascular remodeling and/or repair, and plaque stability through mechanisms such as RNA interference [49].

To date, the effect of 9p21.3 on expression of ANRIL still remains inconsistent, regarding most data that have been collected from circulating blood cells [50, 51]. However, expression analyses in human atherosclerotic plaques showed that variants on 9p21.3 region resulted in changes in the expression of the ANRIL transcripts, which may affect ANRIL expression and neighboring protein-coding genes. Alternatively, ANRIL might only be a bystander of a completely different effect of the 9p21.3 locus that still needs to be determined. This notion is supported by the relatively low expression levels and relative expression differences of ANRIL transcripts [43].

Although variants at the 9p21.3 regions have been shown to have an effect in ANRIL expression, it is also important to point out that considerable evidence revealed the presence of several regulatory enhancer elements at this locus [29]. Further, the 9p21.3 locus is of particular interest because the CAD-associated SNPs fall within an intergenic region that is so far considered a gene desert: 53kb flanked by CDKN2B (130KB upstream) and DMRTA (370KB downstream), indicating that the functional

variants underlying the association are probably in regulatory elements. Furthermore, a previous study has identified regulatory elements in the 9p21 gene desert by examining transcription factor binding and chromatin modification profiles in human cells, which also suggested that variants at this locus disrupt consensus transcription-factor binding sites in the predicted regulatory elements [41], while another study showed that a linked polymorphism found nearby to the putative STAT1 binding sequence may disrupt a putative TEAD consensus sequence binding site (CATTCCG>CATTCTG) [52]. Therefore, it will be of interest to investigate the function of those regulatory enhancer elements in this region.

1.5 Characteristics of TEAD regulatory enhancer elements at the 9p21.3 region

This family of proteins is a group of 4 highly homologous TEAD transcription factors in mammals. It is also known to be a small group that not only regulates expression of genes in the proximity of TEAD-binding regulatory promoter/enhancer but also regulates distal ones [53, 54].

TEAD is a family of proteins that has a TEA domain that enables them to interact with specific DNA sequences and a trans-activation domain that interacts with co-activators [55, 56]. For instance, they are known to be evolutionarily conserved proteins and their orthologs are widely found in all vertebrates, invertebrates, and even single-celled eukaryotes. Furthermore, TEAD transcription factors are proteins that are implicated in different processes such as: cardiogenesis, myogenesis, myogenic signalling propagation, development of neural crest, and trophoectoderm, and cancer [57, 58, 59, 60, 61].

TEAD proteins have emerged in invertebrates as transcription factors implicated in the specific activation of muscle genes through their binding to the M-CAT motif found in the regulatory region of numerous muscle- lineage specific genes [62]. Moreover, this family of proteins is known to be upregulated developmentally and pathologically by several important genes [63]. TEADs as nuclear proteins have the ability to bind and activate transcription from the SV40 enhancer. A study has reported that 4 homologous TEAD genes (TEAD1-TEAD4) are widely expressed during development [64]. The downstream functionality of TEAD proteins depends widely on their interaction with cofactors such YAP, TAZ, and vestigial like protein [65, 66, 67, 68, 69].

1.6 TEAD and its Co-Activators

TEAD transcription factors interact with multiple co-activators, because on their own this family of proteins is unable to activate transcription. Therefore, they require the help of co-activators [63]. For instance, co-activators do not bind DNA; they pair with transcription factors and activate transcription. Moreover, they generally have activation domains that facilitate their interaction with the basal transcription or chromatin remodeling machinery [70].

TEAD interacts with several co-activators that can be classified into multiple groups: YAP, TAZ, and vestigial like proteins. For instance, Yes-associated protein (YAP) transcriptional co-activator is a key regulator of organ size and a candidate human oncogene inhibited by the Hippo tumor suppressor pathway. TEAD binds directly to and mediates YAP-induced gene expression [71].

Studies have revealed that only TEAD family of transcription factors has been demonstrated to be important for the growth-promoting function of YAP [64, 66]. It has also been shown that knockdown of TEAD aborts expression of the majority of YAP-inducible genes and largely attenuates YAP-induced overgrowth, epithelial-mesenchymal transition (EMT), and oncogenic transformation [66]. Recent studies have shown that TEADs and their co-activators may play a role in the progression of various cancers including the difficulty to treat glioblastoma, liver and ovarian cancer [72], and also may have a key role in cancer stem cell pluripotency, self renewal regulation [72].

1.7 TEAD transcription factors and its Role in Gene Regulation

As pointed out previously, TEAD family of proteins regulates development of various tissues, including heart, skeletal muscles, neural crest, notochord and trophoectoderm [65]. Considerable evidence has shown that modulation of TEAD activity alter cell proliferation and cell death [64], and also as downstream target of MyoD1 and MyoG, TEAD are required to regulate muscle structural genes [73]. Further, studies suggest that scalloped (Sd) interacts with Yki and mediates signalling [74, 75, 76]. Similarly, Max a nuclear phosphoprotein that forms a heterodimer with MYC, and poly (ADP-ribose) polymerase, a chromatin bound protein, can also bind TEAD and stimulate expression of cardiac muscle specific genes [76]. In previous years, experimental evidence has suggested that TEAD binds to human papillomavirus-16 (HPV-16) enhancer and activates the HPV-16 oncogenes E6 and E7 [77].

TEAD family of proteins also interacts with Smad3 and cooperates in TGFβ signalling [78]. For instance, TGFβ signalling activates SMAD transcription factors to

regulate gene expression [79]. TGFβ responsive genes can be activated directly by SMADs through SMAD-responsive elements, or by other transcription factors via their interaction with SMAD proteins [80]. Further, TEADs transcription factors interact with SMAD3 and mediate TGFβ-dependent gene activation [78, 81]. For instance, (TGF-β) is involved in the pathogenesis process of multiple cardiovascular diseases including atherosclerosis. Studies have shown that low concentrations of TGF-β may increase smooth muscle cell proliferation with the fact that lows concentrations of plasma TGF-beta are associated with a poor outcome in CAD [82,83]. Inversely, at higher concentrations, TGF-Beta inhibits both the migration [84], and proliferation [85] of VSCMs in cell culture.

1.8 GATA factors in the 9p21.3 region

GATA factors also known as GATA-binding proteins are known to be an ancient family of transcription factors that recognize a target site conforming to the consensus WGATAR (W=A or T and R=A or G) [86]. A common feature of all GATA proteins is their high-affinity binding to a sequence motif conforming to the consensus T/A (GATA) A/G. For instance, motifs with this particular property have been found in various regulatory regions such as: promoters of genes expressed in erythroid, megakaryocytic, and endothelial cells: globin and T-cell receptor α and δ-chain gene enhancers; and α-and β-globin locus control regions [87]. Further, DNA recognition is achieved through novel zinc fingers that can be present once or twice in the protein [88, 89, 90]. This family of transcription factors consists of six proteins (GATA1-6).

Various members of this family of proteins have the abilities to recognize a closely

related, but not identical, DNA sequence element that raises interesting possibilities in the way that differential gene regulation can be accomplished in cells expressing more than one GATA protein [86]. The GATA factors are known to play role in processes such as differentiation, growth and survival [91, 92, 93, 94, 95]. Thus, the involvement of this family of proteins in cell formation renders mutations in family members, such as GATA2, potential candidates for cardiac malformations and smooth muscle dysfunction. Indeed, many studies have examined associations of defects in genes coding for some of these transcription factors with cardiac and congenital malformation [96, 97, 98].

Some GATA factors also function as a transcriptional partner or involved in transcriptional pathways with several other important cardiac transcription factors, including MEF2C (Myocyte enhancer factor 2C). This transcription factor physically interacts with GATA factors to activate the expression of other genes [99]. Very little is known about the role it might play in CAD. Only few studies available today were able to postulate its association with both early-onsets of CAD [100,101,102], and sporadic disease [103], while other studies were unable to establish any relationship with CAD [104].

1.9 Examining the Functional regulation of the SNPs at the 9p21.3 Locus and TEAD transcription factors.

Although studies have revealed the presence of several enhancer elements present at the 9p21.3 locus, which also suggested that variants at this locus may disrupts consensus transcription factor-binding sites [41,42], the functional mechanisms between the identified SNPs and the TEAD consensus sequence found in this locus still remain

unknown. Here, I report the functional regulation of these SNPs identified in the 9p21.3 region, and the TEAD sequences that alter the 9p21.3 locus associated with the risk of CAD.

I demonstrate that these genetic polymorphisms may play a role in the mechanisms of this locus, by disrupting specific TEAD regulatory sequences, and also by activating and/or deactivating the non-risk or the risk. My findings show the functionality of these genetic variants in the functional analysis of the 9p21.3 locus, and also show the role of these SNPs in the risk of CAD.

1.10 Hypothesis and Aims of the thesis

We hypothesize for this research project that: Some genetic polymorphisms at the 9p21.3 locus are functional and will disrupt the specific regulatory sequences within enhancers. We tested our hypothesis through the following aims:

The specific aims of this project include:

I. To determine how polymorphisms affect TEAD-dependent gene regulation at the 9p21.3 locus.

II. To investigate how polymorphisms affect GATA factor-dependent regulation at the 9p21.3 locus.

Thesis rationale as it relates to human health

Globally, people suffering from coronary artery disease (CAD) have an increased burden of the traditional risk factors that contribute to this disease. To date, CAD is a major concern in the developing and developed world. Although, CAD has many complications due to unstable angina, myocardial infarction (MI) and sudden death [7,8], there is compelling evidence from epidemiological studies that have shown that genetic variants contribute to the disease progression [22]. Thus, examination of the genetic polymorphisms in the 9p21.3 locus may help to understand the phenotypic heterogeneity of coronary artery disease and reveal novel therapeutic strategies for the treatment of CAD.

2 Materials and Methods

Experimental Analyses

2.1 Bioinformatics Scan analysis

We undertook a bioinformatics scan analysis of the 59 linked SNPs at the 9p21.3 to identify genetic variants that would disrupt or create transcription factor binding sites in the 9p21.3 coding region covering 53,202bp, using the ENCODE chromatin immunoprecipitation (http://encodeproject.org/ENCODE) and JASPARCORE (http://jaspar.genereg.net) databases. Encode is an online tool which consists of building a comprehensive parts list of functional elements in the human genome, including those that act at the protein and RNA levels, and regulatory elements that control cells and its genetic environment. JASPARCORE is a collection of transcription factor DNA-binding preferences modeled as matrices that contains a collection of experimentally defined transcription factor binding sites for multi-cellular eukaryotes.

2.2 Genotyping

DNA from HEK293, HAoSMCs, and HeLa cells was genotyped for the 9p21.3 locus; these are the cells that were sequenced in method 2.3. A DNA fragment that contains SNP rs10811656 was obtained using primers 5-CGG TGT GGT CAT TCC GGT AAG CAG CGC-3 and 5-TCG AGC GCT GCT TAC CGG AAT GAC CAC ACC GGT AC-3 for the non-risk oligonucleotides, and 5-TCG AGC GCT GCC TAC CAG AAT GAC CAC CAC CGG TAC-3 for the risk oligonucleotides, and another fragment containing SNP rs4977757 was obtained using primers 5-CTA TCT TTG TGG CAT TCT CTG TAT TTC-3 and for the non-risk oligonucleotides 5-TCG AGG AAA TAC AGA GAA TGC CAC AAA

GAT ACG GTA C-3 and 5-TCG AGG AAA TAC AGA GAA CGC CAC AAA GAT ACG GTA C-3 for the risk allele.

2.3 DNA Sequencing

The risk and non-risk of each SNP were amplified by PCR using Taq DNA polymerase. Each 25μL of reaction contained 100-300 ng of plasmid DNA template, 5X sequencing buffer, big dye solution, forward and reverse primers at 1:100 dilution, and H_2O. PCR reactions were run on a Thermocycler using the following cycling parameter: a hold at 4^oC for 30 sec, 25 cycles of 95^oC for 30 sec, annealing at 50^oC for 15 sec, and extension at 60^oC for 4 min. Later, after amplification, PCR reaction was used as a template for DNA sequencing. Each 25 μL of sequencing reaction contained 1 μL of PCR reaction template, 1 μL of sequencing primer, 2 μL of Big Dye sequencing buffer. Sequencing reactions were cleaned using Agencourt CleanSEQ beads and ethanol.

2.4 SNP Constructs

Generation of Luciferase reporter constructs

We amplified a 50 bp known genotype by PCR fragment containing the rs10811656 and rs4977757 polymorphisms from HEK293 cells homozygous for the risk and non-risk alleles. SNPs with their corresponding nucleotides were flanked with KpnI and XhoI sites and then subcloned into pGL3-promoter luciferase reporter vector.

We purchased complementary 30bp oligonucleotides containing the rs9632885 and rs10757269 polymorphisms for the risk and non-risk alleles from IDT (Integrated DNA Technologies, Iowa, USA). SNPs with their respective sequences were flanked with KpnI

and XhoI sites and then subcloned into pGL3-promoter luciferase reporter vector, and sequence verified.

