Functional Analysis of the *TRIB1* Locus in Coronary Artery Disease (CAD)

By Adrianna Douvris

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of
Master of Science

Department of Biochemistry, Microbiology and Immunology
Faculty of Medicine, University of Ottawa
Ottawa, Ontario, Canada

Abstract

Functional Analysis of the *TRIB1* Locus in Coronary Artery Disease (CAD)

By Adrianna Douvris

The *TRIB1* locus (8q24.13) is a novel locus associated with plasma TGs and CAD risk. Trib1 is a regulator of MAPK activity, and has been shown to regulate hepatic lipogenesis and VLDL production in mice. However, the functional relationship between common SNPs at the *TRIB1* locus and plasma lipid traits is unknown; *TRIB1* has not been identified as an eQTL. This cluster of SNPs falls within an intergenic region 25kb to 50kb downstream of the *TRIB1* coding region. By phylogenetic footprinting analysis and DNA genotyping, we identified an evolutionarily conserved region (CNS1) within the risk locus that harbours two common SNPs in tight LD with GWAS risk SNPs and significantly associated with CAD. We investigated the regulatory function of CNS1 by luciferase reporter assays in HepG2 cells and demonstrate that this region has promoter activity. In addition, the rs2001844 risk allele significantly reduces luciferase activity, suggesting that altered expression of the EST-based gene may be associated with plasma TGs. We identified an EST within the risk locus directly downstream of CNS1. We performed 5'/3' RACE using HepG2 RNA, identified multiple variants of this EST-based gene, and confirmed its transcription start site within CNS1. We hypothesize that this EST is a long noncoding RNA due to low abundance, poor conservation, and absence of significant ORF. Over-expression of a short variant implicates its function in the regulation of target gene transcription, although the mechanism of action remains unknown. We conclude that the risk locus at 8q24.13 harbours a novel EST-based gene that may explain the relationship between GWAS SNPs at this locus and plasma lipid traits.

Acknowledgements

I would like to thank my supervisor, Dr. Ruth McPherson, for her guidance and for providing me with the opportunity to work on an exciting novel project in her laboratory. I would also like to thank all members of the Atherogenomics Laboratory: Paulina Lau and Olga Jarinova for their ideas and guidance throughout my MSc project and Thet Naing for her technical assistance. I would also like to extend a thank you to all members of the lab for making my two years here very pleasant and enjoyable. It was a great experience working with each and every one of them. I would also thank my Thesis Advisory Committee members - Drs. Alexandre Stewart, Michael McBurney and Balwant Tuana - for their help expertise. I would further like to thank Dr. Alexandre Stewart and his laboratory for helping with our experiments by generously providing us with expression vectors for TEAD-1 and TEAD-4, as well as an antibody against TEAD-1. Lastly, I would like to acknowledge the patients and volunteers of the Ottawa Heart Study who have made this work possible.

Contents

List of Tables

List of Figures

List of Abbreviations

ACC1 - Acetyl-coA carboxylase
ACOX1 - Acyl-coA oxidase I
Apo - Apolipoprotein
APOB - Apolipoprotein B
BLAST - Basic Local Alignment Search Tool
BMI - Body mass index
CAD - Coronary artery disease
CHD - Coronary heart disease
CM - Chylomicron
CNS - Conserved sequence
DNA - Deoxyribonucleic acid
eQTL - Expression quantitative trait locus
EST - Expressed sequence tag
FAS - Fatty acid synthase
FH - Familial hypercholesterolemia
GWAS - Genome-wide association study
HDL - High-density lipoprotein
HTG - Hypertriglyceridemia
IHD - Ischemic heart disease
LD - Linkage disequilibrium
LDL - Low-density lipoprotein
LDLR - Low-density lipoprotein receptor
LPL - Lipoprotein lipase
MAF - Minor allele frequency
MI - Myocardial infarction
MTP - Microsomal triglyceride transfer protein
ncRNA - non-coding RNA
OHS - Ottawa Heart Study
OR - Odds ratio
PCR - Polymerase chain reaction
RACE - Rapid amplification of cDNA ends
RNA - Ribonucleic acid
SCD1 - Stearoyl-coA desaturase 1
siRNA - Short interfering RNA
SNP - Single nucleotide polymorphism
SQLE - Squalene synthase
SREBP - Sterol regulatory element binding protein
TG - Triglyceride
TC - Total cholesterol
TRIB1 - Tribbles homolog 1
VLDL - Very low-density lipoprotein
ZNF572 - Zinc finger 572

1 Introduction

1.1 Coronary Artery Disease and Established Risk Factors

Coronary artery disease (CAD) is a leading cause of death and disability in Western countries, and its prevalence is increasing within the developing nations as well [1]. Atherosclerosis is a complex underlying pathology of CAD leading up to myocardial infarction (MI) and its manifestation is characterized by lipid accumulation, inflammation, apoptosis, and fibrosis within coronary arteries [2]. Plasma LDL-c and HDL-c are established independent risk factors for CAD. Briefly, LDL transports cholesterol from the liver to the tissues, and its accumulation in coronary arteries contributes to the development of atherosclerosis and CAD. HDL is involved in reverse cholesterol transport, which is the mechanism by which cholesterol returns from the tissues to the liver for catabolism. Evidence strongly shows LDL-c as being pro-atherogenic. Statin treatment, which inhibits HMG-coA reductase - thereby reducing hepatic cholesterol synthesis and plasma LDL-c levels - has been proven to reduce atherogenesis and CAD risk [2]. In contrast, HDL has a protective role, thus, high plasma HDL-c levels are anti-atherogenic [2].

1.2 Hypertriglyceridemia is a Biomarker and Independent Risk Factor for Coronary Artery Disease (CAD)

Plasma triglycerides (TGs) as additional independent risk factors for CAD have been controversial, and as a result, were established more recently [3, 4]. The controversy around hypertriglyceridemia as an independent CAD risk factor arose from the fact that elevated plasma TG levels are associated with other factors including metabolic syndrome [3, 4]. A syndrome is defined as a cluster of factors that occur together too often to be attributable to chance alone [5]. Further, its cause is uncertain, but it is thought that insulin resistance and obesity may contribute to its development [5]. The recent global definition of metabolic syndrome includes a cluster of risk factors for CAD and type 2 diabetes: hypertension

(>130 systolic or 85 diastolic), atherogenic dyslipidemia manifested by elevated plasma TGs (>150mg/dl) and small dense LDL, as well as reduced levels of HDL cholesterol (<40mg/dl in men and <50mg/dl in women), high fasting blood glucose levels (>100mg/dl), and central obesity [5]. Moreoever, individuals with metabolic syndrome also manifest prothrombotic and proinflammatory states, both of which are linked to CAD [5]. Interestingly, evidence suggests that insulin resistance itself is sufficient to induce dyslipidemia and atherosclerosis. In the postprandial state, dietary fat is packaged into intestinally-derived chylomicrons which transport TGs to the peripheral tissues [4]. Consequently, insulin activates LpL and inhibits HSL (hormone-sensitive lipase) in adipose tissue, thereby promoting the uptake of free fatty acids into adipocytes for storage within lipid droplets [3]. Insulin also stimulates decreased hepatic VLDL secretion through increased apoB degradation [6]. Insulin resistance, however, dramatically changes the lipid profile and increases the concentration of apoB-containing lipoproteins in the bloodstream. For instance, a study that used liver insulin receptor knock out mice as their model of pure hepatic insulin resistance noted that these mice developed atherosclerosis and their lipid profile included reduced HDL cholesterol and increased cholesterol-enriched VLDL levels as a result of increased apoB secretion and reduced apoB clearance [7]. Further, in adipose tissue insulin resistance, HSL inhibition is relieved and the liver experiences an increased supply of fatty acids from adipocyte lipolysis [3]. Fatty acids are ligands for several nuclear receptors [3]; they affect cellular lipid metabolism and have been shown to block apoB degradation [6]. Thus, enhanced hepatic TG levels increase VLDL assembly and secretion [3]. Consequently, the hypertriglyceridemic state associated with insulin resistance and metabolic syndrome and the fact that individuals with elevated plasma TGs also have decreased HDL-c levels have greatly complicated our understanding of the role of TGs as independent CAD risk factors.

Moreoever, the diversity of hyperlipoproteinemias adds to the complexity of plasma TG levels as independent risk factors for CAD. Individuals with familial chylomicronemia syndrome have severe hypertriglyceridemia but are not at increased risk for atherosclerosis [8];

it is thought that their plasma lipoproteins are too large to move from the bloodstream into the arterial intimima [9], hence large TG-rich lipoproteins appear to be less atherogenic than smaller ones [8]. In contrast, familial combined hyperlipoproteinemia (FCHL) is a pro-atherogenic condition with a lipid profile of increased VLDL, LDL and small dense LDL from apoB overproduction and reduced lipoprotein clearance [10]. Further, individuals with moderate hypertriglyceridemia from conditions such as familial hypertriglyceridemia (FHTG), where lipid profiles show elevated VLDL levels from increased hepatic TG synthesis [10], are at increased risk for CAD [8]. A 20-year study compared CAD mortality risk between FCHL and FHTG using subjects from 101 American families affected by these conditions [11]. They found that CAD mortality was increased in FCHL but not in FHTG and that this relationship was independent of baseline plasma TG levels in FCHL. Although FTGH was not associated with CAD mortality itself, in FHTG, baseline TGs were significantly associated with CAD mortality independently of total cholesterol [11].

However, given the accumulation of evidence on this question today, the consensus is that both fasting and, more recently, non-fasting TGs [8, 12] are biomarkers for CAD risk prediction as well as independent risk factors, although there exist conflicting findings between studies on the latter issue. Non-fasting TGs reflect the post-prandial state which is accompanied by increased levels of cholesterol from chylomicron remnants [4]. In addition, increased baseline TGs would also translate to increased non-fasting TGs [4]. Consequently, this review will focus on baseline TGs. A 1996 meta-analysis of 17 studies including 46413 men and 10864 women that examined the association between baseline plasma TGs and CAD risk found that the relative risk for CAD in men and women was 1.32 and 1.76, respectively [13]. This indicates that a 1mmol/L TG elevation is accompanied by a 32% increased risk in men, and a 76% increased risk in women. After adjustment for HDL-c and other risk factors, the relative risk for men and women remained statistically significant but was reduced to 1.14 and 1.37, respectively [13]. The Copenhagen Male Study, which examined the relationship between fasting TGs and ischemic heart disease over an 8 year follow-up

period, found that in comparison to the lowest third baseline TG level, the relative risks were 1.5 and 2.2 for the middle and highest third baseline TG levels after risk factor adjustment [14]. Furthermore, this study stratified baseline TGs by HDL levels and found a clear gradient of ischemic heart disease risk in association with increasing baseline TGs within each and every HDL level [14]. Similar findings were also noted in more recent studies. A study using the Reykjavik and EPIC-Norfolk cohorts compared CAD risk between individuals with baseline TGs in the top third of the population with those in the bottom third. They found that the association between baseline plasma TGs and CAD risk was largely attenuated after adjustment for established CAD risk factors, but that it remained significant nonetheless, with odds ratios of 1.76 (Reykjavik) and 1.57 (EPIC-Norfolk) [15]. Moreoever, their meta-analysis, comprising 29 studies including the Reykjavik and EPIC-Norfolk cohorts and totalling 262,525 participants, showed similar results [15]. Similarly, a meta-analysis of 26 prospective studies in the Asia-Pacific region comprising 96,224 participants also found that individuals within the top fifth of TG levels had a 70% increased risk of CAD death and 80% increased risk of developing CAD compared with individuals within the bottom fifth [16]. A more recent meta-analysis of 68 long-term prospective studies in Europe and North America, however, contradicts these previous findings. Instead, they showed that the odds ratio for baseline TGs and CAD was 1.37 after adjustment for non-lipid CAD risk factors, but reduced to 0.99 after adjustment for HDL and non-HDL cholesterol [17]. Thus, although baseline TGs can be viewed as biomarkers for CAD risk, controversy still remains regarding the role of baseline TGs as independent risk factors for CAD.

1.3 Lipoprotein Metabolism and de novo Hepatic Fatty Acid Synthesis Highlight Key Players That Could Likely Affect Plasma TG Concentrations

Plasma TGs are derived intestinally from diet or from hepatic de novo lipogenesis [18]. Intestinally-derived chylomicrons and liver-derived VLDL provide energy to peripheral tissues in the form of triglycerides (TGs). Dietary fat is hydrolyzed within the small intestine by pancreatic lipase; the free fatty acids are taken up by enterocytes and converted back into TGs by the enzyme DGAT2 [3]. The microsomal triglyceride transfer protein (MTP) packages intestinally-derived TGs into chylomicrons, whose production also requires phospholipids and apolipoproteins including apoB-48, apoA-I. apoA-IV, apoE, and apoCs [4]. TG-rich Chylomicrons are subsequently secreted into the lymphatic system and drained into the systemic circulation for delivery to the tissues, thereby bypassing the liver [4]. Their delivery from the lymphatic system to the bloodstream is accompanied by the exchange of surface apolipoproteins, namely, ApoA-IV dissociates and chylomicrons become enriched in ApoC-II - an activator of lipoprotein lipase [4]. LPL-mediated lipolysis of chylomicrons within the capillaries of metabolic tissues results in TG hydrolysis, generating smaller lipid-poor chylomicron remnants that are cleared from the circulation by endocytosis via hepatic low-density lipoprotein receptors (LDLR) [4].

Hepatic de novo lipogenesis includes TG synthesis from de novo synthesized or extra-hepatic sources of fatty acids for VLDL production [19]. De novo fatty acid synthesis is nutritionally regulated and involves acetyl-coA production for fatty acid synthesis from simple carbohydrates. Briefly, after synthesis, TGs are produced from the esterification of fatty acids onto a glycerophosphate backbone and packaged into VLDL particles for export from the liver [19]. Specifically, ingestion of a low fat/high carbohydrate diet causes a marked induction of glycolytic enzymes such as glucokinase and liver pyruvate kinase for acetyl-coA production. This is followed by the induction of enzymes ATP citrate lyase,

acetyl coA carboxylase (ACC), and fatty acid synthase (FAS) for lipogenesis, long-chain elongase (ELOVL6) and stearoyl-coA desaturase for fatty acid chain elongation and desaturation, and mitochondrial glycerol 3-phosphate acyltransferase (GPAT) and diacylglycerol acyltransferase (DGAT) for TG synthesis [19]. Moreoever, increased levels of malonyl-coA - the product of ACC - inhibit carnitine palmitoyltransferase I (CPT I), the rate limiting enzyme of fatty acid β-oxidation, thereby shifting fatty acids from the oxidation pathway to the esterification pathway for TG synthesis [19]. Packing of TGs into VLDL requires the actions of microsomal triglyceride transfer protein (MTP). MTP transfers neutral lipids to apoB-100, thereby allowing apoB to adopt its proper fold and assemble into a spherical lipoprotein with a neutral lipid core [20]. VLDL are secreted from hepatocytes into the circulation and, like chylomicrons, are lipolysed by lipoprotein lipase (Lpl) within the capillaries of adipose tissue and skeletal muscle [3]. VLDL TG lipolysis by LpL generates TG-poor VLDL remnants. Approximately half of these VLDL remnants are taken up by the liver, and the remainder are converted to LDL, which is a smaller, denser apoB-100 containing cholesterol-rich lipoprotein [3, 4]. LDL is cleared from the circulation via cell surface low-density lipoprotein receptors (LDLRs). LDLR binds apoB; the LDLR-lipoprotein complex undergoes clathrin-mediated endocytosis and migrates to endosomes. The complex separates upon acidification of the endosomes, thereby releasing LDL for degradation in the lysosomes while the LDLR is returned to the cell surface [21].

1.4 Environmental Factors that Affect Cellular Lipid Metabolism: Fatty Acids and Insulin

Fatty acids affect lipid metabolism in a variety of ways through several pathways. It is widely established that the synthesis of apoB in the hepatocytes is regulated by oleate [22, 23]. Under serum-free conditions, HepG2 cells (human hepatoblastoma-derived cell line) secrete apoB-100-containing lipoproteins that are similar in size to LDL but TG-rich and cholesterol-poor [24]. Although they do not secrete much VLDL, they are used as a

model of apolipoprotein synthesis and lipoprotein production [23]. Incubation of HepG2 cells with oleic acid results in enhanced apoB synthesis while apoB mRNA levels remain unchanged [23]. Thus, oleic acid treatment increases the number of VLDL particles secreted by hepatocytes. Oleic acid may also have transcriptional effects. For instance, long-term treatment of HepG2 cells with oleic acid upregulates microsomal triglyceride transfer protein (MTP) mRNA, and oleic acid has also been shown to activate the *MTP* promoter in luciferase reporter assays [25]. Consequently, oleic acid may also modulate the size of the secreted VLDL particles.

Other fatty acids, such as n-3 polyunsaturated fatty acids (n-3 PUFA), regulate lipogenesis through the control of hepatic gene expression [26]. Once they enter the hepatocyte, nonesterified fatty acids (NEFA) - including n-3 PUFA - are transported either to intracellular compartments for fatty acid oxidation, or to the nucleus where they can interact with transcription factors to regulate hepatic lipogenesis [27]. Thus, dietary fat composition influences the expression of genes relevant to lipid metabolism. For instance, n-3 PUFAs, which are negative regulators of hepatic lipogenesis [26], bind PPARα, which induces fatty acid oxidation [26], and regulate the expression of PPARα target genes [28]. Interestingly, unsaturated fatty acids of various lengths vary in their ability to activate PPARα, hence, the composition of the NEFA pool influences hepatic lipid metabolism at the transcriptional level [28]. Moreoever, PUFAs can inhibit SREBP-1c promoter activity - SREBP-1c induces fatty acid synthesus [26] - and this inhibition requires the presence of intact LXR-responsive elements (LXREs) within the SREBP-1c promoter region. Interestingly, this suppression is specific to polyunsaturated fatty acids as saturated (palmitic acid) and monounsaturated (oleic acid) fatty acids have only minimal effects [29].

Hepatic fatty acid gene transcription is also under hormonal control. Under normal physiological conditions, insulin stimulates glucose uptake and decreases gluconeogenic rates, thereby reducing blood glucose concentration [19]. Glucose taken up by the liver will first

be converted into glycogen, but, after glycogen stores have been saturated, the remaining glucose is used for the synthesis of fatty acids, which are esterified to form TGs and packaged into VLDL for export to peripheral tissues [19]. Consequently, insulin stimulates the expression of *SREBP-1c* mRNA, resulting in increased nuclear accumulation of SREBP-1c and increased levels of SREBP-1c target genes including fatty acid synthase (FAS) and acetyl-coA carboxylase (ACC) [30].

1.5 Genetic Contributors to CAD: Genome-wide Association Studies, Linkage Disequilibrium and Haplotypes

Although environmental factors such as diet, exercise, and smoking affect individual lipid profiles, it is widely established that nearly 50% of the variation in individual lipid profiles, including LDL-c, HDL-c, and TGs, can be attributed to genetics [31, 32]. As LDL-c has already been strongly associated with CAD risk, it is not surprising that mutations that result in increased plasma LDL-c levels are also predictive of CAD risk. For instance, there is considerable genetic heterogeneity in familial hypercholesterolemia (FH). FH can result from mutations with the gene encoding the LDL receptor, thereby leading to receptor-negative or receptor-defective FH [33]. Alternatively, FH can arise from familial defective apoB100 caused by mutations in the region of apoB that binds the LDL receptor [34]. More recently, another cause of FH has been attributed to gain of function mutations in PCSK9, which has been shown to promote the degradation of the LDL receptor by an endocytic mechanism [35].

While family-based genetic linkage studies have been useful for both single-gene Mendelian disorders such as the causes of the hypercholesterolemias discussed above, they are a limited tool to grasp the genetic complexity behind common diseases [36]. Consequently, a more recent approach - the genome-wide association study (GWAS) - has emerged on the basis of the common disease - common variant hypothesis: common diseases, such as CAD, are

caused by allelic variants present in more than 1-5% of the population [36] that, alone, have a small effect, but along with others confer high risk [37]. Many rarer variants with minor allele frequencies between 1-5% also exist, although these are not as easily analyzed by GWAS. Yet, these low frequency variants remain of interest because it is likely that they are of intermediate genetic effect [36]

.

GWAS rely on arrays that capture a large portion of the common variation throughout the human genome. The variants captured on GWAS chips are single nucleotide polymorphisms: a single nucleotide difference between two individuals at the same position within their genome. The common GWAS platforms include the Affymetrix and Illumina systems, which provide whole genome SNP panels that contain up to 1 million SNPs distributed throughout the genome [37]. The human genome contains many more SNPs on top of those provided on commercially available chips, thus the GWAS SNPs are markers, and not necessarily functional or causative. GWAS are therefore hypothesis-free tests of association between markers (SNPs) across the genome and the disease or phenotype of interest [37]. GWAS rely on a large number of statistical analyses for common SNPs in large sample sizes, thus, there is a high rate of false positive signals. Consequently, genome-wide statistical significance is set at a cut-off of $p<5.0x10^{-8}$.

Given the genetic diversity within the human genome, several SNPs will fall in the regions around the GWAS chip SNPs. Since GWAS SNPs are not necessarily functional, we must use the concept of linkage disequilibrium (LD) - the non-random association of alleles at different loci - to identify other SNPs that are associated with a particular marker and therefore also associated with the trait or disease of interest. If a surrounding SNP is in LD with a GWAS marker, this means that the alleles do not segregate independently and that they have a tendency to be inherited together. As a result, if allele A of the GWAS marker is associated with CAD and is always inherited together with allele A of the surrounding SNP, then one can infer that allele A of the surrounding SNP is also associated

with CAD due to linkage disequilibrium. The degree of genetic linkage can be assessed statistically if SNP genotypes are available by determining the values of linkage coefficients D'[38] and r^2 [39]. These range between values of 0 and 1, indicating that two SNPs are completely independent of one another or inherited together, respectively. D' is a better representation of linkage disequilibrium between two SNPs of unequal allele frequencies, whereas r^2 does not account for differences in allele frequencies [38, 39]. However, when comparing two SNPs, r^2 is more useful in determining disease association [40]. For example, if r^2 is found to be 1 between the tag SNP and an unknown SNP, this indicates that the SNPs are in strong linkage disequilibrium and have identical minor allele frequencies. Hence, without functional studies, one cannot determine which of the two SNPs is causative.

A GWAS locus is defined as a genomic region marked by a common SNP significantly associated with the trait or complex disease of interest [41]. Thus, a GWAS locus is a region of genomic DNA that harbours SNPs associated with the trait or disease of interest. These common variants 'tag' additional variants by linkage disequilibrium, and, as a result, one can more specifically define a GWAS locus as containing blocks of SNPs in LD with one-another over a large genomic distance. Therefore, GWAS loci often harbour elements including additional SNPs in LD with GWAS-identified SNPs, genes, and regulatory elements such as promoters and enhancers [41]. GWAS are unable to directly determine causation; rather, they implicate new genomic regions for further analysis [41]. It is important to note that the strength and effect size of significantly associated loci does not necessarily reflect the potential biological importance of the causal gene. More accurately, it suggests that common variation at a particular locus affects the trait of interest. Thus, the establishment of biochemical and molecular mechanisms are critical to fully comprehend the importance of GWAS signals [41].

There are several tools available to aid with the identification of a functional variant after an association has been found. As discussed previously, SNPs used in GWAS are not

necessarily causative or even functional. One useful approach to locate a functional SNP involves the haplotype block - a genomic region of strong LD and low haplotype diversity [42]. The distribution of alleles for each polymorphism within a haplotype is non-random [37]. This, strong LD within a block means that knowing the allele at one position will allow one to predict the allele at another polymorphic site within the haplotype block with reasonable accuracy. The International HapMap project has created a public database of common variation within the human genome [43]. To date, HapMap has obtained LD data across several ethnic populations for 3.5 million common SNPs, however, it is limited in its ability to capture rarer SNPs with minor allele frequenices of under 5% [44]. Nonetheless, this is an extremely useful tool for determining the linkage disequilibrium between a trait-associated SNP and other common SNPs within a GWAS risk locus of interest. Hence, using HapMap LD data if available, one can infer that a common SNP in tight LD with a trait-associated SNP is itself associated with that trait, thereby eliminating the need for manual genotyping of the surrounding SNPs. More recently, the 1000 Genomes Project has added to the public catalogue of human genetic variation by reporting the location, allele frequency and haplotype structure of 15 million SNP, as well as 1 million insertions/deletions and 20,000 structural variants, the latter of which were primarily novel discoveries [45]. Whereas HapMap collected data on common genetic variations (>5%), the 1000 Genomes Project aimed to characterize over 95% of SNPs with minor allele frequencies greater than 1%, thereby complementing the HapMap project [45].

Lastly, genetic studies across multiple ethnicities are also useful approaches in the identification of causal variants. Linkage disequilibrium of alleles from polymorphic sites across a genomic region forms a haplotype; thus, a set of alleles in tight LD are inherited as a unit. The size of haplotype blocks, however, has been a topic of debate. It has been found that LD blocks can extend to regions as large as or larger than 100kb [46] but the average of size of LD blocks appears to be a population-specific phenomenon. An earlier study compared LD blocks between a population of North European descent and Nigerian population. In-

terestingly, they concluded that LD blocks within Europeans spanned an average of 60kb, whereas LD blocks are typically much smaller in Africans [47]. This dramatic difference in LD pattern between the two populations means that a SNP strongly associated with a trait in one ethnic population may show either an association of the same strength of smaller, or the SNP may lose association all together in a population of different ethnicity. Consequently, if the genetics that contribute to a trait or complex disease are a global phenomenon, then this approach may help narrow down a GWAS signal from a large LD block containing numerous SNPs in tight LD to a signal that contains fewer SNPs with different strengths of disease-association.

1.6 From GWAS Signal to Functional Analysis

GWAS have shown that disease or trait-associated variants are over-represented within protein-coding regions, which together make up approximately 2% of the human genome [48]. Yet, over 80% of disease or trait-associated variants fall found outside of protein-coding regions, within introns or intergenic DNA [48, 36]. Thus, one cannot neglect non-coding DNA when searching for functional disease-associated SNPs. Since noncoding regions are not well characterized, GWAS have the potential to identify novel loci of unknown function that were never considered in relation to the complex disease of interest. For example, GWAS for CAD have identified two loci that elegantly illustrate the transition from the GWAS signal to functional analysis: the 1p13 and 9p21 loci.

