Université de Montréal

Identifying the Causal SNP(s) Determining Dalcetrapib Responses

par Magdalena Burchert

Département de Biologie moléculaire

Faculté de médecine

Mémoire présenté

en vue de l'obtention du grade de

Maîtrise en Biologie Moléculaire

Option Générale

Février 2019

RÉSUMÉ

Introduction: Le dalcétrapib est un inhibiteur de la protéine de transfert des esters de cholestérol (CETP) qui augmente le niveau du cholestérol-HDL. Des études d'association pangénomiques ont révélé une association entre les polymorphismes du gène adénylate cyclase de type 9 (*ADCY9*) et les réponses au dalcétrapib. Le but de cette étude était d'identifier le polymorphisme nucléotidique (SNP) causal, ce qui pourrait mener à comprendre le mécanisme moléculaire modifiant les effets du dalcétrapib sur les bénéfices cardiovasculaires.

Méthodes: Des essais d'EMSA (electrophoretic mobility shift assay) ont été réalisés afin d'analyser les effets modificateurs de douze SNPs candidats sur la liaison de protéines nucléaires, provenant de cellules monocytaires THP-1. Ensuite, des essais de transfections avec un gène rapporteur ont été utilisées pour évaluer l'effet transcriptionnel de ces SNPs. La liaison des protéines au SNP rs12920508 a par la suite été étudiée par des chromatographies d'affinité d'ADN suivies par des spectrométries de masse et par MC-EMSA (multiplexed competitor EMSA).

Résultats: Sept sur douze SNPs ont démontré une liaison spécifique à un allèle qui n'a pas été influencée par l'exposition des cellules au dalcétrapib. Le résultat des transfections de vecteurs rapporteurs dans les cellules THP-1 a montré que les constructions plasmidiques portant les variants rs1967309 et rs12920508 augmentaient l'activité transcriptionnelle. Onze protéines ont été identifiées comme des candidats potentiels pouvant se lier à la région du SNP rs12920508. De plus, la région contenant les deux variants rs1967309 et rs12920509 a présenté une activité transcriptionnelle accrue et significativement plus élevée pour l'haplotype délétère.

Conclusion: Le polymorphisme rs1967309 semble causer la majorité des effets fonctionnels dans la lignée cellulaire THP-1. Cependant, une interaction avec le SNP rs12920508 ou la présence de la région de ce SNP pourrait être nécessaire pour l'activité optimale de rs1967309. Des travaux supplémentaires sont nécessaires pour élucider le lien entre le SNP potentiellement causal et les réponses cardiovasculaires induites par le dalcétrapib.

Mots clés: Dalcétrapib, *ADCY9*, SNP, EMSA, maladie cardiovasculaire, HDL-C

ABSTRACT

Introduction: Dalcetrapib is a cholesteryl ester transfer protein (CETP) inhibitor that increases the circulating level of HDL-cholesterol. Genome-wide association studies have revealed an association between polymorphisms found in the adenylate cyclase type 9 (*ADCY9*) gene and responses to dalcetrapib, including its cardiovascular benefits. The purpose of this study was to identify the causal single nucleotide polymorphisms (SNP) which could lead to understand the molecular mechanisms altering dalcetrapib effects on cardiovascular outcomes.

Methods: Electrophoretic mobility shift assays (EMSA) were performed to analyze the allele-specific effects of the best causal SNP candidates on binding with nuclear proteins obtained from a THP-1 monocytic cell line. Afterwards, a dual luciferase reporter assay was used to assess the effect of selected genetic variants on gene transcription. Protein binding to SNP rs12920508 was investigated by DNA-affinity chromatography followed by mass spectrometry and multiplexed competitor EMSA.

Results: Seven out of 12 SNPs demonstrated allele-specific protein binding, which was not influenced by dalcetrapib exposure of the cells. Results from dual luciferase reporter assay showed that plasmid constructs bearing variants rs12920508 and rs1967309 increased transcriptional activity when transfected into THP-1 undifferentiated monocytic cells. Eleven proteins were identified as potential candidates binding to region of SNP rs12920508. Additionally, region containing both SNPs rs1967309 and rs12920508 displayed increased transcriptional activity with significantly higher activity for deleterious haplotype.

Conclusion: Polymorphism rs1967309 seems to be causing most functional effects in the THP-1 monocytic cell line. However, an interaction with rs12920508 or presence of the DNA region of this SNP may be necessary for optimal activity of rs1967309. Further work is required to elucidate the link between potentially causal SNPs and cardiovascular responses induced by dalcetrapib.

Keywords: Dalcetrapib, *ADCY9*, SNP, EMSA, cardiovascular disease, HDL-C

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAV8 : adeno-associated virus serotype 8