2.5 Cell maintenance

HEK293 cells were maintained at 5% CO_2 in 1X Minimal Essential Medium (1XMEM) containing 4.5g/L glucose, and supplemented with 10% FBS and 1% L-glutamine and 1% Pen-Strep (Penicillin-Streptomycin). HAoSMCs (Human primary aortic smooth muscle cells were purchased from Cell Applications (San Diego, CA) and were maintained at 5% CO_2 in SmBM (Smooth Muscle Cell Basal Medium), containing 10% FBS, 0,025% GA (Gentamicin Sulfate Amphotericin-B), 0,025% rhEGF (Epidermal Growth factor Human Recombinant), 0,05% rhFGF-B (Human Fibroblast Growth factor-B) 0,025% Insulin Recombinant Human cell culture Tested (The culture medium for HAoSMC was purchased from Lonza Walkersville, MD) (SmGM-2 Bulletkit, Cat# CC-3182). Both cell lines were subcultured when they attained a confluency of 80%, and were plated at a minimum density of 25%. In the case of HAoSMCs, the number of passages was limited to 9.

2.6 Electrophoretic mobility shift assay (EMSA)

Nuclear protein extracts were prepared from cultured HAoSMCs and HeLa cells (as described by Farrance et al. [106]. All subsequent procedure was carried out at room temperature. Nuclear protein concentrations were determined by Bradford assay. EMSA using SNP-specific oligonucleotides (**Table 5**) was performed using double-stranded synthetic oligonucleotides of the 35-pGL3 promoter sequence that was used as probe. T4

polynucleotide kinase and [γ-^{32}ATP] (6000 Ci/mmol) were used to label the 5' end of the upper strand oligonucleotide. To maintain virtually 100% of probe in the double-stranded form, labeled coding strand oligonucleotide was annealed to a 2-fold molar excess of unlabeled noncoding strand oligonucleotide. Binding reactions were initiated by incubating nuclear extracts (10-15 μg of protein) with 750 ng of poly (dI – dC) for 15 min at room temperature. Thirty fmol of double-stranded probe was then added, and incubations were continued for another 30 min at room temperature. Final binding reaction (10μl) contained 20 mM Hepes at pH 7.8 70 mM KCL, 0.5 mM DTT, 0.4 mM EDTA, 0.1 mM EGTA, and 10% glycerol. For competition EMSA, a 100-fold molar excess of various competitor oligonucleotides was included from the initial incubation. The binding reactions were loaded onto a 4% native polyacrylamide gel running at 11 V/cm containing 45 mM Tris borate at pH 8.3 and 1.25 mM EDTA. After running for 2 h at 4° C to separate free and complexed probe, the gel was dried and exposed to a storage phosphor screen (GE Healthcare Life Sciences). Where available, we have used commercial antibodies known to supershift transcription complexes. (Method was performed by Naif Almonthashiri). TEAD3 was a mouse polyclonal antibody and TEAD4 a mouse monoclonal, and both are from (Abnova). GATA4 antibody was provided by Dr. Mona Nemer's lab.

2.7 Transient Transfection of HEK293 and HAoSMC cells

Seeding cells for transient transfection

10% HEK293 and HAoSMC cells were grown in 10 cm dishes and seeded upon reaching 80% confluency. HEK293 and HAoSMCs were plated at density of 25% and 40% of one 10cm dish per 6-well plate, respectively. 96 hours after seeding, both cell types reached

at least 75% of confluency and were ready to be transiently transfected.

2.8 Transient transfection for Luciferase reporter assays

The cells were transfected 96 hours after seeding, by using the Invitrogen Lipofectamine 2000 transfection reagent. The main goal of this experiment was to determine whether members of the TEAD family of transcription factors would drive the expression of luciferase reporters containing the risk and non-risk alleles of rs10811656 and rs4977757. This experiment had 2 controls: one used the empty pGL3 luciferase vector for co-transfections with each of the expression plasmids to control for the presence of cryptic TEAD sites within the pGL3-luciferase plasmid that might have been activated by TEAD transcription factors. The other used an empty pCMV expression plasmid to control for promoter competition. The pGL3-luciferase reporter plasmids bearing the 50 bp non-risk or risk alleles of rs10811656 and rs4977757 were co-transfected with expression plasmids bearing the cDNA sequences of each TEAD family member (TEAD1, TEAD2, TEAD3, and TEAD4) under the control of the human cytomegalovirus (CMV) enhancer. As a control, the pCMV expression vector bearing no cDNA insert was co-transfected with the pGL3-luciferase reporters. As mentioned above, the SNPs (rs10811656, rs4977757) with their respective sequences **(as seen in Table 3**) from risk and non-risk alleles were subcloned into the pGL3-promoter luciferase reporter vector. The pGL3-promoter vector contains an SV40 promoter, a cDNA sequence encoding a luciferase transcriptional unit, a multiple cloning site that can be digested with a number of restrictions enzymes, an origin of replication, and a marker for ampicillin resistance. This vector does not contain an

enhancer, enabling DNA fragments containing putative enhancer elements to be inserted either upstream or downstream of the promoter-luciferase transcriptional unit. Co-transfection was carried out in HEK293 cells. All assays were performed in triplicates and duplicates. Each well was transfected with 1µg total DNA (200 ng pGL3-promoter luciferase expression vector, and 800ng pGL3-rs10811656, rs4977757 reporter vectors). For co-transfection with TEAD1, TEAD2, TEAD3, TEAD4 expression vector using 6.5µl of DNA for the expression vector and 26 µl of DNA for the reporter vector.

Plasmid DNA was diluted with TE buffer from Invitrogen at a concentration of 100ng/µl in all luciferase assay transient transfections, and Lipofectamine 2000 from Invitrogen at a concentration of 1mg/mL was added to each well at a volume of 2µl per well. The incubation time at room temperature after mixture was always 20 min for the samples. After incubation, 500µl of mixture (DNA+Lipofectamine) was added to each well. The cells were harvested for the luciferase reporter assay for 48 hours post-transfection.

2.9 Luciferase Reporter Assay

HEK293 and HAoSMC cells were harvested 48 hours post-transfection. Cells were washed with 1X cold PBS and harvested using the cell scraper by the addition of 1000µl of 1X cold PBS +EDTA per well. After addition of this buffer solution the cells were harvested and spun down for 5 min at 13,000 rpm at 4^{o}C to pellet cell debris. Later, the pellet was resuspended with-100µl of TCEP (Tris 2-carboxyethyl phosphine) buffer by vortexing each tube. Following resuspension, the eppendorf tubes with lysates were quick frozen 3 times for 10 min at -80°C in ethanol at a concentration of 100%. And later the

lysates were centrifuged again for 5 min at 13,000rpm at 4°C. Post centrifugation; the supernatants were transfected to clean 1.5ml eppendorf tubes.

The luciferase reporter assay was performed using a single-tube luminometer (Berthold). First 300 µl of lux buffer was added to the single-tube luminometer, and prior to measure the activity of each respective lysate, 100 µl of luciferin was mixed with the 300 µl of lux buffer. Next, 20 µl of cleared lysate was laterly added into the same tube, and luciferase activity was recorded. This sequence was repeated for each single lysate until all lysates were measured for luciferase reporter activities.

2.10 Bacterial Culture and Plasmid Preparation

Chemically competent *E.coli* TOP10 (Invitrogen) were transformed by heat shock at 42°C and incubated in a rotary shaker at 37°C and at 250 rpm for 90 mins. The transformation reaction were plated on LB-agar plates and incubated overnight at 37°C. After transformation, *E.coli* TOP10 was plated on LB-agar containing 100mg/mL ampicillin for colony screening. For maxi-prep of plasmid DNA, single colonies were inoculated in 100 mL of LB medium containing 100mg/mL ampicillin and incubated in a rotary shaker at 250 rpm overnight for 14 to 16 hours. Later plasmid DNA was purified using the Invitrogen High Pure Plasmid Isolation kit for maxi-prep.

2.11 Transient Transfection for Immunoblotting

HAoSMC cells were transfected 96 hours after seeding, by using Invitrogen Lipofectamine transfection reagent. Each 10 cm plate was transfected with 30 µg of total

DNA in 1.5 mL of free smooth muscle medium and 45 µL of lipofectamine in 1.5mL of free smooth muscle medium in order to bring the total up to 3 mL the volume of the plate. Later, cells were incubated for 6 hours. After incubation, cells were harvested within 48 hours.

2.12 Immunoblotting

- **Preparation of Whole Cell Extracts**

HAoSMC cells (10 cm plate) were washed with 1X cold PBS and harvested by lysis with 150 µl of RIPA (Radioimmunoprecipitation assay buffer). 1 tablet of phosphatase inhibitor + 1 tablet of protease inhibitor were dissolved in 150 µl of RIPA, followed by vortexing for 20 secondes. Later, cells were incubated on ice for 10 minutes, and then vortexed again for 20 secondes. After the last vortexing, cells were spun down for 10 min at 4°C at 13,000 rpm. Later the supernatant was added in new 1.5 mL eppendorf tubes for future analysis of the lysates concentration by Bradford assay normalizing the protein concentration to BSA standard curve.

- **SDS-Polyacrylamide Gel Electrophoresis and Protein Transfer to PVDF (Polyvinylidene difluoride) Membrane**

4X SDS-PAGE loading buffer and 1M DTT were added to 30µg of protein lysate. All samples were boiled for 5 minutes, and later loaded onto a 10% polyacrylamide gel with 4% stacking gel. The electrophoresis was performed at 80-90V initially for 15 minutes, later was increased to 120 V till bands enter resolving gel and the loading dye disappeared. The proteins were transferred to a PVDF membrane at 150V for 60 to 90

minutes using chilled 1X transfer buffer containing 20% methanol.

- **Blotting**

After the proteins were transferred, the PVDF membrane was rinsed out briefly with 0, 05% of TBST (Tris Buffered Saline with Tween) and blocked in 5% non-fat milk in TBST for 1 hour at room temperature with gentle shaking. The blot was later incubated overnight at 4 °C with primary antibody TEAD1 mouse monoclonal, TEAD4 mouse polyclonal) at a 1:1000, 1:5000 dilution in blocking buffer with 0.05% TBST. The blot was washed for 4x15 min with 0.05% TBST) and incubated at room temperature with secondary antibody at 1:15000 blocking buffer containing 0.05% TBST for 1 hour with gentle shaking. After incubation, the blot was rinsed out again for 4x15 min in TBST. Later, the blot was decanted by getting rid of the TBST to later covering it with chemoluminescence in order to detect the signal using the Film Processor SRX-101A.

2.13 Statistical Analysis

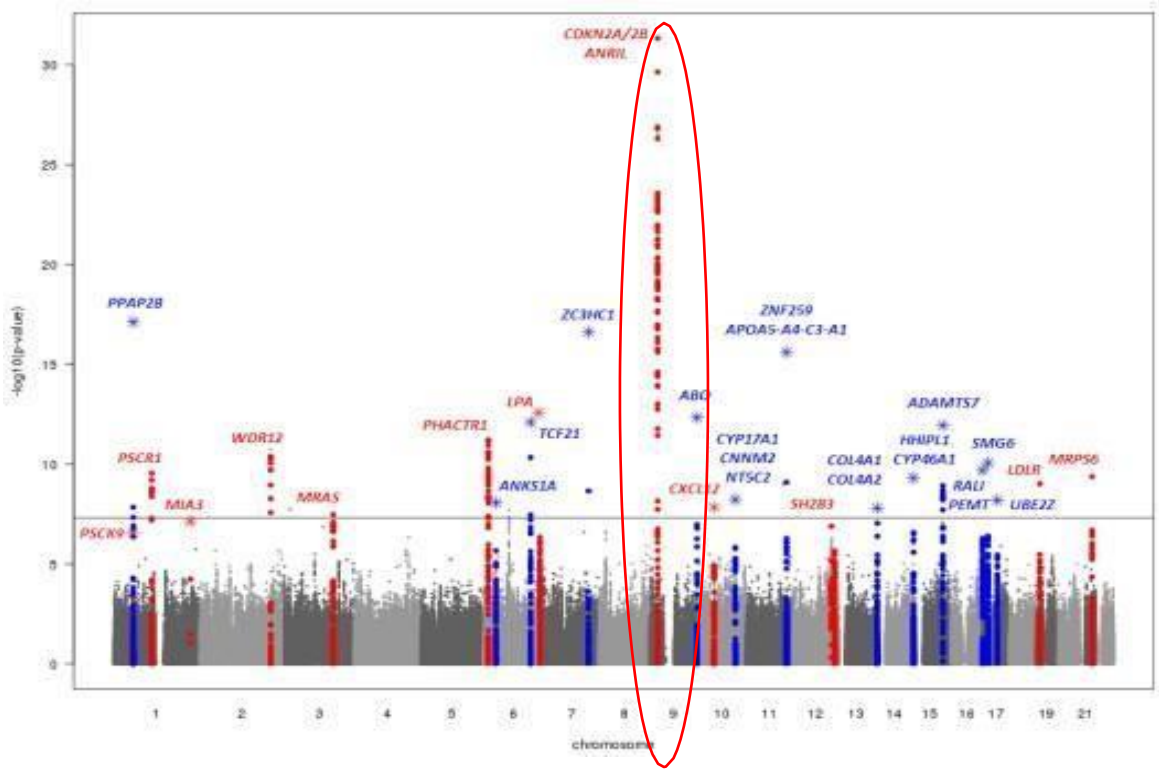
Each luciferase assay was performed in duplicate and luciferase activities of the reporters were expressed as fold of a negative control (luciferase activity of the empty reporter vector or the luciferase reporter co-transfected with an empty expression plasmid) set at 1 fold. Luciferase activities for each experimental treatment were averaged from multiple independent experiments and expressed as Mean±SEM. One-way ANOVA and post-hoc analysis were carried out to determine differences in mean values between groups. The Bonferroni method was used to correct for multiple comparisons. A $P<0.05$ was considered significant. All the statistical analysis in this study was carried out by my colleague **Ragnar Oli Vilmundarson**.