The chromosome 1p13 locus harbours SNPs that are strongly associated with LDL-c, CAD, and myocardial infarction (MI) [49, 50, 51]. The SNPs that have the strongest association with LDL-c are in strong linkage disequilibrium in the Caucasian population and are found in a non-coding region between two genes of unknown function: *CELSR2* and *PSRC1*. As non-coding regions may harbour regulatory elements that alter the expression of nearby genes, a study by Musunuru *et al* - published in *Nature* last year - explored

the possibility that the non-coding variants may alter gene expression in *cis* and found that the SNP minor allele increased hepatic expression of *CELSR2, PSRC1* and *SORT1* mRNA [51]. Moreoever, the minor haplotype of the intergenic region between *CELSR2* and *PSRC1* had enhanced transcriptional activity in comparison to the major haplotype. Further fine-mapping of the non-coding region, including examining the associations between these SNPs and LDL-c in African Americans, identified the causal SNP - rs12740374 - and determined that the SNP minor allele created a C/EBP transcription factor binding site which is responsible for altered hepatic gene expression. Lastly, knockdown and overexpression studies in mice determined that the *SORT1* protein product sortilin regulates VLDL secretion, thereby modulating plasma VLDL and LDL levels [51].

The 9p21.3 risk locus is strongly associated with CAD but not associated with established risk factors such as plasma lipids [52]. Thus, this GWAS signal identified a novel biological mechanism - cell cycle regulation - that contributes to the development of CAD. The risk locus spans 58kb and contains several SNPs in tight linkage disequilibrium that fall within an intergenic region near the CDKN2A and CDKN2B genes, which encode tumour suppressors $p16^{INK4A}$ and $p15^{INK4B}$, respectively [52]. In addition, the risk locus overlaps the long non-coding *ANRIL*, whose 5' end overlaps and is antisense to CDKN2B [53]. It has been shown in whole blood RNA that the expression of *ANRIL* is altered in individuals homozygous for the risk allele compared to those homozygous for the reference allele [53]. Fine-mapping of the risk locus for functional SNPs involved phylogenetic footprinting analysis to identify regions of evolutionary conservation. Enhancer activity was found in one such region, and the risk allele of the SNP within this conserved region - termed CNS3 - increased enhancer activity relative to the reference allele [53]. A second study identified another enhancer and they showed that the risk alleles of two SNPs destroy a STAT1 binding site which, when intact, inhibits *ANRIL* expression [54]. Further, it has been shown that *ANRIL* inhibits the expression of CDKN2B by recruiting polycomb repressive complex 2 - a transcriptional silencer - to this locus through a direct interaction with SUZ12 -

a component of PRC2 [55].

1.7 Genome-Wide Association Studies for Plasma TGs

As discussed in the above examples, genome-wide association studies (GWAS) have identified novel and common genetic variants associated with CAD. These GWAS signals can be independent of known risk factors - the 9p21.3 locus - or can be associated with both CAD and established risk factors such as LDL-c - the 1p13 locus. Further, some GWAS signals are associated with a particular lipid trait but not with CAD [4]; this may arise due to pleiotropy, where the GWAS locus is associated with multiple traits [41].

GWAS have also been successful at identifying common SNPS significantly associated with baseline plasma TGs at known and novel loci. The earlier most robust signals have been for *LPL* at 8p21, and the *ZNF259/APOA5/APOA4/APOC3/APOA1* gene cluster at 11q23 [50, 49, 56]. The protein products at these loci play critical roles in lipoprotein metabolism as discussed previously. The *LPL* gene encodes lipoprotein lipase, which is primarily expressed in cardiac muscle, skeletal muscle, and adipose tissue [4]. Lpl is secreted in the vasculature that supplies these tissues where it binds glycosylphosphatidylinositol-anchored HDL-binding protein 1 (GPIHBP1) [41]. Cardiac muscle Lpl knockout in mice results in hypertriglyceridemia [57], which is reversed by transgenic expression of Lpl in the heart [58]. Moreoever, factors that reduce Lpl expression or activity have been shown to cause substantial hypertriglyceridemia. For instance, *ANGPTL3* and *ANGPTL4* encode angiopoietin-like proteins 3 and 4, both of which inhibit Lpl [4] and have also been identified as significant loci in TG-associated GWAS [50, 49, 56]. The lead TG-associated *LPL* SNP is rs12678919 [56], located in an intergenic region downstream of the *LPL* gene that also harbours several SNPs in LD with eachother within the *LPL* risk locus. Although this signal has been validated, the functional variants themselves have yet to be determined [41]. While the *LPL* is most significantly associated with plasma TGs and CAD risk, it is

an example of a pleiotropic signal because it is also associated with reduced plasma HDL-c concentrations [50, 49, 56]. This is not a surprising observation because it has previously been established that Lpl-mediated lipolysis of TG-rich lipoproteins releases ApoC particles that are transferred to HDL. This transfer is accompanied by the movement of lipids [59]. Further, cholesteryl ester transfer protein (CETP) contributes to reduced plasma HDL-c by promoting the exchange of cholesteryl ester for TG in TG-rich lipoproteins, thereby generating small TG-rich HDL particles that are readily catabolized [60, 61]. Thus, these events partially accounts for the observation that Lpl inhibition - reduction in TG-rich lipoprotein lipolysis - significantly reduces circulating HDL-c levels [62]. The GWAS locus at 11q23 - the *ZNF259/APOA5/APOA4/APOC3/APOA1* gene cluster - is most strongly associated with plasma TGs and CAD risk, but is yet another example of a pleiotropic signal because it is also associated with elevated LDL-c and reduced HDL-c [56]. The 'tag' SNP - rs964184 [56] - falls within the 3'UTR of *ZNF259*, upstream from the 5' end of *APOA5*. This gene cluster encodes several apolipoproteins, however, several functional studies have implicated ApoA5 as an important regulator of plasma TG levels. Briefly, it has been shown that overexpression of ApoA5 in mice reduces plasma TG concentrations [63]. In addition, human ApoA5 transgenic mice exhibit increased chylomicron and VLDL catabolism while hepatic VLDL and intestinal chylomicron production remain unaffected [64]. Another study noted that ApoA5 reduced VLDL-TG production without affecting VLDL numbers, which suggests that ApoA5 may also impair the lipidation of apoB during hepatic VLDL production [65]. It has been shown that ApoA5 interacts directly with Lpl, and that this interaction increases Lpl activity [63] - thereby increasing the rate of lipolysis. Moreover, the ApoA5-Lpl interaction requires the presence of proteoglycans; thus, it appears as though ApoA5 reduces plasma TG levels by positioning VLDL and chylomicrons to proteoglycan-bound Lpl for lipolysis [66]. Lastly, the GWAS association between the *APOA5* risk locus and reduced plasma HDL-c may be an indirect effect through the modulation of Lpl activity by ApoA5.

Novel TG-associated loci that have been identified by earlier GWAS from Caucasian cohorts include 2p23 near *GCKR*, 7q11 near *TBL2, MLXIPL*, 8q24 near *TRIB1*, 1q42 within *GALNT2*, 1p31near *ANGPTL3* [50, 49]. These TG-associated GWAS loci have been confirmed by the recent study from the Global Lipids Genetics Consortium (GLGC) [56] - a meta-analysis of over 100,000 subjects spanning multiple ethnic groups that gathered plasma lipid and CAD phenotypes, as well as expression quantitative loci [56]. The GLGC study identified 32 TG-associated loci, that, together, cumulatively explain 9.6% of the total variation in plasma TG concentration, thus constituting nearly 30% of the genetic contribution to plasma TG variability [41]. Further, this study identified for the first time another 21 TG-associated loci, which add about 2.2% to the total plasma TG variation [41]. Interestingly, these percentages indicate that a large source of genetic variability remains unaccounted for. This issue is currently being addressed by resequencing TG-associated GWAS-identified genes for rare variants with individual large effects [67].

The GLGC compared GWAS signals across different ethnic groups and also provided evidence - in vitro and in vivo - in support of the association between specific genes and plasma TG concentrations [56]. To address the global relevance of the GWAS-identified loci, analyses were performed in European, East Asian, South Asian and African American cohorts and it was determined that many loci could contribute to the genetics of plasma lipid variabilities across the global population [56]. This analysis can be extremely useful because it constitutes an additional approach for fine-mapping a GWAS locus to the causal SNP [56]. The approach is effective because LD blocks vary between ethnic populations. For instance, the SNPs within the 1p13 locus associated with LDL-c and CAD were in tight linkage disequilibrium in Caucasians, hence they were equally associated with the phenotype of interest [51]. However, the LD block distribution for this locus differs in African Americans as compared to Europeans, resulting in different association strengths for each SNP. Consequently, analyzing the same SNP associations in African Americans revealed that one SNP within the 6.1kb genomic region clearly had the best statistical association

with LDL-c and CAD - the causal SNP rrs12740374 whose function has been described above [51]. The GLGC also searched for expression quantitative trait loci (eQTLs) to further validate the relationship between genes located near and/or within GWAS loci and plasma lipid traits. Specifically, the study employed RNA expression profiling of numerous transcripts to determine whether GWAS SNPs affected the expression of the nearby genes in liver and adipose tissue - tissues relevant to lipoprotein metabolism [56]. Altered transcript expression at a GWAS locus suggests that a causal SNP may be located within a regulatory region of DNA such as a promoter or an enhancer. Lastly, they performed in vivo mouse model studies to validate the associations between plasma lipids and three novel protein-coding genes that were eQTLs in liver [56]. Thus, eQTLs can facilitate the identification of causal genes within GWAS loci that harbour multiple genes [56].

In addition, the GLGC identified TG-associated loci that significantly increased CAD risk [56]. The strongest CAD associations for TG loci include the *ZNF259/APOA5/APOA4/APOC3/APOA1* locus, the *NAT2* locus which encodes N-acetyltransferase 2, and the *TRIB1* and *LPL* loci [56]. But, there is still some uncertainty regarding an independent association between plasma TGs and CAD risk because some of these loci are associated with multiple traits. For instance, the *ZNF259/APOA5/APOA4/APOC3/APOA1* strongly influences plasma TG concentrations, however, this gene cluster also affects other lipoproteins and thus carries a secondary association with total cholesterol (TC), LDL-c and HDL-c levels [56]. SNPs at the *TRIB1* locus on 8q24.13 are primarily associated with plasma TGs, but have secondary associations with LDL-c and HDL-c. Interestingly, pleiotropy at a GWAS locus does not necessarily result in increased CAD risk. For instance, the rs1260326 SNP within the *GCKR* locus, which encodes glucokinase regulatory protein - a regulator of the glycolytic enzyme glucokinase [68] - is robustly associated with plasma TGs ($p=6x10^{-133}$) but has no association with CAD risk [56]. However, this lead SNP is in tight LD with a common missense variant - rs780094 - within the *GCKR* gene. This SNP is associated with increased fasting TGs, reduced fasting glucose, and reduced risk of diabetes [68]. Thus, in

this situation, the lack of CAD association may be explained by the opposing effect observed on CAD risk factors [4].

1.8 The *TRIB1* Locus

The *TRIB1* locus (8q24.13) is a novel GWAS-identified locus harbouring common genetic variants associated with plasma triglycerides (TGs), LDL-c, and CAD risk [56] and increased risk of myocardial infarction (MI) [69]. In the Ottawa Heart Study, we have demonstrated that the relationship between this locus and CAD risk is entirely explained by the association with plasma lipids - primarily TGs. Other recent studies prior to the Global Lipids Genetics Consortium performed meta-analyses using data from several GWAS and obtained findings consistent with our results that the *TRIB1* locus reaches genome-wide significance for plasma TGs [70, 50, 49]. Another study determined that this locus is also associated with waist circumference in a Spanish population of familial hypercholesterolemia (FH) heterozygotes [71]. Further, significant associations between these risk SNPs and polygenic traits including severe hypertriglyceridemia (HTG) [72] and hyperlipoproteinemias (HLP) [73] have also been found in patients of European ancestry.

These findings are not limited solely to Europeans. SNPs at the *TRIB1* locus are significantly associated with LDL-c levels along with increased risk of CAD and other forms of cardiovascular disease in a Malay population [74], and TG and LDL-c levels in a Japanese population [75]. However, the *TRIB1* locus does not carry a robust signal in African Americans. The Jackson Heart Study, which sought to compare differences between the genetic determinants of lipid profile phenotypes in European and African Americans, did not detect a significant association between SNPs at the *TRIB1* and plasma lipid traits [76]. The *TRIB1* locus was also not significant for CAD and CAD risk factors in the NHLBI CARe Project - a GWAS of CAD and CAD risk factors in African Americans [77]. In addition, this locus was also not significant in African Americans for lipid traits and CAD from the

Global Lipids Genetics Consortium GWAS [56].

The lead GWAS SNPs at this locus - rs2954029 and rs17321515 - are in tight LD and fall within an intergenic region between 25-50kb downstream of the nearest gene - *TRIB1*, which encodes the tribbles homolog 1 protein. Intergenic regions are known to harbour regulatory elements, thus, it is surprising that the Global Lipids Genetics Consortium study did not identify an eQTL at this locus in human liver and adipose tissue [56]. This suggests that the causal SNP does not influence *TRIB1* transcript levels. This risk locus is of particular interest because the LD block of TG-associated SNPs falls within an intergenic region that is more specifically considered a gene desert: a region greater than 500kb that is devoid of genes [78]. Moreoever, the common TG-associated SNPs are not in LD with SNPs that fall within the *TRIB1* coding region. Consequently, this raises the possibility that the GWAS SNPs may not directly affect *TRIB1*, but rather, are associated with plasma lipid traits and CAD risk through a completely novel unknown mechanism. Thus, it will be of interest to begin to elucidate the function of this intergenic region with respect to TG levels.

1.9 Characteristics of Tribbles Proteins

The cluster of TG-associated SNPs falls within a large intergenic region of unknown function 25kb to 50kb downstream of the *TRIB1* coding region - Chr8: 126,442,563 - 126,450,647. Tribbles proteins are evolutionarily conserved, which suggests that they are ancient members of signalling pathways involved in fundamental cellular processes [79]. They have been implicated in embryonic development, as well as in various human diseases, including cancers, atherosclerosis, and insulin resistance [79]. Drosophila tribbles is a key regulator of cell cycle progression in morphogenesis [80, 81, 82] as well as cell migration during drosophila oogenesis [83]. Drosophila tribbles inhibits the activity of mitotic activator String/Cdc25 by inducing its proteolytic degradation via the proteasome [81, 80]. Further, slbo, a homolog of the mammalian C/EBP transcription factor, is required for border cell migration during

drosophila oogenesis [83]. Tribbles negatively regulates Slbo expression by specifically targeting it for proteasomal degradation [83].

Mammalian tribbles are encoded for by three separate genes. Structurally, Trb1-3 have a proline-rich N-terminal domain, a center kinase-like domain, and a C-terminal domain rich in charged amino acids [84] The NT domain contains potential phosphorylation sites [84]. It also contains predicted PEST regions responsible for the control of protein half-life [84]; this prediction supports studies showing tribbles to be a labile protein [85] and also fits its emerging role as a regulator of cell signalling cascades [86, 87] and cell division [81]. The NT domain is also required for Trib1 and Trib3 nuclear localization [88]. The kinase-like domain is more evolutionarily conserved than the NT and CT domains [79, 84], yet, *in vitro* kinase assays to date have not been able to detect any kinase activity and further studies are required to determine this domain's function [89]. One plausible mechanism - consistent with reports demonstrating that Trib proteins, when overexpressed, act as negative regulators of protein-kinase mediated signalling pathways - is that Trib competes with kinases for their substrates through its kinase-like domain [79].

1.10 Known Functions of Trib1, Encoded for by the *TRIB1* Gene

Trib1 is a known regulator of MAPK signalling [84]. Studies suggest that Trib1 regulates signalling at the level of MAPK phosphorylation and activation via interactions with MAPKKs, specifically MEK-1 (ERK activator) and MKK4 (JNK activator) [86]. It has also been shown that Trib1 has an inhibitory effect on MAPK signalling at higher levels, suggesting that it acts in a dose-dependent manner [86]. In addition, Trib1 activation and its ability to regulate MAPK signalling has been shown to be cell-type specific [87]. Interestingly, a study aimed at building a database for mRNA half life of 19,977 genes found that the mean estimated half-life of a transcript was 7.1 hours [90]. However, *TRIB1* was among the few genes that was found to have a short half life of under 1 hour [90]. Consequently,

the short-lived message and its PEST domain that controls protein half-life make it suited for its role as a regulator of MAPK signalling cascades.

Thus far, Trib1 has been implicated in processes including cell proliferation and inflammation as well as in conditions such as atherosclerosis, and myeloid leukemogenesis [87, 91, 92]. Trib1 involvement in acute myeloid leukemogenesis (AML) requires interactions via its CT domain with MEK-1, resulting in ERK phosphorylation and activation [92]. Moreoever, Trib1 can function as an adapter protein, recruiting E3 ubiquitin ligases to promote the degradation of C/EBPα which is required to induce AML [93]. C/EBPα degradation has been shown to depend on the Trib1-MEK-1 interaction [92]. In addition, Trib1 has been shown to function as a negative regulator of the retinoic acid receptor (RAR) and suppresses RAR target gene expression [94]. It appears as though Trib1 directly regulates the transcriptional activity of the RAR/RXR heterodimer via a direct interaction through its kinase-like domain [94]. Further, this interaction required the ligand-binding domain of RARα. This could be another potential mechanism by which Trib1 influences cancer progression, as it has been established that loss of normal RAR function - even in the presence of physiological retinoic acid concentrations - is associated with cancer progression [95]. For example, acute promeylocytic leukemia (APL) - a subtype of AML - results from an abberant chromosomal translocation that juxtaposes the *PML* gene and the *RAR*α, generating a chimeric protein which retains the DNA-binding and ligand-binding domains of RARα [96] and this retains its function. A recent report also showed that Trib1 is potentially relevant to atherosclerosis; Trib1 interacts with MKK4 to control vascular smooth muscle cell proliferation and chemotaxis through the JNK pathway and that Trib1 expression is elevated in human atherosclerotic arteries [91]. Trib1 expression is also upregulated in inflamed white adipose tissue, where it has been found to induce proinflammatory cytokine production [97]. Interestingly, *TRIB1* haploinsufficiency in mice protected them against high fat diet-induced obesity under low-grade inflammatory conditions but had no effect on serum TG levels [97].

Although the link between Trib1, inflammation, and atherosclerosis may be a plausible explanation for the association of this locus with CAD risk, the functional relationship between SNPs downstream of *TRIB1* identified by the GWAS approach and serum TGs remains unknown. A recent study demonstrated a relationship between Trib1 and hepatic lipogenesis. Hepatic Trib1 overexpression in mice decreased VLDL production, which resulted in decreased plasma TG and total cholesterol (TC) levels [98]. Hepatic overexpression of Trib1 was accompanied with decreased TG synthesis along with reduced levels of *ACC, FAS*, and *SCD1* mRNA, while *DGAT2, MTP*, and *CPT1A* mRNA levels remained unchanged. *Trib1-/-* mice had a 30% increase in TG, TC, VLDL and LDL-c, and exhibited an increased TG production rate, all of which were rescued by hepatic Trib1 overexpression. They also reported increased *ACC, FAS, SCD1* mRNA levels, decreased *ACOX1, CPT1A, CPT2* mRNA levels, no change in *DGAT2, MTTP, APOB, LDLR* mRNA levels, and no change in *SREBP1c* mRNA [98]. In addition, they noted that Trib1 was exclusively localized to the nucleus of HepG2 cells by immunofluorescence [98]. This study strongly implicates Trib1 as a regulator of hepatic lipogenesis, and fits the plasma lipid profile observed in the numerous GWAS that have been conducted. Yet, the Global Lipids Genetics Consortium did not identify *TRIB1* as an eQTL. Given the extremely short half-life - under 1 hour [90] - of this transcript, it still remains a possibility that *TRIB1* is the causal gene and that the causal SNP within this risk locus alters transcript abundance.

Further, the relationship between the MAPK signalling pathways and hepatic de novo fatty acid synthesis has not been studied extensively. Studies in HepG2 cells - which have an overactive MEK-ERK signalling pathway - have shown that ERK1/2 inhibition using a MEK1/MEK2 inhibitor resulted in increased apoB secretion [99]. Additionally, ERK inhibition in the presence of oleic acid resulted in the formation of VLDL, thereby correcting the defective VLDL assembly normally seen in HepG2 cells [99]. ERK1/2 inhibition also led to increased cellular microsomal TG mass and increased expression of *DGAT2* mRNA

[99]. Also, SREBP-1c is a direct phosphorylation substrate of ERK1/2 *in vitro* [100]. As a result, MAPK signalling may also be linked to insulin-induced gene expression through SREBP-1c activation [100]. Since Trib1 has been shown to interact with the ERK activator MEK1, it is possible that Trib1 may be acting through this pathway. Moreoever, p38 MAPK signalling has an inhibitory effect on hepatic lipogenesis [101]. More specifically, p38 appears to inhibit hepatic lipogenesis through an inhibitory role in the transcription of *SREBP-1* and *PGC-1β* [101]. Finally, although Trib1 is now known to regulate hepatic lipogenesis - which appears to be affected by MAPK signalling - the mechanism by which the *TRIB1* locus SNPs modulate TG levels still remains a mystery.

1.11 Intergenic Regions and Long Non-Coding RNA

HapMap data shows that these intergenic GWAS risk SNPs, located between Chr8:126477153 - 126499878, are not in LD with SNPs in the *TRIB1* coding region (Chr8:126442563 - 126450647). This means that allele segregation between between *TRIB1* coding region and the risk locus is random that these two regions are genetically independent of one another. This again raises the possibility that the intergenic GWAS risk SNPs may not function through *TRIB1* directly, if at all, but rather, may influence a novel pathway.

Although it was traditionally believed that only 2% of the human genome is expressed, new research methods including cDNA libraries and genomic tiling arrays have discovered that likely over 90% of the human genome is transcribed [102, 103]. Consequently, in addition to regulatory sequences, it is also possible that this large intergenic region may harbour non-coding RNAs (ncRNA). Generally, many long ncRNA transcripts are low abundance and of unknown functional significance because, unlike most protein coding genes, they are not highly conserved between species, they contain short open reading frames (ORFs) of under 50-100 amino acids, and can span large genomic distances [104, 103]. However, this does not necessarily indicate lack of function; rather, ncRNAs are complex molecules

with various functions. To date, known functions of long ncRNA include chromatin modification, transcriptional regulation, and post-transcriptional regulation [104, 103]. Further, long ncRNA transcripts with short ORFs may also encode peptides [104, 103]. For instance, the *polished rice* sORF in *Drosophila* has been shown to encode small bioactive peptides between 11-32 amino acids in length that control epidermal differentiation [105].

Since ncRNAs are being implicated in the regulation of fundamental cellular processes such as gene expression, they are gaining interest with respect to the molecular mechanisms behind complex diseases including cancers and atherosclerosis. This section will review some examples highlighting the diverse functions of ncRNAs.

There are many examples that highlight the roles of ncRNA in transcriptional regulation. Long ncRNAs can modulate gene transcription via the recruitment of RNA binding proteins [106]. For example, low copy number ncRNA transcripts located near the 5' end of *CCND1* - whose gene product is cyclin D1 - are induced in response to DNA damage. These transcripts recruit the RNA binding protein TLS to the *CCND1* promoter, thereby repressing target gene transcription [106]. Non-coding RNA has also been shown to regulate RNA polymerase II transcription. This mechanism of regulation is illustrated by the SINE-encoded mouse B2 RNA, which is induced in response to cellular heat shock and binds RNA polymerase II, thus associating with the transcription pre-initiation complex and preventing RNA polymerase II from properly associating with the template DNA strand [107]. Similar mechanisms have also been established in humans. For instance, a ncRNA transcribed from an upstream region of the *DHFR* locus - encoding dihydrofolate reductase - forms a triplex within the *DHFR* promoter which blocks the binding of the transcriptional co-factor TFIID [108]. In general, short interspersed repetitive elements (SINEs) - primarily *Alu* elements in humans - are normally transcribed by RNA polymerase III machinery and regulate the transcription of RNA polymerase II target genes by interacting with components of RNA polymerase II transcription machinery [109]. Interestingly, it has also been shown that small

cytoplasmic *Alu* RNA is initially transcribed as a longer primary transcript that undergoes post-transcriptional processing [110].

Long ncRNAs can also influence epigenetic modifications via the recruitment of chromatin modification complexes to specific genomic loci [104]. For instance, the mouse X chromosome inactivation center harbours the gene encoding the long ncRNA XIST; accumulation of XIST transcripts is associated with chromatin changes [111]. XIST modulates X-chromosome inactivation in females by recruiting Polycomb Repressive Complex 2 (PRC2) - a histone H3K27 methylase which consists of several components including Eed, Suz12, RbAp48, and the catalytic subunit Ezh2 [112]. RNA immunoprecipitation using antibodies against these complexes revealed that XIST was consistently bound to the Suz12 and Ezh2 subunits [111]. Epigenetic changes are also critical in development. The *HOX* loci are essential for specifying the positional identities of cells during development; *HOX* expression is also regulated by epigenetic changes [113]. The long ncRNA HOTAIR is transcribed from the *HOXC* locus. The 5' end of HOTAIR binds PRC2 [113, 114], and the 3' end has recently been shown to bind another repressive complex - the LSD1/CoREST/REST complex [114]. Moreover, the noncoding RNA ANRIL - encoded for at the 9p21 locus - altered in several tumours [115] and also identified in CAD association GWAS [52]- represses expression of the $p15^{INK4B}$ locus by directly interacting with Suz12 in the recruitment of PRC2 [55].

1.12 Determining the Functional Relationship Between SNPs at the *TRIB1* Locus and Elevated Plasma TGs

Although Trib1 has been shown to regulate hepatic lipogenesis at the level of hepatic de novo fatty acid synthesis, the relationship between GWAS-identified SNPs at this locus and plasma TG levels remains unknown. Thus, this research project focuses on the functional analysis of the intergenic region harbouring TG-associated SNPs. It does not focus

on the mechanism by which Trib1 modulates hepatic lipogenesis - possibly via regulation of MAPK signaling - because there is no evidence of an association between the GWAS signal and Trib1. The aims of this project include the identification and characterization of regulatory elements and/or novel genes within the 8q24.13 gene desert within the focus of TG-associated SNPs.

2 Materials and Methods

2.1 Phylogenetic Footprinting Analysis

SNPs were identified in evolutionarily conserved regions within the *TRIB1* risk locus by phylogenetic footprinting analysis using PipMaker, an online tool which computes the alignment of similar regions between multiple sequences. The Ensembl Genome Browser was used to obtain the sequence of the *TRIB1* coding region and the flanking 100,000bp on either side in humans. It was also used to collect the sequence of orthologous regions in the following species: chimpanzee, macaque, cow, rat, mouse, elephant, opossum, anolis lizard, finch, and xenopus. The human sequence was analyzed by Repeat Masker - an online tool that screens DNA sequences for regions of low complexity and interspersed elements - to generate a file for PipMaker that masked repeats and low complexity regions. Files for PipMaker input were also constructed to annotate the position of exons and SNPs within the given sequence.