ABCA-1 : ATP-binding cassette transporter 1

ABCG-1 : ATP-binding cassette sub-family G 1

AC : adenylyl cyclase

ACS : acute coronary syndrome

ADCY9 : adenylate cyclase type 9

ApoA-I : Apolipoprotein AI

ApoA-II : Apolipoprotein AII

ApoB: Apolipoprotein B

Ago2 : Argonaute 2

ARNT : aryl hydrocarbon receptor nuclear translocator

ATP : -adrenergic receptor

β2AR : β2-adrenergic receptor

BMI : body mass index

cAMP : cyclic adenosine monophosphate

CCR2 : C-C chemokine receptor type 2

CCL2 : C-C motif chemokine ligand 2

cDNA : complementary DNA

cd5/p35 : cyclin-dependent protein kinase 5/p35 complex

CE : cholesteryl esters

CETP : cholesteryl ester transfer protein

ChIP : chromatin immunoprecipitation

CK1 : casein kinase 1

CoQ10 : ubiquinone

CRISPR : clustered regularly interspaced short palindromic repeats

CVD : cardiovascular diseases

DGAT2 : hepatocyte microsomal diacylglycerol

DHSs : DNase I hypersensitivity site

EMSA : electrophoretic mobility shift assay

eNOS : endothelial nitric oxide synthase

eQTL : expression quantitative trait loci

GSPx : glutathione selenoperoxidase

GTP : guanosine-5'-triphosphate

GWAS : genome-wide association study

HDL : high-density lipoproteins

HL : hepatic lipase

HMG-CoA : 3-hydroxy-3-methyl-glutaryl-coenzyme A

HOMER1 : homer scaffolding protein 1

hs-CRP : high-sensitivity C reactive protein

Hsp20 : heat shock protein 20

ICAM : Intercellular adhesion molecule I

IGF-1 : insulin-like growth factor 1

IL-1β : interleukin 1 β

IL-6 : interleukin 6

IMT : intima-media thickness

LCAT : lecithin-cholesterol acyltransferase

LD : linkage disequilibrium

LDL : low-density lipoproteins

LDLR : low-density lipoprotein receptor

lincRNA : long intergenic non-coding RNA

lncRNA : long non-coding RNA

MCP-1 : monocyte chemotactic protein-1

M-CSF : macrophage colony stimulating factor

miRNA : micro RNA

mRNA : messenger ribonucleic acid

mTOR2 : rapamycin complex 2

Myo II : myosin II

ncRNA : non-coding RNA

NO : nitric oxide

PAF : platelet-activating factor

PCSK9 : proprotein convertase subtilisin/kexin type-9

PDGF : platelet-derived growth factor

PGI2 : prostacyclin

PIC : transcription preinitiation complex

PKA : protein kinase A

PKC : protein kinase C

PLTP : phosholipid transfer protein

Pol II : RNA polymerase II

PON-1 : paraoxonase 1

PPARα : peroxisome proliferator-activated receptor-α

pre-mRNA : precursor mRNA

pri-mRNA : primary RNA

PSL : piglet splay leg

P-TEFb : positive transcription elongation factor b

qPCR : quantitative polymerase chain reaction

RBP : RNA-binding proteins

RCT : reverse cholesterol transport

RISC : RNA-induced silencing complex

ROS : reactive oxygen species

SEC : super elongation complex

SMC : smooth muscle cells

SNP : single nucleotide polymorphism

SR-B1 : scavenger receptor class type 1

TAE : Tris-acetate buffer

TBE : Tris-borate buffer

TBP : TATA-binding protein

TF : transcription factor

TG : triglyceride

TM : transmembrane

TNFα : tumor necrosis factor α

TSS : transcription start site

UTR : untranslated region

VCAM-1 : vascular cell adhesion protein 1

VLDL : very-low-density lipoprotein

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation towards all the great people I encountered during my master studies. First of all, I would like to thank my director Dr Jean-Claude Tardif and my co-director Dr Eric Rhéaume for giving me the opportunity to work on such an interesting and revolutionary project. Their extensive knowledge and valuable advices guided me throughout my master studies. I would also like to thank members of the Dr Tardif's laboratory, in particular Dr Rocio Sanchez and Gabriel Théberge-Julien for teaching me new laboratory techniques. Moreover, I have to thank former master students Julien Renaud and Matthieu Blanchard for helpful tips about various laboratory techniques and administrative part of master program.

Outside of our laboratory, I would like to thank Isabel Gamache for all the lunches eaten together and long discussions about ADCY9 and favorite SNPs.

Finally, I would like to thank my parents and my fiancé for their support. Their continuously encouraged me during my studies and never stopped believing in me.

1 INTRODUCTION

1.1 Atherosclerosis

Cardiovascular disease is the leading cause of human death in North America and other developed countries. In most cases, the major cause of the cardiovascular event is atherosclerosis. Atherosclerosis is a chronic inflammatory disease characterized by the accumulation of lipids, inflammatory cells and fibrous elements in medium-sized and large arteries. In the heart, atherosclerosis can result in myocardial infarction and heart failure caused by coronary artery stenosis. In the brain, rupture of atherosclerotic plaques or stenosis may induce a stroke [1-3].

1.1.1 Risk factors for atherosclerosis

Numerous risk factors for atherosclerosis were identified over the past years. Some of them, called modifiable risk factors, can be controlled in order to delay or prevent atherosclerosis progression. To the modifiable factors we can include: elevated levels of cholesterol and low-density lipoprotein (LDL), reduced levels of high-density lipoprotein (HDL), cigarette smoking (>10/day), hypertension, diabetes, severe obesity (>30% overweight) and inactive lifestyle [4]. The other risk factors such as sex, age, ethnicity and family history cannot be regulated and are therefore called non-modifiable risk factors.

The majority of enumerated risk factors are not independent and contribute to global risk profile. For example, diabetic patients often exert elevated levels of low-density lipoprotein-cholesterol (LDL-C) and hypertension that might contribute to atherogenesis. Also, active lifestyle is known to have protective effect on atherosclerosis, likely due to reduced blood pressure and body weight, increase in high-density lipoprotein-cholesterol (HDL-C) levels and decrease in LDL-C. Therefore, understanding the global risk profile of a patient could provide improved predictive power for atherosclerosis [5-7].