3 Results

3.1 Bioinformatics scan analysis identified 15 SNPs within the 9p21.3 locus

The 9p21.3 locus was the first to yield to genome-wide association studies (GWAS) seeking common genetic variants predisposing to increased risk of CAD [42]. Although it has been shown that the 9p21.3 locus represents the most replicated marker of CAD and MI [105], the functional mechanisms between those genetic variants at 9p21.3 locus still remain elusive. A mutagenesis study has clearly defined TEAD consensus sequences (5'- CATTCCT-3') in the context of muscle-specific promoters [106]. We examined the polymorphisms at the 9p21.3 locus; we performed bioinformatics scan analysis using the Encode chromatin immunoprecipitation and JASPAR Core within the 9p21.3 locus to identify single nucleotide polymorphisms. Our bioinformatics scan using the ENCODE chromatin immunoprecipitation and JASPAR CORE conducted within the 9p21.3 locus identified genetic variants **(as shown in Figure 1**) that could disrupt or create transcription factors binding sites in the 9p21.3 coding region and the flanking 53,202 bp.

This design study identified 15 SNPs as potentially disruptive for transcription factor binding. A cluster of single nucleotide polymorphisms was identified within this locus as potentially disruptive for transcription factor binding. Four of the 59 SNPs (rs10757271, rs4977757, rs1333043, and rs10811656) disrupt TEAD factors consensus sites (5'- CATTCCT-3').

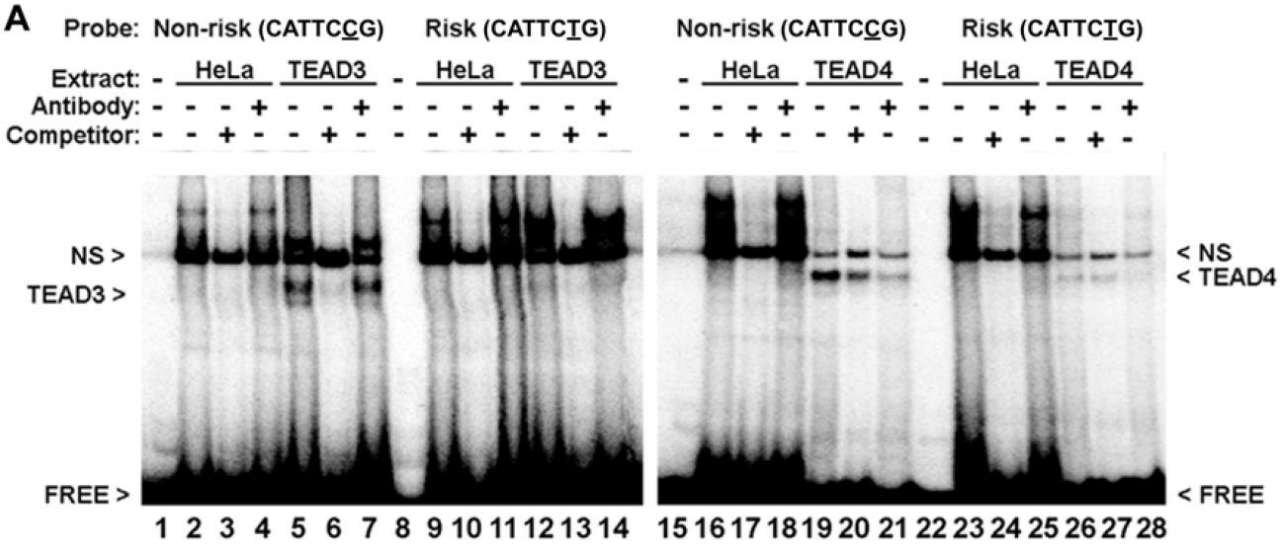


Schunkert et al. Nature Genetics 2011

Figure 1: The 9p21.3 locus is associated with coronary artery disease in the genome-wide association studies. The Manhattan plot of GWAS where the x-axis represents the genome in physical order, the y-axis shows $-\log_{10}$ P values for all SNPs. Data from discovery phase are shown in circles and data from the combined discovered and replication phases in stars. Genes at the significant loci are listed above the signals. Know loci are shown in red and novel loci are shown in blue. The oval outlines the short arm of chromosome 9 and the 9p21.3 CAD risk locus [107].

3.2 The Functional Analysis of the Genetic Variants at the 9p21.3 Locus

Variants at the 9p21.3 locus have been established as amongst the strongest common genetic factors associated with the risk of CAD in people of European continental ancestry [23]. GWAS have revealed a strong association between DNA sequence variation on chromosome 9p21 and the risk of CAD [21, 22, 23]. Many different groups confirmed the evidence of those studies [25, 26, 27, 28]. In order to determine whether the genetic variants at the 9p21.3 play role in the function of this locus, we performed EMSA to test for allele-specific transcription factor binding. Although our bioinformatics scan analysis revealed 15 SNPs that could potentially disrupt transcription factor binding sites, EMSA was used to exclude SNPs that show no differential binding between the risk and non-risk alleles. **Figures 2A** and **B** show that the SNPs rs10811656 and rs4977757 disrupt TEAD3-4 regulatory sequences within enhancers' activity for the risk alleles. In contrast, the non-risk alleles for those SNPs were binding detected in the presence of TEAD3-4 regulatory sequences, suggesting that the SNPs rs10811656 and rs4977757 may be functional within the 9p21.3 locus. These EMSA were performed in HeLa cells.


A
Probe: Non-risk (CATTCCG) Risk (CATTCTG) Non-risk (CATTCCG) Risk (CATTCTG)
Extract: - HeLa TEAD3 - HeLa TEAD3 - HeLa TEAD4 - HeLa TEAD4
Antibody: - - - + - - + - - - + - - + - - - + - - + - - - + - - +
Competitor: - - + - - + - - - + - - + - - - + - - + - - - + - - +
NS >
TEAD3 >
FREE >
< NS
< TEAD4
< FREE
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

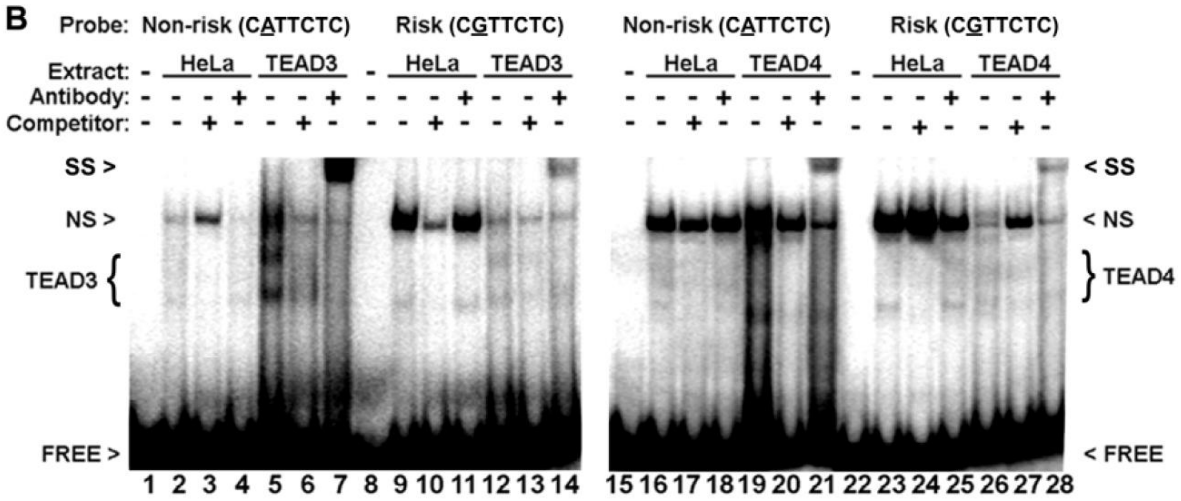

B
Probe: Non-risk (CATTCTC) Risk (CGTTCTC) Non-risk (CATTCTC) Risk (CGTTCTC)
Extract: - HeLa TEAD3 - HeLa TEAD3 - HeLa TEAD4 - HeLa TEAD4
Antibody: - - - + - - + - - - + - - - + - - - + - - - + - - - +
Competitor: - - + - - + - - - + - - + - - - + - - + - - - + - - + -
SS >
NS >
TEAD3
FREE >
< SS
< NS
TEAD4
< FREE
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

Figure 2 A-B: Determining the Functionality of SNPs rs10811656 and rs4977757 for TEAD transcription factor binding in HeLa cells. EMSA showed that the SNPs rs10811656 (**A**) and rs4977757 **(B**) disrupted TEAD3-4 binding for the risk alleles, (where lanes 2 to 4 show HeLa cells as control and lanes 5 to 7 show the over-expression of TEAD3-4 in HeLa for both genotypes). Nuclear extracts from HeLa cells were transiently transfected with TEAD3 and TEAD4 using empty pGL3 promoter expression plasmid and oligonucleotide probes that contains the risk and non-risk alleles for both SNPs (rs10811656 and rs4977757). The non-risk alleles were functional, forming specific DNA-TEAD complexes. However, the risk allele for both SNPs (rs10811656 and rs4977757) disrupted the TEAD binding site to an extent of $1/10^{th}$ compared to the non-risk allele. n=3 independent experiments. (**Experiments performed by Naif Almontashiri)**

3.3 The Regulatory Functions of TEAD transcription factors for these SNPs.

Since TEAD family of proteins are known to regulate expression of genes in the proximity of TEAD-binding regulatory promoter/enhancer elements, and are also required for cardiogenesis, myogenesis and for the development of others organs [57, 58, 59, 60, 61], genetic variants (rs10811656 and rs4977757) were tested for enhancer activity using luciferase reporter assays.

Although our bioinformatics scan analysis identified a marked over representation of the TEAD family of transcription factor binding sites in the 53k bases pair regions at the 9p21.3, four of the 59 SNPs (rs10757271, rs4977757, rs1333043, and rs10811656) disrupt TEAD factor consensus sites (5`CATTCCT-3`), and only 2 of these SNPs were tested for the functional significance of the TEAD consensus sequences in participating in genes regulation by the 9p21.3 locus, because the other 2 SNPs did not show a gel shift (data not shown). I performed luciferase reporter assays to test whether the non-risk alleles for SNPs rs10811656 and rs4977757 respond to transcriptional activation by different members of the TEAD family of transcription factors and to determine whether the risk allele affects this activation. Both SNPs (rs10811656, rs4977757) from their risk and non-risk alleles were tested with the TEADs in order to examine whether or not they are functional.

As observed in Fig 3A, the activity of the pGL3-luciferase reporter was not different whether it carried the non-risk (black bars) or the risk (red bars) alleles of rs10811656 when co-transfected with the empty pCMV expression vector. When over-expressing TEADs in HEK293 cells in the presence of the non-risk allele of rs10811656,