2.2 A Database Search for Novel Genes Within the Risk Locus

The Ensembl Genome Browser - which aligns entries from various databases with their respective genomic position - was used to identify novel genes within the risk locus. The databases activated within Ensembl included those for protein-coding genes, non-coding genes, and human expressed sequence tags (EST). Upon identification of an EST from human lens tissue, a BLAST search was performed with its sequence in Open Biosystems and a construct was identified and purchased: clone fs35c04 human lens cDNA. The construct was fully sequenced using sequencing methods described under *DNA sequencing* with the exception that 300ng of plasmid DNA was used as template rather than a PCR product from genomic DNA. The next step involved determining its expression in more relevant tissue types - HepG2 cells and whole blood - by qRT-PCR. Further, 5'/3' RACE was performed to map the EST and identify transcript variants in HepG2 cells.

2.3 Genotyping SNPs Within Conserved Regions by DNA Sequencing

(i) Subjects

CHD Cases

Subjects for CHD cases included Caucasian men and women (from the Ottawa Heart Study) with severe premature coronary heart disease (CHD) that resulted in coronary revascularization. In all CHD cases, the documented onset was under 60 years of age. Diabetics and individuals with plasma cholesterol levels indicative of monogenic hypercholesterolemia (>280mg/dl) - both of which contribute to the development of severe premature CHD - were excluded from the study. Written and informed consent was obtained from all participants.

Controls

For SNP genotyping at the *TRIB1* locus, controls included healthy Caucasian men (>65 years) and women (>70 years) with no symptoms and no previous personal or family history of CHD. For *TRIB1* mRNA expression analysis in whole blood, the chosen control subjects were not on lipid-altering medication, and were matched for rs17321515 genotype (10AA, 10AB, 10BB, where B is the minor allele and the effect allele), age, gender, and BMI. All characteristics are summarized in Table 1. Written and informed consent was obtained from all participants.

(ii) DNA Sequencing

5 SNPs were genotyped by DNA sequencing. The first amplicon - the first region of conservation within the risk locus (CNS1) - had four SNPs, and the second amplicon (CNS2) contained 1 SNP. Both regions were amplified by PCR using the Roche Fast-Start Taq DNA Polymerase dNTP pack. Each 25μL reaction contained 100-300ng genomic DNA template, 1X PCR buffer with 2mM $MgCl_2$ and 1X GC-rich solution, forward and reverse primers at 0.4μM each (Appendix A), 0.2mM of each dNTP, and

1U FastStart Taq DNA polymerase. PCR reactions were run on the Applied Biosystems (ABI) GeneAmp® PCR System 9700 using the following cycling parameters: a hold at 95°C for 4min, 45 cycles of 95°C for 1min, annealing at 57°C for 1min, and extension at 72°C for 1min30sec, and, at the end, a further extension at 72°C for 8min. After amplification, the PCR reaction itself was used directly as a template for DNA sequencing, performed according to the Applied Biosystems BigDye Terminator v3.1 Cycle Sequencing Kit. Each 20μL sequencing reaction contained 1μL of PCR reaction template, 0.16μM sequencing primer (Appendix A), 0.625X BigDye Sequencing Buffer and 0.25X Ready Reaction (RR) Premix. Sequencing reactions were run on the ABI GeneAmp® PCR System 9700 using the following cycling parameters: 96°C for 30sec, 50°C for 15sec, and 60°C for 4min for 25 cycles. Sequencing reactions were purified using the Montage SEQ_{96} Sequencing Reaction Cleanup Kit (Millipore). Electrophoresis of purified DNA samples was performed on the ABI Prism 3730 DNA Analyzer and sequences were analyzed for SNP genotyping using DNA Sequencing Analysis Software v5.2 (ABI).

2.4 Generation of Luciferase Reporter and Expression Constructs

(i) Luciferase Reporter Constructs

CNS1 luciferase enhancer constructs

Risk and reference haplotypes of evolutionarily conserved region CNS1 were generated by PCR with primers containing Nhe1 and Xho1 restriction sites (Appendix A). Reference and risk CNS1 haplotypes were amplified by PCR using genomic DNA (100-300ng) from reference and risk subjects as templates. Reference and risk haplotypes were determined by genotyping. CNS1 was amplified by PCR using the *PfuUltra*II system from Stratagene (Agilent Technologies). Each 50μL reaction contained 1X *PfuUltra*II reaction buffer with 2mM $MgCl_2$, forward and reverse primers at 0.2μM each, 0.25mM of each dNTP, and 1μL of *PfuUltra*II Fusion HS DNA polymerase.

PCR reactions were run on the ABI GeneAmp® PCR System 9700 using the following cycling parameters: an inital 2min denaturation at 95°C, 40 cycles of 95°C for 30sec, annealing at 57°C for 30sec, and extension at 72°C for 20sec, and, at the end, a further extension at 72°C for 7min. Next, 2.5U of Taq polymerase (Roche) was added and the reactions were incubated for another 20 minutes at 72°C to add A' overhangs to each PCR product. The final PCR products were TOPO®-cloned into the pCR®2.1-TOPO® vector (Invitrogen) and the TOPO® Cloning reaction was transformed into chemically competent TOP10 *E.coli*. To subclone the CNS1 inserts into the luciferase reporter vector pGL4.23 (Promega), the inserts were first excised from the pCR®2.1-TOPO® vector using the NheI and XhoI restriction sites, purified from 1% agarose gel (Qiagen Gel Extraction Kit), and ligated into the pGL4.23 minimal promoter vector by T4 DNA ligase using the NheI and XhoI restriction sites. The ligation reactions were transformed into chemically competent *E.coli* DH5α. All clones were screened by restriction digest of mini-prep plasmid DNA (Roche) using NheI and XhoI. Positive clones that contained an insert were further verified by DNA sequencing with the same conditions described in section 2.2.2, with the exception that 300ng plasmid DNA was used as the template for the sequencing reaction. All plasmids are summarized in Appendix A.

CNS1 luciferase promoter constructs

Several CNS1 promoter constructs were generated (Appendix A). The first four, however, were constructed for the reference haplotype only. The largest promoter construct was 2kb and the other three were 500bp deletion derivates from the 5'end (1.5kb, 1kb, 0.5kb). Consequently, the smallest of the four - exactly 557bp - comprises the entire CNS1 region flanked on either end with minimal additional sequence. These regions were amplified by PCR using the *PfuUltra*II system from Stratagene (Agilent Technologies) with genomic DNA (100-300ng) from a subject who has been genotyped and identified as homozygous reference for CNS1 SNPs as the template. PCR conditions were nearly identical to those described above, with the following exceptions:

the extension time was increased to 40sec and A' overhangs were not added, thus the last 20min hold at 72°C was also removed.

The remaining CNS1 luciferase promoter constructs and all cloning primers are given in Appendix A. Each insert was generated by PCR using the *PfuUltra*II system from Stratagene (Agilent Technologies). In each case, 30ng of plasmid DNA was used as the template. Specifically, the templates were either the risk or reference CNS1-pGL4.23 (Promega) luciferase reporter vectors described under *CNS1 luciferase enhancer constructs*. Each PCR reaction included 30ng plasmid DNA template, 1X *PfuUltra*II reaction buffer with 2mM $MgCl_2$, forward and reverse primers at 0.2μM each, 0.25mM of each dNTP, and 1μL of *PfuUltra*II Fusion HS DNA polymerase. PCR reactions were run on the ABI GeneAmp® PCR System 9700 using the following cycling parameters: an inital 2min denaturation at 95°C, 40 cycles of 95°C for 30sec, annealing at 57°C for 30sec, and extension at 72°C for 15sec, and, at the end, a further extension at 72°C for 7min.

All PCR products were purified (Qiagen PCR Purification) and incubated with enzymes MluI and BglII (NEB) for at least 1 hour at 37°C for restriction digest. Following restriction digest, the products were purified (Qiagen PCR Purification) and ligated into the pGL3 basic luciferase reporter vector (Promega) by T4 DNA ligase. The ligation reactions were transformed into chemically competent *E.coli* DH5α. All clones were screened by restriction digest of mini-prep plasmid DNA (Roche) using MluI and BglII. Positive clones that contained an insert were further verified by DNA sequencing with the same conditions described in section 2.2.2, with the exception that 300ng plasmid DNA was used as the template for the sequencing reaction.

(ii) **Mammalian Expression Constructs**

EST variants were identified by 3'RACE (Roche), as detailed in the section *Mapping the EST by 5'/3' RACE*. All primer sequences are given in Appendix A. Briefly, for cloning applications, all EST inserts contain a PflF1 restriction site near the 5' end, and a MluI site at the 3' end. PCR products were TOPO®-cloned into the pCR®2.1-TOPO vector (Invitrogen), followed by transformation of chemically competent *E.coli* TOP10, blue/white colony screening, and plasmid DNA preparation as detailed previously. All clones were sequenced to verify that no mutations had been introduced in the PCR steps. Of all variants that were screened, only one was found free of mutations and was selected for cloning into the pCMV5 mammalian expression vector. PCR was used to generate the 5' end of the EST transcript since this EST insert from 3' RACE does not contain the full 5' sequence. The missing 5' fragment was generated by PCR using the *PfuUltra*II system from Stratagene (Agilent Technologies). The PCR template was plasmid DNA that included the full 5' end of the EST as identified by 5' RACE. The PCR reaction contained 30ng of plasmid DNA template, forward and reverse primers at 0.2μM each (with BglII and PflF1 restriction sites, respectively; Appendix A), 1X *PfuUltra*II reaction buffer with 2mM $MgCl_2$, 0.25mM of each dNTP, and 1μL of *PfuUltra*II Fusion HS DNA polymerase. PCR reactions were run on the ABI GeneAmp® PCR System 9700 using the following cycling parameters: an inital 2min denaturation at 95°C, 40 cycles of 95°C for 30sec, annealing at 57°C for 30sec, and extension at 72°C for 15sec, and, at the end, a further extension at 72°C for 7min. The PCR product was purified (Qiagen PCR Purification) and digested by restriction enzymes BglII and PflF1; restriction digest was followed by another purification (Qiagen). The pCR®2.1-TOPO® vector containing the 3' RACE EST insert was digested using restriction enzymes PflF1 and MluI. The pCMV5 plasmid was digested using restriction enzymes BglII and MluI. Following restriction digests, the digest reactions were analyzed on 1% agarose gel. Linearized pCMV5, and the EST insert were excised and gel-purifed (Qiagen Gel Extraction Kit). Next, the 5' PCR

product, the 3' EST insert, and linearized pCMV5 were mixed with T4 DNA ligase in a triple-ligation reaction, which was incubated at 16°C overnight, followed by heat shock transformation (42°C) of chemically competent *E.coli* DH5α with the ligation reaction. All clones were screened by restriction digest of mini-prep plasmid DNA (Roche) using BglII and MluI restriction enzymes. Positive clones that contained an insert were further verified by DNA sequencing with the same conditions described in section 2.2.2, with the exception that 300ng plasmid DNA was used as the template for the sequencing reaction. The expression vector is summarized in Appendix A.

2.5 Bacterial Culture and Plasmid Preparation

Chemically competent *E.coli* DH5α (NEB) or TOP10 (Invitrogen) were transformed by heat shock at 42°C and incubated in a rotary shaker at 37°C and at 250rpm for 1 hour. The transformation reactions were plated on LB-agar plates and incubated overnight at 37°C. After transformation, *E.coli* DH5α were plated on LB-agar containing 100μg/mL ampicillin, and *E.coli* TOP10 were plated on LB-agar containing 50μg/mL kanamycin and 40μL of 40mg/mL X-gal (in DMF) for blue/white colony screening. For mini-prep of plasmid DNA in all cloning applications, single colonies were inoculated in 3mL of LB medium containing 100μg/mL ampicillin and incubated in a rotary shaker at 250rpm overnight at 37°C. For large-scale preparation of plasmid DNA, single colonies were inoculated as described above, and incubated in a rotary shaker (250rpm, 37°C) for 8 hours. Next, 100μL of fresh culture was inoculated into 100mL of LB medium containing 100μg/mL ampicillin, and incubated overnight at 37°C and 225rpm. Plasmid DNA was purified using the Roche High Pure Plasmid Isolation Kit for mini-prep, or the Qiagen Plasmid Midi Kit for midi-prep.

2.6 Cell Maintenance

Cos-7 cells were maintained at 5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM), containing 4.5 g/L glucose, and supplemented with 10% FBS and 1% L-glutamine. HepG2 cells were maintained at 5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM), containing 1g/L glucose and 1mM sodium pyruvate, and supplemented with 10% FBS and 1% L-glutamine. HepG2 cells were subcultured when they attained a confluency of 80%, were plated at a minimum density of 25%, and were discarded by the 11th passage number.

2.7 Transient Transfection of Cos-7 and HepG2 Cells

Seeding Cells for Transient Transfection

Cos-7 and HepG2 cells were grown in 10cm dishes and seeded upon reaching 80% confluency. Cos-7 cells and HepG2 cells were seeded at a density of 50% and 75% of one 10cm dish per 12-well plate, respectively. The day after seeding, both cell types had reached at least 80% confluence and were ready for transient transfection.

Enhancer assays

The cells were transfected 24 hours after seeding, using the Fugene HD transfection reagent (Roche). All assays were performed in triplicates. Each well was transfected with 1μg total DNA (980ng CNS1-pGL4.23 luciferase reporter and 20ng pGL4.74 *Renilla* luciferase (Promega)).

Transient transfection for luciferase promoter assays

The cells were transfected 24 hours after seeding, using the Fugene HD transfection reagent (Roche). All assays were performed in triplicates. Each well was transfected with 1μg total DNA (980ng CNS1-pGL3-basic reporter vector and 20ng pRL-SV40 *Renilla* luciferase (Promega)). For co-transfection with TEAD1 or TEAD4 expression vector, each well was also transfected with 1μg total DNA (500ng CNS1-pGL3-basic luciferase reporter, 10ng pRL-SV40 *Renilla* luciferase, 0-25ng TEAD1/TEAD4, 465-490ng pCMV5).

In all luciferase assay transient transfections, DNA was diluted with OptiMem (Invitrogen) at a concentration of 50μL/μg of DNA, and Fugene HD (Roche) was added at a concentration of 3μL/μg of DNA. The samples were incubated at room temperature for 15 minutes. After the incubation, 50μL of complexed DNA was added to each well. The cells were harvested for the luciferase reporter assay 24 hours post-transfection.

Transient transfection for EST and TEAD1 overexpression

HepG2 cells were transfected 24 hours after seeding, using the Fugene HD transfection reagent (Roche). Each well was transfected with 500ng DNA (either empty pCMV5, pCMV5-EST, or pXJ40-TEAD1/4) using the same conditions described previously. The cells were harvested for RNA or protein analysis 24 hours post-transfection.

2.8 siRNA-Mediated Knockdown of the EST in HepG2 Cells

Four siRNA oligos were generated using the Dharmacon siDESIGN Center (Dharmacon RNAi Technologies). These siRNA oligos were designed to specifically target EST exons 1 and 2, which are common to all EST transcript variants identified in HepG2 cells by 5'/3' RACE. Of the 4 custom-designed oligos, only two achieved significant - yet small - knockdown of the EST transcripts in initial optimization experiments. These two oligos were pooled together in equimolar amounts to produce a 100μM siRNA stock.

HepG2 cells at (10cm dish at 80% confluency) were seeded at 30% in 12-well plates to achieve a density of 50% after 24 hours for next-day transfection. Transfection of siRNA was performed in triplicate. Non-target siRNA was used as a negative control (NT-1 from Dharmacon). First, the 100μM siRNA stock (EST and non-target (NT) was diluted to 10μM in RNase-free water. Secondly, 7.5μL of 10μM siRNA (x3.5 = 26.25μL 10μM siRNA) was diluted in 92.5μL OptiMem (Invitrogen) (x3.5 = 323.75μL). In a separate tube, 4μL Dharmafect-1 (Dharmacon) (x3.5 = 14μL) was mixed with 96μL OptiMem (x3.5 = 336μL).

Both tubes were incubated for 5 minutes at room temperature. Next, the siRNA dilution in tube 1 as added to the Dharmafect-1 dilution in tube 2; the contents were mixed and incubated for 20 minutes at room temperature. Lastly, 800μL of media was added to each well of the 12-well plate, and 200μL of siRNA complexes was added to each well for a final siRNA concentration of 75nM. RNA was isolated 72 hours post-transfection using the TRIzol method (as outlined in section 2.11) for cDNA synthesis and quantitative RT-PCR.

2.9 Luciferase Reporter Assay

HepG2 and Cos-7 cells were harvested 24 hours post-transfection. The cells were washed with 1X PBS and lysed by the addition of 200μL of 1X Passive Lysis Buffer (Promega) per well. After addition of lysis buffer, the cells were incubated at room temperature for 15 minutes with gentle shaking. To ensure complete lysis of cells, lysates were mixed with a micro-pipette, transferred to 1.5mL microcentrifuge tubes, and centrifuged for 30 seconds at 10,000rpm at 4°C to pellet cell debris. Following centrifugation, the supernatants were transferred to clean 1.5mL microcentrifuge tubes. The luciferase reporter assays were conducted using a single-tube luminometer (Berthold). 20μL of cleared lysate was first added to the tube. 50μL of LARII (Promega) was mixed with the lysate and luciferase activity was recorded. Next, 50μL of Stop & Glow® reagent (Promega) was mixed into the same tube, and renilla activity was recorded. This sequence was repeated until all lysates were measured for firefly and renilla luciferase activities.

2.10 Immunoblotting

(i) **Preparation of Whole Cell Extracts**

HepG2 cells (12-well plate) were washed with 1X PBS and harvested by trypsinization with 200μL of 1X trypsin, followed by an incubation for 5 minutes at 37°C. The cells were pelleted at 1000rpm for 5 minutes and the supernatant was discarded. The cell

pellets were resuspended in 200μL of RIP buffer (25mM Tris-HCl pH 7.4, 150mM KCl, 5mM EDTA, 0.5% NP-40, 0.5mM DTT, 1X Complete Protease Inhibitor Cocktail Tablets (Roche)) and incubated on ice for 15 minutes to swell. To ensure complete lysis, the cells were sheared on ice using a dounce homogenizer. The lysed cells were centrifuged at 13,000rpm for 5 minutes to pellet cell debris, and supernatants were transferred to clean 1.5mL microcentrifuge tubes. Protein concentrations were determined by BCA protein assay (Pierce).

(ii) **SDS-Polyacrylamide Gel Electrophoresis and Protein Transfer to Nitrocellulose Membrane**

5X SDS-PAGE loading buffer and β-mercaptoethanol were added to 30μg of protein extract for a final concentration of 1X and 5% (v/v), respectively. Samples were boiled for 5 minutes, loaded onto a 10% polyacrylamide gel (with 4% stacking gel), and electrophoresis was performed at 150V until the loading dye completely disappeared. The proteins were transferred to a nitrocellulose membrane at 70V for 1 hour using chilled 1X transfer buffer containing 20% methanol.

(iii) **Western Blotting**

The nitrocellulose membrane was rinsed briefly with 1X PBS and blocked in Odyssey® Blocking Buffer (LI-COR®) for 1 hour at room temperature with gentle shaking. The membrane was incubated overnight at 4°C with primary antibody (TEAD1 mouse monoclonal) at a 1:2000 dilution in Odyssey® Blocking Buffer (LI-COR®) with 0.02% sodium azide. The membrane was washed for 3x10min with PBS-T (1X PBS, 0.1% Tween20) and incubated in the dark with IRDye® 800CW goat anti-mouse IgG secondary antibody at a 1:15000 dilution in Odyssey® Blocking Buffer containing 0.01% SDS for 1 hour at room temperature with gentle shaking. The addition of SDS reduces

background upon detection. The membrane was washed again for 3x10min in PBS-T, followed by a final wash for 5 minutes in 1X PBS to remove traces of Tween20. The signal was detected using the Odyssey® Infrared Imaging System (LI-COR®).

2.11 RNA extraction and cDNA synthesis

Isolation of total RNA from HepG2 cells

Total RNA was extracted from HepG2 cells using the TRIzol method with the TriPure Isolation Reagent (Roche). Briefly, the media was removed and cells were immediately lysed with 0.5mL of TriPure reagent per well for a 12-well plate. The lysates were transferred to 1.5mL microcentrifuge tubes and incubated for 5 minutes at room temperature. Next, 0.1mL of chloroform was added to each sample, followed by vigorous shaking for 15 seconds and a room temperature incubation for 10-15 minutes. The samples were centrifuged at 12,000rpm for 15 minutes at 4°C, and the aqueous phases (top layer) were transferred to clean, RNase-free 1.5mL microcentrifuge tubes. 0.25mL of isopropanol was mixed with each sample and the sample was incubated for 10-15 minutes at room temperature, followed by centrifugation at 12,000rpm for 8 minutes at 4°C. The supernatant was discarded, the RNA pellet was washed with 0.5mL of 75% ethanol, and the sample was centrifuged at 8000rpm for 5 minutes at 4°C. The supernatants were discarded, and the pellet was allowed to air-dry for 5-10 minutes. The pellets were dissolved in RNase-free water by a 5 minute incubation at 50°C and RNA concentrations were determined. For most applications, the RNA was ready at this stage for cDNA synthesis.

However, RNA isolated from HepG2 cells after pCMV5-EST overexpression required DNAseI digestion to remove the contaminating plasmid DNA. Each 100μL DNAseI digestion reaction contained HepG2 RNA contaminated with plasmid DNA, 1X DNaseI incubation buffer (Roche), 20U DNaseI recombinant (10U/μL, Roche), 10U Protector RNase inhibitor (40U/μL, Roche), and 20U EcoRI (20U/μL, NEB). The digestion reactions were

incubated at 37°C for 30 minutes, after which the RNA was purified and concentrated using the RNeasy MinElute Cleanup Kit (Qiagen). Final RNA concentrations were determined.

Isolation of RNA from whole blood

RNA was isolated from whole blood of Ottawa Heart Study Control subjects using the PAXgene® Blood RNA Kit (PreAnalytiX). The isolated RNA is further purified and concentrated using the RNeasy MinElute Cleanup Kit (Qiagen). Final RNA concentrations were determined.

Reverse transcription

1μg of RNA (HepG2 and blood) was used for cDNA synthesis (Transcriptor First Strand cDNA Synthesis Kit, Roche). Reverse transcription with this kit is a two step process. The first step includes incubating the RNA with oligo-dT and random hexamer primers for a final volume of 13μL in a thin-walled 0.2mL PCR tube. The second step includes the addition of cDNA synthesis buffer, dNTPs, Protector RNase inhibitor, and Transcriptor Reverse Transcriptase, for a final reaction volume of 20μL as outlined in the instruction manual. The cDNA synthesis reaction was run on the ABI GeneAmp® PCR System 9700 using the following conditions: a 10 minute incubation at 65°C after the addition of the first components, and incubations of 10 minutes at 25°C followed by 60 minutes at 50°C after the addition of the remaining components in the second step.

2.12 Mapping the EST by 5'/3' RACE

3' RACE

Total RNA was isolated from HepG2 cells using the TRIzol method outlined previously. 3' RACE was performed using the 5'/3' RACE Kit, 2nd Generation (Roche). All primer sequences are listed in Appendix A. For first-strand cDNA synthesis, each 20μL reaction was prepared in a thin-walled 0.2mL PCR tube and contained 1X cDNA synthesis buffer, 1mM of each dNTP, 20U Protector RNase Inhibitor (Roche), oligo dT-anchor primer at

a concentration of 1.875μM, 1μg of HepG2 total RNA, and 25U of Transcriptor Reverse Transcriptase. The cDNA synthesis reaction was run on the ABI GeneAmp® PCR System 9700 using the following conditions: 55°C for 60 minutes followed by 5 minutes at 85°C.

EST transcript amplification from 3' RACE cDNA required two rounds of PCR using the Roche FastStart Taq DNA Polymerase, dNTPack. For the first round of PCR, each 50μL reaction contained a gene-specific forward primer against exons 1 or 2 at a concentration of 0.25μM, a modification of the PCR anchor primer at a concentration of 0.25μM (PCR anchor primer supplied with the Roche 5'/3' RACE Kit; modified primer to decrease the T_m for compatibility with gene-specific primers), 1X PCR buffer with 2mM $MgCl_2$ and 1X GC-rich solution, 0.2mM of each dNTP, 1μL of undiluted 3' RACE cDNA template, and 2.5U of FastStart Taq DNA polymerase (Roche).

The first round PCR products were diluted 1:20 for the next round of PCR. The 50μL reactions for the second round of PCR contained the same reagents used for the first round with the following exceptions: 1μL of diluted first round PCR product as template, and, in all cases, the forward primer was a nested gene-specific primer at a concentration of 0.25μM.

All PCR reactions for 3' RACE were run on the ABI GeneAmp® PCR System 9700 using the following conditions: (1) an initial denaturation for 2 minutes at 95°C; (2) 10 cycles of (i) 95°C for 15sec, 57°C for 30sec, 72°C for 40sec; (3) 25 cycles of (i) 95°C for 15sec, 57°C for 30sec, 72°C for 40sec (+20sec/cycle for the 25 cycles); (3) a final elongation at 72°C for 7 minutes.

5' RACE

Total RNA was isolated from HepG2 cells using the TRIzol method outlined previously. 5' RACE was performed using the 5'/3' RACE Kit, 2nd Generation (Roche). All primer sequences are listed in Appendix A. For 5' RACE, the cDNA synthesis reaction was set up in 0.2mL thin-walled PCR tubes. Each 20μL reaction contained 1X cDNA synthesis buffer, 1mM of each dNTP, 20U Protector RNase Inhibitor (Roche), a gene-specific reverse primer

(SP1) at a concentration of 0.625μM, 1μg of HepG2 total RNA, and 25U of Transcriptor Reverse Transcriptase. The cDNA synthesis reaction was run on the ABI GeneAmp® PCR System 9700 using the following conditions: 55°C for 60 minutes followed by 5 minutes at 85°C. The cDNA was purified using reagents from the High Pure PCR Purification Kit (Roche) but according to instructions from the 5'/3' RACE Kit, 2nd Generation (Roche). The purified cDNA was used directly for the two-step poly(A)-tailing reaction. The 25μL poly(A)-tailing reaction was set up in a thin-walled 0.2mL PCR tube and contained purified 5' RACE cDNA, 1X reaction buffer (supplied with the RACE Kit), 0.2mM dATP, and 80U of Terminal Transferase. The poly(A)-tailing reaction was run on the ABI GeneAmp® PCR System 9700. In the first step, the cDNA, reaction buffer, and dATP were mixed together and incubated at 94°C for 3 minutes. The tube was chilled on ice, 80U of Terminal Transferase was added, and the reaction was incubated at 37°C for 30 minutes followed by a 10 minute incubation at 70°C to heat inactivate the enzyme.