In addition, total risk profile of atherosclerosis is complicated even more by presence of genetic factors. One of the first genetic factors affecting cardiovascular diseases was discovered in a patient with hypocholesterolemia and his mother. Both patients possessed a 5-kb deletion in low-density lipoprotein receptor (LDLR) gene that result in reduced production of LDLR on the cell surface. Decreased levels of LDLR subsequently impair receptor–mediated hepatic uptake of LDL and lead to elevated levels of circulating cholesterol [8]. Another family-based study

identified mutations in two additional genes, ApoB and PCSK9, which cause hypercholesterolemia and elevated levels of LDL [9, 10].

1.1.2 Pathogenesis of atherosclerosis

Atherosclerosis is a chronic inflammatory condition, which remains unnoticed during its first phases of development. Formation of lesions begins with endothelial dysfunction, endothelial cell activation, accumulation of atherogenic LDL, inflammation and may lead to an acute clinical event induced by plaque rupture and thrombosis.

Endothelial dysfunction

The basic structure of blood vessel is divided into three morphologically distinct layers. The outermost layer called adventita is built from connective tissues with scattered fibroblasts and smooth muscle cells (SMCs). The middle layer, media, consists essentilally of SMCs. The innermost layer, which is in direct contact with the blood flow is called the intima. On the luminal side, the intima consists of a monolayer of endothelial cells, which play an important role in development of atherosclerosis [2]. The endothelium regulates vascular homeostasis, serving as a selectively permeable barrier between blood and tissues. The endothelium also produces a vast range of factors, which regulate processes affecting creation of atherosclerotic lesions, such as vascular tone, adhesion of cells, thromboresistance, proliferation of SMCs, and inflammation of vessel wall. One of the most important factors released by endothelium is a potent vasodilatator called nitric oxide (NO). Nitric oxide opposes the effect of endothelium-derived vasoconstrictors and inhibits oxidation of low-density lipoproteins. Defect in its production or activity leads to endothelial dysfunction [11, 12]. Oxidative stress, which may originate from smoking, hypertension, hyperlipidemic states and diabetes, can also interfere with the production and activity of nitric oxide and as a result provoke an inflammatory response, vascular remodeling and higher permeability of endothelium [3].

Monocyte adhesion and accumulation of LDL

Beside causing increased permeability of endothelium, endothelial dysfunction is also implicated in the inflammatory response, known as endothelial cell activation. Factors inducing endothelial cell activation are: certain bacteria and viruses, pro-inflammatory cytokines such as interleukin 1 and tumour necrosis factor, physical and oxidative stress, oxidised low density lipoproteins.

During endothelial cell activation, the endothelium express cell surface molecules, such as vascular cell adhesion protein 1 (VCAM-1), intercellular adhesion molecule 1 (ICAM-1), and endothelial leukocyte adhesion molecule (E-selectin). Expression of these surface molecules facilitates the recruitment and attachment of circulationg monocytes and T-lymphocytes to the vessel wall [13, 14] .

Increased permeability of arterial wall also favors accumulation of atherogenic lipoproteins, mainly LDL, which extravasate through the leaky and defective endothelium into the subendothelial space. Retained lipoprotein particles undergo oxidative modifications and become cytotoxic, proinflammatory and proatherogenic. The molecules responsible for such modifications are reactive oxygen species (ROS) as well as myeloperoxidases and lipoxygenases released by inflammatory cells and responsible for the production of HOCl and oxygenated lipids, respectively [15, 16].

Plaque progression

Oxidized lipids trigger secretion of chemokines (chemotactic cytokines) by overlying endothelial cells. Next, chemokines direct transmigration of adherent monocytes across endothelium into the intima. Monocyte chemotactic protein-1 (MCP-1) is the most important atherogenic chemoatractant. Its receptor on monocytes C-C chemokine receptor type 2 (CCR2) may be upregulated significantly during plaque development. Monocytes located within subendothelial space, differentiate into macrophages. The macrophages express scavenger receptors that bind oxidized lipoproteins. The expression of scavenger receptors is influenced by other important cytokines secreted by endothelial cells, namely macrophage colony stimulating factors (M-CSF) [2]. These receptors are not down-regulated by cholesterol accumulation, thus the macrophages internalize continously modified LDL, what leads to foam cell formation [1]. Foam cells trapped within arterial intima ultimately die and contribute to the formation of destabilizing lipid-rich core in the plaque. Transformation of the macrophages to foam cells may be inhibited by removal of excess cholesterol by HDL in a process called reverse cholesterol transport (RCT) [2].

Macrophages and T cells located within lesions produce growth factors, such as insulin-like growth factor 1 (IGF-1) and platelet-derived growth factor (PDGF), which induce smooth muscle cells (SMC) migration to arterial intima [17]. Subsequently, SMCs proliferate and produce collagen-rich matrix which cause plaque increase in size. Advanced lesions can grow sufficiently

large to block completely blood flow [2]. Nevertheless, smooth muscle cells confer stability to plaques, protecting them from more important consequences such as plaque rupture that can lead to thrombosis [15].

Additionally, cholesterol crystals can form in foam cells and cause release of interleukin 1 β (IL-1β), which further provokes SMC to produce interleukin 6 (IL-6). Both inteleukins are pro-inflammatory, and induce synthesis of C-reactive protein (CRP) in the liver. Increased production of CRP results in its secretion into the circulation and subsequent rise in CRP serum concentration. Since the concentration of CRP in the serum is known to correlate with the occurence of cardiovascular disease, CRP levels are considered as a risk marker [18].