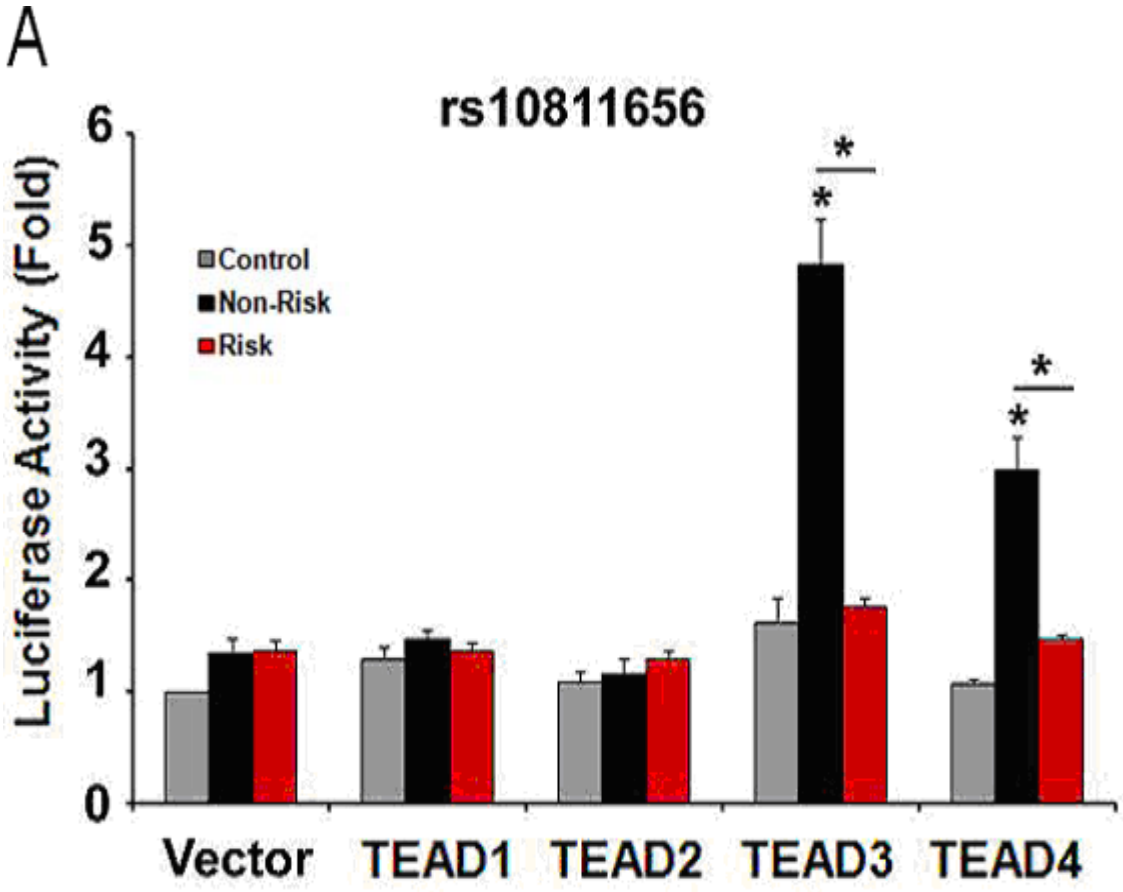
TEAD3 and TEAD4 produced a robust transcriptional activation while TEAD1 and TEAD2 showed no activity compared to the empty pCMV expression plasmid. The risk allele for rs10811656 showed markedly and significantly reduced transactivation by TEAD3 and TEAD4 compared to the non-risk allele. In Fig 3B the pGL3-luciferase reporter bearing the non-risk allele of rs4977757 (black bar) was significantly more active than the empty pGL3-luciferase reporter (gray bar) or the pGL3-luciferase reporter bearing the risk allele of rs4977757 (red bar) when co-transfected with the empty pCMV expression plasmid. In both Fig. 3A and 3B, the empty pGL3-luciferase reporter (gray bars) showed similar activities when co-transfected with TEAD expression plasmids, indicating that no cryptic TEAD element is contained within the empty pGL3-luciferase reporter.

As in Fig. 3A, neither TEAD1 nor TEAD2 had an effect on these reporters compared to empty pCMV expression vector. The non-risk allele was activated by TEAD3 and by TEAD4 over-expression and the risk allele of rs4977757 markedly impaired the transactivation.

In summary, the luciferase reporter assay revealed that the 9p21.3 locus contains enhancer sequences that respond to TEAD factors in HEK293 cells (**Figure 3 A-B**) and that the risk alleles significantly interfere with this response.

To confirm that the pCMV-TEAD4 expression plasmid could over-express TEAD4, the protein level of TEAD4 was also examined in transiently transfected HAoSMCs by

Western Blot analysis in order to confirm its expression in primary cell type that is pathologically relevant to CAD (**As shown in Figure 4).** This western blot analysis was used to confirm the level of expression of the TEAD construct, and as the results show: an increased amount of TEAD4 was found when over-expressing this construct in HAoSMC cells.


A
rs10811656
Luciferase Activity (Fold)
0
1
2
3
4
5
6
Control
Non-Risk
Risk
*
*
*
*
Vector
TEAD1
TEAD2
TEAD3
TEAD4

B

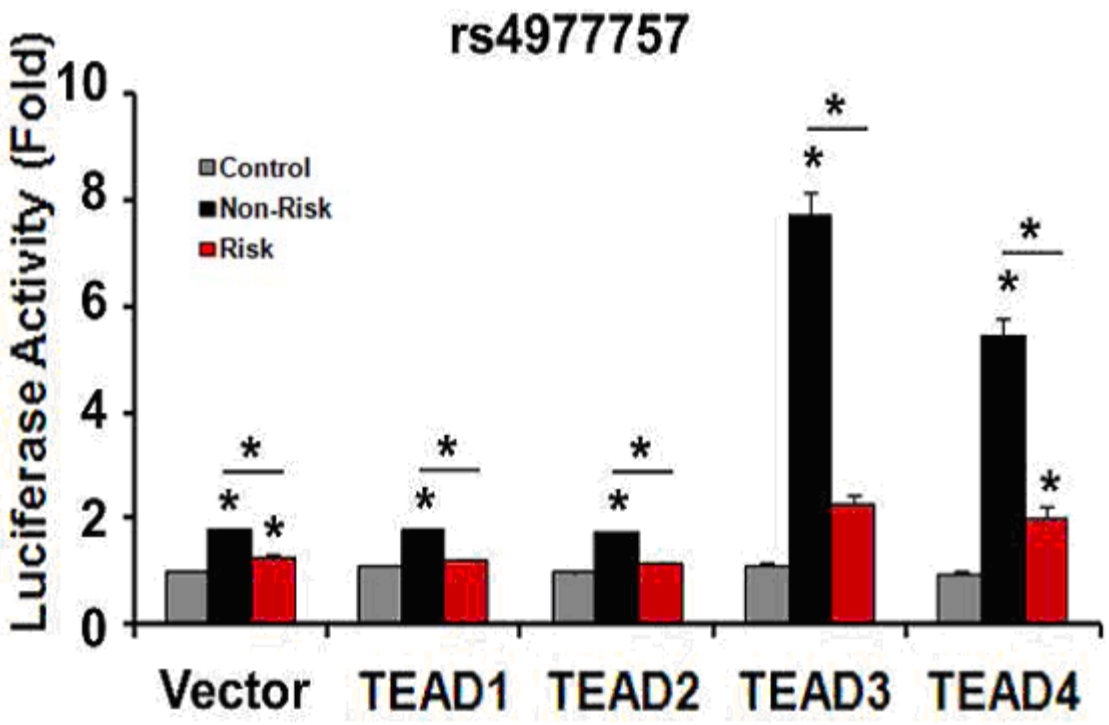

rs4977757
Luciferase Activity (Fold)
10
8
6
4
2
0
Control
Non-Risk
Risk
Vector
TEAD1
TEAD2
TEAD3
TEAD4

Figure 3: Testing the effect of TEAD transcription factors on a promoter bearing the risk and non-risk alleles of rs10811656 (A) and rs4977757 (B) in HEK293 cells. The pGL3-promoter, bearing an SV40 minimal promoter and a luciferase reporter, was used to test the TEAD-dependent enhancer function of rs10811656 and rs4977757 risk and non-risk alleles. The pGL3-vectors were co-transfected together with empty pCMV expressing plasmid or pCMV plasmids expression TEAD1, TEAD2, TEAD3, and TEAD4. n=4 independent experiments in duplicate. For normalization, the luciferase activity was expressed as fold relative to luciferase activity of the empty pGL3 promoter vector co-transfected with the empty pCMV expression vector. One-way ANOVA and post-hoc pairwise comparisons were used, correcting for multiple testing using Bonferroni. *P<0.05. Fold is expressed as Mean ±SEM. Statistical Analysis was carried out by **Ragnar Oli Vilmundarson**.

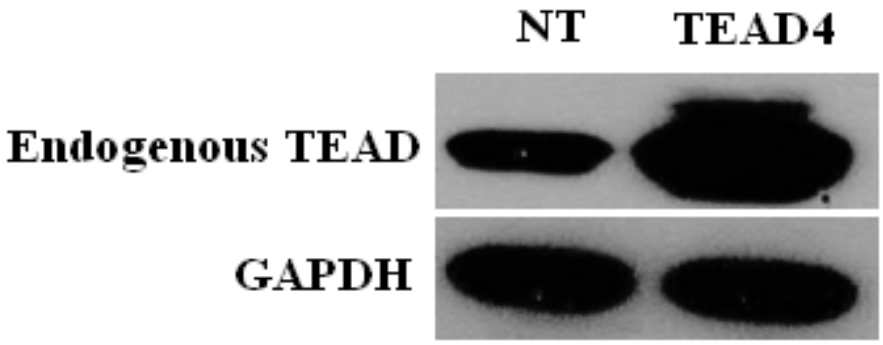
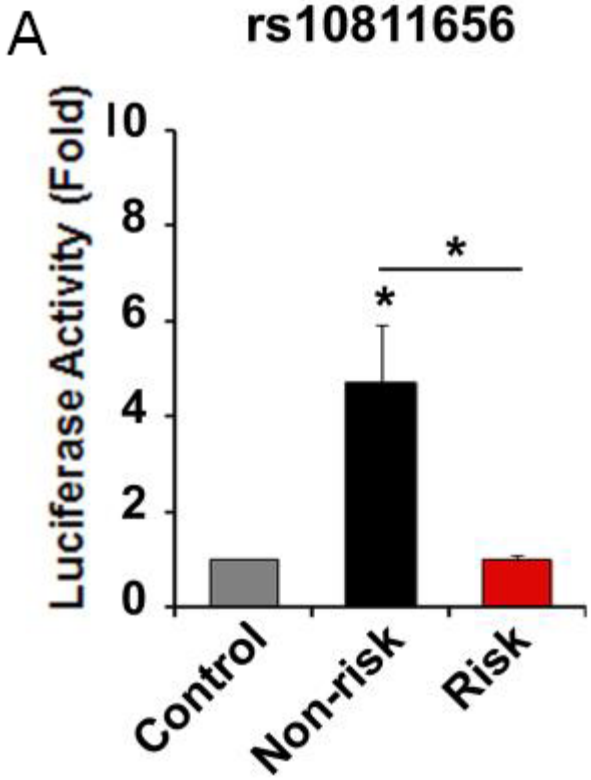

NT
TEAD4
Endogenous TEAD
GAPDH

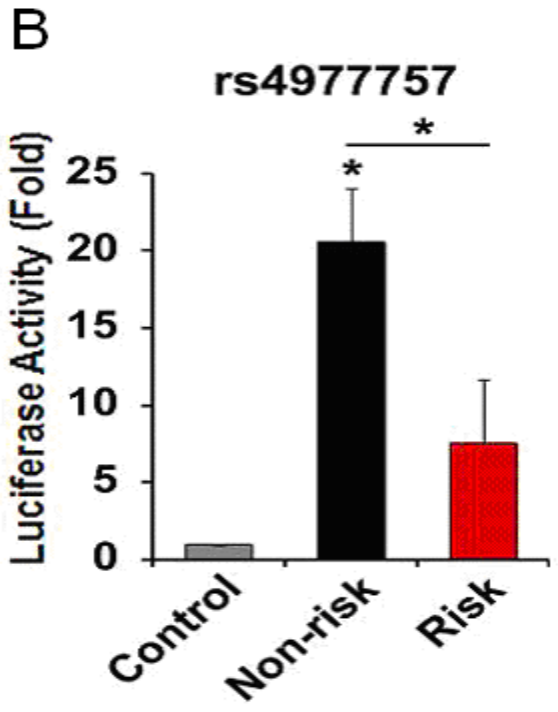
Figure 4: Over-expressing TEAD4 transcription factor in HAoSMCs. Western blot analysis showed that when TEAD4 is over-expressed in HAoSMCs, and endogenous TEAD as the control, the expression of TEAD4 is highly significant compared to the endogenous TEAD. It is representative blot from 4 independent experiments. **NT= Non-Transfected=pCMV.**

3.4 Endogenous TEAD in HAoSMCs for the risk and non-risk alleles of rs10811656 and rs4977757

Studies have revealed that multiple different cell types, including macrophages, lymphocytes, endothelial cells, and smooth muscle cells (SMC), play a role in atherosclerotic lesion formation [108,3]. Although one of the major roles of smooth muscle cells is to produce extracellular matrix, which accumulates over the course of lesion progression [109,110], SMCs are also known to be the major producers of connective tissue both in the healthy and atherosclerotic vessels [111].

Since vascular SMCs are known to make up a large part of atherosclerotic lesions [112], we examined the possibility that the risk and non-risk genotypes of both SNPs rs10811656 and rs4977757 are functional in primary human aortic smooth muscle cells (HAoSMCs). Indeed, we further tested the functionality of endogenous TEAD consensus sequences by transiently transfecting HAoSMCs with the rs10811656 and rs4977757 constructs in the pGL3 luciferase reporter vector compared with empty pGL3 promoter vector. This empty vector is ideally suited for enhancer studies because of its low background (**Figure 5 A-B**).


A
rs10811656
Luciferase Activity (Fold)
10
8
6
4
2
0
*
*
Control
Non-risk
Risk


B
rs4977757
Luciferase Activity (Fold)
25
20
15
10
5
0
*
*
Control
Non-risk
Risk

S

Figure 5A-B: Determining enhancer activity for endogenous TEAD in rs10811656 and rs4977757 in HAoSMCs. HAoSMCs upon the binding of endogenous TEAD were transfected with empty pGL3 promoter vector or the indicated rs10811656 and rs4977757 luciferase reporter constructs. n=5 independent experiments in triplicates. For normalization, the luciferase activity was expressed as fold change relative to empty pGL3 promoter vector. One-way ANOVA and post-hoc pair wise comparisons were used, correcting for multiple testing using Bonferroni. *P<0.05. Fold is expressed as Mean ±SEM. The statistical analysis was carried out by **Ragnar Oli Vilmundarson**.