EST transcript amplification from 5' RACE dA-tailed cDNA required two rounds of PCR using the Roche FastStart Taq DNA Polymerase, dNTPack. For the first round of PCR, each 50μL reaction contained 1X PCR buffer with 2mM $MgCl_2$, 0.2mM of each dNTP, oligo dT-anchor primer at a concentration 0.75μM, a nested gene-specific reverse primer at a concentration of 0.25μM, dA-tailed cDNA for template, and 2.5U of FastStart Taq DNA Polymerase. The PCR product was diluted 1:20 to be used as a template for the second round of PCR. The second PCR reaction (50μL) contained 1X PCR buffer with 2mM $MgCl_2$, 0.2mM of each dNTP, a nested forward primer at a concentration of 0.25μM (modification of the PCR anchor primer supplied with the Roche 5'/3' RACE Kit; modified primer to decrease the T_m for compatibility with gene-specific primers), a nested gene-specific reverse primer at a concentration of 0.25μM, diluted PCR product as template, and 2.5U of FastStart Taq DNA Polymerase.

All PCR reactions for 5' RACE were run on the ABI GeneAmp® PCR System 9700 using the following conditions: (1) an initial denaturation for 2 minutes at 95°C; (2) 45 cycles of (i) 95°C for 15sec, 57°C for 1min, 72°C for 1min30sec.

Lastly, gene-specific PCR was used to complement the 5' RACE findings. Several primers were designed upstream from one another to "chase" the 5' end of the EST transcripts. All PCR reactions contained the same gene-specific reverse primer but each had a different gene-specific forward primer. PCR was performed using the Roche FastStart Taq DNA Polymerase, dNTPack, with HepG2 cDNA (First Strand cDNA Synthesis Kit; Roche) as a template. In addition to the cDNA template and primers at a concentration of 0.4μM each, all 25μL PCR reaction contained 1X PCR buffer with 2mM $MgCl_2$, 0.2mM of each dNTP, and 2.5U of FastStart Taq DNA Polymerase. The 5' "chase" only required one round of PCR which was performed with the same equipment and under the same cycling conditions as those detailed for the 5' RACE. PCR products were visualized by 1% agarose gel and were verified by DNA sequencing.

Analysis of RACE products

5'/3' RACE PCR products were first visualized by 1% agarose gel. PCR products were subsequently TOPO®-cloned into the pCR®2.1-TOPO® vector as described previously. TOPO®-TA clones were verified for the presence of an insert by restriction digest, and all inserts were sequenced to map the 5' and 3' ends of EST transcripts, and to map the various transcript variants that were amplified.

2.13 RNA Expression Analysis by Quantitative Real-Time RT-PCR

After RNA extraction from HepG2 cells and whole blood and reverse transcription detailed above, the cDNA was used as a template for qRT-PCR. Primers for qRT-PCR were designed to generate amplicons ranging from 150-300bp in length (Appendix A). Each 20μL PCR reaction was set up in white 96-well plates and contained 5μL cDNA at a 1:5 dilution, forward and reverse primers at 0.5μM each, and 1X SYBR Green I Master (Roche). Quantitative PCR was performed using the LightCycler® 480 SYBR Green I Master(Roche) system with the LightCycler 480II instrument (Roche). All quantitative PCR experiments

included these four steps: (i) pre-incubation at 95°C for 10 minutes (ii) amplification and quantification for 45 cycles (iii) melting and (iv) cooling at 40°C for 30 seconds.

The conditions for the 45-cycle amplification of *EST, TRIB1* and *PPIA* mRNA were as follows: 95°C for 10sec, annealing at 60°C (touchdown program; annealing temperature changes by -0.5°C/cycle) for 30sec, an extension at 72°C for 1min, and a signal acquisition at 82°C for 5sec. All other genes were amplified for 45 cycles using the following conditions: 95°C for 10sec, annealing at 65°C (touchdown program; annealing temperature changes by -0.5°C/cycle) for 30sec, and extension with signal acquisition at 72°C for 20sec. The one exception to these conditions is *ZNF572*, whose signal was acquired at 82°C.

Upon completion of 45 cycles, all PCR products were melted at the following conditions: 95°C for 10sec, 40°C for 1min, 60°C for 10sec, and from there, melting was achieved by increasing the temperature continuously until it reached 95°C. A ramp rate of 0.11°C/sec was used, and signals were acquired at 5 acquisitions/°C.

PCR products were initially verified by melting curves, visualization on 1% agarose gel, and DNA sequencing to ensure specific amplification of the correct transcript. After initial verification, each PCR product was verified by its signature melting curve. Relative copy numbers of each target transcript were normalized to the expression levels of the abundantly expressed house-keeping gene cyclophilin A (PPIA).

2.14 RNA Immunoprecipitation

This RNA immunoprecipitation protocol was adapted from Rinn *et al.*, *Cell*, (2007) [113]. 6x10cm dishes of HepG2 cells were grown to confluency as described in *Cell Maintenance*. The cells were trypsinized at confluency and resuspended with 1X PBS. 2x10cm dishes were pooled into one 15mL Eppendorf tube and the tubes were centrifuged at 1000rpm for 5 minutes to pellet the cells. The PBS was discarded and each cell pellet was resuspended in 2mL 1X PBS, 2mL Nuclear Isolation Buffer (1.28M sucrose; 40mM Tris-HCl pH 7.4; 20mM $MgCl_2$; 4% Triton X-100) and 6mL of water on ice for 20 minutes with frequent

mixing. The nuclei were pelleted by centrifugation at 2500xg for 15 minutes at 4°C. Each nuclear pellet was resuspended in 1mL RIP Buffer (150mM KCl; 25mM Tris-HCl pH 7.4, 5mM EDTA, 0.5mM DTT, 0.5% NP-40, 1X Complete Protease Inhibitor Cocktail Tablet (Roche), 100U Protector RNase Inhibitor (Roche)). The resuspended nuclear pellets were pooled together, mixed, and split into 500μL fractions. The nuclei were sheared using a dounce homogenizer with 15-20 strokes. Nuclear membrane and debris were pelleted by centrifugation at 13,000rpm for 10 minutes at 4°C. An antibody was added to each supernatant (6μg for Suz12/Ezh2/H3K27/CoREST (Abcam); 10μg rabbit IgG (Abcam)) and incubated for two hours at 4°C with gentle rotation. Next, 40μL of protein A/G beads (Roche) were added to each supernatant and incubated for 1 hour at 4°C with gentle rotation. The beads were pelleted by centrifugation at 2500rpm for 30 seconds at 4°C, the supernatant was discarded, and the beads were resuspended in 500μL RIP. These steps were repeated for a total of three RIP washes followed by one wash in 1X PBS. The washed beads were resuspended in 1mL of TriPure Isolation Reagent (Roche) and co-precipitated RNA was isolated using the TRIzol method. The isolated RNA was quantified and equal amounts were used for cDNA synthesis (Transcriptor First Strand cDNA Synthesis Kit, Roche). Quantitative RT-PCR (LightCycler480 SYBR Green I Master) was used to detect the EST transcripts. Relative copy numbers of the EST were normalized to the presence of U1 RNA.

3 Results

3.1 Phylogenetic Footprinting Analysis Identifies SNPs Within Intergenic Evolutionarily Conserved Regions

Although Trib1 is a known regulator of MAPK activity [86, 91] and hepatic lipogenesis [98], the functional relationship between SNPs at the *TRIB1* locus and plasma lipid traits remains unknown. Moreover, HapMap data indicates that these common SNPs are not in LD with SNPs within the *TRIB1* coding region. Since the cluster of TG-associated SNPs identified by GWAS is located within an intergenic region between 25-50kb downstream of the *TRIB1* coding region and are not in LD with SNPs within the coding region, we performed phylogenetic footprinting analysis using PipMaker to identify any evolutionarily conserved sequences - potential regulatory sequences - within or nearby this risk locus. This approach identified two evolutionarily conserved regions - CNS1 and CNS2 - located 28kb and 53kb downstream of the *TRIB1* coding region, respectively, as shown in figure 1. CNS1 is 284bp in length, and is located at Chr8: 126478696 - 126478979. CNS2 is 350bp in length and spans Chr8: 126503634 - 126503983.

The PipMaker output represents the orthologous regions in the other species as percent sequence identity to the human sequence. The *TRIB1* coding region - particularly exons - is highly conserved between species. Yet, the downstream intergenic region, excluding CNS1/2, has very poor sequence conservation. Moreover, none of the TG-associated SNPs identified from GWAS were found within conserved regions. SNPs within CNS1 and CNS2 were identified using the Ensembl Genome Browser. CNS1 contains three SNPs: rs2001844 (126478745), rs7015677 (126478758), and rs57863956 (126478790). A fourth SNP - rs2001845 - is located just upstream of CNS1 as position Chr8:126478644. CNS2 contains one SNP - rs55921265 - located at position Chr8:126503915.

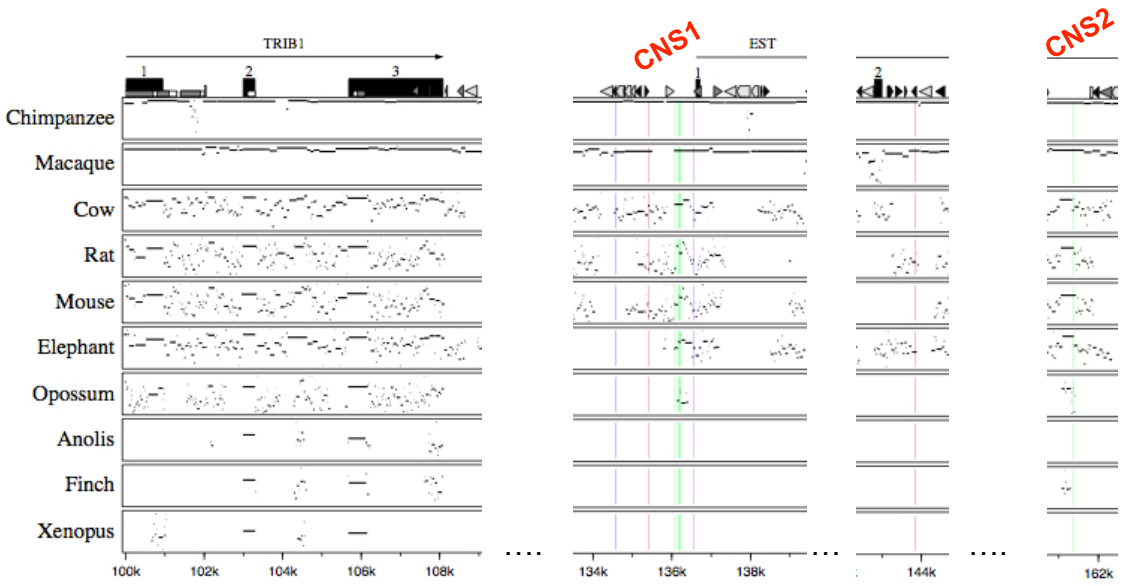

TRIB1
1
2
3
CNS1
EST
CNS2
Chimpanzee
Macaque
Cow
Rat
Mouse
Elephant
Opossum
Anolis
Finch
Xenopus
100k
102k
104k
106k
108k
134k
136k
138k
144k
162k

Figure 1: Phylogenetic footprinting analysis of the *TRIB1* locus by PipMaker. The intergenic region downstream of the *TRIB1* harbours two evolutionarily conserved regions, CNS1 and CNS2. GWAS TG-associated SNPs (blue and red) fall outside the conserved regions. SNPs located within CNS1 and CNS2 were not identified by GWAS (green).

3.2 Minor Allele Frequencies and Odds Ratios for SNPs Within CNS1 and CNS2 Genotyped in CHD Cases and Controls from the Ottawa Heart Study

To determine whether SNPs within evolutionarily conserved regions were in LD with TG-associated GWAS SNPs and/or associated with CAD themselves, these regions were genotyped by DNA sequencing using OHS Caucasian CHD cases and elderly controls. Control subjects included men and women over 65 and 70 years of age, respectively, with no personal or family history of CAD. Cases included men and women with severe premature CHD with documented onset under 60 years of age.

HapMap did not contain any data on SNPs within CNS1 and CNS2 but did have LD data for rs2001845, which is located just upstream of CNS1. HapMap data for this SNP confirmed statistical findings obtained from genotyping. The 8q24.13 GWAS TG-associated SNPs are very common; data from the Ottawa Heart Study shows that SNPs of the first haplotype have minor allele frequencies of 0.48 (eg. rs2001945, rs17321515 [50]), and those of the second haplotype (eg. rs2954018) have minor allele frequencies of 0.30. To determine allele frequencies, LD data, and CAD association for SNPs within regions of evolutionary conservation, approximately 300 CHD cases and controls and 200 CHD cases and controls were genotyped for CNS1 and CNS2, respectively. The results from SNP genotyping are shown in Table 1.

The SNP just upstream from CNS1 - rs2001845 - has a minor allele frequency of 0.28 in controls; the frequency increased to 0.37 in CHD cases (p=0.0062) and the odds ratio (OR; 95% CI) was found to be 1.46. In addition, rs2001845 is in tight LD with GWAS SNP rs2954018 at 8q24.13 (D'=1; r^2=0.99). Although this SNP is located directly upstream from CNS1 and is significantly associated with CAD and in tight LD with one haplotype of GWAS TG-associated SNPs, it is very poorly conserved between species. The first SNP within CNS1 - rs2001844 - is well conserved between species and also associated significantly with CAD (p=0.004). This SNP has a minor allele frequency of 0.48 in controls; the fre-

SNP	Chr8 location	Sample Size	AA	AB	BB	MAF	Allelic p-value	Odds Ratio 95% CI	LD with GWAS SNPs	
									rs2954018	rs2001945
rs2001845	*CNS1*	Cases (n=266)	106	122	38	0.37	0.0062	1.46	D′=1	D′=1
	126478644	Controls (n=271)	145	98	28	0.28		(1.13,1.87)	r^2=0.99	r^2=0.52
rs2001844	*CNS1*	Cases (n=266)	47	125	94	0.59	0.004	1.50	D′=0.99	D′=1
	126478745	Controls (n=271)	74	131	66	0.48		(1.18,1.90)	r^2=0.42	r^2=0.81
rs7015677	*CNS1*	Cases (n=266)	185	74	7	0.16	0.65	1.16	D′=1	D′=1
	126478758	Controls (n=271)	198	67	6	0.14		(0.83,1.62)	r^2=0.09	r^2=0.2
rs57863956	*CNS1*	Cases (n=266)	247	19	0	0.04	0.35	0.73	D′=1	D′=1
	126478790	Controls (n=271)	245	26	0	0.05		(0.39,1.34)	r^2=0.02	r^2=0.04
rs55921265	*CNS2*	Cases (n=172)	108	56	8	0.2	0.25	1.07	D′=1	D′=0.36
	126503915	Controls (n=176)	109	64	3	0.2		(0.73,1.56)	r^2=0.11	r^2=0.04

Table 1: Allele frequencies, CAD association, and LD data for SNPs within evolutionarily conserved regions genotyped in CHD cases and controls from the Ottawa Heart Study Minor allele frequencies (MAF) were compared between CHD cases and controls for each SNP. Statistical analysis was based on a univariate linear regression model; biochemical and physical parameters such as plasma lipid levels and BMI were not considered in the analyses. Allelic p-values were determined using Fisher's Exact Test and Odds Ratios (OR) were determined with 95% confidence intervals (CI). Linkage disequilibrium (LD) between SNPs was determined using R Program 2.10.1. A=major allele; B=minor allele.

quency increased to 0.59 in CHD cases and the OR (95% CI) was found to be 1.50. Further, rs2001844 is in tight LD with the second haplotype of GWAS TG-associated SNPs, such as rs2001945 and rs17321515. D' and r^2 with rs2001945 and rs17321515 were determined as being 1, 0.81 and 1, 0.98, respectively.

The last two SNPs within CNS1 - rs7015677 and rs57863956 - were much less common. Moreover, the minor allele frequencies of these SNPs did not change significantly between CHD cases and controls. Interestingly, not a single homozygote for the rs57863956 minor allele was detected in 300 CHD cases and 300 controls. Finally, rs55921265 within CNS2 has a minor allele frequency of 0.20, which did not change between CHD cases and controls and thus was not significant for CAD. Although the number of subjects genotyped for these SNPs was low, these findings were validated by data from the 1000 Genomes Project [45], which became available only recently.

3.3 The Regulatory Functions of Evolutionarily Conserved Regions CNS1 and CNS2

Gene deserts have been known to harbour regulatory elements, and these are often marked by evolutionary conservation [78, 53, 54]. Consequently, the genomic regions containing CNS1 and CNS2 were initially tested for enhancer activity using luciferase reporter assays (Promega). Although CNS1 harbours four SNPs, only the two common SNPs with significant CAD odds ratios were pursued in functional studies. Briefly, amplicons containing CNS1 and CNS2 were amplified by PCR from subjects homozygous for the reference and risk alleles, respectively. The PCR products were subcloned into the pGL4.23 luciferase reporter vector (Promega). This vector is ideally suited for enhancer studies because it contains only a minimal promoter and gives low background. Variations in transfection efficiency between samples was controlled for by co-transfection of the pGL4.74 *Renilla* luciferase vector (Promega). Figure 2 shows that CNS1 does not have enhancer activity and that the presence of reference or risk alleles for rs2001845 and rs2001844 do not alter CNS1 activity. In contrast, the presence of CNS2 upstream from the minimal promoter reduced luciferase reporter activity, suggesting that CNS2 may be a repressor. These studies were first performed in Cos-7 and these findings were also validated in HepG2 cells.

To further investigate the repressor activity of CNS2, the insert containing CNS2 was subcloned from pGL4.23 into the pGL3-SV40 luciferase reporter vector (Promega), which has a strong promoter. Since the pGL3-SV40 promoter will give stronger luciferase activity, the pRL-SV40 *Renilla* luciferase was used as the appropriate internal control for transfection efficiency. Figure 3 shows that CNS2 functions as a repressor in (a) Cos-7 and (b) HepG2 cells. Interestingly, the rs55921265 'A' allele further reduced luciferase activity in HepG2 cells ($p<0.01$), but not in Cos-7 cells. CNS2 was also a stronger repressor in HepG2 cells as compared to Cos-7 cells.

Since CNS1 showed neither enhancer nor repressor activity, this region was assayed for

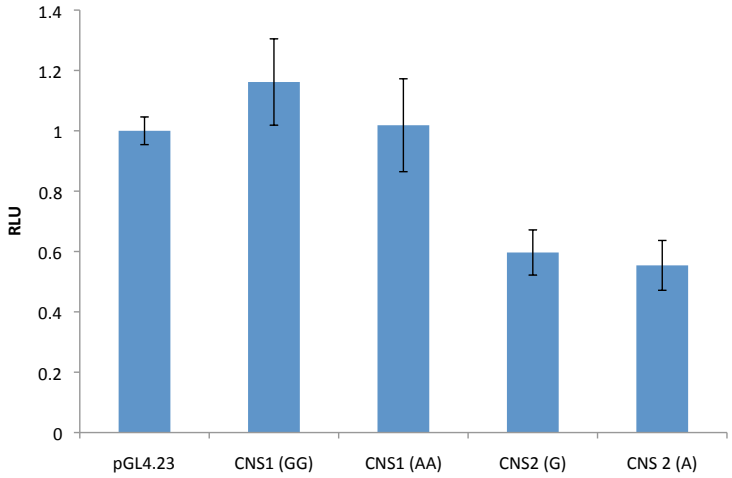

1.4
1.2
1
0.8
0.6
0.4
0.2
0
RLU
pGL4.23
CNS1 (GG)
CNS1 (AA)
CNS2 (G)
CNS 2 (A)

Figure 2: Determining enhancer activity in CNS1/CNS2 Cos-7 cells were transfected with empty pGL4.23 minimal promoter vector or the indicated CNS1/CNS2-pGL4.23 luciferase reporter constructs, and pGL4.74 *Renilla* luciferase. n=6 independent experiments.

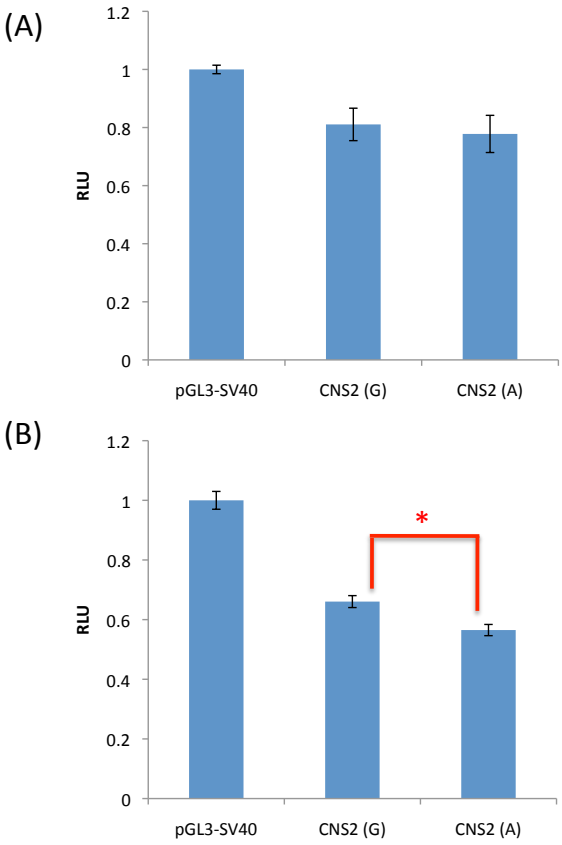
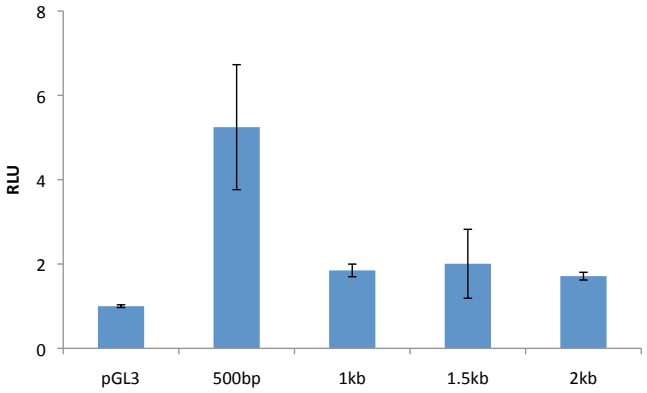

(A)
1.2
1
0.8
0.6
0.4
0.2
0
RLU
pGL3-SV40
CNS2 (G)
CNS2 (A)
(B)
1.2
1
0.8
0.6
0.4
0.2
0
RLU
*
pGL3-SV40
CNS2 (G)
CNS2 (A)

Figure 3: Determining repressor activity in CNS2 (A) Cos-7 and (B) HepG2 cells were transfected with empty pGL3-SV40 promoter vector or the indicated CNS2-pGL3-SV40 luciferase reporter constructs, and pRL-SV40 *Renilla* luciferase. n=4 independent experiments. (*$p<0.01$)

promoter activity. All luciferase reporter constructs were generated by PCR using genomic DNA from a individual who was homozygous reference for GWAS TG-associated 8q24.13 SNPs and for CNS1 SNPs. The largest region analyzed for promoter activity spanned 2kb in size. The remaining regions were 500bp deletion derivates of the 2kb region, generating inserts of 1.5kb, 1kb, and 500bp, respectively. All inserts include CNS1 at their 3'; sequences extend upstream of CNS1. The smallest insert - 500bp - contains the entire CNS1 region (approximately 300bp in size) with an additional 100bp flanking either end. The rs2001845 SNP lies just upstream of the region of conservation and thus falls within the 100bp flanking sequence included in the 500bp luciferase reporter construct. The other conserved common SNP - rs2001844 - falls within the conserved sequence. These regions were inserted into the pGL3-basic luciferase reporter vector (Promega). This vector is ideally suited for promoter studies because it itself is promoter-less. Thus, observed luciferase activity above background levels indicates that the sequence inserted upstream from the *luciferase* transcription start site has promoter activity itself. Figure 4 clearly shows that the 500bp region - CNS1 with minimal additional sequence - has promoter activity in both (a) Cos-7 and (b) HepG2 cells. Interestingly, although all inserts contain CNS1, the larger regions are unable to activate luciferase expression. Since promoted activity was abolished by the addition of 500bp directly upstream of CNS1 (1kb construct), this effect may be attributed to sequences within this region that are targeted by factors that repress transcription.

(A)


8
6
4
2
0
RLU
pGL3
500bp
1kb
1.5kb
2kb

(B)

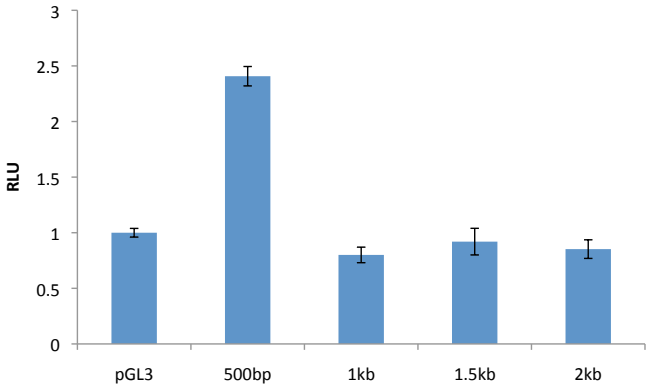

3
2.5
2
1.5
1
0.5
0
RLU
pGL3
500bp
1kb
1.5kb
2kb

Figure 4: Determining promoter activity in CNS1 (A) Cos-7 and (B) HepG2 cells were transfected with empty pGL3-basic promoter-less vector or the indicated CNS1-pGL3-luciferase reporter constructs, and pRL-SV40 *Renilla* luciferase. n=2 independent experiments.

3.4 Risk Alleles of TG-associated SNPs do not Alter *TRIB1* mRNA Expression in Whole Blood

Since GWAS TG-associated risk SNPs are located within an intergenic region and the nearest protein-coding gene is *TRIB1*, we investigated the possibility that the risk allele of rs17321515 - a GWAS SNP in tight LD with rs2001844 - alters *TRIB1* mRNA expression in whole blood from Ottawa Heart Study healthy elderly controls. Control subjects were carefully matched for rs17321515 genotype, age, gender, and BMI and the group characteristics are summarized in Table 2. Control subjects used in this study were also not on any type of lipid-altering medication.