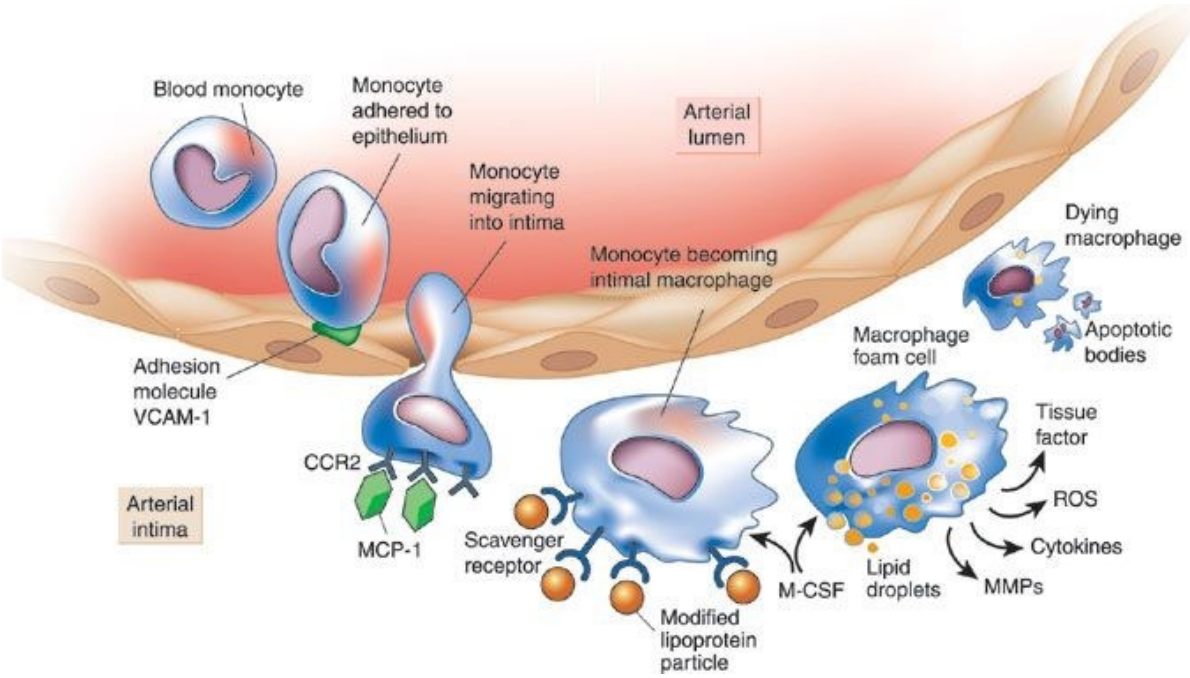

Figure 1 : Monocytes in atherogenesis. Monocytes are recruited to defective endothelium by adhesion molecules. Adherent monocytes migrate into the intima and differentiate into macrophages. Macrophages internalize oxidized low-density lipoproteins and become foam cells. Trapped foam cells eventually die and contribute to atherosclerotic plaque formation. Figure presented with permission (Peter Libby, 2002) [19].

1.2 Existing treatments

Prevention is an important part of atherosclerosis treatment. Change of lifestyle including modification of diet, stimulation of physical activity and suspension of smoking may greatly reduce progression of the disease. However, if the modification of lifestyle is not sufficient, introduction of medical therapy is neccesary.

The most popular therapies reducing cardiovascular events focus on lowering the levels of atherogenic LDL cholesterol. Although LDL-C-lowering medications significantly improve cardiovascular health of the patients, remaining risk needs to be targeted by modulation of other factors, such as triglycerides (TGs) and HDL-C. In addition, there is more and more evidence that therapy outcomes vary depending on the patient genotype. Therefore, the future of cardiovascular treatment should focus on personalized medicine to deliver right treatment to the patients, taking into account their genetic profiles [20].

LDL-C-lowering treatments

Until now 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) reductase inhibitors (**statins**) are the most effective and the most often prescribed medicines for atherosclerosis. Statins competitively inhibit HMG-CoA reductase and cause an elevated expression of LDL receptors on the hepatocytes surface. Upregulated expression of LDL receptors then leads to decreased plasma levels of LDL-C and other apo B-containing lipoproteins. Apart from their lipid lowering action, statins have been reported to exert anti-inflammatory activity by lowering serum levels of C-reactive protein [21]. Statins are generally safe and well-tolerated in patients. At the maximum dosage, the most effective statins reduce LDL-C levels by 55% to 60% [22]. Clinical trials with patients with and without coronary heart disease, reported that statins decrease the relative risk of major coronary events by ~30%. Also, patients with higher baseline risk, benefit more from statin treatment [23].

Although statins greatly reduce cardiovascular risk, patients often stop statin therapy due to adverse effects. Two causally related adverse effects of statins are myopathy and impairement of insulin resistance. Myopathy is the most common adverse effect of statins, and is caused by depletion of muscle ubiquinone (CoQ10), what results in impairment of mitochondrial function. In relation to insulin resistance, statins increase the risk of incident diabetes by ≈ 9% to 28% [24].