3.5 GATA transcription factors do not alter the risk and non-risk alleles in 9p21.3 locus

Our bioinformatics scan analysis revealed that GATA transcription factors are located within the intergenic region of the 9p21.3 locus. Since GATA-factor was reported to have relationship with early-onset cardiovascular diseases [100], we asked whether those GATA sites found along the 53,202 bp region of this locus were associated with the risk and non-risk genotypes of the 9p21.3 locus.

First, we tested these GATA sites using gel mobility shift assays to investigate the functionality of those sites (**Figure 6**). The examination of these SNPs rs9632885 and rs10757269 revealed a *de novo* binding site for GATA complex at rs10757269. Note that a GATA4 specific antibody (provided by Dr. Mona Nemer) failed to supershift the GATA complex (lanes 3 and 12), suggesting that other GATA family members (i.e. GATA5 and GATA6) are likely to make up the GATA complex in smooth muscle cells.

Second, to further investigate the function *in vitro* of these SNPs rs9632885 and rs10757269, we performed luciferase reporter assay experiments, using the insert containing rs9632885 and rs10757269 to be subcloned into the pGL3 luciferase reporter vector. The enhancer activity carrying the risk and non-risk alleles for the SNPs revealed no difference in the activity of the luciferase reporter assay in HAoSMCs between risk and non-risk (**Figure 7 A-B**).

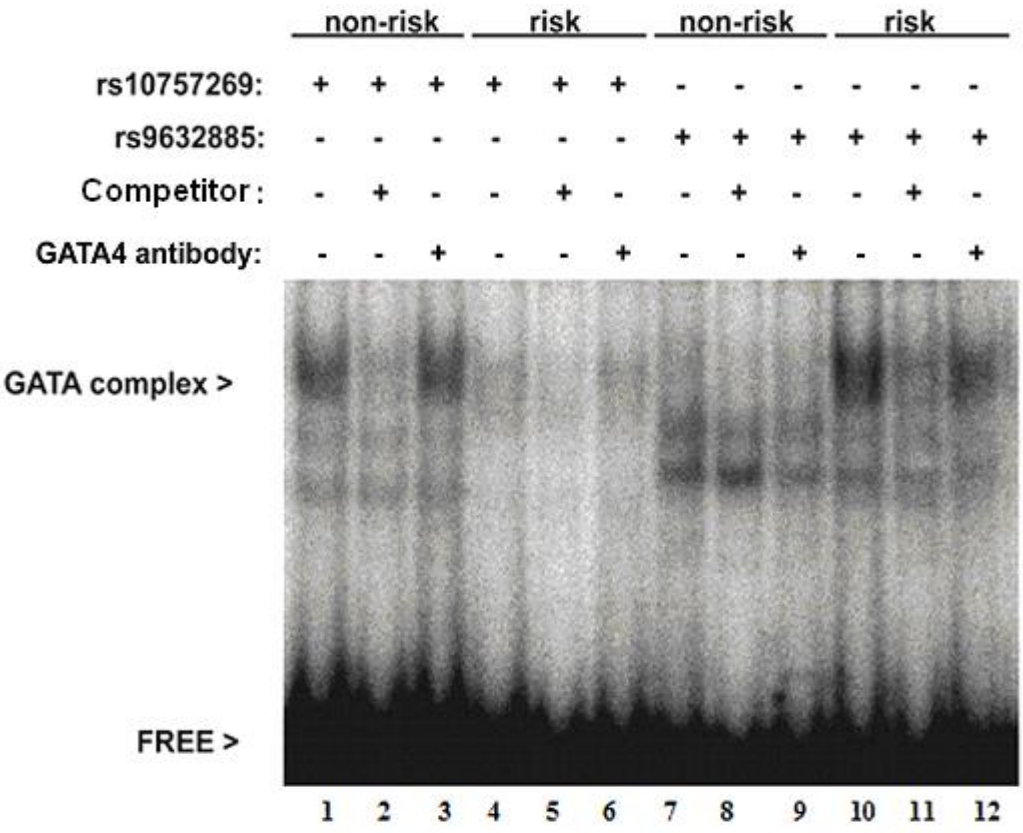
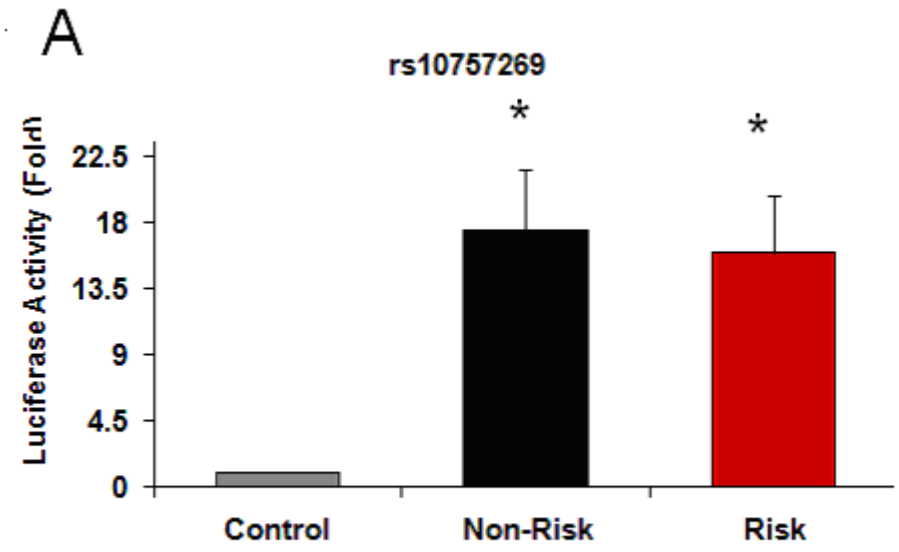

non-risk
risk
non-risk
risk
rs10757269: + + + + + + - - - - - -
rs9632885: - - - - - - + + + + + +
Competitor : - + - - + - - + - - + -
GATA4 antibody: - - + - - + - - + - - +
GATA complex >
FREE >
1 2 3 4 5 6 7 8 9 10 11 12

Figure 6: Binding of GATA transcription factors is affected by 9p21.3 risk alleles. Gel mobility shift assay using nuclear extracts from HAoSMCs showed differential GATA binding sites for the risk alleles compared to the non-risk alleles for the SNPs rs9632885 and rs10757269. For rs9632885 (lanes 1-6), the risk allele disrupts GATA factor binding (lanes 4-6) whereas for rs10757269 (lanes 7-12), the risk allele creates a GATA site resulting in a new shifted complex (lanes 10-12). Note that a GATA4-specific antibody (provided by Dr. Mona Nemer) failed to supershift the GATA complex (lanes 3 and 12), suggesting that other GATA family members (like GATA5 or GATA6) are likely to make up the GATA complex in smooth muscle cells. (**Experiment performed by Naif Almontashiri**)


A
rs10757269
*
*
Luciferase Activity (Fold)
22.5
18
13.5
9
4.5
0
Control
Non-Risk
Risk

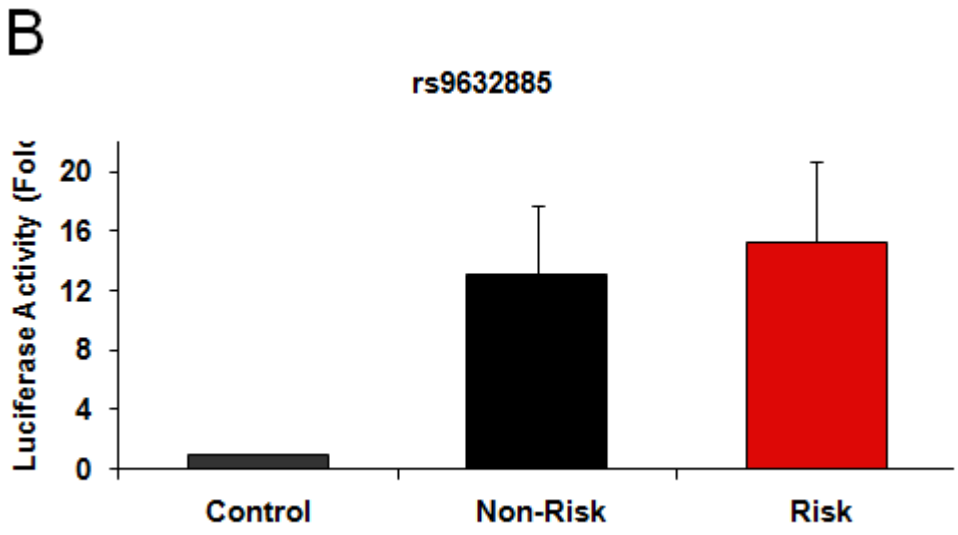

B
rs9632885
Luciferase Activity (Fold)
20
16
12
8
4
0
Control
Non-Risk
Risk

Figure 7A-B: The rs9632885 and rs10757269 SNPs that alter GATA binding do not significantly alter enhancer activity in HAoSMCs. HAoSMCs were transfected with empty pGL3 promoter or the indicated non-risk or risk alleles of rs9632885 and rs10757269-pGL3 luciferase reporter constructs. n=5 independent experiments in triplicates. One-way ANOVA and post-hoc pair wise comparisons were used, correcting for multiple testing using Bonferroni. *P<0.05. Fold is expressed as Mean ±SEM. The statistical analysis was carried out by **Ragnar Oli Vilmundarson**.

4 Discussion

I sought to determine the functional regulation at the 9p21.3 genetic risk locus in CAD. The focus for the investigation was to identify transcription factors whose regulation of enhancers at the 9p21.3 locus would be disrupted by CAD risk alleles. We determined how polymorphisms affect TEAD-dependent regulation, and further evaluated whether or not GATA transcription factors play a role in the 9p21.3 risk and non-risk genotypes.

4.1 The Functional Relationship between the Polymorphisms identified at the 9p21.3 locus and the TEAD transcription factors.

The 9p21.3 locus is known to be the first genetic risk factor to be identified by different independent groups of GWAS in 2007 [20,21,22,23]. It has already been shown that the localization of the risk locus to a region of known protein coding genes implicates a novel gene or regulatory element that promotes atherosclerosis independently of established risk factors [51]. To explore the functional regulation at the 9p21.3 genetic risk locus in CAD, we focused our studies on further determining how polymorphisms affect TEAD-dependent regulation. TEADs are a group of transcription factors that are known to be involved in several processes such as myogenesis, development of neural crest etc [58, 60]. Therefore, TEAD factors might be relevant to atherosclerosis since they are important regulatory factors in SMCs, and also SMCs are known to be affected by the deletion of 9p21.3 in terms of reduced senescence and increased proliferation [29,113,114].

Consequently, we identified two functional SNPs (rs10811656 and rs4977757) within the 53,202bp intergenic region of the 9p21.3 locus that disrupt binding of TEAD3 and TEAD4 transcription factors in HEK293 cells and HAoSMCs. This was a relevant finding since the risk allele was shown to be higher frequency in patients in the Ottawa Heart Genomics Study with severe premature atherosclerosis as compared to subjects with incidents CAD events [115].

In an attempt to determine how polymorphisms at the 9p21.3 locus affect TEAD-dependent gene regulation, we carried out a bioinformatics scan analysis of the 53,202 bp region of the 9p21.3 locus to elucidate the functional consequences of some SNPs on gene regulation by TEAD factors. Our data demonstrate that 4 of the 59 SNPs disrupt consensus sites for TEADs. Several enhancer elements have been shown to greatly influence the 9p21.3 region [51]. Also, a previous study showed that M-CAT (consensus sequence: 5'-CATTCCT-3'), which is the consensus binding site for TEAD, plays a major role in enhancer regulation of the SV40 virus [106]. Given that TEAD`s role in enhancer regulation is known then the disruption of TEAD consensus sequences by these 4 SNPs could be mediating CAD risk through an effect on enhancers at the 9p21.3 region. Additionally, the functional variants underlying the 9p21.3 association with CAD have been shown to likely reside in regulatory elements, further reinforcing this mechanism [41]. A recent study has suggested that TEAD transcription factors can regulate gene expression by binding to proximal promoter sequences or have long-range effects on chromatin structure [78, 79, 81]. Thus, the combination of GWAS association data and the functional regulation of the 9p21.3 locus have given strength to this idea that his gene

contributes to CAD.

We examined the function of these 4 polymorphisms (rs10811656, rs4977757, rs9632885, rs10757269) present in the 9p21.3 locus and demonstrated that the TEAD consensus sequences act as regulatory elements to enhance reporter gene expression when performed in HAoSMCs and HEK293 cells. TEAD3 and TEAD4 disrupting the functional binding sites for the risk alleles showed significantly lower levels of enhancer activity compared to the non-risk allele for both SNPs.

In addition, the difference in the non- risk allele observed in the EMSA for 2 of these SNPs (rs10811656 and rs4977757) was 10 times stronger than to the risk allele containing sequence. Consequently, our study shows that TEAD3 and TEAD4 act as transcriptional enhancer elements increasing the level of the non-risk allele expression. Consistent with our hypothesis, these findings show that some genetic variants at the 9p21.3 locus are functional and disrupt the specific regulatory sequences within enhancers.