The blood from OHS controls for RNA analysis was stored in PAXgene® tubes and isolated using the PAXgene® Blood RNA Kit (PreAnalytiX). The RNA was further purified and concentrated using the RNeasy MinElute Cleanup Kit (Qiagen), and 1μg was used for cDNA synthesis (Transcriptor First Strand cDNA Synthesis Kit, Roche). Quantitative PCR for the expression of *TRIB1* was performed using the LightCycler® 480 SYBR Green I Master (Roche) system and melting peaks were verified to ensure specific amplification of the correct product. Relative copy numbers of *TRIB1* transcripts were normalized to the levels of constitutively-expressed house-keeping gene *PPIA* and the results are shown in figure 5. *TRIB1* expression was highly variable between subjects within and among the genotype categories. Consequently, there was no association between GWAS SNP risk alleles and *TRIB1* expression at the mRNA level. Further, the GWAS from Global Lipids Genetics Consortium failed to identify *TRIB1* as an eQTL with their large sample size of various tissues: 960 liver samples, 741 omental fat samples, and 609 subcutaneous fat samples [56].

Trait	Homozygous for rs17321515 Reference Allele (A)	Heterozygous for rs17321515	Homozygous for rs17321515 Risk Allele (B)
Number of subjects	11	10	10
Age, years	75 +/- 3.69	73.9 +/- 2.68	76 +/- 3.23
Gender	6M / 5F	5M / 5F	5M / 5F
BMI , kg/m2	26.02 +/- 1.79	25.86 +/- 1.66	26.91 +/- 1.67
Cholesterol, mmol/L	5.54 +/- 0.56	5.77 +/- 1.03	5.31 +/- 0.74
Triglyceride, mmol/L	0.98 +/- 0.32	1.03 +/- 0.45	1.22 +/- 0.46
LDL-c, mmol/L	3.60 +/- 0.40	3.65 +/- 0.72	3.50 +/- 0.67
HDL-c, mmol/L	1.50 +/- 0.34	1.64 +/- 0.44	1.25 +/- 0.20
Glucose, mmol/L	5.24 +/- 0.36	5.45 +/- 0.69	5.45 +/- 0.32

Table 2: Clinical characteristics of Ottawa Heart Study Controls for RNA expression analysis Whole blood RNA was isolated from a total of 31 carefully matched healthy elderly controls for qRT-PCR. Clinical characteristics are provided as Mean +/- SD.

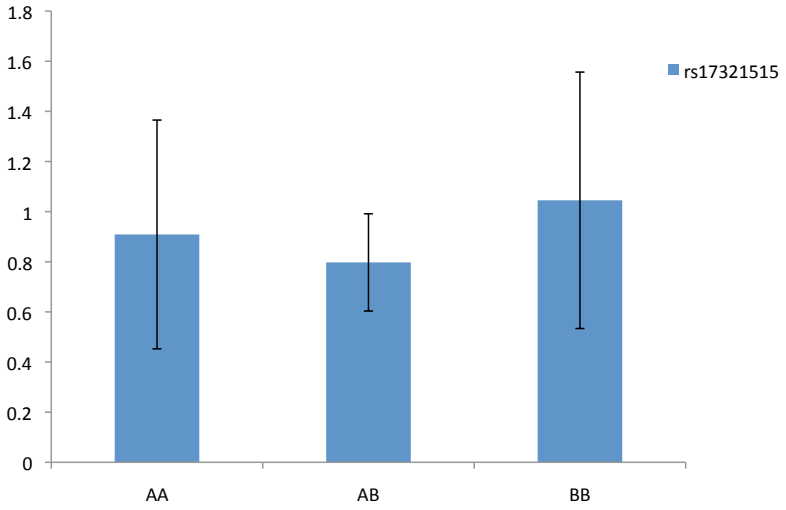

1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
AA
AB
BB
rs17321515

Figure 5: Effect of risk alleles of GWAS-identified SNPs on *TRIB1* mRNA expression in whole blood from Ottawa Heart Study Controls Whole blood RNA was isolated from a total of 31 carefully matched healthy elderly controls for qRT-PCR. Quantitative PCR for the expression of *TRIB1* was performed using the LightCycler® 480 SYBR Green I Master (Roche) system and melting peaks were verified to ensure specific amplification of the correct product. Relative copy numbers of *TRIB1* transcripts were normalized to the levels of constitutively-expressed house-keeping gene *PPIA*.

3.5 A Search of the Risk Locus for Novel Genes Identifies an EST-Based Gene

The luciferase reporter assays demonstrating that CNS1 has promoter activity also suggest that the intergenic risk locus may harbour a novel gene whose transcription start site would be located nearby the CNS1 region. A database search for novel genes within the risk locus was accomplished using the Ensembl Genome Browser, which organizes information from various databases according to chromosomal position. An expressed sequence tag (EST) - cloned from human lens cDNA - from the human EST database aligned to the risk locus very near CNS1. A BLAST search with this EST sequence in OpenBiosystems identified a construct for purchase: clone fs35c04 human lens cDNA. The construct was fully sequenced, and the exons are shown with respect to chromosome position in figure 6. The complete sequence of the clone identified Exon1 as beginning just 118bp downstream of CNS1. Further, this EST spans a very large genomic distance greater than 120kb from the presence of two long introns. CNS2 is also located within an intron of this EST.

Next, primers were designed for various lens EST exons to determine its expression in relevant tissues: HepG2 for functional analyses and whole blood for RNA expression analysis. Figure 7 shows the chromosomal position of EST exons 1 and 2 relative to CNS1, as well as the PCR products obtained from HepG2 and blood cDNA. Interestingly, only exons 1 and 2 were detected in these two tissues; the 274bp PCR product was verified by DNA sequencing. Moreover, this transcript is expressed at a much lower level in whole blood as compared to HepG2 cells despite attempting to optimize PCR conditions (figure 7). This finding was further validated by qRT-PCR, thus, we were unable to determine if risk alleles of GWAS SNPs alter its expression in whole blood from Ottawa Heart Study controls.

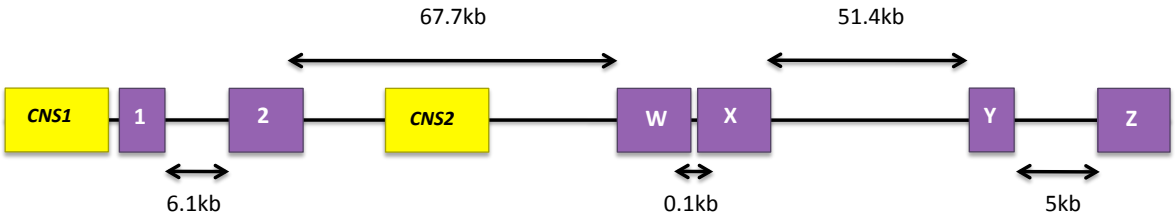
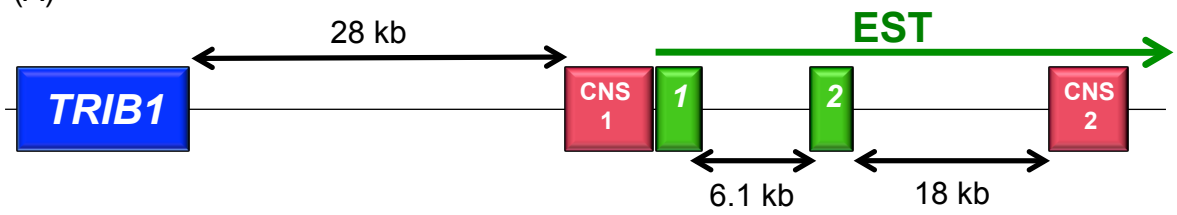

67.7kb
51.4kb
CNS1
1
2
CNS2
W
X
Y
Z
6.1kb
0.1kb
5kb

Figure 6: Human lens cDNA clone mapped to the 8q24.13 risk locus The human lens cDNA clone fs35c04 was purchased from OpenBiosystems and sequenced to obtain a full map of the exons. Exon1 begins 118bp downstream of CNS1. The EST spans a large genomic distance due to very long introns of 68kb and 51kb.

(A)


28 kb
EST
TRIB1
CNS 1
1
2
CNS 2
6.1 kb
18 kb

(B)

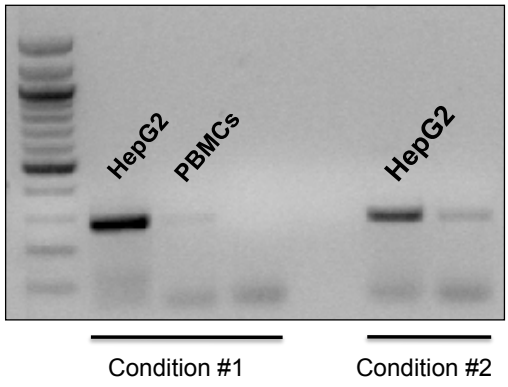

HepG2
PBMCs
HepG2
Condition #1
Condition #2

Figure 7: Human lens EST exons1,2 are expressed in HepG2 cells and whole blood The human lens cDNA clone fs35c04 was purchased from OpenBiosystems and sequenced to obtain a full map of the exons. Exon1 begins 118bp downstream of CNS1. The EST spans a large genomic distance due to very long introns of 68kb and 51kb.

3.6 Mapping the EST in HepG2 Cells by 5'/3' RACE

The observation that only exons 1 and 2 were detected in HepG2 cells and whole blood suggests that there exist tissue-specific EST transcript variants. To address this possibility, 5'/3' RACE (Roche) was performed with HepG2 cDNA from total RNA, using gene-specific primers for exons 1 and 2. Figure 8 is a result from 3' RACE. Lanes 1 to 3 are positive controls for the various steps, and lanes 4 to 13 are 3'RACE products. Two rounds of PCR were required for amplification of EST transcripts from 3'RACE. The second round of PCR used a first round PCR product as the template, the same reverse primer, but a nested gene-specific forward primer. Lanes 4-8 are products from the same template but amplified with different nested forward primers, resulting in PCR products that are gradually smaller in size; specifically, their size difference corresponds to the distance between their respective nested forward primers. Lanes 9-10 and 11-13 were amplified using the same strategy. The pool of PCR products was TOPO®-cloned and positive clones were screened by DNA sequencing. Figure 9 is a summary of the EST transcript variants detected thus far by 3'RACE in HepG2 cells and Table 3 summarizes all exon lengths, chromosomal positions and total lengths of the EST transcripts shown in figure 9. EST transcripts range from 700bp to 1400bp in size. Interestingly, all variants lack significant open reading frame (ORFs), suggesting that they do not code for protein. Further, UCSC Genome Browser shows that exon 4 and the last 200bp of exon 5 contain many repetitive elements a BLAST search with Ensembl shows that they align with several other regions of the genome. All other exons appear to be unique to this EST.

A combination of gene-specific PCR and 5' RACE was used to map the 5' end of EST transcripts. Results from both these approaches are detailed in figure 10. For PCR 5' chase, all PCR reactions used the same reverse primer on exon 2 (figure 10a), while exon 1 primers were designed upstream from one another. This approach continued until PCR products were no longer detected. Originally, exon 1 of the human lens cDNA EST was 95bp in

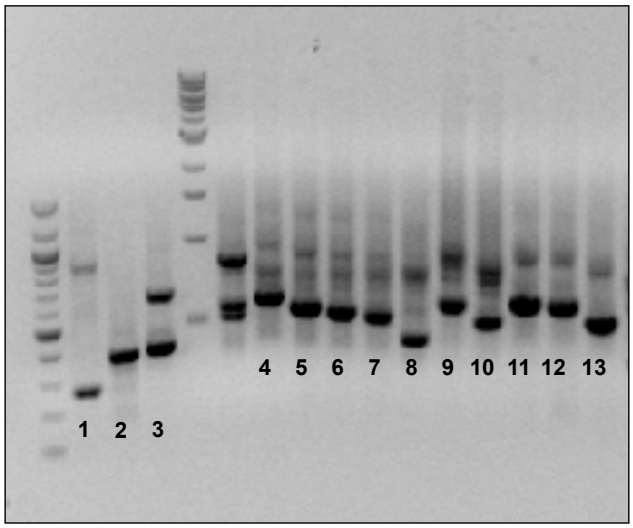

1
2
3
4
5
6
7
8
9
10
11
12
13

Figure 8: Multiple EST variants identified from 3' RACE in HepG2 cells. 3'RACE PCR products visualized on 1% agarose gel. Lanes 1-3 are controls for the various steps in the 3'RACE protocol. The pool of products in lanes 4-13 are EST variants in HepG2 cells identified by 3'RACE.

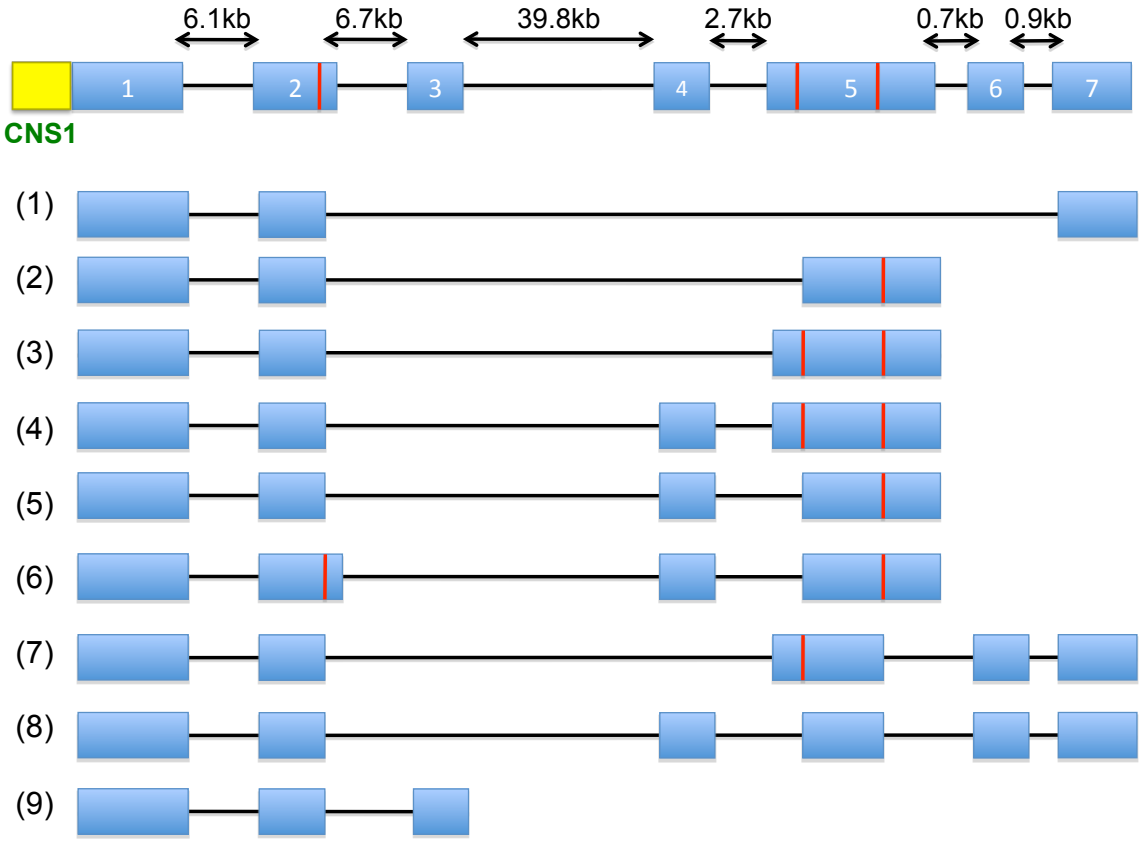

6.1kb
6.7kb
39.8kb
2.7kb
0.7kb
0.9kb
1
2
3
4
5
6
7
CNS1
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)

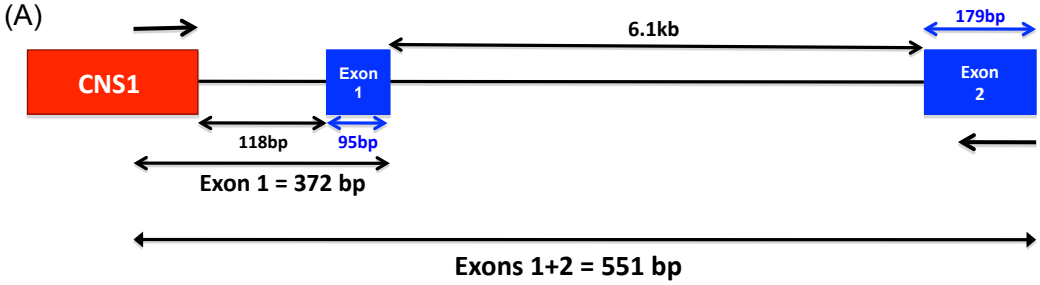
Figure 9: Map of EST transcript variants identified in HepG2 cells by 3'RACE 3'RACE of the EST in HepG2 cells has identified 9 transcript variants to date that range in size from approximately 700bp to 1300bp.

EST transcript	Exons	Chr8 location	Exon Length (bp)	Total Size (bp)
1	1	126478904 – 126479275	372	772
	2	126485378 – 126485556	179	
	7	126537340 – 126537560	221	
2	1	126478904 – 126479275	372	1162
	2	126485378 – 126485556	179	
	5	126534985 – 126535595	611	
3	1	126478904 – 126479275	372	1218
	2	126485378 – 126485556	179	
	5	126534929 – 126535595	667	
4	1	126478904 – 126479275	372	1341
	2	126485378 – 126485556	179	
	4	126532149 – 126532271	123	
	5	126534929 – 126535595	667	
5	1	126478904 – 126479275	372	1285
	2	126485378 – 126485556	179	
	4	126532149 – 126532271	123	
	5	126534985 – 126535595	611	
6	1	126478904 – 126479275	372	1338
	2	126485378 – 126485609	232	
	4	126532149 – 126532271	123	
	5	126534985 – 126535595	611	
7	1	126478904 – 126479275	372	1204
	2	126485378 – 126485556	179	
	5	126534929 – 126535255	327	
	6	126536330 – 126536434	105	
	7	126537340 – 126537560	221	
8	1	126478904 – 126479275	372	1271
	2	126485378 – 126485556	179	
	4	126532149 – 126532271	123	
	5	126534985 – 126535255	271	
	6	126536330 – 126536434	105	
	7	126537340 – 126537560	221	
9	1	126478904 – 126479275	372	681
	2	126485378 – 126485556	179	
	3	126492248 – 126492377	130	

Table 3: Exons, chromosome positions, and lengths of EST transcript variants identified in HepG2 cells by 5'/3' RACE

length and located 118bp away from CNS1. The PCR approach used to chase the 5' end determined that exon 1 in HepG2 cells extended into CNS1 (figure 10a). These findings were validated by 5' RACE. Several 5' RACE products were screened after TOPO®-cloning; all 5' ends extended into CNS1 - which is depicted in red - and not a single product extended beyond CNS1 - depicted in black. Hence, the largest product that was screened indicates that exon 1 in HepG2 cells is, at maximum, 372bp long and is represented in the sequence (figure 10b) in red, bold, italics. Further, 5' RACE results suggested that this EST has two possible transcription start sites, highlighted within the nucleotide sequence in green and turquoise (figure 10b). Finally, the observation that the EST does not extend upstream beyond CNS1, and that the 500bp region containing CNS1 exhibited promoter activity in luciferase reporter assays, suggests that CNS1 contains the EST promoter.

(A)



(B)

GGAGACTGAGGCAGGAGAATCGCTGAAGCTGAGGTGTGGCAGCTTCAGCGTTTGGCTCCATTATTCCCT
CAGGCATGAAATCTTCCTCTCTAGCCTTGTTTAAATCTCAA*CACTTCCCTTTCCAAATG*ATTTCTAAGGGGA
AACATAAAAGGAATGTTCCAGAATACTCTGTTGGCCCAGGGGTGATTTGCCAGAGCAGTTTCCTGGATTA
CGGCCTCTTCCCCCATCAGCAAATCATGTGAAACAAACTACTGTCAGCTGTAAACCCTGAATGTGGGAAA
AAGAGCTGACACACAGGGCCCACTTTCG*TAAAGCTGACATTTCTCATGAATAGCGAATTCCTGAGTGTG*
GGCAGTAGACCAAGGCCAGGAATTCTCATTAGCTGGTGATAATCAGGCATGACTTGGACTGTGGAAGG
CAGCTGGGCGGAGAGGTCATAGGTATGGGCAGGGGTGAGCCACAGCTGATAGCTGGGCGCGATCTGC
CTGGATTGTTACTGAAATACTAACATGCTCTTTATGGGTTAGTAACTCCCTTCACCTCCCCAACGGCTCCT
ACCCGCTGGGACCTTCACCTCCCCATGATC .....

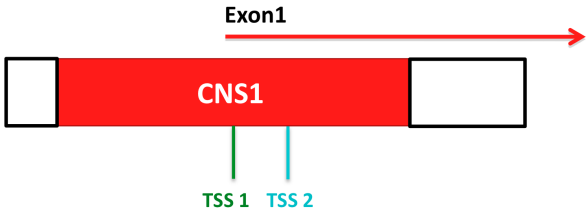


Figure 10: Mapping the EST by 5' RACE in HepG2 cells (A) Gene-specific PCR to chase the 5' end of EST transcripts found that the 5' end extends into CNS1 but not beyond in HepG2 cells. (B) 5' RACE in HepG2 cells extended the 5' end of the EST into CNS1 and confirmed that the 5' end does not extend beyond CNS1 (region shown in red). The conserved region is shown in red. Sequence in black is outside the region of conservation. Underlined sequences indicate primers used for luciferase construct cloning (Appendix A), and the underlined blue sequence corresponds to the 3' end of the largest promoter construct (the original 500bp construct). This approach further confirmed PCR findings that exon 1 of the EST begins - at the earliest - where the nucleotides are depicted in red, bold italics. Green and turquoise highlighting indicates two possible transcription start sites as identified by 5' RACE.

3.7 The Effect of the rs2001844 Risk Allele on EST Promoter Activity

The initial 500bp CNS1-containing sequence that demonstrated promoter activity in luciferase reporter assays was further refined from both the 5' and 3' ends to map the minimal promoter sequence. Briefly, all sequence outside of the region of conservation was removed from the 5' end. Since EST exon 1 has been mapped to the first transcription start site, a luciferase reporter construct was generated by truncating the 3' end to remove all exon 1 sequence. Figure 11 demonstrates that the 5' end truncation to remove the non-conserved sequences retains its promoter activity (construct 1). However, removal of all EST exon1 from the construct completely abolished promoter activity (construct 2). Two additional constructs were generated to address the observation that the EST promoter contains two transcription start sites. Construct 3 was generated by a 3' truncation that removed the second start site, leaving only the first one. As a result, construct 3 contains an additional 40bp at its 3' end in comparison to the non-active construct 2. Construct 4 contains both identified transcription start sites. The first transcription start site (construct 3) was sufficient for promoter activity. Thus, those 40bp around the first start site are absolutely required for EST promoter activity. Further, the presence of both start sites significantly increased promoter activity. Extending the 3' end farther into exon1 - including sequence outside of CNS1 - significantly reduces promoter activity to the level observed with one transcription start site. Moreoever, these EST promoter constructs harbour a common SNP - rs2001844 - which has been shown to be in tight LD with GWAS TG-associated SNPs, and significantly associated with CAD. Both risk (rs2001844 - A allele) and reference (rs2001844 - G allele) EST promoter constructs were generated for constructs 3 and 4 and, in both cases, the rs2001844 risk allele (A) significantly reduced promoter activity as shown in figure 12. Consequently, we anticipate that the expression of this transcript is reduced in carriers of the risk versus reference alleles.