Ezetimibe is another medicine that reduces levels of LDL-C by inhibiting cholesterol absorption. No significant adverse effects were observed for ezetimibe thus it is ofen used as second-line therapy in patients with contraindications to statins [20]. Also, therapy combining ezetimibe with statins was found to lower the risk of cardiovascular events in high-risk patients in comparison to statin therapy alone [25].

Bile acid sequestrants can produce a reduction in LDL-C of 18-25%, when used at the highest dose. But their use is limited due to gastrointestinal secondary effects and interactions with other often prescribed drugs [20].

Recently discovered proprotein convertase subtilisin/kexin type-9 **(PCSK-9) inhibitors** are promising medicines, which significantly decrease LDL-C levels. PCSK9 was found to regulate LDL receptor degradation. Its inhibition prevents degradation of LDL receptors, improves absorption of LDL-C particles by liver and therefore decreases LDL-C plasma concentration. The limiting factors for PCSK-9 inhibitors are their current relatively high costs and lack of solid outcomes in large randomised controlled trials [26].

TG-lowering treatments

Beside LDL-lowering ability, **statins** also reduce plasma levels of TG-rich particles throughout inhibition of HMG-CoA reductase [21].

Another type of drugs which reduce concentrations of triglycerides are the agonists of peroxisome proliferator-activated receptor-α (PPARα), called **fibrates**. PPARα is a nuclear hormone receptor, which binds to DNA and modulates the transcription of genes involved in lipid and glucose metabolism. Fibrates decrease the plasma levels of TG by stimulating the catabolism of TG-rich lipoproteins [27].

HDL-C-increasing treatments

Treatment with **fibrates** is also associated with a moderate increase in HDL-C concentrations. Raise of HDL-C levels by fibrates results from the increased synthesis of the major HDL apolipoproteins, apoA-I and apoA-II , mediated by PPARα. Most studies conducted with fibrates reported coronary risk reduction after fibrate treatment [28]. However, the response of hypercholesterolemic patients to fibrates was found to vary. For example, patients with higher body mass index (BMI>29) and with higher probability of having the metabolic syndrome

benefited more from fibrate treatment. Also, fibrate treatment was associated with significantly less ischemic heart disease in patients with diabetes [29, 30]. The Helsinki Heart Study analysed the baseline levels of HDL-C in patients treated with fibrate (gemfibrozil) and found that patients with lower levels of baseline HDL-C showed greatest benefit from treatment [31]. However, there is controversy as to wheter the beneficial effects of fibrates are related to the increase of HDL-C or to an absolute reduction in LDL-C [32]. Surprisingly, in some patients with dyslipidemia, fibrate treatment resulted in increase of LDL-C levels [33]. Therefore, treatment with fibrates should be prescribed with consideration of patient profile.

Niacin increases the levels of HDL-C throught non-competitive inhibition of enzyme involved in triglyceride synthesis, called hepatocyte microsomal diacylglycerol (DGAT2). Additionally, niacin raise HDL-C by selective inhibition of apoA-I uptake. In two recent clinical trials (AIM-HIGH and HPS2-THRIVE) niacin failed to significantly reduce cardiovascular events. In addition, niacin treatment is limited by adverse side-effects [34].

Another promising group of medications increasing levels of HDL-C are the cholesteryl ester transfer protein **(CETP) inhibitors**. CETP inhibitors increase HDL-C levels by inhibition of cholesteryl esters transfer from HDL to apoB-containing proteins, and are described in detail in chapter 1.3.

1.3 HDL and their atheroprotective properties

Although statins constitute an essential part of the standard of care forin cardiovascular disease (CVD) secondary prevention, there remains a significant CV risk in these patients. Therefore, development of new therapies with different targets is necessary [35]. Low plasma concentrations of high-density lipoprotein cholesterol are commonly associated with higher risk for cardiovascular disease in patients. Thus, raising the levels of HDL-C became a target of many novel therapies. Disappointingly, the results from the majority of clinical trials demonstrate so far that despite significant rises of HDL-C levels in patients, HDL-raising drugs show no significant improvement in cardiovascular events. Such results emphasize the complexity of high-density lipoprotein molecule structure, metabolism and function. It is important for further studies in HDL field to consider different subclasses of HDL particles and their various involvement in atheroprotective mechanisms, such as reverse-cholesterol transport [34].

1.3.1 The HDL hypothesis and epidemiological studies

Early studies in 1951, reported by Barr et al. showed that patients with coronary artery disease have reduced plasma levels of high-density lipoprotein cholesterol [36]. A subsequent Framingham study of Gordon et al. in 1977, demonstrated that low level of HDL-C is a risk factor for coronary heart disease [37]. These discoveries initiated multiple HDL studies, which reported numerous beneficial effects of increased levels of HDL-C and led to formulation of the HDL hypothesis. This hypothesis states that the intervention to raise plasma levels of HDL-C protects against atherosclerosis. Following this belief, strategies to raise plasma HDL levels were developped but they did not bring the expected results so far. Therefore, the HDL hypothesis started to be questioned.

In the last few decades many epidemiological studies supported the HDL hypothesis [38]. A meta-analysis of four large studies, including in total 15252 individuals, showed that each $1\,mg/dL$ raise of HDL-C levels was associated with a 2-3% decrease in risk of cardiovascular disease [39]. However, the fact that many factors affect both CVD risk and HDL-C levels brings uncertainty to the causality of HDL-C alone. For example, it was shown that women have on average higher levels of HDL-C than men [40]. Smokers show 14% lower levels of HDL-C than nonsmokers [41]. Also, it was reported that regular aerobic exercise raise HDL-C levels by around $2.5\,mg/dl$ [42]. Moreover, the abdominal obesity is associated with lower HDL-C levels, whereas weight loss is associated with raise of HDL-C levels [43, 44]. Finally, systemic inflammation that is a known risk factor for CVD was shown to be associated with low levels of HDL-C [45].