Common genetic variants at the 9p21.3 region have been shown to influence disease risk by regulating gene expression through a cis-effect [116]. Thus, by mapping regulatory elements, we were able to evaluate the functional relevance of some polymorphisms at 9p21.3 locus.

4.2 Bioinformatics analysis of the 9p21.3 risk locus also identified 2 genetic variants of GATA factors.

Our bioinformatics scan analysis at the 9p21.3 risk locus also identified 2 SNPs that could affect GATA sites in the 9p21.3 region that might affect the function of this locus. For instance, the GATA factors exhibited a high degree of evolutionary conservation, consistent with a role as a regulator of genes [117]. Thus, functionally important GATA binding sites have been identified and may play a role within the promoters of multiple cardiac-specific genes [118,119,120].

In an attempt to investigate these GATA factors found in the 9p21.3 region, EMSA was performed to examine these SNPs (rs10757269, rs9632885) that affect potential GATA complex in HAoSMCs cells, also we examined whether the GATA4 forms part of this complex. EMSA confirmed that the risk allele of rs10757269 disrupts a functional GATA site, whereas the risk allele of the other SNP, rs9632885, creates a functional GATA binding site. It should be pointed out that the net effect of these linked alleles would be to change the location of a functional GATA site (from the position at rs10757269 to the position of rs9632885). GATA factors, important for gene regulation in SMCs [97], might contribute to atherosclerosis, as suggested by previous studies that reported GATA2 variants associated with early onset CAD [100]. However, another study has reported that GATA2 variants do not mediate risk for CAD [104].

Consequently, we pursued functional studies with the GATA factors to determine

whether these SNPs are functional and could have some potential role within the 9p21.3 region. Notably, the regulatory function of each SNP was examined using luciferase reporter assay in HAoSMCs. Both SNPs were shown to have enhancer activity upon the binding of endogenous GATA in HAoSMCs. However, no significant difference was shown between the risk and the non-risk alleles. We found that both genotypes affect the GATA complex. Therefore, we are unsure which GATA factor forms the complex as GATA4 was hypothesized to be part of this complex but this was shown by EMSA not to be the case. As the EMSA result provides evidence of a GATA complex forming at the tested consensus site then this leaves other GATA factors present in HAoSMCs (GATA5 and GATA6) as the remaining candidates for the factor behind this complex.

4.3 Study Strengths and Weaknesses

This work is one of the few systematic examinations of polymorphisms in the 9p21.3 locus in relation to the functional mechanisms of this locus. The strengths of our work include the careful and detailed analysis of polymorphisms in the 9p21.3 region, and the functionality of a newly studied subset of polymorphisms combines SNP data across multiple studies. Because the 59 SNPs at the 9p21.3 locus are co-inherited, the precise nature of the functional SNPs has been difficult to elucidate.

A potential limitation of my study is that in Fig 5, I showed that when transfecting only the constructs of both SNPs (rs10811656 and rs4977757) in HAoSMCs by luciferase reporter assay, what we observed is due to endogenous TEADs. Specifically I attributed the enhancer activity in both SNPs (rs10811656, rs4977757) for the risk and

non-risk alleles in HAoSMCs to endogenous TEADs (TEAD3 and TEAD4). It was assumed that the endogenous TEADs were responsible for this effect as TEAD consensus sequences was used in the experiment. These results are not definitive, however, further evidence of this effect being driven by endogenous TEADs was observed in an experiment carried out by my colleague Naif Almontashiri (from his paper 9p21.3 coronary artery disease risk variants disrupt TEAD transcription factor-dependent TGFβ regulation of p16 expression in human aortic smooth muscle cells ***Circulation accepted, soon will be published***). From this research, he showed that overexperession of TEAD3 and TEAD4 or both TEAD3/TEAD4 in HAoSMCs leads to an increased expression of p16 (CDKN2A) for the non-risk but not the risk allele (homozygous non-risk vs. risk). However, when knocking down TEAD3 and TEAD4, or both TEAD3/TEAD4 by the lentiviral shRNA, there was a downstream effect on p16 that was observed in HAoSMCs heterozygous for 9p21.3. An ideal experiment would have been: to knockdown TEAD3, TEAD4, or both by introducing the shRNAs to examine whether or not the pGL3-luciferase reporters would still have enhancer activity of TEADs family in HAoSMC.

As a summary, these findings support our results indicating that TEAD3 and TEAD4 proteins were responsible for the difference in enhancer activity between non-risk and risk alleles.

Another limitation of my study is in Fig 4 by the western blot analysis, the same experiment should also have been carried out for TEAD3, but because of the lack of a good antibody, it was left out. Obviously, the same experiment for TEAD4 should have

been better carried out in HEK293 cells in order to examine the TEAD4 expression construct. For instance, we examined the TEAD4 expression construct in HAoSMCs instead, because we asked whether endogenous TEADs would be sufficient to elicit the same luciferase response differences seen in HEK293 cells. This proves that there is a difference in enhancer activity caused by endogenous TEADs differentially binding between the non-risk and risk alleles in the primary cell line, thus indicative of what could be occurring *in vivo*.

The findings suggest that determining how polymorphisms affect TEAD-dependent gene regulation may provide subsequent or additional information in understanding disease association with GWAS variants. We also have found several limitations to our study in term of: our sample size that was in general small, which may limit our statistical analysis for identifying significant increases of the enhancer's activity of the reporter genes. In addition, the number of cell types in which we performed our experiments was not so varied, which may have limited our ability to detect additional functional variants. Our study is also limited by the fact that the 9p21.3 locus is likely to exert multiple effects in different cell types. This was recently outlined in our review which states that some variants at the 9p21.3 locus are pleiotropic in that they are associated with several diseases [42].

5 Conclusions

This thesis aimed to investigate the functional regulation of the 9p21.3 genetic risk locus in CAD. Our study demonstrated that the intergenic region of the 9p21.3 locus harbours regulatory elements (TEAD consensus sequences) that show enhancer activity in luciferase reporter assays. Further, we identified 2 functional SNPs that show an effect for the non-risk alleles in HEK293 and HAoSMCs. Bioinformatics scan analysis identified 15 SNPs in the approximate 53,202bp in length of the 9p21.3 region, but only 2 of them appear to be functional in mediating TEAD factor-dependent regulation of the intergenic region at 9p21.3.

In addition, rs10811656 and rs4977757 SNPs constructs including the risk and non-risk alleles were clearly shown by EMSA to disrupt TEAD binding site, suggesting the functionality of these SNPs for the non-risk alleles. Therefore, we were able to demonstrate by luciferase reporter assays that TEAD3 and TEAD4 activate enhancer activity for the non-risk but not the risk allele.

In terms of the over-expression studies of the transcription factors that activate the non-risk but not the risk alleles, we found a highly increased enhancer activity of TEAD3 and TEAD4 in different cells types. Taken together, our data suggest that rs10811656 and rs4977757 are functional and disrupt specific TEAD regulatory sequences within enhancers. Thus, enhancer sequences that respond to TEAD factors are disrupted by the risk alleles at the 9p1.3 locus.

In summary, our study reports findings that may provide insights to unravel the phenomenon of cardiovascular disease. This study is a step forward in the understanding of the functional mechanisms of the 9p21.3 locus. However, further studies such as: testing other genetic variants of the 9p21.3 will be required for a deeper exploration of the function of this locus in CAD. Although our study shows the functionality of these genetic polymorphisms, the mechanisms of the 9p21.3 locus remain elusive. However, it is expected that such studies will help to elucidate novel diagnostic and therapeutic approaches that will alter CAD.

REFERENCES

[1] Suwaidi J A, Hamasaki S, Higano T S, Nishimura A R, Holmes R D, Lerman A. Long-term Follow-Up of Patients With Mild Coronary Artery Disease and Endothelial Disfunction.*Circulation.*2000*;*101:948-954

[2] Lopez A D, Murray C C.The global burden of disease, 1990-2020. *Nat Med.* 1998*;* 4(11):1241-1243

[3] Doran C A, Meller N, Mcnamara A.C. Role of Smooth Muscle Cells in the Initiation and Early Progression of Atherosclerosis. *Arterioscler Thromb Vasc Biol.* 2008; 28:812-819

[4] Incalcaterra E, Accardu G, Balistreri CR, Caimi G, Candore G, Caruso M, Caruso C. *Curr Atherosclerosis Rep. 2013* Jun; 15(6):329. Doi: 10.1007/s11883-013-0329-5

[5] American Heart Association. Heart and Stroke Statistics: 2003 Update. Dallas, Texas. *American Heart Association*; 2002

[6] Wang Q. Advances in the Genetic Basis of Coronary Artery Disease. *Curr Atherosclerosis Rep*. 200; 7(3): 235-241

[7] Weissberg P L. Atherogenesis: current understanding of the causes of atheroma. *Heart* .2000; 83:247

[8] Rauch U, Osende J I, Fuster V, Badimon J J, Fayad Z, Chesebro J H. Thrombus formation on atherosclerotic plaques: Pathogenesis and clinical consequences.*Ann Intern Med.* 2001; 134:224–238

[9] Lusis J A, Fogelman M A, Fonarow C G. Genetic Basis of Atherosclerosis. *Circulation*. 2004; 110:1868-1873

[10] Breslow J L. Genetics of lipoprotein abnormalities associated with coronary artery disease susceptibility. *Annu Rev Genet.* 2000; 34:233-254

[11] Shepherd J, Cobbe S M, Ford I, Isles C G, Lorimer AR, Macfarlane P W, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med*. 1995 Nov 16; 333(20):1301-8

[12] Downs J R, Clearfield M, Weis S, Whitney E, Shapiro D R, Beere PA, et al., for the AFCAPS/TexCAPS Research Group.Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels. *JAMA*. 1998. 27; 279(20):1615-22

[13] Scheuner T M et al.Genetic evaluation for coronary artery disease. *Genet Med* 2003:5(4):269-285

[14] Gordon T, Kannel W B. Multiple risk functions for predicting coronary heart disease: the concept, accuracy, and application. *Am Heart J*. 1982; 103:1031–1039

[15] Gordon T, Castelli W P, Hjortland M C, Kannel W B, Dawber T R. Diabetes, blood lipids, and the role of obesity in coronary heart disease risk for women. *Ann Intern Med.* 1977; 87:393–397

[16] Scheuner M T. Genetic predisposition to coronary artery disease. *Curr Opin C a r d i o l* 2001; 16:251–260

[17] Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes T L, Thompson J, Ingelsson E, Saleheen D, Erdmann J, Goldstein B A, Stirrups K, Koning I, Cazier J-B, Johansson A, Hall A S, Lee J-Y, Willer C, Chambers J, Esko T, Folkersen L ,et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genet*. 2013. Doi: 10.1038/ng.2480

[18] Davies, R W, Dandona S. 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

[19] Roberts R. Coronary artery disease: An Update. *Methodist Debakey Cardiovasc J*. 2014; 10(1): 7-12

[20] Welcome_Trust_Case_Control_Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007. 447(7145):661-678

[21] McPherson R, Pertsemlidis A, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007. 316(5830): 1488-1491

[22] Helgadottir A, Thorleifsson G et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007, 316(5830): 1491-1493

[23] Samani, N, Erdmann J., et al.Genome-wide association analysis of coronary artery disease. *N Engl J Med* 2007, 357(5): 443-453

[24] Munier et al. The association of the 9p21.3 locus with coronary atherosclerosis: a systematic review and meta-analysis. *BMC Medical Genetics*. 2014.15:16.