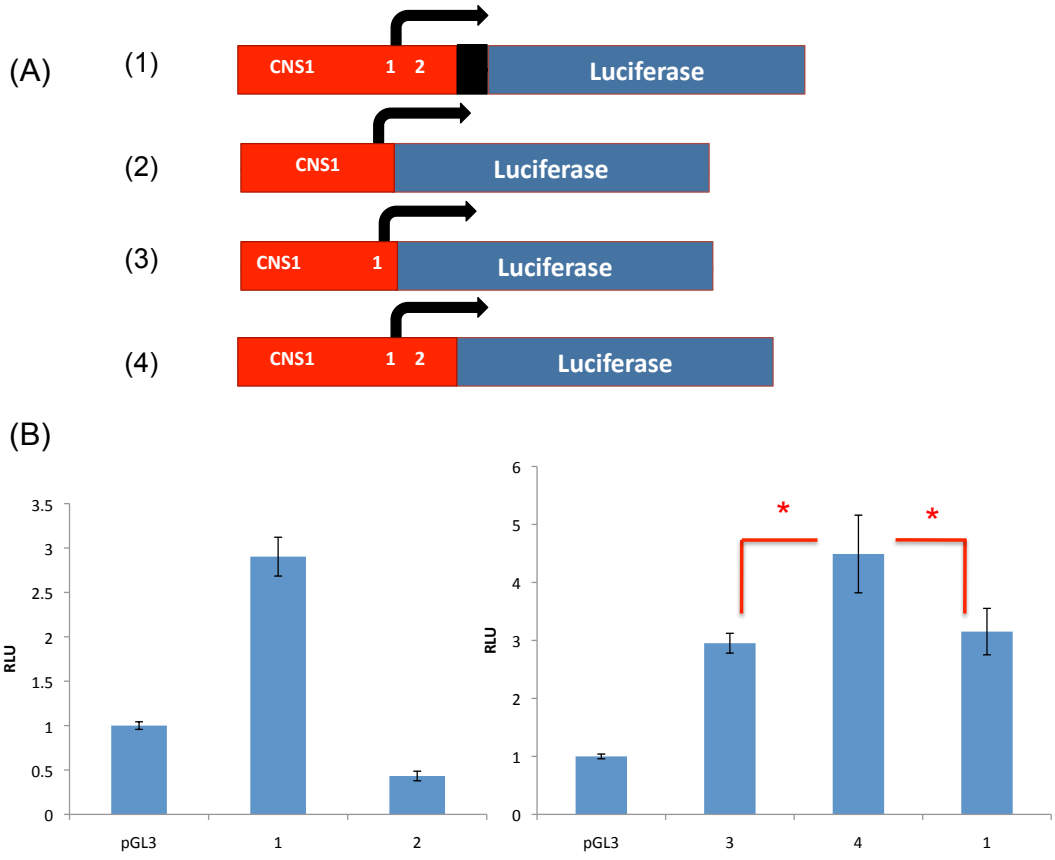
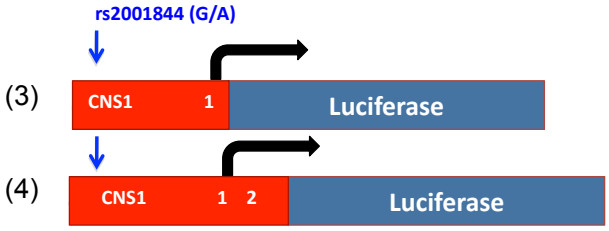

(A)
(1)
CNS1
1
2
Luciferase
(2)
CNS1
Luciferase
(3)
CNS1
1
Luciferase
(4)
CNS1
1
2
Luciferase
(B)
RLU
3.5
3
2.5
2
1.5
1
0.5
0
pGL3
1
2
RLU
6
5
4
3
2
1
0
*
*
pGL3
3
4
1

Figure 11: The EST minimal promoter requires the first 40bp of exon 1, including the first transcription start site. (A) Schematic representation of the CNS1 luciferase reporter constructs. Red indicates sequence within CNS1, black represents sequence outside CNS1. There are two transcription start sites labeled in white as 1 and 2, respectively. (B) HepG2 cells were transfected with pGL3 basic or the indicated luciferase reporter constructs, and pRL-SV40 as an internal control for transfection efficiency. n=2 independent experiments. (*$p<0.01$)


rs2001844 (G/A)
(3)
CNS1
1
Luciferase
(4)
CNS1
1
2
Luciferase

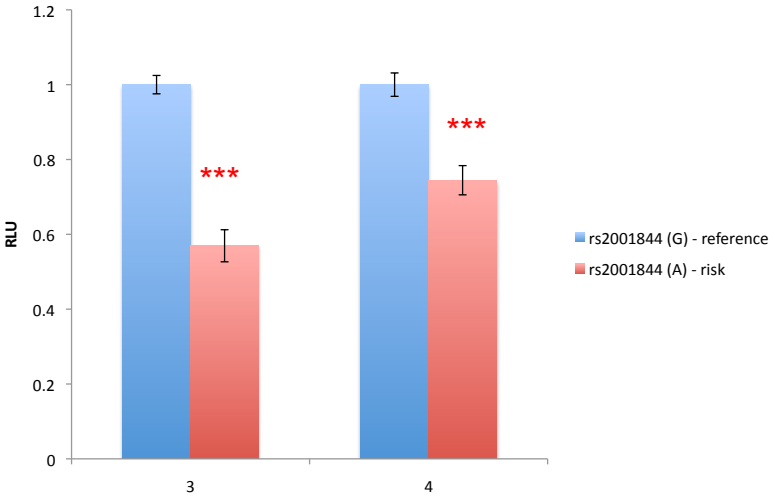
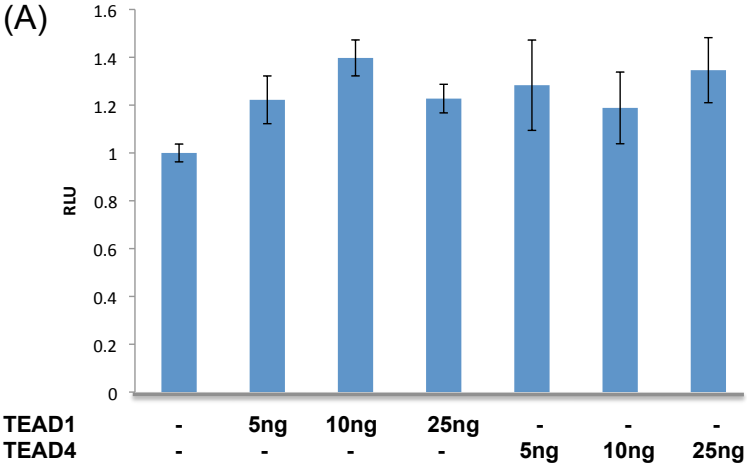

1.2
1
0.8
0.6
0.4
0.2
0
RLU
***
***
3
4
rs2001844 (G) - reference
rs2001844 (A) - risk

Figure 12: Effect of the rs2001844 risk allele on EST promoter activity. (Top) Schematic representation of the CNS1 luciferase reporter constructs. Red indicates sequence within CNS1, black represents sequence outside CNS1. There are two transcription start sites labeled in white as 1 and 2, respectively. (Bottom) HepG2 cells were transfected with pGL3 basic or the indicated luciferase reporter constructs, and pRL-SV40 as an internal control for transfection efficiency. n=4 independent experiments. (***$p<1x10^{-4}$)

3.8 Transcription Factor Binding Sites Predicted Around the rs2001844 SNP

MatInspector and TRANSFAC were used to predict transcription factor binding sites that include the rs2001844 SNP to determine if the risk allele of this SNP alters a binding site, thereby explaining the observed reduced luciferase activity. The transcription factors predicted from these tools include Krueppel-like transcription factors (KLFs), STAT3, STAT1, C/EBPδ, and TEAD1. Expression vectors for TEAD1 and TEAD4, as well as an antibody against TEAD1 were generously given from Dr. Alexandre Stewart's laboratory (University of Ottawa Heart Institute). TEAD1 and TEAD4 are the first transcription factors attempted on the CNS1 promoter. Briefly, HepG2 cells were co-transfected with a CNS1-luciferase reporter construct (reference allele; construct 3), pRL-SV40, and TEAD1 or TEAD4 to determine if TEAD1/4 increase promoter activity in a dose-dependent manner. Figure 13(a) shows that both TEAD1 and TEAD4 are unable to increase CNS1 promoter activity in a dose-dependent manner. Below, the Western blot shows that TEAD1 was indeed over-expressed, and that TEAD1 is present at very low levels endogenously in HepG2 cells. Thus it can be concluded that the TEAD family of transcriptional enhancers do not enhance the transcriptional activity of the CNS1 promoter and therefore are not responsible for the observed SNP effect. These experiments will be pursued with the other transcription factors that were mentioned.

(A)


1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
RLU
TEAD1 - 5ng 10ng 25ng - - -
TEAD4 - - - - 5ng 10ng 25ng

(B)

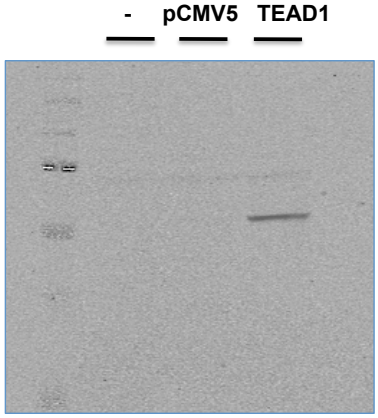

- pCMV5 TEAD1

Figure 13: TEAD1 and TEAD4 do not enhance CNS1 promoter activity. (A) HepG2 cells were co-transfected with CNS-luciferase reporter construct 3, pRL-SV40, increasing doses of TEAD1 or TEAD4, and pCMV5 to ensure all transfection cocktails contain the same amount of DNA. (B) Western blot of TEAD1 shows that TEAD1 was overexpressed in HepG2 cells. n=2 independent experiments.

3.9 Overexpression and siRNA-mediated Knockdown of the EST in HepG2 Cells

The most common variant identified by 3' RACE was 772bp long and contained exons 1, 2, and 7. Consequently, transcripts of this variant were most easily detected and screened for mutations introduced by Taq polymerase in the PCR steps. The 3' RACE product was combined with the additional 5' sequence to generate a full-length cDNA in the pCMV5 expression vector. This construct was over-expressed in HepG2 cells, RNA was isolated 24 hours post-tranfection, and gene expression was assessed by qRT-PCR. The results are shown in figure 14. The top panel includes genes located near *TRIB1*, and the bottom panel includes the genes encoding fatty acid synthase and apoB. EST overexpression had no effect on the expression of *KIAA0196* and *SQLE*, two genes located near the *TRIB1* gene on chromosome 8q24.13. However, EST over-expression significantly reduced *TRIB1, ZNF572, FAS* and *APOB* mRNA levels by 7%, 23%, 30%, and 20%, respectively. These findings suggest that this EST may have an effect on hepatic lipogenesis at the level of de novo fatty acid synthesis through fatty acid synthase - the *FAS* protein product. However, further studies are required to address the specificity of this effect.

We attempted siRNA-mediated knockdown of the EST in HepG2 cells. We optimized assay conditions to obtain a significant EST knockdown (40% knockdown) but this required a 72 hour transfection. The preliminary results from 2 successful independent experiments are shown in figure 15. A 40% EST knockdown increases *TRIB1* and *ACC1* mRNA levels by a statistically significant 15%. Changes in *APOB, SCD1* and *ACOX1* mRNA levels were not significant (not shown). However, the siRNA findings are questionable in terms of specificity of the response because in other siRNA optimization experiments, expression of genes in opposing pathways - fatty acid biosynthesis (*ACC1*) versus fatty acid β-oxidation (*ACOX1*) - followed the same pattern. Further, the most recent data suggests that these changes may not be consistent in all genes within a biochemical pathway, thus affecting the reproducibility of these results. Lastly, siRNA knockdown experiments have not been consistent with the over-expression findings, hence, they require further optimization.

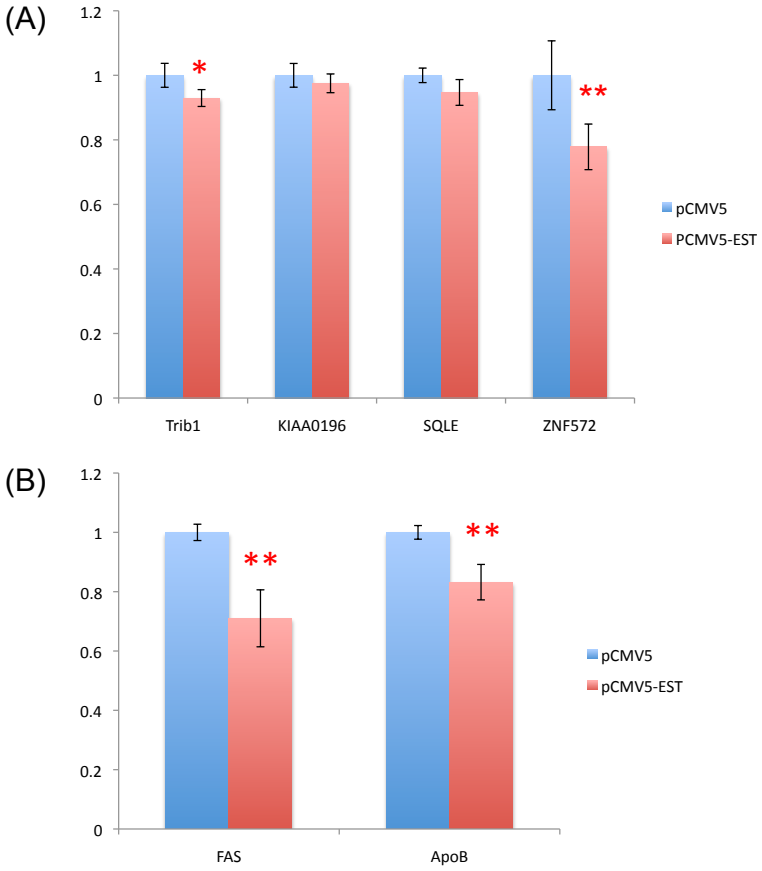

(A)
1.2
1
0.8
0.6
0.4
0.2
0
*
**
Trib1
KIAA0196
SQLE
ZNF572
pCMV5
PCMV5-EST
(B)
1.2
1
0.8
0.6
0.4
0.2
0
**
**
FAS
ApoB
pCMV5
pCMV5-EST

Figure 14: Effect of EST over-expression on hepatic lipogenesis and on the expression of *TRIB1* and adjacent genes. Empty pCMV5 or pCMV5-EST were transiently transfected in HepG2 cells. Total RNA was isolated 24 hours post-transfection and gene expression was assessed by qRT-PCR. (A) Effect of EST over-expression on the expression of *TRIB1* and adjacent genes at the mRNA level (B) Effect of EST over-expression on transcript levels of fatty acid synthase and apoB. n=3 independent experiments (**$p<0.01$; *$p<0.05$)

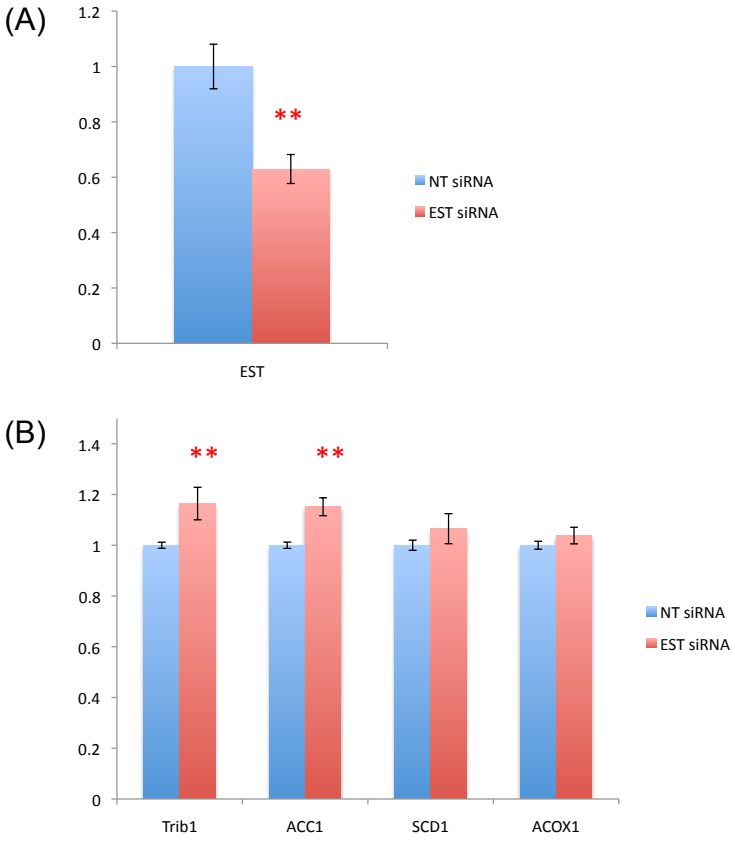

(A)
1.2
1
0.8
0.6
0.4
0.2
0
**
NT siRNA
EST siRNA
EST
(B)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
**
**
NT siRNA
EST siRNA
Trib1
ACC1
SCD1
ACOX1

Figure 15: siRNA-mediated EST knockdown increases *TRIB1* and *ACC1* expression. EST siRNA and non-targeting (NT) siRNA were transfected into HepG2 cells using Dharmafect-1 (Dharmacon). Total RNA was isolated 72 hours post-transfection and gene expression was assessed by qRT-PCR. (A) EST siRNA reduces EST expression by 40% (B) Effect of siRNA-mediated EST knockdown on transcript levels of genes in fatty acid biosynthesis (*ACC1, SCD1*) and β-oxidation (*ACOX1*) (**p<0.001)

3.10 The EST Does Not Bind Polycomb Repressive Complex PRC2

EST overexpression reduced transcript levels of *ZNF572, FAS* and *APOB*. We also observed a small but significant reduction in *TRIB1* mRNA. Further, siRNA-mediated EST knockdown of 40% resulted in increased transcript levels of *TRIB1* and *ACC1*. It recently been shown that up to 20% of intergenic long ncRNAs are bound by Polycomb Repressive Complex 2 (PRC2) and regulate gene expression [116]. Given that one EST transcript variant appears to affect gene expression, and such a large amount of intergenic long ncRNAs share the same mechanism of action, we investigated the possibility that the EST also binds PRC2 using RNA immunoprecipitation [113]. HepG2 nuclear extracts were incubated with antibodies against H3K27, EzH2, and Suz12 - components of PRC2 - and IgG as a negative control (Abcam). RNA was isolated from the beads, reverse transcribed and analyzed for EST expression by qRT-PCR. Relative copy numbers of the EST were normalized to the expression of small nuclear RNA U1, which does not bind PRC2 and results are expressed fold changes over the IgG fraction in figure 16. The graph clearly shows that the EST does not physically associate with any components of PRC2. Thus, its mechanism of action does not include repression of gene expression through recruitment of the chromatin modification complex PRC2.

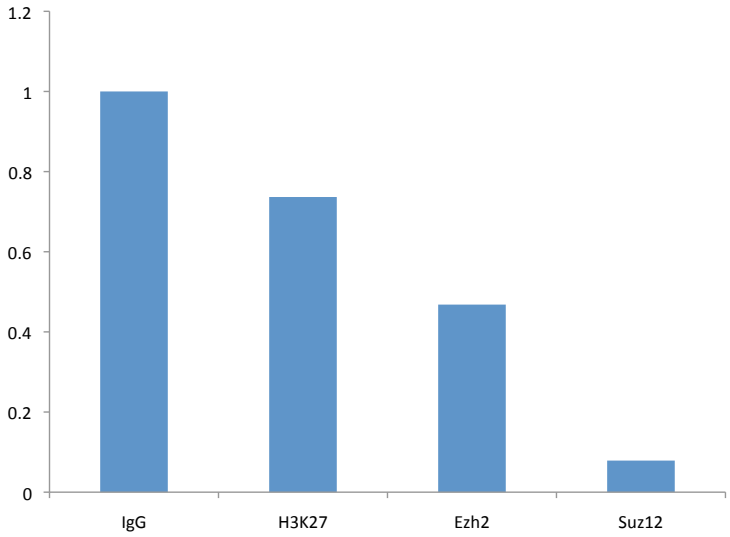

1.2
1
0.8
0.6
0.4
0.2
0
IgG
H3K27
Ezh2
Suz12

Figure 16: The EST does not physically associate with PRC2. Nuclear extracts were prepared from HepG2 cells. RNA was immunoprecipitated using antibodies against PRC2 components (H3K27, Ezh2, Suz12) and IgG as a negative control (Abcam). RNA was isolated from the beads, reverse transcribed and analyzed for EST expression by qRT-PCR. Relative copy numbers of the EST were normalized to the expression of small nuclear RNA U1, which does not bind PRC2. Results are expressed fold changes over the IgG fraction. n=2 independent experiments.

4 Discussion

4.1 The Functional Relationship Between GWAS-identified SNPs at the *TRIB1* Locus and Plasma Lipid Traits Remains Elusive

Recently, *TRIB1* has grown in interest because it has been shown to regulate hepatic lipogenesis in mice after over-expression and liver-specific knockout [98]. Specifically, altering *TRIB1* expression affected transcript levels of genes involved in fatty acids synthesis, resulting in increased VLDL production without changing *APOB* and *MTP* expression levels. This was an interesting finding because it has been well established that TGs - arriving to the liver as NEFAs - within cytosolic lipid droplets in hepatocytes contribute most to VLDL production regardless of metabolic state [3] and make up approximately 60% of liver TG content [117]. Yet, again regardless of metabolic state, it has also recently been shown that fatty acids from de novo lipogenesis make up approximately 30% of liver TG content and contribute up to 20% of the TG for VLDL production [117]. Consequently, the combination of GWAS association data and *TRIB1* functional studies have added strength to this suggestion that de novo lipogenesis could significantly contribute to hypertriglyceridemia.

Several GWAS, including the Global Lipids Genetics Consortium, have identified and replicated the association between *TRIB1* locus SNPs, plasma TGs, and CAD risk [49, 70, 56]. In the Ottawa Heart Study, the relationship between SNPs at the *TRIB1* locus and CAD is entirely mediated by effects on plasma lipids, notably plasma TGs. The secondary effects on HDL-c and LDL-c have also been replicated [56]; these associations do not reach genome-wide significance. In addition, the *TRIB1* locus SNPs are associated with severe hypertriglyceridemia; thus, the use of extreme phenotypes is a powerful tool in understanding the relationship between genetics and disease [106, 10]. Since a large amount of association data has been generated and replicated by various studies, the Global Lipids Genetics Consortium also searched for eQTLs at significant loci. Although they measured *TRIB1* transcript levels in liver (n=960), omental fat (n=741), and subcutaneous fat (n=609),

they did not detect differences in *TRIB1* mRNA levels between individuals homozygous reference, heterozygous, and homozygous risk for the trait-associated SNPs [56]. We investigated whether *TRIB1* risk locus SNPs alter the expression of *TRIB1* in whole blood from Ottawa Heart Study elderly healthy control subjects. Although our sample size was much smaller, no trend was detected between rs17321515 genotype and *TRIB1* mRNA levels; mRNA levels were also quite variable within each genotype group. These results could be due to the extremely short half life of the *TRIB1* message [90], or could indicate that the relationship between SNPs at the *TRIB1* locus and plasma lipid traits is independent of *TRIB1* expression levels.

Interestingly, the TG-associated SNPs from GWAS are located within the intergenic region 25-50kb downstream of the *TRIB1* coding region. In addition, this intergenic region can be termed a 'gene desert'[78] because it spans approximately 500kb in length and there are no nearby genes upstream from the SNP cluster; the nearest gene in the vicinity is *TRIB1*. GWAS have not identified TG-associated SNPs within the *TRIB1* coding region, and, HapMap LD data suggests that the GWAS signal at 8q24.13 is independent of SNPs within *TRIB1*. Hence, according to the common disease - common variant hypothesis, the functional SNP(s) - either from GWAS or in LD with GWAS SNPs - should be located somewhere within the haplotype block 25-50kb downstream of *TRIB1*. However, GWAS have not been able to explain a large part of the heritability behind complex disease [36]. For instance, results from the Global Lipids Genetics Consortium GWAS explain, in total, approximately 30% of the genetic variance for plasma TGs [56]. Other hypotheses have been proposed to account for the missing heritability behind complex diseases. Some of these include low frequency variants of intermediate effect sizes that have yet to be identified, or common variants with even smaller effect sizes that do not reach genome-wide significance and thus are missed by GWAS, but, in combination with other variants, do contribute to the overall genetic variance of a complex trait [36]. In effect, it is possible that the GWAS signal at the *TRIB1* locus is not coming from the identified common SNPs,

but, rather, a functional, less frequent SNP within the *TRIB1* coding region that has yet to be discovered. We may expect to obtain more information on this issue from the 1000 Genomes Project, which aims to cover a larger proportion of allelic variation in the human genome by identifying SNPs with minor allele frequencies between 1% and 5% [45]. Further, resequencing of candidate genes initially identified in GWAS for lipid traits has been successful in identifying a significant number of rare variants [67]. This approach may prove useful in finding the missing contributors to the genetic variance of complex traits, however, has not been attempted with *TRIB1*. Consequently, the functional TG-associated SNP (or SNPs) at the *TRIB1* locus remain elusive.

Despite the interest that has been centered on *TRIB1* since it has been shown to regulate hepatic lipogenesis [98], there remains no indication that the GWAS signal for plasma lipids and CAD risk is directly attributable to the *TRIB1* protein product. Further, a study examining genome characteristics of trait-associated SNPs found that a 88% of trait-associated SNPs were in non-coding regions; 45% and 43% fell within intronic and intergenic DNA, respectively [48]. These findings suggest non-coding SNPs may have larger roles than expected in human complex diseases [48]. Given that the GWAS SNPs are located in an intergenic region and are in very tight LD with one another, we chose to pursue the identification of functional elements within the intergenic region in an attempt to explain the functional relationship between SNPs at the *TRIB1* locus and plasma lipid traits. Although this study did not determine a definitive molecular mechanism linking TG-associated SNPs to plasma lipid traits, the novel results that have been obtained have added to our understanding of poorly characterized intergenic regions and opened new directions for further research in this area.

4.2 Functional Analysis of the 8q24.13 Risk Locus Identifies Novel Regulatory Elements and an EST-based Gene

Phylogenetic footprint analysis of the 8q24.13 risk locus identified two regions - each approximately 300bp in length - that had a high degree of conservation between species in comparison to the the remainder of the intergenic region. The *TRIB1* coding region, on the other hand, exhibited a high degree of evolutionary conservation, consistent with its role as a regulator of fundamental cellular processes, such as MAPK signaling cascades [86, 91] and cell cycle regulation in *Drosophila* [80, 81, 79]. The two regions of conservation identified within intergenic DNA, termed CNS1 and CNS2, are located 28kb and 53kb downstream of the *TRIB1* coding region, respectively. The regulatory function of each region was determined using various luciferase reporter assays in HepG2 cells. CNS2 was shown to have repressor activity, but it was not pursued in further functional studies because the rs55921265 SNP within this sequence was poorly linked to GWAS TG-associated SNPs (rs2954018, rs2001945, rs17321515) and also was not significantly associated with CAD. This was further confirmed in a larger sample size with additional data from Ottawa Heart Study CHD case and control subjects.

In contrast, CNS1 harboured one SNP - rs2001844 - in very tight LD with GWAS SNP rs17321515 (D' and r^2 of 1); rs2001845, located just upstream of CNS1, was determined to be in tight LD with the GWAS SNP rs2954018. Consequently, we pursued functional studies with CNS1 to determine whether these SNPs are functional and could explain the relationship between the GWAS signal and plasma lipid traits. Luciferase reporter assays identified this region as a weak promoter. In the initial promoter assays, we compared of four regions of increasing size from 500bp to 2kb; each region was extended by 500bp at the 5' end. The 500bp region showed promoter activity in both HepG2 cells and Cos-7 cells while the larger regions had absolutely no activity (figure 4) despite the fact that they all contained the CNS1 sequence at the 3' end. This was unexpected because previous studies have used luciferase reporter constructs larger than 2kb. For instance, an enhancer element

within the 9p21 risk locus - specifically within the *ANRIL* coding region - was discovered by luciferase reporter assays using a region that was greater than 2kb in size [53]. The discovery of an enhancer at the 1p13 locus near the *SORT1* gene also followed similar methodology. This study used the entire 6kb region - including 3' UTRs and intergenic DNA - between genes *CELSR2* and *PSRC1* in their luciferase reporter assays for the detection of enhancer activity [51]. One difference between our findings and these other studies is that the 8q24.13 locus is a gene desert, thus, the region upstream of CNS1 is devoid of both protein-coding and possibly non-coding genes. Perhaps sequences upstream of CNS1 recruit factors that repress transcription. Since the only difference in our luciferase reporter constructs is the addition of sequential 5' end extensions, it is not possible to determine if more upstream sequences within the 1.5 and 2kb constructs are also responsible for this repressive effect. It suggests, however, that there is an element within the 500bp directly upstream of CNS1 that can repress transcription. This would require further validation by determining its ability to repress transcription from a strong promoter, similar to the methodology used to determine the repressor activity of CNS2 in figure 3.

The most active CNS1-luciferase reporter construct - consisting of solely conserved sequence and both potential transcription start sites - exhibited approximately a 5-fold increase in luciferase activity as compared to the empty pGL3 basic luciferase reporter vector (figure 11b). Interestingly, online promoter prediction tools (Promoter 2.0 [118]) identified one weak match with RNA polymerase II transcription start sites within the CNS1 sequence, just upstream of of the 5' end that was mapped by 5' RACE. Yet, tools for identifying putative transcription factor binding sites (MatInspector, TRANSFAC) were unable to predict the presence of an RNA polymerase II binding site within CNS1. However, the novel EST-based gene is likely transcribed by RNA polymerase II because we detected EST transcript variants by 3' RACE using oligodT primers for cDNA synthesis, and it has been well-established that RNA polymerase II is necessary for polyadenylation [119, 120]. Taken together, these findings suggest that CNS1 is a real RNA polymerase II promoter, but

there remain aspects of RNA polymerase II promoters that have not yet been extensively characterized.