Studies investigating human genetics of HDL led to better understanding of HDL metabolism, but also suggested that HDL-C may not be significantly associated with cardiovascular risk. Three mendelian disorders causing the loss of the capacity to produce mature HDL, were identified up to date [46]. These include the apoA-1 structural mutations, ABCA1 deficiency (Tangier disease), and familial LCAT deficiency. Although, studying these disorders helped identification of genes and proteins playing important role in regulation of HDL-C levels, they did not demonstrate convincing association with increased risk of CVD.

The other evidence causing doubt in HDL hypothesis are the clinical trials with HDL-raising drugs, niacin and CETP inhibitors [47]. Niacin is a medication used to raise HDL-C levels and

lower LDL-C and TGs. A trial called Coronary Drug Project reported that niacin reduced cardiovascular events in hypercholesterolemic men [48]. Two more recent trials AIM-HIGH and HPS2-THRIVE, which studied treatment of patients with cardiovascular heart disease who had well-controlled LDL-C levels, failed to demonstrate significant reduction of cardiovascular events. However, there is uncertainty associated with interpretations of these two trials. The number of individuals taking part in AIM-HIGH study was relatively low, and there was a modest increase in HDL-C levels in the treated group compared to placebo [49]. The HPS2-THRIVE trial was larger but included a additional drug called laropiprent, which reduces the flushing associated with nacin [50]. Although HPS2-THRIVE study failed to demonstrate significant association of niacin therapy with lower cardiovascular risk, a tendency was observed with significant benefit in a group of patients with higher baseline LDL-C levels.

Another HDL-raising group of drugs, which failed to confirm HDL hypothesis is called CETP inhibitors. Torcetrapib was the first CETP inhibitor to enter in a phase III clinical trial called ILLUMINATE [51]. In the ILLUMINATE study, torcetrapib raised HDL-C levels by 72.1 %, but this trial was prematurely terminated due to elevated rate of cardiovascular events and mortality in the torcetrapib group. Such unexpected results were further explained by the fact that torcetrapib has off-target effect on blood pressure and aldosterone that may explain the increased cardiovascular risk [52]. What is more, group of researchers investigating effects of torcetrapib in animal species, claimed that the blood pressure response caused by torcetrapib is independent on CETP inhibition [53]. They showed that torcetrapib increases blood pressure to an equivalent extent in normal and CETP transgenic mice. Also, the same group demonstrated that treatment of CETP transgenic mice with another CETP inhibitor, anacetrapib results in an equal rise of HDL-C like with torcetrapib treatment, but has no influence on blood pressure. Explanations of failure of niacin and torcetrapib trials and the fact that HDL has been demonstrated with many anti-atherogenic properties (described in section “Atheroprotective properties of HDL”) bring hope for HDL-raising therapies.

1.3.2 HDL composition and structure

Plasma HDL represents an heterogenous group of small discoidal and spherical particles. The HDL particles distinguish themselves from other lipoprotein classes by they small particle size (7-12 nm diameter) and high protein content (30-70% by weight). The HDL proteins may be divided into four groups: apolipoproteins, enzymes, lipid transfer proteins and minor proteins that

make up <5% of total HDL proteins. Apolipoproteins and enzymes play an important role in HDL metabolism and function. Additionally, minor proteins, such as acute phase response proteins and proteins involved in regulation and protection against infectious disease gained increased attention in the past few years. The major HDL structural apolipoprotein is apolipoprotein (apo) A-I. The ApoA-I accounts for around 70% of total HDL proteins, and is significantly involved in HDL biogenesis, and function. Apolipoprotein (apo) A-II is the second main HDL apolipoprotein, and accounts for around 20% of total HDL proteins. Less aboundant HDL apolipoproteins include among others: apoA-IV, apoA-V, apoC-I and apoE. Some of the enzymes carried by HDL particles are involved in important mechanisms. For example, lecithin/cholesterol acyltransferase (LCAT) plays and important role in lipid metabolism. Other enzymes, paraoxygenase 1 (PON1), platelet-activating factor-acetyl hydrolase (PAF-AM) and glutathione selenoperoxidase (GSPx) have antioxidative activities. The HDL lipidome comprises 40-60% of phospholipids, 30-40% of cholesteryl esters, 5-12% of triglycerides and 5-10% of free cholesterol. Additionally, more than 200 individual lipid species were identified in HDL particles due to advances in lipidomic analyses. HDL was also found to carry multiple copies of microRNAs that may be delivered to cells and tissues [54].

The subclasses of plasma HDL particles differ significantly in their physicochemical properties, metabolism and biological function. The heterogeneity in HDL particles size and structure is caused by different conformations of apoA-I, induced by different quantity of attached lipids [55]. The smallest class of HDL particles is called pre-β-1 HDL. This subclass is represented by lipid-poor (lipid content $\leq 30\%$), $\leq 8\,nm$ diameter discoid particles. Pre-β-1 HDL particles mainly contain apoA-I together with small amounts of phospholipids and free cholesterol. Larger class of HDL particles, called spherical HDL are > $8\,nm$ in size and contain hydrophobic core of cholesteryl esters and triglycerides. Two subclasses of spherical HDL can be recognised, HDL2 and HDL3. The HDL2 are larger, lipid-rich particles, containing: 9% of TGs, 18% of CE, 6% of free cholesterol and 25% of PLs, of total HDL mass. The HDL3 particles are smaller and denser than HDL2, and contain lipid-poor HDL subfraction, including: 7% of TGs, 14% of CEs, 3% of free cholesterol and 23% of PLs of total HDL mass [54].