[25] Broadbent H M, Peden J F, Lorkowski S, Goel A, Ongen H, Green F, Clarke R, Collins R, Franzosi M G, Tognoni G. et al. Susceptibility to coronary artery disease and diabetes is encoded by 8 Human Molecular Genetics, 2012 distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum. Mol. Genet*.2008. 17, 806–814

[26] Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet D.A, Mangino M, Linsel-Nitschke P, Cambien F, Hengstenberg C, Stark K. et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease.*Circulation*. 2008.117, 1675–1684

[27] Dandona S, Stewart A R F, Chen L, Williams K, So D, O'Brien E, Glover C, Lemay M, Assogba O, Vo L. et al. Gene dosage of the common variant 9p21 predicts severity of coronary artery disease. *J. Am. Coll. Cardiol*.2010 56, 479–486

[28] Patel R S, Su S, Neeland I J, Ahuja A, Veledar E, Zhao J, Helgadottir A, Holm H, Gulcher J R, Stefansson K, et al. The chromosome 9p21 risk locus is associated with angiographic severity and progression of coronary artery disease. *Eur. Heart J*.2010. 31, 3017–3023

[29] Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, Blow MJ, Cohen J C, Rubin EM, Pennacchio LA: Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature*. 2010, 464(7287):409–412

[30] Gil J, Peters G. Regulation of the INK4b-ARFINK4a tumour suppressor locus: all for one or one for all. *Nat Rev Mol Cell Biol.* 2006; 7:667–77

[31] Yang XR, Liang X, Pfeiffer RM, Wheeler W, Maeder D, Burdette L, Yeager M, Chanock S, Tucker MA, Goldstein AM .Associations of 9p21variants with cutaneous malignant melanoma, nevi, and pigmentation phenotypes in melanoma-prone families with and without CDKN2A mutations. *Fam Cancer*.2010b. 9 (4): 625–633)

[32] Samani N J, Schunkert H. Chromosome 9p21 and cardiovascular disease: the story

unfolds. *Circ Cardiovasc Genet.* 2008; 1: 81-84
[33] Yap K L, Li S, Munoz-Cabello A M, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh M J, Zhou M M. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol cell*. 2010. 38 (5): 662–674
[34] Pasmant E, Sabbagh A, Vidaud M, Bieche I .ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J*. 2011. 25 (2): 444–448
[35] Stark G R, Kerr I M, Williams B R, Silverman R H, Schreiber R D. How cells respond to interferons. *Annu Rev Biochem*.1998 67: 227–264
[36] Lyssenko V, Burtt N P, de Bakker P I, Chen H, Roix J J, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*.2007. 316: 1331–1336
[37] Cugino D, Gianfagna F, Santimone I, de Gaetano G, Donati MB, et al. Type 2 diabetes and polymorphisms on chromosome 9p21: A meta-analysis. *Nutr Metab Cardiovasc Dis*. 2012. 22: 619–625
[39] Roberts R. Personalized Medicine: An idea Whose Time Is Approaching. *Methodist DeBakey Cardiovascular Journal*. 2007. Doi.10.14797
[40] Chen H H, Stewart A.F.R.Making sense of GWAS: Integrating genetic variation with gene expression to derive functional mechanisms underlying disease risk.*Circulation.* 2014. Doi: 10.1161
[41] Harimendy O, Notani D, Song X, Rahim G. N, Tanasa B, Heintzman N, Ren B, Fu D-X, Topol J. E, Rosenfeld G. M, Frazer A. K**.** 9p21 DNA variants associated with coronary artery disease impair interferon-y signaling response.*Nature*.2011. Doi: 10.1038/nature09753
[42] Chen H H, Almontashiri N A, Antoine D, Stewart A F. Functional genomics of the 9p21.3 locus for atherosclerosis: clarity of confusion?*Curr Cardiol Rep*. 2014. 16(7):502. Doi: 10.1007/s11886-014-05502-7
[43] Zhang W, Chen Y, Liu Peng, Chen J, Song L, Tang Y, Wang Yuyao, Liu J, Hu B. F. Variants on Chromosome 9p21.3 Correlated with ANRIL Expression Contribute to Stroke Risk and Recurrence in a Large Prospective Stroke Population. *Stroke* .2012.43:14-21
[44] Bishop D T, Demenais F, Iles M M, Harland M, Taylor J C, Corda E, Randerson-Moor J, Aitken J F, Avril M F, Azizi E, Bakker B, Bianchi-Scarra G, Bressa c-de Paillerets B, Calista D, Cannon-Albright L A, Chin A W T, Debniak T, Galore-Haskel G, Ghiorzo P, Gut I, Hansson J, Hocevar M, Hoiom V, Hopper J L, Ingvar C, Kanetsky P A, Kefford R F, Landi M T, Lang J, Lubinski J, Mackie R, Malvehy J, Mann G J, Martin N G, Montgomery G W, Van Nieupoort F A, Novakavic S, Olsson H, Puig S, Weiss M, Van W, Zelenika D, Brown K M, Goldstein A M, Gillanders E M, Boland A, Galan P, Elder D E, Gruis N A, Hayward N K, Lathrop G M, Barrett J H, Bishop J A. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet*. 2009. 419(8):920-925
[45] Falchi M, Bataille V, Hayward N K, Duffy D L, Bishop J A, Pastinen T, Cervino A, Zhao Z Z, Deloukas P, Soranzo N, Elder D E, Barrett JH, Martin N G, Bishop D T, Montgomery G W, Spector T D. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet* .2009.41 (8): 915–919

[46] Pasmant E, Laurendeau I, Heron D, Vidaud M, Vidaud D, et al. 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: 3963–3969

[47] Wilusz J E, Sunwoo H, and Spector D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*.2009 23, 1494-1504

[48] Mercer T R, Dinger M E, and Mattick, J S. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*.2009 .doi:10, 155-159

[49] Holdt. M, Sass K, Gilbel G, Berger H, Thierry I, Teupser D. Expression of Ch9p21 genes CDKN2A (p16 INKa), p14(ARF), and MTAP in human atherosclerotic plaque. *Atherosclerosis*. 2011; 214:264-270

[50] Liu Y, Sanoff H K, Cho H, Bund C E, Torrice C, Mohike K L, et al. INK4/ARF transcript expression to associated with chromosome 9p21 variants linked to atherosclerosis. *Plos One*. 2009; 4:e5027

[51] Jarinova O, Stewart A F, Roberts R, Wells G, Lau P. Naing T. et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Atheroscler Thromb Vasc Biol*. 2009; 29:1671-1677

[52] Almontashiri A M N, Fan M, Cheng M L B, Chen H-H, Roberts R, Stewart F RA. Interferon-y Activates Expression of p15 and p16 Regardless of 9p21.3 Coronary Artery Disease Risk Genotype. *JACC*. 2013. Doi: 10.1016

[53] Cao X, Pfaff S L, Gage F H. Yap regulates neural progenitor cell number via the TEA domain transcription factor. *Genes Dev.* 2008. 22:3320-34

[54] Blatt C, DePamphilis M L. Striking homology between mouse and human transcription enhancer factor-1 (TEF-1). *Nucleic Acids Res.* 1993; 21:747

[55] Jacquemin P, Hwang J J, Martial J A, Dolle´ P, and Davidson I. A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS DNA binding domain. *J. Biol. Chem*. 1996. 271, 21775- 21785

[56] Kaneko K J. and DePamphilis M L. Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. *Dev Genet*.1998 22, 43-55

[57] Chen Z, Friedrich G A & Soriano P. Transcriptional enhancer factor 1 disruption b a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev*. 1994. 8: 2293-2301

[58] Yoshida T. MCAT elements and the TEF-1 family of transcription factors in muscle development and disease. *Arterioscl Thromb Vasc Biol*.2008. 28:8-17, doi ATVBAHA. 107.155788

[59] Blais A et al. An initial blueprint for myogenic differentiation.*Genes Dev*.2005. 19:553-569. doi: gad. 1281105

[60] Milewski R C et al. Identification of minimal enhancer elements sufficient for Pax3 expression in neural crest and implication of TEAD2 as a regulator of Pax3. *Development*. 2004. 131: 829-837. Doi: 10.1242

[61] Yagi R et al. Transcription factor TEAD specifies the trophectoderm lineage at the beginning of mammalian development. *Development*.2007.134: 3827-3836. doi: dev. 010223

[62] Larkin S B, and Ordahl C P. Multiple layers of control in transcriptional regulation by MCAT elements and the TEF-1 protein family. *Heart Development*. 1999. 307- 329

[63] Pobbati V A, and Hong W. Emerging roles of TEAD Transcription factors and its coactivators in cancers. *Cancer Biolog & Therapy*. 2013. 14:5; 390-398
[64] Mitsunori O, Hiroshi S. Mammalian TEAD proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signalling. *Development*. 2008.135, 4059-4069. doi:10.1242/dev.027151
[65] Maeda T, Chapman L D, Stewart A R F. Mammalian Vestigial-like 2, a Cofactor of TEF-1 and MEF2 Transcription Factors That Promotes Skeletal Muscle Differentiation. *Journal of Biological Chemistry*. 2002doi: 10.1074/jbc.M206858200
[66] Zhao B, Ye X, Yu J, Li L, Li W, et al. TEAD mediates YAP-dependent gene induction and growth control. Genes Dev.22:1962-1971
[67] Zhang W, Gao Y, Li P, Shi Z, Guo T, Li F, Han X, Feng Y, Zheng C, Wang Z, Li F, Chen H, Zhou Z, Zhang L, Ji H. VGLL4 functions as a new tumor suppressor in lung cancer by negatively regulating the YAP-TEAD transcriptional complex.*Cel*l *Research*. 2014. 24:331-343
[68] Mahoney M W Jr, Hong Ho-J, Yaffe B M, Farrance G K I. The transcriptional co-activator TAZ interacts differentially with transcriptional enhancer factor-1(TEF-1) family members. *Biochem J*. 2005.15:338:217-225
[69] Lamar M J, Stern P, Liu H, Schindler W J, Jiang G Z, Richard O. The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. *Cell Biology*. 2012. doi:10.1073/pnas.1212021109
[70] Gray L T, Fong K K, Pavelitz T, Weiner A M. Tethering of the conserved piggyback transposase fusion protein CSB-PGBD3 to chromosomal AP-1 proteins regulates expression of nearby genes in humans. *Plos Genet*. 2012; 8; e1002972
[71] Ze Li, Bin Zhao, Ping Wang, Fei Chen, Zhenghong Don, Huirong Yang, Kun-Liang Guan, and Yanhui Xu.Structural insights into the YAP and TEAD complex. *Genes Dev*. 2010. 24: 235-240 doi: 10.1101/gad.1865810
[72] Yan X, Yin-L Z, Chao Y, Ting C, Heng-Y F. Yap/TEAD Co-Activator Regulated Pluripotency and chemoresistance in Ovarian Cancer Initiated Cells. *Plos One*. 2014. 11:e109575. doi:10.1371/journal.pone.0109575
[73] Benhaddou A, Keime C, Ye T, Morlon A, Michel I, Jost B, Mengus G, Davidson I. Transcription factor TEAD4 regulates expression of myogenin and the unfolded protein response genes during C2C12 cell differentiation. *Cell Death Differ*. 2012. 19(2):220-31. Doi: 10.1038
[74] Goulev Y, Fauny J D, Gonzalez-Marti B, Flagiello D, Silber J and Zider, A. SCALLOPED interacts with YORKIE, the nuclear effector of the hippo tumor- suppressor pathway in Drosophila. *Curr. Biol*. 2008.18, 435-441
[75] Wu S, Liu Y, Zheng Y, Dong J. and Pan D. The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth regulatory pathway. *Dev Cell*. 2008. 14, 388-398
[76] Mar J H, and Ordahl C P. M-CAT binding factor, a novel trans-acting factor governing muscle-specific transcription. *Mol Cell Biol.*1990. 10(8): p.4271-83
[77] Ishiji T, Lace M J, Parkkinen S, Anderson R D, Haugen T H, Cripe T P, Xiao J H, Davidson I, Chambon P, Turek L P. Transcriptional enhancer factor (TEF-1) and its cell-specific co-activator activate human papillomavirus-16 e6 and e7 oncogene transcription in keratinocytes and cervical carcinoma cells. *Embo J*. 1992; 11:2271- 2281
[78] MacLellan W R, Lee T C, Schwartz R J, Schneider M D. Transforming growth factor-

response elements of the skeletal –actin gene.Combinatorial action of serum response factor, YY1, and the SV40 enhancer-binding protein, TEF-1. *J Biol Chem*.1994.269:16754-60

[79] Leask A, Holmes A, Black C M, Abraham D J. Connective tissue growth factor gene regulation.Requirements for its induction by transforming growth factor-beta 2 in fibroblasts. *J Biol Chem*. 2003; 278:13008-15

[80] Brown K A, Pietenpol J A, Moses H L. A tale of two proteins: differential roles and regulation of SMAD2 and SMAD3 in TGFβ signaling. *J Cell Biochem*. 2007; 101:9- 33

[81] Fujii M, Toyoda T, Nakanishi H, Yatabe Y, Sato A, Matsudaira Y, et al. TGFβ synergizes with defects in the Hippo pathway to stimulate human malignant mesothelioma growth. *J Exp Med*. 2012; 209:479-94

[82] Tashiro H, Shimokawa H, Sadamatu K. & Yamamoto K. Prognostic significance of plasma concentrations of transforming growth factor-beta in patients with coronary artery disease. *Coron Artery Dis.*2002 13, 139-143

[83] Berk B C. Vascular smooth muscle growth: autocrine growth mechanisms. *Physiol Rev*.2001. 81, 999-1030

[84] Kojima S, Harpel P.C& Rifkin D B. Lipoprotein (a) inhibits the generation of transforming growth factor b: an endogenous inhibitor of smooth muscle cell migration.*J cell Biol*.1991. 113, 1439-1441

[85] Grainger D J, et al. Proliferation of human smooth muscle cells promoted by lipoprotein (a). *Science*.1993 260, 1655-1658

[86] Merika M and Orkin H. S. DNA-Binding Specificity of GATA Family Transcription Factors. *Mol and Cel Biology*. 1993.3990-4010

[87] Orkin S H. GATA-binding transcription factors in hematopoietic cells. *Blood*.1992 80:575-581

[88] Martin D I K, and Orkin S H. Transcriptional activation and DNA binding by the erythroid factor GF-1/NFEl/Eryfl. *Genes Dev*.1990. 4:1886-1898

[89] Spieth J, Y.-H.Shim, K. Lea, R. Conrad, and T. Blumenthal. 1991. elt-1, an embryonically expressed Caenorhabditis elegans gene homologous to the GATA transcription factor family. *Mol. Cell. Biol*. 11:4651-4659

[90] Yang, H.-Y., and T. Evans. Distinct roles for the two cGATA-1 finger domains. *Mol Cell Biol*. 1992. 12:4562-4570