In addition to the lack of enhancer activity in our luciferase reporter assays, the classification of CNS1 as a promoter was further supported by the discovery of an EST-based gene in the region. 5'/3' RACE in HepG2 cells extended the 5' end of the EST into CNS1 but not beyond it (figure 10), strongly suggesting that the regulatory element within CNS1 was responsible for driving EST transcription. However, a recent study that found RNA polymerase II transcription start sites overlapping enhancers further complicates this area [121]. Firstly, promoters and enhancers have distinct chromatin signatures. For instance, promoters are generally enriched in CpG islands and H3K5me3 while enhancers have high levels of H3K4me1, low or no H3K4me3, and are frequently bound by histone acetyltransferase p300 [122]. Interestingly, the UCSC Genome Browser shows that the chromatin signature of CNS1 follows the latter pattern for enhancers, yet, the CNS1 region was unable to enhance luciferase expression from a minimal promoter in luciferase reporter assays. The UCSC Genome Browser also includes regulatory activities in multiple cell types. It shows that the CNS1 region is a weak promoter in HepG2 cells, but, unlike our luciferase reporter results, it shows that the region upstream from CNS1 is a strong enhancer. Experimentally, we found that region directly upstream from CNS1 (the 1kb construct and beyond) repressed transcription from the CNS1 promoter. Future studies should include analyzing the function of the region upstream from CNS1 in subsets to determine the regulatory of each one individually. Another reason for this discrepancy could arise from methodology. Our results have been experimentally determined, while the features from UCSC are predicted functional elements.

The 5'RACE results suggested the presence of two transcription start sites located approximately 70bp apart (figure 10). The latter observation was also supported by luciferase assays which demonstrated that the presence of both transcription start sites - in compari-

son to the first site only - yielded a significantly more active promoter construct (figure 11, right panel). Further, complete removal of EST exon 1 resulted in complete loss of CNS1 promoter activity (figure 11, left panel), and, including the first 40bp of exon 1 as mapped by 5' RACE restored CNS1 promoter activity (figure 11, right panel; construct 3). Thus, one can conclude that these 40bp are necessary for the regulatory activity of CNS1. The tight clustering of transcription start sites is a feature more commonly found with promoters rather than with enhancers [121].

The 3' RACE identified 9 transcript variants of this EST-based gene to date in HepG2 cells (figure 9; table 3). Results suggest that there is an abundance of large transcripts from this EST-based gene; the smallest and largest variants were found to be 681bp and 1341bp, respectively. However, individual variants have proven difficulty to study given their extremely low abundance. EST variant amplification by PCR has been possible from HepG2 cells, however, it has not been possible to amplify specific variants in whole blood and has proven extremely difficult to amplify the region corresponding to exons 1 and 2, which are found within multiple variants (figure 7). Consequently, although it was attempted in blood as described earlier, we have not determined whether the rs17321515 risk allele is associated with altered expression of the EST-based gene because the samples available for RNA isolation include only whole blood from Ottawa Heart Study controls. An interesting feature was noted in certain transcripts identified by 3' RACE. Transcript sequences obtained from DNA sequencing of 3' RACE products were mapped for specific exon locations by BLAST search using the Ensembl Genome Browser because it organizes its sequence information with respect to chromosome position. One striking feature of all transcripts was the poor sequence conservation between species; exons are also not conserved between humans and rodents. One view on sequence conservation is that a non-conserved transcript itself has no biological function, but that the act of its transcription promotes or represses the transcription of an adjacent gene, thereby acting in cis as regulators of gene transcription [123]. Other experimental studies on long ncRNA function have shown that several long ncRNAs

can also regulate gene expression in *trans* [124]. In both cases, sequence conservation is generally poor. Lastly, given that there exist several variants of this EST-based gene, it is less likely that the RNA itself has no biological relevance. If this were so, transcripts would expected to be be shorter, more unstable, less likely to be detected in full form by 3' RACE, and there would be expected to be multiple 5' ends spaced farther apart from each other [121]. Further, the EST-based gene is located 28kb downstream from *TRIB1*, the nearest protein-coding gene and is transcribed on the forward strand, away from *TRIB1*.

Most EST transcript sequences were unique and aligned specifically to their position in chromosome 8 (Table 3). However, exons 4 and 5 happened to have nearly identical sequence similarity to many regions across the human genome. Further analysis of these regions using the UCSC Genome Browser revealed that exon 4 and the last 250bp of exon 5 consist entirely of SINEs, which, in humans, are primarily *Alu* sequences. The presence of these repetitive elements in some transcripts but not in others suggests that the variants may have different functions. Some *Alu* RNAs have been characterized and have been shown to have a common mechanism of action [124]. Human small cytoplasmic *Alu* RNAs have been shown to regulate gene transcription in *trans* by binding RNA polymerase II [109]. Moreoever, earlier studies found that these RNAs are transcribed by RNA polymerase III as larger nuclear, polyadenylated precursors [110]. Consequently, poly(A)-tails can be generated from both RNA polymerase II and III transcription - a finding that questions whether RNA polymerase II is truly responsible for transcription from the CNS1 promoter, or whether its transcription is mediated by RNA polymerase III, whose transcriptome is ever-expanding [125]. Additional studies will be required to further characterize the CNS1 regulatory element. Although the EST transcripts are larger than the *Alu* RNA precursors that have been identified, the presence of these repetitive elements suggests that these transcripts could function in a similar manner. Consequently, other unknown mechanisms must direct the specificity of these ncRNAs' mechanisms of action. For instance, ncRNA cellular trafficking must be controlled to ensure that the ncRNA is exerting its function

at the appropriate target gene promoter. This is especially important if several ncRNAs interact with RNA polymerase II machinery, which is responsible for a large portion of the expression of the eukaryotic transcriptome. Thus, future studies can address the possibility that the transcripts containing exons 4 and 5 may act in *cis* or in *trans* by binding RNA polymerase II and the transcription pre-initiation complex. The other transcript variants that contain only unique exons would be expected to have a different mechanism of action. Future studies include over-expressing the different variants to determine if any are functional and what their mechanism of action is. We do not expect that all transcripts will function via the same mechanism.

We sought to address the function of the most abundant EST variant identified by 3' RACE by over-expression in HepG2 cells. Since ncRNAs can act in *cis* or in *trans*, we explored the possibility that over-expression of a specific variant would affect the mRNA expression of nearby genes, or genes related to the phenotype of interest - plasma triglyceride levels. Our findings show that EST over-expression significantly reduced transcript levels of *ZNF572*, a gene located just upstream of *TRIB1*, and slightly but significantly reduced *TRIB1* expression as well. However, it also affected mRNA expression in *trans* by reducing transcript levels of *FAS* and *APOB* by 30% and 20%, respectively. As a result, it is unclear whether these observed effects are specific, but, nevertheless, these findings to infer that EST transcripts have a role in the regulation of hepatic lipogenesis. In this case, however, it is unclear whether this regulation is occuring in *cis* through *TRIB1* and/or *ZNF572*, or whether it is specifically a *trans* effect at the mRNA level of *FAS* and *APOB*. Interestingly, evidence suggests that long ncRNAs share common mechanisms of action. For instance, approximately 20% of intergenic RNAs physically associate with PRC2, and additional intergenic RNAs are bound by other chromatin-modifying complexes [116]. In addition, *Alu* ncRNAs affect gene transcription by directly binding RNA polymerase II [109]. While the latter function does not fit the over-expression data - we should have observed repression at several other genes but some genes we examined were unaffected -

it is entirely possible that this short variant binds a chromatin-modification complex, and that this interaction is responsible for the mRNA expression changes we observed. Since the most commonly studied interaction appears to be with PRC2, we sought to determine if any EST transcripts can bind PRC2 using RNA immunoprecipitation. The results in figure 16 indicate that EST transcripts do not physically associate with PRC2, thus, additional chromatin-modification complexes will be considered in future studies.

Next, the function of EST transcripts was examined by siRNA-mediated knockdown of all variants in HepG2 cells. The results from siRNA knockdown were much less clear than the over-expression data. After much optimization, a 40% knockdown was achieved in two independent experiments. The results from these two experiments (figure 15) suggest that EST depletion may increase *TRIB1* and *ACC1* mRNA levels. However, the siRNA findings do not consistently complement the over-expression studies because we could not detect significant increases in *ZNF572, FAS* and *APOB* mRNA levels. Studies have had difficulty in knocking down ncRNAs [121], however, the reasoning behind this difficulty is that these transcripts arise from RNA polymerase II transcription of enhancer elements, thereby generating diverse, short, unstable transcripts that are not easily targeted by RNA interference [121]. In our experiments, the ncRNAs are not short ($>$ 600bp in length) and are consistently detected by PCR from HepG2 cDNA, although individual variants are of low abundance. The siRNA approach has been successful for *HOTAIR* - which binds PRC2 - but this long ncRNA is more conserved than the EST-based gene and is expressed at a similar level to protein-coding genes [113, 116, 121]. Since the EST transcripts do not bind PRC2, they must have a different secondary structure and behave via a different mechanism of action. Their expression is also of lower abundance than other more conserved mammalian ncRNAs. One reason that partially explains the inconsistent and perhaps non-specific findings for the EST-based gene may be attributed to the large number of EST variants. This siRNA attempts to knock down all variants, but may not be targeting the same variants in the same proportions between experiments. If variants have different ef-

fects, then this could be reflected in these siRNA knockdown results. In addition, siRNA may be unsuccessful as a result of ncRNA cellular localization. Although the EST variants are polyadenylated, this does not necessarily mean that they localize to the cytoplasm. For instance, one long polyadenylated ncRNA - Gonafu - was found to localize exclusively to the nucleus, within 'nuclear speckles', and does not colocalize with any known nuclear marker [126]. Perhaps the nuclear RNA is not accessible to the siRNA and RNAi machinery and thus its expression cannot be reduced this way. These EST transcripts will be assessed for cellular localization by Northern blot of cytoplasmic and nuclear HepG2 RNA fractions. Further, a 3-day siRNA transfection is required to achieve a statistically significant knockdown. These experimental conditions could have other effects on the HepG2 cells, which may also explain the inconsistencies between experiments. Consequently, further understanding of EST localization and transcript variants is needed, as well as additional optimization of experimental conditions.

The CNS1 luciferase reporter assays have provided additional interesting findings. The minimal promoter region in HepG2 cells was mapped to the CNS1 region exclusively at either end, and to the first 40bp of exon 1, corresponding to the first transcription start site (figure 12; construct 3). We also generated a construct that contained both transcription start sites (figure 12; construct 4). This promoter region harbours one common SNP - rs2001844 - hence, luciferase reporter constructs were generated using genomic DNA from Ottawa Heart Study control subjects who were either homozygous reference or homozygous risk for this SNP. Interestingly, the rs2001844 risk allele reduced promoter activity by 40% from construct 3 in comparison to the reference allele. It also reduced promoter activity by 30% from construct 4 in comparison to the reference allele. We are pursuing the characterization of this apparently functional SNP by determining which transcription factor binding site it alters. We used MatInspector and TRANSFAC to predict transcription factor binding sites within the sequence that contains the rs2001844 SNP. However, upon comparing the predictions between these tools, it was evident that they both predicted

different transcription factor sites. Since there was no overlap between predictions within the sequence of interest, we systematically began searching for the transcription factor that bind this region. The first transcription factors that were attempted include TEAD-1 and TEAD-4. However, co-transfection of these expression vectors with the CNS1-luciferase reporter vector in HepG2 cells showed that both TEAD-1 and TEAD-4 were unable to enhance transcription from the CNS1 promoter (figure 13a). This finding, however, is not entirely surprising. Firstly, Western blot analysis of HepG2 whole cell lysate (figure 13b) for TEAD-1 expression shows that hepatocytes express very low levels of TEAD-1 protein. TEAD-1 has been shown to be abundantly expressed in heart and skeletal muscle [127]. Moreoever, Northern blot analysis in this same study did not detect TEAD-1 or TEAD-4 levels above background [127]. Thus, these transcriptional enhancers are not enriched in liver, therefore it is less likely that one of these two factors controls EST expression. All members of the TEAD family of transcriptional enhancers bind the same target sequence [128], thus, it is possible that TEAD-2 or TEAD-3 can enhance CNS1 promoter transcription. Further, TEAD factors have been shown to have different transcriptional effects. A study that compared the effects of TEAD factors on the promoter activity of the mouse mammary tumour virus long terminal repeat (MMTV LTR) found that TEAD-1 over expression squelched LTR promoter activity, whereas TEAD-2 over-expression transactivated the LTR [129]. Consequently, it will be of interest to determine the expression of the various TEAD factors in HepG2 cells and to examine the effects of the remaining TEAD factors on EST promoter activity. Moreover, the other predicted transcription factor binding sites will be examined. These studies will be pursued by luciferase reporter assays similar to what was described for TEAD-1 and TEAD-4, and by electromobility shift assays (EMSA).

Overall, however, the finding that the risk allele of the rs2001844 SNP significantly reduces EST promoter activity has functional implications for the GWAS signal. For instance, it is well-established that the GWAS SNPs are associated with plasma TGs at a level that reaches genome-wide significance [56], indicating that the functional risk allele(s) somehow

modulate plasma TG levels. The effect of the rs2001844 risk allele on EST promoter activity suggests that altered expression of this EST-based gene is itself associated with elevated plasma TGs. The EST itself could have a biological relevance, thereby altering biochemical pathways leading to the TG phenotype identified in GWAS, or, the act of EST transcription at the risk locus could affect the expression of neighbouring genes, which suggests that the EST itself has no biological relevance [121]. Given that both our RNA expression analysis in whole blood and the Global Lipids Genetics Consortium mRNA expression analyses in large sample sizes of a variety of tissues have been unable to determine a relationship between *TRIB1* mRNA and genotype of TG-associated SNPs [56], the latter hypothesis is found to be questionable. Although *TRIB1* mRNA has a surprisingly short half-life [90], the sample size from this recent lipids genetics study was very large and the GWAS SNPs are very common, with a minor allele frequency of 0.47 [56]. Thus, it would be expected that regulation of neighbouring gene expression would be reflected by changes in the mRNA levels of the nearest gene - *TRIB1*. Further, although functional studies on the *TRIB1* protein product in mice have shown that it is a regulator of hepatic lipogenesis [98], this pathway may be subject to other levels of regulation that are specific to higher mammals. The observation that this EST-based gene is poorly conserved between humans and rodents does not necessarily imply lack of function [130]; alternatively, this may suggest that it adds another layer of complexity to the regulation of hepatic lipogenesis in humans. Thus, future studies are required to determine the biological function - if any - of the various EST transcripts.

5 Conclusion

The 8q24.13 risk locus contains a haplotype block of SNPs that are associated with plasma TG levels with genome-wide significance. These GWAS findings have been replicated in several studies including the Ottawa Heart Study. The cluster of risk SNPs is located within a gene desert, 25kb-50kb downstream from *TRIB1* - the nearest coding region. Consequently, our studies indicate that the intergenic region downstream from the *TRIB1* coding region harbours a regulatory element - CNS1 - that shows promoter activity in luciferase reporter assays. Further, we have identified and mapped an EST-based ncRNA to the risk locus in HepG2 cells. 3' RACE has identified several long variants - ranging between approximately 700bp-1400bp in length - but they each appear to be of low abundance. 5' RACE has mapped the transcription start site of this EST-based gene within CNS1. Interestingly, this approach has identified two possible transcription start sites, located in close proximity of one another at 70b apart. Luciferase reporter assays further confirmed these findings. Complete removal of the EST exon 1 resulted in complete loss of EST promoter activity. The addition of the first 40bp of exon 1 - containing the first transcription start site - restored promoter activity. Further, the presence of the second transcription start site further enhanced the measured transcriptional activity of this region. In addition, the CNS1 promoter constructs include the TG-associated common SNP rs2001844. Luciferase reporter assays clearly show that the risk allele of this SNP significantly reduces promoter activity, thereby suggesting that the rs2001844 risk allele may modulate the expression of this EST-based gene. However, we were unable to determine this using whole blood due to the extremely low expression of this EST in whole blood as compared to HepG2 cells. Liver samples would be a more appropriate tissue for the study for this genotype-dependent RNA expression analysis.

In terms of the functional significance of this EST-based gene, siRNA-mediated knockdown requires further optimization of experimental conditions and characterization of the

EST-based gene transcripts. For instance, the large number of variants and their cellular localization could all affect the knockdown experiments. In contrast, over-expression studies of a short variant - that does not contain exons rich in repetitive elements - found that increased levels of this transcript variant were accompanied by reduced mRNA expression of *ZNF572* and to a lesser extend *TRIB1*, both of which are located just upstream from this intergenic region. The over-expression study also affected mRNA expression in *trans* by reducing *FAS* and *APOB* transcript levels. Further studies will be performed to determine the effects of over-expression of the remaining transcript variants. Taken together, the data suggests that this EST-based gene generates functional transcripts. Thus, it remains possible that the relationship between the functional TG-associated SNPs from the GWAS signal and plasma TG levels can be explained by this EST-based gene. Further studies, however, will be required for a deeper understanding of the function of this intergenic region.

References

[1] Negi, S. and Anand, A. (2010) Atherosclerotic coronary heart disease - epidemiology, classification and management. *Cardiovasc Hematol Disord Drug Targets*, **10**, 257–261.

[2] Chen, Y., Rollins, J., Paigen, B., and Wang, X. (2007) Genetic and genomic insights into the molecular basis of atherosclerosis. *Cell Metabolism*, **6**, 164–179.

[3] Harchaoui, K., Visser, M., Kastelein, J., Stroes, E., and Dallinga-Thie, G. (2009) Triglycerides and cardiovascular risk. *Current Cardiology Reviews*, **5**, 216–22.

[4] Goldberg, I., Eckel, R., and McPherson, R. (2011) Triglycerides and heart disease: Still a hypothesis? *Arterioscler Thromb Vasc Biol*, **31**, 00–00.

[5] Alberti, K., et al. (2009) Harmonizing the metabolic syndrome: A joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; american heart association; world heart federation; international atherosclerosis society; and interntional association for the study of obesity. *Circulation*, **120**, 1640–1645.

[6] Ginsberg, H. and Fisher, E. (2009) The ever-expanding role of degradation in the regulation of apolipoprotein B metabolism. *J. Lipid Res.*, **50**, S162–S166.

[7] Biddinger, S., Hernandez-Ono, A., et al. (2008) Hepatic insulin resistance is sufficient to produce dyslipidemia and susceptibility to atherosclerosis. *Cell Metabolism*, **7**, 125–134.

[8] Nordestgaard, B., Benn, M., Schnohr, P., and Tybjaerg-Hansen, A. (2007) Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. *JAMA*, **298**, 299–308.

[9] Nordestgaard, B. and Zilversmit, D. (1988) Large lipoproteins are excluded from the arterial wall in diabetic cholesterol-fed rabbits. *J. Lipid Res.*, **29**, 1491–1500.

[10] Hegele, R. A. and Pollex, R. L. (2009) Hypertriglyceridemia: phenomics and genomics. *Mol Cell Biochem*, **326**, 35–43.

[11] Austin, M. A., McKnight, B., Edwards, K. L., Bradley, C. M., McNeely, M. J., Psaty, B. M., Brunzell, J. D., and Motulsky, A. G. (2000) Cardiovascular disease mortality in familial forms of hypertriglyceridemia: A 20-year prospective study. *Circulation*, **101**, 2777–2782.

[12] Langsted, A., Freiberg, J., Tybjaerg-Hansen, A., Schnohr, P., Jensen, G., and Nordestgaard, B. (2010) Nonfasting cholesterol and triglycerides and association with risk of myocardial infarction and total mortality: the Copenhagen City Heart Study with 31 years of follow-up. *J Intern Med*, **269**, 1–11.

[13] Hokanson, J. and Austin, M. (1996) Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. *J Cardiovasc Risk*, **3**, 213–219.

[14] Jeppesen, J., Hein, H. O., Suadicani, P., and Gyntelberg, F. (1998) Triglyceride concentration and ischemic heart disease: An eight-year follow-up in the Copenhagen Male Study. *Circulation*, **97**, 1029–1036.

[15] Sarwar, N., Danesh, J., Eiriksdottir, G., Sigurdsson, G., Wareham, N., Bingham, S., Boekholdt, S., Khaw, K., and Gudnason, V. (2007) Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation*, **115**, 450–458.

[16] Patel, A., Barzi, F., Jamrozik, K., Lam, T., Ueshima, H., Whitlock, G., and Woodward, M. (2004) Serum triglycerides as a risk factor for cardiovascular disease in the Asia-Pacific region. *Circulation*, **110**, 2678–2686.

[17] Di, A., et al. (2009) Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*, **302**, 1993–2000.

[18] Yen, C., Stone, S., Koliwad, S., et al. (2008) Thematic review series: Glycerophospholipids, DGAT enzymes and triacylglycerol biosynthesis. *J. Lipid Res.*, **49**, 2283–301.

[19] Postic, C. and Girard, J. (2008) Contribution of de novo fatty acid synthesis to hepatic steatosis and insulin resistance: lessons from genetically engineered mice. *J. Clin. Invest.*, **118**, 829–838.

[20] Gordon, D. and Jamil, H. (2000) Progress towards understanding the role of microsomal triglyceride transfer protein in apolipoprotein-B lipoprotein assembly. *Biochimica et Biophysica Acta*, **1486**, 72–83.

[21] Jeon, H. and Blacklow, S. (2005) Structure and physiologic function of the low-density lipoprotein receptor. *Annu. Rev. Biochem.*, **74**, 535–562.

[22] Rash, J. M., Rothblat, G. H., and Sparks, C. E. (1981) Lipoprotein apolipoprotein synthesis by human hepatoma cells in culture. *Biochimica et Biophysica Acta*, **666**, 294–298.

[23] Moberly, J. B., Cole, T. G., Alpers, D. H., and Schonfeld, G. (1990) Oleic acid stimulation of apolipoprotein B secretion from HepG2 and Caco-2 cells occurs post-transcriptionally. *Biochimica et Biophysica Acta*, **1042**, 70–80.

[24] Thrift, R. N., Forte, T. M., Cahoon, B. E., and Shore, V. G. (1986) Characterization of lipoproteins produced by the human liver cell line, hep g2, under defined conditions. *J. Lipid Res.*, **27**, 236–250.

[25] Qiu, W., Taghibiglou, C., Avramoglu, R.-K., van Idetstine, S. C., Naples, M., Ashrafpour, H., Mhapsekar, S., Sato, R., and Adeli, K. (2005) Oleate-mediated stimulation of microsomal triglyceride transfer protein (MTP) gene promoter: implications for hepatic MTP overexpression in insulin resistance. *Biochemistry*, **44**, 3041–3049.

[26] Jump, D. B., Botolin, D., Wang, Y., Xu, J., Christian, B., and Demeure, O. (2005) Fatty acid regulation of hepatic gene transcription. *J. Nutr.*, **135**, 2503–2506.

[27] Wolfrum, C., Borrmann, C., Borchers, T., and Spener, F. (2001) Fatty acids and hypolipidemic drugs regulate PPAR α- and γ-mediated gene transcription via liver fatty acid binding protein: a signalling path to the nucleus. *Proc. Natl. Acad. Sci. USA*, **98**, 2323–2328.

[28] Pawar, A. and Jump, D. B. (2003) Unsaturated fatty acid regulation of peroxisome-proliferator-activated receptor α activity in rat primary hepatocytes. *J Biol Chem*, **278**, 35931–35939.

[29] Yoshikawa, T., et al. (2002) Polyunsaturated fatty acids suppress sterol regulatory element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding to LXR response elements. *J Biol Chem*, **277**, 1705–1711.

[30] Shimomura, I., Bashmakov, Y., Ikemoto, S., Horton, J. D., Brown, M. S., and Goldstein, J. L. (1999) Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes. *Proc. Natl. Acad. Sci. USA*, **96**, 13656–13661.

[31] Pollin, T. I., Hsueh, W.-C., Steinle, N. I., Snitker, S., Shuldiner, A. R., and Mitchell, B. D. (2004) A genome-wide scan of serum lipid levels in the Old Order Amish. *Atherosclerosis*, **173**, 89–96.

[32] Pilia, G. et al. (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.*, **2**, e132.

[33] Goldstein, J. L., Dana, S. E., Brunschede, G. Y., and Brown, M. S. (1975) Genetic heterogeneity in familial hypercholesterolemia: Evidence for two different mutations affecting functions of the low-density lipoprotein receptor. *Proc. Nt. Acad. Sci. USA*, **72**, 1092–1096.

[34] Breslow, J. L. (2000) Genetics of lipoprotein abnormalities associated with coronary heart disease susceptibility. *Annu. Rev. Genet.*, **34**, 233–254.

[35] Nassoury, N., Blasiole, D. A., Oler, A. T., Benjannet, S., Hamelin, J., Poupon, V., McPherson, P. S., Attie, A. D., Prat, A., and Seidah, N. G. (2007) The cellular trafficking of the secretory proprotein convertase PCSK9 and its dependance on the LDLR. *Traffic*, **8**, 718–732.

[36] Manolio, T. A. et al. (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

[37] Hardy, J. and Singleton, A. (2009) Genomewide association studies and human disease. *N Engl J Med*, **360**, 1759–1768.

[38] Lewontin, R. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49–67.

[39] Hill, W. and Robertson, A. (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet*, **38**, 226–231.

[40] Neale, B. M. (2010) Introduction to linkage disequilibrium, the HapMap, and imputation. *Cold Spring Harb Protoc*.

[41] Johansen, C. T., Katherisan, S., and Hegele, R. A. (2011) Genetic determinants of plasma triglycerides. *J. Lipid Res.*, **52**, 189–206.

[42] Cardon, L. R. and Abecasis, G. R. (2003) Using haplotype blocks to map human complex trait loci. *TRENDS in Genetics*, **19**, 135–140.

[43] International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.

[44] International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.

[45] The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.

[46] Abecasis, G. R., et al. (2001) Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.*, **68**, 191–197.

[47] Reich, D. E., et al. (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

[48] Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., and Manolio, T. A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA*, **106**, 9362–9367.

[49] Willer, C. J. et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genetics*, **40**, 161–169.

[50] Katherisan, S. et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genetics*, **40**, 189–197.

[51] Musunuru, K. et al. (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*, **466**, 714–721.

[52] McPherson, R., et al. (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science*, **316**, 1488–1491.

[53] Jarinova, O., et al. (2009) Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol*, **29**, 1671–1677.

[54] Harismendy, O., et al. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature*, **470**, 264–270.

[55] Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011) Long non-coding RNA *ANRIL* is required for the prc2 recruitment to and silencing of p15 tumor suppressor gene. *Oncogene*, **30**, 1956–1962.