In-depth understanding of different properties of HDL subclasses is crucial for comprehension of HDL protective properties. A recent meta-analysis of two large studies (Framingham Offspring

Study and Jackson Heart Study) showed that HDL2 and HDL3 subclasses were differently associated with cardiovascular events. Only HDL3 was found to significantly reduce cardiovascular risk [56].

1.3.3 HDL metabolism

The main function of HDL proteins is to transport cholesterol and other lipids between circulating cells, lipoproteins, tissues and organs. During this process, HDL particles undergo dynamic remodeling (Figure 2) [54]. At first small precursors of mature HDL, lipid-free apoAI and lipid-poor pre-β-1 HDL particles are synthesized by liver and intestine. Next, these particles collect increasing quantities of free cholesterol from peripheral tissues, via ATP binding cassette transporter 1 (ABCA-1) mediated efflux. Afterwards, lecithin/cholesterol acyltransferase (LCAT) transfers fatty acid residues from lecithin to the hydrozyl group of cholesterol, what results in formation of cholesteryl esters. Cholesteryl esters have hydrophobic properties, therefore they migrate into the hydrophobic core of HDL particles, converting discoidal HDL into lighter and larger spherical HDL3. Further, acquired cholesterol can be converted to cholesterol esters by LCAT enzyme and create less dense spherical HDL2 particles. Phospholipids for the LCAT reaction are transfered from VLDL into HDL by phospholipid transfer protein (PLTP) [57]. Spherical HDL particles can promote cholesterol efflux from peripheral cells, such as macrophage foam cells, by aqueous diffusion, ATP-binding cassette sub-family G member 1 (ABCG-1) or scavenger receptor class type 1 (SR-B1) [58]. HDL2 then transfer cholesterol esters to the liver through SR-B1 receptor. After HDL2 particles return CE to liver they become HDL3 particles again. The other pathway which involves HDL2 and HDL3 is the CETP-mediated transfer of cholesteryl esters from spherical HDL to apoB-containing lipoproteins, such as LDL and VLDL. Throughout heteroexchange of cholesteryl esters and triglycerides between HDL and apoB-containing lipoproteins, TG-rich particles are created. These TG-rich HDLs can be hydrolysed by hepatic lipase (HL) to small TG-rich particles. Next, upon action of both CETP and HL, the HDL size is reduced and lipid-poor HDL particles are generated, which then can interact with ABCA1 in next lipidation cycle [59]. Therefore, the HDL lipids are catabolised in the liver upon selective uptake via scavenger receptor class type 1 (SR-B1), or upon CETP-mediated transfer to apoB-containing particles.