[91] Patient R K, McGhee J D. The GATA family (vertebrates and invertebrates*). Curr Opin Genet Dev*. 2002; 12:416 – 22

[92] Ohneda K, Yamamoto M. Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage. *Acta Haematol* . 2002; 108:237 – 45

[93] Cantor A B, Orkin S H. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene*. 2002; 21:3368 – 76

[94] Charron F, Nemer M. GATA transcription factors and cardiac development. Semin Cell Dev Biol 1999; 10:85 – 91

[95] Molkentin J D. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue specific gene expression. *J Biol Chem*. 2000; 275:38949 – 52

[96] Brewer A, Pizzey J. GATA factors in vertebrate heart development and disease. *Expert Reviews in Molecular Medicine*.2006. 8 (22), 1–20

[97] Nemer, M., 2008. Genetic insights into normal and abnormal heart development. *Cardiovascular Pathology*. 2008. 17 (1), 48–54

[98] Tsai F Y, Keller G, Kuo F C, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*.1994. 371 (6494), 221–226

[99] Morin S, Charron F, Robitaille L, Nemer M. GATA-dependent recruitment of MEF2 proteins to target promoters. *Embo J*. 2000; 19:2046–2055

[100] Connelly J J, Wang T, Cox J E, et al. GATA2 is associated with familial early-onset coronary artery disease. *PLoS Genetics*.2006. 2 (8), e139

[101] Hauser E R, Crossman D C, Granger C B, et al. A genome wide scan for early-onset coronary artery disease in 438 families: the GENECARD Study. *American Journal of Human Genetics*.2004. 75 (3), 436–447

[102] Shah S H, Kraus W E, Crossman D C, et al. Serum lipids in the GENECARD study of coronary artery disease identify quantitative trait loci and phenotypic subsets on chromosomes 3q and 5q. *Annals of Human Genetics*. 2006. 70 (Pt 6), 738–748

[103] Alberti K G, Zimmet P Z, 1998. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic Medicine*. 1998. 15 (7), 539–55

[104] Dandona S, Chen L, Fan M., et al. The transcription factor GATA-2 does not associate with angiographic coronary artery disease in the Ottawa Heart Genomics and Cleveland Clinic Gene Bank Studies. *Human Genetics*.2009. 127 (1), 101–105

[105] Holdt L.M and Teupser D. Recent Studies of the Human Chromosome 9p21 Locus, Which is Associated with Atherosclerosis in Human Populations. *Arterioscler Thromb Vasc Biol*. 2012:32:196-206

[106] Farrance I K G, Ordahl C.P. The Role of Transcription Enhancer Factor-1(TEF-1) Related Proteins in the Formation of M-CAT Binding Complexes in Muscle and Non-muscle Tissues. *J of Biological Chem*. 1996. 8266-827

[107] Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anan SS, Andersen K et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011; 43(4):333-8. Doi: 10.1038.

[108] Lusis A J. Atherosclerosis. *Nature*. 2000; 407:233–241.

[109] Geng Y J, Libby P. Progression of atheroma: a struggle between death and procreation. *Arterioscler Thromb Vasc Biol*. 2002; 22:1370 –1380

[110] Engelberg H. Endogenous heparin activity deficiency: the 'missing link' in atherogenesis? *Atherosclerosis.* 2001; 159:253–260

[111] Falk E. Pathogenesis of atherosclerosis. *J Am Coll Cardiol*. 2006; 47: C7–C12

[112] Allahverdian S, Chehroudi A C, McManus B M, Abraham T, Franciss G A. Contribution of intimal smooth muscle cells to cholesterol accumulation and macrophage-like cells in human atherosclerosis. *Circulation*. 2014:129; 1551-1559

[113] Gan Q, Yoshida T, Li J, Owens G K. Smooth muscle cells and myofibroblasts use distinct transcriptional mechanisms for smooth muscle α-actin expression. *Circ Res*. 2007; 101:883-892

[114] Pasquet S, Naye F, Faucheux C, Bronchain O, Chesneau A, Thiebaud P, Theze N. Transcription enhancer factor-1-dependent expression of the alpha-tropomyosin gene in the three muscle cell types. *J Biol Chem*. 2006; 281:34406-34420

[115] Robert R, Alexandre F R S, George A W, Kathryn A W, Nihan K, Ruth M. D R.

Identifying genes for coronary artery disease: An idea whose time has come. *Can. J of Cardiol.* 2007. V.23 PMCID: PMC2787000.

[116] Gu F, Pfeiffer R M, Battacharjee S, Han S S, Taylor P R, Berndt S, Yang H, Sigurdson A J, Toro J, Mirabello L, Greene M H, Freedman N D, Abnet C C, Dawsey S M, Hu N, Qiao Y-L, Ding T, Brenner A V, Garcia-Closas M, Hayes R, Brinton L A, Lissowska J, Wentzensen N, Kratz C, Moore L E, Ziegler R G, Chow W-H, Savage S A, Burdette L, Yeager M, Chanoch S J, Chatterjee N, Tucker M A, Goldstein A M, and Yang X R. Common genetic variants in the 9p21.3 region and their associations with multiple tumours. *British Journal of Cancer*. 2013. 108; 1378- 1386 doi:10.1038/bjc.2013.7

[117] Lowry A J, Atchley R W. Molecular Evolution of the GATA family of Factors: Conservation within the DNA-Binding Domain. *J Mol Evol.* 2000. 50:103-115 doi: 10.1007/s00239991001

[118] Morrisey E E, Ip H S, Lu M M, Parmacek M S. Gata-6: A zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol*. 1996; 177:309-322

[119] Morrissey E E, Ip H S, Tang Z, Lu M M, Parmacek M S. Gata-5: A transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev Biol*. 1997; 183:21-36

[120] Suzuki Y J, Day R M, Tan C C, Sandven T H, Liang Q, Molkentin J D, Fanburg B L. Activation of gata-4 by serotonin in pulmonary artery smooth muscle cells. *J Biol Chem*. 2003; 278:17525-17531

List of Tables

GENE	Primers
TEAD1	hTEAD For: 5'-CCACATGGTGGATAGATAGC-3' hTEAD Rev: 5'-GATCAACTTCATCCACAAGC-3'
TEAD2	mTEAD For: 5'-CTGCCATCACCTCCAACGCC-3' mTEAD Rev: 5'-GGTTAATTTCCTGCACAAGC-3'
TEAD3	mTEAD For: 5'-GCTTTCAGGAGGCCCTGGCC-3' mTEAD Rev: 5'-ACGGCCGCTTCGTGTACCGC-3'
TEAD4	hTEAD For: 5'-GTCCGACAGGATGATTTTGC-3' hTEAD Rev: 5'-GATCAACTTCATCCACAAGC-3'

Table 1: Primers for TEAD1-4 DNA sequencing

Gene	SNPs	Primers	
GATA	**rs9632885**	Non-risk	For: 5'- CATGTTATTAGATAATATAGTCTCAGC-3' Rev: 5' GCTGAGACTATATTATCTAATAACATG-3'
		Risk	For: 5'- CATGTTATTAGATGATATAGTCTCAGC-3' Rev : 5' GCTGAGACTATATCATCTAATAACATG-3'
GATA	**rs10757269**	Non-risk	For: 5'-CAGGTTCTTTTAGATAATTTTTTTATC-3' Rev: 5'-GATAAAAAAATTATCTAAAAGAACCTG-3'
		Risk	For: 5'- CAGGTTCTTTTAGGTAATTTTTTTATC-3' Rev : 5'- GATAAAAAAATTACCTAAAAGAACCTG-3'

Table 2: Primers were designed according to the bioinformatics scan analysis for PCR and sequencing for the 2 GATA SNPs.

SNP	Non-risk oligonucleotides	Risk oligonucleotides
rs10811656	5’GGTGTGGTCATTCCGGTAAGCAGCG 3’ 5’CGCTGCTTACCGGAATGACCACACC 3'	5’GGTGTGGTCATTCTGGTAGGCAGCG3’ 5’CGCTGCCTACCAGAATGACCACACC3’
rs4977757	5’GAGTATCTTTGTGGCATTCTCTGTATTTCC3’ 5’AGGAAATACAGAGAATGCCAAAAGATAC3’	5’GAGTATCTTTGTGGCGTTCTCTGTATTTCCT 3’ 5’AGGAAATACAGAGAACGCCACAAAGATACTC3

Table 3: Oligonucleotides used for the SNP constructs rs10811656 and rs4977757 for the risk and non-risk alleles.

SNP	Binding element (non-risk)	Binding element (risk)	Transcription factors	Source of nuclear extracts	Predict risk Allele affect	Testing method	
rs10811656	GGTCATTCCGG	GGTAGGCAG	TEAD4 / TEAD3	HAoSMCs /HeLa cells	Disrupted binding	EMSA(+)	Reporter Assay(+)
rs4977757	TGGGAATGCT	TGGGAAAGCTT	TEAD4 / TEAD3	HAoSMCs /HeLa cells	Disrupted binding	EMSA(+)	Reporter Assay(+)

Table 4: Summary showing the sequence for the binding elements for the risk and non-risk alleles with the transcription factors involved in the disruptive binding examined by EMSA which functionally was further analyzed by luciferase reporter assay.

SNP	Non-Risk Oligonucleotides	Risk Oligonucleotides
rs10811656	5’ GGTGTGGTCATTCCGGTAAGCAGCG 3’ 5’ CGCTGCTTACCGGAATGACCACACC 3’	5’ GGTGTGGTCATTCTGGTAGGCAGCG 3’ 5’ CGCTGCCTACCAGAATGACCACACC 3’
rs4957777	5’ GAGTATCTTTGTGGCATTCTCTGTATTTCCT 3’ 5’ AGGAAATACAGAGAATGCCACAAAGATACTC 3’	5’ GAGTATCTTTGTGGCGTTCTCTGTATTTCCT 3’ 5’ AGGAAATACAGAGAACGCCACAAAGATACTC 3’
rs9632885	5’ ATGTTATTAGATAATATAGTCTCAG 3’ 5’ CTGAGACTATATTATCTAATAACAT 3’	5’ ATGTTATTAGATGATATAGTCTCAG 3’ 5’ CTGAGACTATATCATCTAATAACAT 3’
rs10757269	5’ AGGTTCTTTTAGATAATTTTTTTAT 3’ 5’ATAAAAAAATTATCTAAAAGAACCT 3’	5’ AGGTTCTTTTAGGTAATTTTTTTAT 3’ 5’ ATAAAAAAATTACCTAAAAGAACCT 3’

Table 5: Summary of oligonucleotides used for EMSA (Electrophoretic mobility shift assay) to test transcription factor binding to different SNPs.

Appendix A-Curriculum Vitae

Darlène Antoine

University of Ottawa Heart institute

Education

- Master of Science in Biochemistry with Specialization in HMG(2013-Present) University of Ottawa, Ottawa, ON, Canada

- Bachelor of Science , in Biology (2002-2007)
  Universidad de Oriente, Santiago de Cuba,
  Cuba

- High School and Baccalaureate Diploma
  (2000) College St-Louis, Jérémie, Haiti

Conferences

- 3rd Annual Ottawa Heart Research Conference (April 2015) Shaw Center, Ottawa, ON, Canada
  Poster Presentation

- University of Ottawa Research Day (March 2014)
  University of Ottawa, Faculty of Medicine, RGN, Ottawa, ON, Canada Oral presentation

Scholarship and Awards

- University of Ottawa Admissions Scholarship (2013-2015)
- Haitian Government Full Scholarship for Undergraduate Studies (2002-2007)

Research internships

- University of Ottawa Heart Institute, Ottawa, ON (January 2013)
  Supervisor: Dr. Alexandre F.R. Stewart
  Research Project: Functional Regulation at the 9p21.3 genetic risk locus in coronary artery disease (CAD)

- Universidad de Oriente, Santiago de Cuba, Cuba (Nov 2006-June 2007)
  Supervisor: Dr. Luis Enrique Almaguer Mederos
  Research Project: Allelic interaction and its influence in clinical phenotype in spino cerebellar Ataxia type 2

Appendix-B List of Publications

1. Chen H H, Almontashiri N A M, **Darlène Antoine**, Stewart A R F. Functional Genomics of the 9p21.3 locus for Atherosclerosis«; Clarity or Confusion?

2. Almontashiri N A M , **Darlene Antoine**, Zhou X, Vilmundarson R O, S Zhang S X, Hao N K, Chen H H, Stewart A R F, PhD* **Circulation** (**Accepted**) 9p21.3 coronary artery disease risk variants disrupt TEAD factor-dependent TGF-β regulation of p16 expression in human aortic smooth muscle cells.

STATEMENT

I, Darlène Antoine, declare to have fully performed all the experiments presented here in this manuscript, excluding the genotyping of the cultured cells, the construction of the reporter plasmids and the electrophoresis mobility shift assays performed by my colleague Naif Almontashiri, and to have redacted this thesis under my thesis Director supervision: Dr. Alexandre F.R. Stewart, and followed the guidelines of the Department of Biochemistry, of the Faculty of Graduate and Postgraduate Studies of the University of Ottawa.


Online supplement new changes highligh


CIRCULATIONAHA.114015023R3.doc