[56] Teslovich, T. N. et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.

[57] Augustus, A., Yagyu, H., Haemmerle, G., Bensadoun, A., Vikramadithyan, R., Park, S., Kim, J., Zechner, R., and Goldberg, I. (2004) Cardiac-specific knockout of lipoprotein lipase alters plasma lipoprotein triglyceride metabolism and cardiac gene expression. *J Biol Chem*, **279**, 25050–25057.

[58] Levak-Frank, S., Hofmann, W., Weinstock, P., Radner, H., Sattler, W., Breslow, J., and Zechner, R. (1999) Induced mutant mouse lines that express lipoprotein lipase in cardiac muscle, but not in skeletal muscle and adipose tissue, have normal plasma triglyceride and high-density lipoprotein-cholesterol levels. *Proc. Natl. Acad. Sci. USA*, **96**, 3165–3170.

[59] Havel, R., Kane, J., and Kashyap, M. (1973) Interchange of apolipoproteins between chylomicrons and high density lipoproteins during alimentary lipemia in man. *J Clin Invest*, **52**, 32–38.

[60] Lamarche, B., Uffelman, K., Carpentier, A., Cohn, J., Steiner, G., Barrett, P., and Lewis, G. (1999) Triglyceride enrichment of HDL enhances in vivo metabolic clearance of HDL apo A-I in healthy men. *J Clin Invest*, **103**, 1191–1199.

[61] Horowitz, B., Goldberg, I., Merab, J., Vanni, T., Ramakrishnan, R., and Ginsberg, H. (1993) Increased plasma and renal clearance of an exchangeable pool of apolipoprotein A-I in subjects with low levels of high density lipoprotein cholesterol. *J Clin Invest*, **91**, 1743–1752.

[62] Bharadwaj, K., Hiyama, Y., Hu, Y., Huggins, L., Ramakrishnan, R., Abumrad, N., Shulman, G., Blaner, W., and Goldberg, I. (2010) Chylomicron-and VLDL-derived lipids enter the heart through different pathways: in vivo evidence for receptor- and non-receptor-mediated fatty acid uptake. *J Biol Chem*, **285**, 37976–37986.

[63] Fruchart-Najib, J., Baugé, E., Niculescu, L.-S., Pham, T., Thomas, B., Rommens, C., Majd, Z., Brewer, B., Pennacchio, L. A., and Fruchart, J.-C. (2004) Mechanism of triglyceride lowering in mice expressing human apolipoprotein A5. *Biochemical and Biophysical Research Communications*, **319**, 397–404.

[64] Merkel, M., Loeffler, B., Kluger, M., Fabig, N., Geppert, G., Pennacchio, L. A., Laatsch, A., and Heeren, J. (2005) Apolipoprotein AV accelerates plasma hydrolysis of triglyceride-rich lipoproteins by interaction with proteoglycan-bound lipoprotein lipase. *J Biol Chem*, **280**, 21553–21560.

[65] Schaap, F. G., Rensen, P. C., Voshol, P. J., Vrins, C., van der Vliet, H. N., Chamuleau, R. A., Havekes, L. M., Groen, A. K., and van Dijk, K. W. (2004) ApoAV reduces plasma triglycerides by inhibiting very low density lipoprotein-triglyceride (VLDL-TG) production and stimulating lipoprotein lipase-mediated VLDL-TG hydrolysis. *J Biol Chem*, **279**, 27941–27947.

[66] Lookene, A., Beckstead, J. A., Nilsson, S., Olivecronat, G., and Ryan, R. O. (2005) Apolipoprotein A-V-heparin interactions: implications for plasma lipoprotein metabolism. *J Biol Chem*, **280**, 25383–25387.

[67] Johansen, C. T. et al. (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridermia. *Nature Genetics*, **42**, 684–688.

[68] Orho-Melander, M. et al. (2008) Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes*, **57**, 3112–3121.

[69] Varbo, A., Benn, M., Tybjaerg-Hansen, A., Grande, P., and Nordestgaard, B. (2011) TRIB1 and GCKR polymorphisms, lipid levels, and risk of ischemic heart disease in the general population. *Arterioscler Thromb Vasc Biol*, **31**, 451–57.

[70] Katherisan, S. et al. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*, **41**, 56–65.

[71] Garcia-Rios, A. et al. (2010) Polymorphisms at the TRIB1 gene modulates plasma lipid levels: Insight from the spanish familial hypercholesterolemia cohort study. *Nutr Metab Cardiovasc Dis*, pp. 1–7.

[72] Wang, J., Ban, M. R., Zou, G. Y., et al. (2008) Polygenic determinants of severe hypertriglyceridemia. *Human Molecular Genetics*, **17**, 2894–2899.

[73] Hegele, R. A., Ban, M. R., Hsueh, N., et al. (2009) A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. *Human Molecular Genetics*, **18**, 4189–4194.

[74] Tai, E. S., Sim, X. L., Ong, T. H., et al. (2009) Polymorphisms at newly identified lipid-associated loci are associated with blood lipids and cardiovascular disease in an Asian Malay population. *J. Lipid Res.*, **50**, 514–520.

[75] Nakayama, K., Bayasgalan, T., Yamanaka, K., et al. (2009) Large scale replication analysis of loci associated with lipid concentrations in a Japanese population. *J Med Genet*, **46**, 370–374.

[76] Deo, R. C., Reich, D., Tandon, A., Akylbekova, E., Patterson, N., et al. (2009) Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: The Jackson Heart Study. *PLoS Genet.*, **5**, e1000342.

[77] Lettre, G., Palmer, C., Young, T., Ejebe, K., Allayee, H., et al. (2011) Genome-wide association study of coronary artery disease and its risk factor in 8,090 African Americans: The NHLBI CARe Project. *PLoS Genet.*, **7**, e1001300.

[78] Nobrega, M. A., Ovcharenko, I., Afzal, V., and Rubin, E. M. (2003) Scanning human gene deserts for long-range enhancers. *Science*, **302**, 413.

[79] Hegedus, Z., Czibula, A., and Kiss-Toth, E. (2006) Tribbles: novel regulators of cell function; evolutionary aspects. *Cell Mol Life Sci*, **63**, 1632–1641.

[80] Mata, J., Curado, S., Ephrussi, A., and Rørth, P. (2000) Tribbles coordinates mitosis and morphogenesis in drosophila by regulating String/CDC25 proteolysis. *Cell*, **101**, 511–522.

[81] Seher, T. C. and Leptin, M. (2000) Tribbles, a cell-cycle brake that coordinates proliferation and morphogenesis during drosophila gastrulation. *Current Biology*, **10**, 623–629.

[82] Johnston, L. A. (2000) Cell cycle: The trouble with tribbles. *Current Biology*, **10**, R502–R504.

[83] Rørth, P., Szabo, K., and Texido, G. (2000) The level of C/EBP protein is critical for cell migration during drosophila oogenesis and is tightly controlled by regulated degradation. *Mol Cell*, **6**, 23–30.

[84] Hegedus, Z., Czibula, A., and Kiss-Toth, E. (2007) Tribbles: A family of kinase-like proteins with potent signalling regulatory function. *Cellular Signalling*, **19**, 238–250.

[85] Wilkin, F., Suarez-Huerta, N., Robaye, B., Peetermans, J., Libert, F., Dumont, J. E., and Maenhaut, C. (1997) Characterization of a phosphoprotein whose mRNA is regulated by the mitogenic pathways in dog thyroid cells. *Eur J Biochem*, **248**, 660–668.

[86] Kiss-Toth, E. et al. (2004) Human tribbles, a protein family controlling mitogen-activated protein kinase cascades. *J Biol Chem*, **279**, 42703–42708.

[87] Sung, H., Francis, S., Crossman, D., and Kiss-Toth, E. (2006) Regulation of expression and signalling modulator function of mammalian tribbles is cell-type specific. *Immunology Letters*, **104**, 171–177.

[88] Kiss-Toth, E. et al. (2006) Functional mapping and identification of novel regulators for the Toll/Interleukin-1 signalling network by transcription expression cloning. *Cellular Signalling*, **18**, 202–214.

[89] Bowers, A., Scully, S., and Boylan, J. (2003) SKIP3, a novel drosophila tribbles ortholog, is overexpressed in human tumours and is regulated by hypoxia. *Oncogene*, **22**, 2823–2835.

[90] Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. (2009) Database for mRNA half-life of 19 977 genes obtained by DNA microanalysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Research*, **16**, 45–58.

[91] Sung, H. Y. et al. (2007) Human tribbles-1 controls proliferation and chemotaxis of smooth muscle cells via MAPK signaling pathways. *J Biol Chem*, **282**, 18379–18387.

[92] Yokoyama, T., Kanno, Y., Yamazaki, Y., Takahara, T., Miyata, S., and Nakamura, T. (2010) Trib1 links the MEK1/ERK pathway in myeloid leukemogenesis. *Blood*.

[93] Dedhia, P. H. et al. (2010) Differential ability of tribbles family members to promote degradation of C/EBPα and induce acute myelogenous leukemia. *Blood*.

[94] Imajo, M. and Nishida, E. (2010) Human tribbles homolog 1 functions as a negative regulator of retinoic acid receptor. *Genes to Cells*, **15**, 1089–1097.

[95] Altucci, L. and Gronemeyer, H. (2001) The promise of retinoids to fight against cancer. *Nat Rev Cancer*, **1**, 181–193.

[96] Saeed, S., Logie, C., Stunnenberg, H., and Martens, J. (2011) Genome-wide functions of PML-RARalpha in acute promyelocytic leukaemia. *British Journal of Cancer*, **104**, 554–558.

[97] Ostertag, A. et al. (2010) Control of adipose tissue inflammation through TRB1. *Diabetes*, **59**, 1991–2000.

[98] Burkhardt, R., et al. (2010) Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. *J Clin Invest*, **120**, 4410–4414.

[99] Tsai, J., Qiu, W., Kohen-Avramoglu, R., and Adeli, K. (2007) MEK-ERK inhibition corrects the defect in VLDL assembly in HepG2 cells: Potential role of ERK in VLDL-apoB100 particle assembly. *Arterioscler Thromb Vasc Biol*, **27**, 211–218.

[100] Kotzka, J., Müller-Wieland, D., Koponen, A., Njamen, D., Kremer, L., Roth, G., Munck, M., Knebel, B., and Krone, W. (1998) ADD1/SREBP-1c mediates insulin-induced gene expression linked to the MAP kinase pathway. *Biochemical and Biophysical Research Communications*, **249**, 375–379.

[101] Xiong, Y., Collins, Q. F., An, J., Lupo, E., Liu, H.-Y., Liu, D., Robidoux, J., Liu, Z., and Cao, W. (2007) p38 mitogen-activated protein kinase plays an inhibitor role in hepatic lipogenesis. *J Biol Chem*, **282**, 4975–4982.

[102] Kapranov, P., Crawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P., and Gingeras, T. R. (2002) Large-scale transcriptional activity in chromosomes 22 and 22. *Science*, **296**, 916–919.

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

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

[105] Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y. (2010) Small pepties switch the transcriptional activity of shavenbaby during drosophila embryogenesis. *Science*, **329**, 336–339.

[106] Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M., Glass, C., and Kurokawa, R. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*, **454**, 126–130.

[107] Espinoza, C. A., Goodrich, J. A., and Kugel, J. F. (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA*, **13**, 583–596.

[108] Martianov, I., Ramadass, A., Barros, A. S., Chow, N., and Akoulitchev, A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, **445**, 666–670.

[109] Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F., and Goodrich, J. A. (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell*, **29**, 499–509.

[110] Maraia, R. J., Driscoll, C. T., Bilyeu, T., Hsu, K., and Darlington, G. J. (1993) Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol. Cell. Biol.*, **13**, 4233–4241.

[111] Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J., and Lee, J. T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**, 750–756.

[112] Ringrose, L. and Paro, R. (2004) Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annu. Rev. Genet.*, **38**, 413–443.

[113] Rinn, J. L., et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.

[114] Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689–693.

[115] Sharpless, N., Bardeesy, N., Lee, K., Carrasco, D., Castrillon, D., Aguirre, A., et al. (2001) Loss of $p16^{INK4A}$ with retention of $p19^{Arf}$ predisposes mice to tumorigenesis. *Nature*, **413**, 86–91.

[116] Khalil, A. M., et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA*, **106**, 11667–11672.

[117] Donnelly, K. L., Smith, C. I., Schwarzenberg, S. J., Jessurun, J., Boldt, M. D., and Parks, E. J. (2005) Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J Clin Invest*, **115**, 1343–1351.

[118] Knudsen, S. (1999) Promoter 2.0: for the recognition of PolII promoter sequences. *Bioinformatics*, **15**, 356–361.

[119] Hirose, Y. and Manley, J. (1998) Rna polymerase II is an essential mRNA polyadenylation factor. *Nature*, **395**, 93–96.

[120] Rigo, F. and Martinson, H. (2009) Polyadenylation releases mRNA from RNA polymase II in a process that is licensed by splicing. *RNA*, **15**, 823–836.

[121] de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010) A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol*, **8**, e1000384.

[122] Heintzman, N., Stuart, R., Hon, G., Fu, Y., Ching, C., et al. (2007) Distinct and predictive chromatin signatures or transcriptional promoters and enhancers in the human genome. *Nature Genetics*, **39**, 311–318.

[123] Marques, A. C. and Ponting, C. P. (2009) Catalogues of mammalian long noncoding RNA: modest conservation and incompleteness. *Genome Biology*, **10**, R124.

[124] Ponting, C. P., Olivier, P. L., and Reik, W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.

[125] Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007) The expanding RNA polymerase III transcriptome. *TRENDS in Genetics*, **23**, 614–622.

[126] Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M., and Nakagawa, S. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J. Cell Sci.*, **120**, 2498–2506.

[127] Stewart, A. F., Richard, C. W., Suzow, J., Stephan, D., Weremowicz, S., Morton, C. C., and Adra, C. N. (1996) Cloning of human RTEF-1, a transcriptional enhancer factor - 1 related gene preferentially expressed in skeletal muscle: evidence for an ancient multigene family. *Genomics*, **36**, 68–76.

[128] Kanecko, K. and DePamphilis, M. (1998) Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. *Dev. Genet.*, **22**, 43–55.

[129] Maeda, T., Maeda, M., and Stewart, A. F. (2002) TEF-1 transcription factors regulate activity of the mouse mammary tumor virus in LTR. *Biochemical and Biophysical Research Communications*, **296**, 1279–1285.

[130] Babak, T., Blencowe, B. J., and Timothy R, H. (2005) A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription. *BMC Genomics*, **6**, 104.

[131] Peng, Y., Lei, T., Yuan, J., Chen, X., Long, Q., Zhan, J., Lei, P., Feng, B., and Yang, Z. (2009) Arachidonic acid induces acetyl-coA carboxylase 1 expression via activation of CREB1. *Endocr*, **36**, 491–497.

Appendix A - Summary of primers and vectors

Region	SNPs	PCR Primers	Sequencing Primer	Length
CNS1	rs2001845 rs2001844 rs7015677 rs57863956	For: 5′- AAATTAGTGGAGCGTGGTGG-3′ Rev: 5′- CCCACATTCAGGGTTTACAG-3′	5′- AAATTAGTGGAGCGTGGTGG-3′	324bp
CNS2	rs55921265	For: 5′-ATTTTGTAGGTGACCCGAGG-3′ Rev: 5′-CTCCTCTCAGGGTAGCATG-3′	5′-CTCCTCTCAGGGTAGCATG-3′	533bp

Table 4: Primers for CNS1 and CNS2 SNP genotyping in CHD cases and controls from the Ottawa Heart Study

Primer	Sequence	For/Rev	Location
Oligod(T) – anchor (Roche)	5′-GACCACGCGTATCGATGTC-GACTTTTTTTTTTTTTTTV-3′	cDNA: rev 5′RACE PCR: forward	Poly(A) tail
PCR anchor (Roche)	5′-GACCACGCGTATCGATGTCGAC-3′	5′RACE PCR: forward 3′RACE PCR: reverse	Anchor sequence on oligod(T) anchor primer
F13EST	5′-GCTGACATTTCTCATGAATAGC-3′	Forward	Exon 1
F12EST	5′-GCAGTAGACCAAGGCCAG-3′	Forward	Exon 1
F11EST	5′-GCCAGGAATTCTCATTAGCTG-3′	Forward	Exon 1
F10EST	5′-AGGCATGACTTGGACTGTG-3′	Forward	Exon 1
F9EST	5′-GCGGAGAGGTCATAGGTATG-3′	Forward	Exon 1
F8EST	5′-GTGAGCCACAGCTGATAGC-3′	Forward	Exon 1
F7EST	5′-GATAGCTGGGCGCGA-3′	Forward	Exon 1
F6EST	5′-GATCTGCCTGGATTGTTACT-3′	Forward	Exon 1
F5EST	5′-GAAATACTAACATGCTCTTTATGGG-3′	Forward	Exon 1
F4EST	5′-GTTAGTAACTCCCTTCACCTCC-3′	Forward	Exon 1
F3EST	5′-CCCAACGGCTCCTACC-3′	Forward	Exon 1
F2EST	5′-TACCCGCTGGGACCTT-3′	Forward	Exon 1
F1EST	5′-ACCTCCCCATGATCCAA-3′	Forward	Exon 1
EST2F	5′-ATGAAGAAATTTTGGTCAGAGAG-3′	Forward	Exon 2
EST2R	5′-CATTCTGGAAGCTCTGTTGAC-3′	Reverse	Exon 2
EST2RSP2	5′-GCTTCTTCAGGATCCCATCT-3′	Reverse	Exon 2
EST1R	5′-GCATCGCAGATCTGTGAAGGGAGTTACTAACCC-3′	Reverse	Exon 1

Table 5: Primers for 5'/3' RACE The oligod(T)-anchor and PCR anchor primers are from the 5'/3' RACE Kit, 2nd Generation (Roche). All other primers were designed according to EST exons 1 and 2. These primers were for PCR to complement and verify RACE results.

Gene	Primers
EST	For: 5′-ACCTCCCCATGATCCAA-3′ Rev: 5′-CATTCTGGAAGCTCTGTTGAC-3′
TRIB1	For: 5′-TTCAAGCAGATTGTCTCCGC-3′ Rev: 5′-AGTGGTGTTGAGGATCTCAG-3′
PPIA	For: 5′-ACCGTGTTCTTCGACATTGC-3′ Rev: 5′-TTCTGTGAAAGCAGGAACCC-3′
U1	For: 5′-ATACTTACCTGGCAGGGGAG-3′ Rev: 5′-CAGGGGGAAAGCGCGAACGCA-3′
FAS	For: 5′-CCAATGCCTTGTTCCC-3′ Rev: 5′-AGGTAGTGGTCAGGAG-3′
ACC1	For: 5′-TGGTAATGCGGTATGGAAGTCG-3′ Rev: 5′-TGTATGTTGTCCCTAAGGATTGTGC-3′
APOB	For: 5′-CTGCAGCTTCATCCTGAAGA-3′ Rev: 5′-CAGGATGTAAGTAGGTTCATC-3′
SCD1	For: 5′-ACCGTGCCCACCACAAGTTT-3′ Rev: 5′-ATACCAGGGCACAAGCGTGG-3′
ACOX1	For: 5′-ACCATTGCCATCCGATACAG-3′ Rev: 5′-GGTCTCCTTCATGTATGCGC-3′
KIAA0196	For: 5′-CCGACCTGCTCAGCGTGTCAC-3′ Rev: 5′-AGGCAACGCGCTTCACAAGC-3′
SQLE	For: 5′-GGAGCAGCTCGAGGCCAGGA-3′ Rev: 5′-ACCACCCGGCTGCAGGAATT-3′
ZNF572	For: 5′-TCTGCCTTCCGGGTTTGAGAGT-3′ Rev: 5′-TGCTCCATCACAATGATCACAGCCA-3′

Table 6: Primers for qRT-PCR *ACC1*[131] and *U1*[113] primers were taken from the literature.

Vector	Description of insert	Restriction sites	Primers
pGL4.23	CNS1 – 497bp	NheI XhoI	For: 5′-AAGCTAGCAAAAATTAGTGGAGCGTGGTG-3′ Rev: 5′-AACTCGAGTGCCCATACCTATGACCTCT-3′
pGL4.23	CNS2 – 508bp	NheI XhoI	For: 5′-AAGCTAGCCCTCAGCATCTTCTGTTTGC-3′ Rev: 5′-AACTCGAGGCTGTCTCCTCTCAGGGTA-3′
pGL3-SV40	CNS2 – 508bp	NheI XhoI	* Subcloned from pGL4.23 into pGL3-SV40 *
pGL3 basic	CNS1 – 2071bp	MluI BglII	For: 5′-GCTTAAACGCGTGAGCTAGACATGGATTACCTC-3′ Rev: 5′-GCATCGTAGATCTGATCATGGGGAGGTGAAGGT-3′
pGL3 basic	CNS1 – 1463bp	MluI BglII	For: 5′-GCTTAAACGCGTTAGATGAGTAAACCAGGGCAC-3′ Rev: 5′- GCATCGCAGATCTGTGAAGGGAGTTACTAACCC-3′
pGL3 basic	CNS1 – 1013bp	MluI BglII	For: 5′-GCTTAAACGCGTCCTGAAGAAAGCTGGCACTA-3′ Rev: 5′- GCATCGCAGATCTGTGAAGGGAGTTACTAACCC-3′
pGL3 basic	CNS1 – 557bp	MluI BglII	For: 5′-GCTTAAACGCGTGTAATCCCAGCTACTTGGGA-3′ Rev: 5′- GCATCGCAGATCTGTGAAGGGAGTTACTAACCC-3′
pGL3 basic	CNS1 – Construct 1	MluI BglII	For: 5′-GCTAGTACGCGTCACTTCCCTTTCCAAATG-3′ Rev: 5′- GCATCGCAGATCTGTGAAGGGAGTTACTAACCC-3′
pGL3 basic	CNS1 – Construct 2	MluI BglII	For: 5′- GCTAGTACGCGTCACTTCCCTTTCCAAATG-3′ Rev: 5′-GTATCGTAGATCTCGAAAGTGGGCCCTGTGT-3′
pGL3 basic	CNS1 – Construct 3	MluI BglII	For: 5′- GCTAGTACGCGTCACTTCCCTTTCCAAATG-3′ Rev: 5′-GCATCGTAGATCTAGGAATTCGCTATTCATGAG-3′
pGL3 basic	CNS1 – Construct 4	MluI BglII	For: 5′- GCTAGTACGCGTCACTTCCCTTTCCAAATG-3′ Rev: 5′-GCATCGTAGATCTCACAGTCCAAGTCATGC-3′
pCMV-EST (795bp)	5′ region – 355bp	BglII PflF1	For: 5′- GCATCGTAGATCTTAAAGCTGACATTTCTCAT-3′ Rev:5′- ATCCAAATCAGCATCAGACCC-3′
	3′RACE insert – 440bp	PflF1 MluI	For: 5′- GAAATACTAACATGCTCTTTATGGG-3′ Rev: 5′-GACCACGCGTATCGATGTCGAC-3′

Table 7: Cloning strategies for luciferase reporter and mammalian expression constructs The construction of the pCMV5-EST expression vector was complex. Briefly, 3'RACE using the gene-specific forward primer and the PCR anchor primer from the Roche 5'/3' RACE Kit (primer sequences provided) generated the 3' RACE insert, which itself contains one PflF1 restriction site. The 5' region was amplified from a 5' RACE product template using the primers given in the table. All inserts were digested with the correct restriction enzymes (given in the table) and ligated together into the pCMV5 vector.

Plasmid	Plasmid Type	Description
pGL4.23 (Promega)	Reporter	Minimal promoter luciferase reporter
pGL4.74 (Promega)	Control reporter	Weak constitutive expression of *Renilla* luciferase
pGL3-SV40 (Promega)	Reporter	Strong promoter luciferase reporter
pRL SV40 (Promega)	Control reporter	Moderate constitutive expression of *Renilla* luciferase
pGL3 basic (Promega)	Reporter	Promoter-less reporter vector; requires the insertion of a promoter to drive luciferase expression
pGL4.23 – CNS(1/2)	Reporter	The conserved sequence insert enhances transcription of luciferase from the minimal promoter
pGL3 – SV40 – CNS2	Reporter	CNS2 represses the transcriptional activity of the SV40 promoter
pGL3 – CNS1	Reporter	CNS1 acts as a promoter to activate the transcription of luciferase
pCMV5	Mammalian expression	No insert; CMV promoter
pCMV5-EST	Mammalian expression	EST (exons 1,2,7) insert; CMV promoter drives EST expression
pXJ40 – TEAD1 (Dr. A. Stewart laboratory; UOHI)	Mammalian expression	CMV promoter drives expression of TEAD1
pXJ40 – TEAD4 (Dr. A. Stewart laboratory; UOHI)	Mammalian expression	CMV promoter drives expression of TEAD4

Table 8: Plasmid types for luciferase reporter assays and mammalian expression

Adrianna Douvris
University of Ottawa Heart Institute

Education

- Master of Science in Biochemistry (2009 - Present)
  University of Ottawa, Ottawa, ON, Canada

- Bachelor of Science (Hons.) in Biochemistry (2005 - 2009)
  University of Ottawa, Ottawa, ON, Canada

- High School and International Baccalaureate Diploma (2001 - 2005)
  Colonel By Secondary School, Ottawa, ON, Canada

Conferences

- University of Ottawa Heart Institute Research Day (May 2011)
  University of Ottawa Heart Institute, Ottawa, ON, Canada
  Oral presentation (2nd prize, Basic Science Category)

- Arteriosclerosis, Thrombosis and Vascular Biology 2011 Scientific Sessions (April 28-30 2011)
  Chicago, Illinois, USA
  Oral abstract presentation

- Canadian Lipoprotein Conference (Oct 21-24 2010)
  Niagara-on-the-Lake, ON, Canada
  Oral abstract presentation

- University of Ottawa Heart Institute Research Day (May 12 2010)
  University of Ottawa Heart Institute, Ottawa, ON, Canada
  Poster presentation

Scholarships and Awards

- Ontario Graduate Scholarship (OGS) (2010-2011)
- Canadian Institute of Health Research Masters Award (CGS) (2009-2010)
- University of Ottawa Excellence Scholarship (2009-2011)

- Dean's List, University of Ottawa (2005-2009)
- University of Ottawa Entrance Scholarship (2005-2009)

Research internships

- University of Ottawa Heart Institute, Ottawa, ON (May-August 2009)
  Supervisor: Dr. Ruth McPherson
  Research Project: Functional Analysis of the *TRIB1* Locus in Coronary Artery Disease (CAD)

- University of Ottawa, Ottawa, ON (May 2008 - April 2009)
  Supervisor: Dr. Robert Haché
  Research Project: The isoform-specific adipogenic potential of the C/EBPβ Transcription Factor