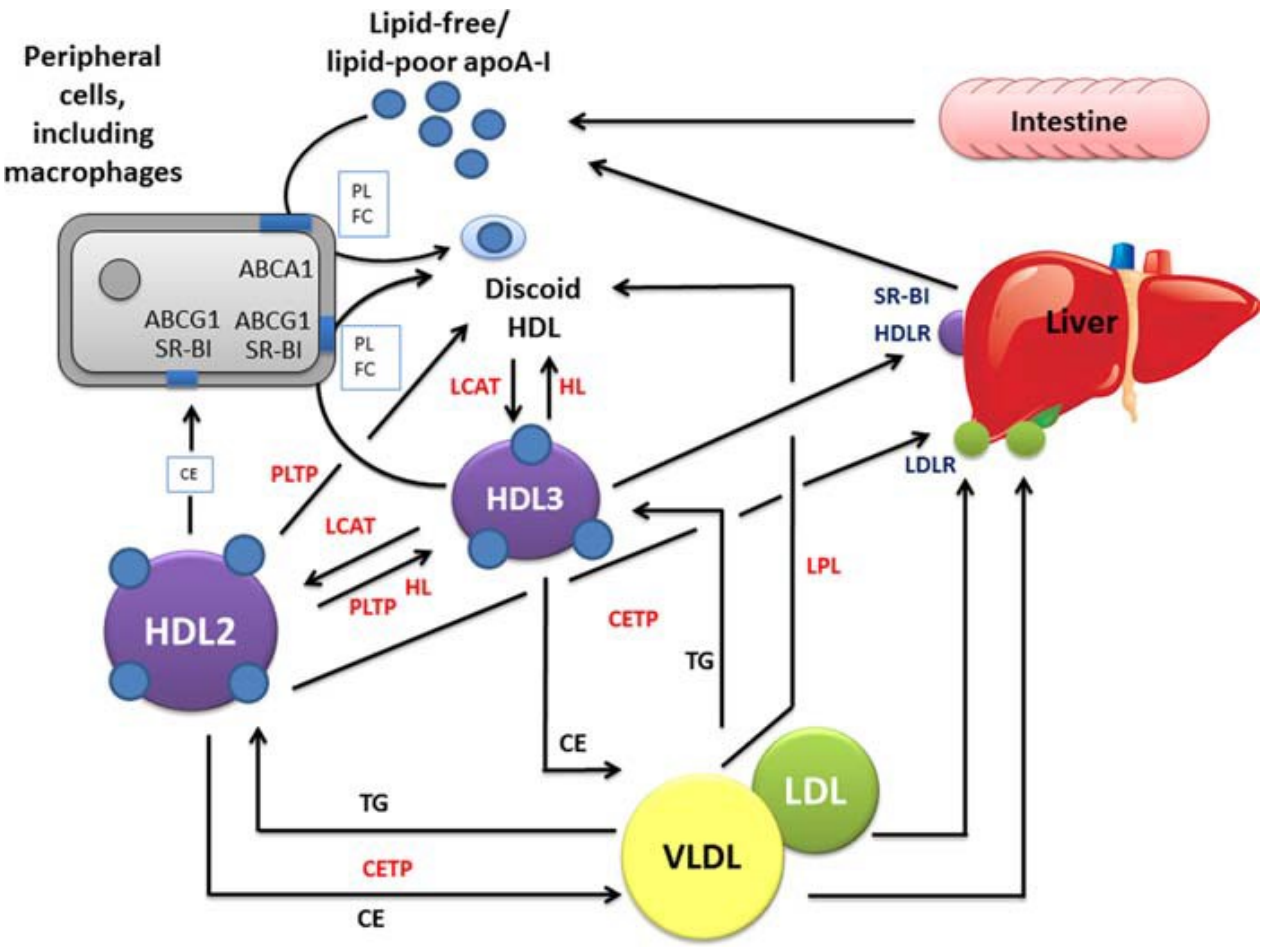

Figure 2: Principal stages of HDL metabolism. HDL discoidal particles, synthesised by the liver acquire cholesterol from peripheral cells by ABCA1-mediated cholesterol efflux. Free cholesterol in discoidal HDL is converted into cholesterol ester by LCAT. Cholesterol esters migrate into hydrophobic core of HDL and generate spherical HDL. Mature HDL can further acquire free cholesterol from peripheral tissues by aqueous diffusion or cholesterol efflux mediated by either SR-B1 receptor or ABCG1. Finally, HDL cholesterol esters and free cholesterol are transferred to the liver, via SR-B1 receptor for elimination. Figure presented with permission (Fabian H. Rached, 2015) [54].

1.3.4 Atheroprotective properties of HDL

The main objective of HDL research nowadays concentrates on understanding the mechanisms by which HDL protects against atherosclerosis. The most important antiatherogenic function of HDL is reverse-cholesterol transport (RCT), a process crucial for HDL particle maturation. In RCT process, HDL particles collect free cholesterol from lipid-rich macrophage foam cells and as a result facilitate plaque regression. Acquired cholesterol is further excreted from organism by the liver and biliary system [58]. Epidemiological studies showed strong inverse association between cholesterol efflux capacity and coronary and peripheral atherosclerosis [34].

Reverse cholesterol transport starts with the cholesterol efflux from macrophages to HDL particles. Four cellular efflux pathways were identified, which involve key proteins, such as SR-B1, ABCG1 and ABCA1. Different subclasses of HDL particles were shown to interact differently with these proteins, possibly influencing the cholesterol efflux capacity. The first pathway that mediates the bidirectional flux of cholesterol between the cell plasma membrane and HDL is called aqueous diffusion. The direction of mass transport of cholesterol is determined by the cholesterol concentration gradient in the donor and acceptor molecules [60]. Since, cholesterol efflux in this pathway is not significantly affected by the size of HDL particle, the efflux by aqueous diffusion is equally effective for all HDL subclasses [61]. In the second pathway, transfer of cholesterol to HDL particles is facilitated by SR-B1 [62]. HDL binds to SR-B1 and forms complex, which possess hydrophobic channel via which cholesterol molecules may diffuse. Larger HDL particles are more effective in mediating cholesterol efflux by this pathway, because they bind better to SR-B1 receptor [63, 64]. In the third pathway, ABCG1 increases the pool of free cholesterol in plasma membrane and reorganises it so it desorbs better into the extracellular medium. Larger HDL2 and smaller HDL3 are equally effective acceptors in this pathway. Discoidal HDL, created by the apoAI/ABCA1 reaction are also an effective acceptors of cholesterol effluxed by ABCG1 [65, 66]. Last pathway for cellular cholesterol efflux is mediated by ABCA1 receptor. In this case pre-β-1 HDL particles are the main acceptors. Subsequently, cholesterol collected by HDL particles is transferred through the plasma compartment by diverse HDL remodeling steps described in the previous section, "HDL metabolism", and is delivered to the liver. The cholesterol uptake by the liver is mediated by SR-B1 receptor expressed at hepatocytes' surface. HDL particles bind to the receptor and free cholesterol and cholesterol esters are diffused into the cell plasma membrane. Large HDL particles (10 nm diameter) bind better to SR-B1 receptor, thus deliver more cholesterol ester than small HDL particles (8 nm diameter) [57].

Other antiatherogenic properties of HDL include its antioxidant, anti-inflammatory and anti-thrombotic functions. HDL exert antioxidant activity thanks mainly to its associated enzyme called paraoxonase-1 (PON1). PON1 reduces formation of lipid peroxide known to oxidize LDL. Oxidized LDL accumulates on endothelium and disrupts its structural integrity and function. Thus, PON1 activity decrease the risk of cardiovascular event [34, 67]. Anti-inflammatory properties of HDL concern the ability of HDL particles to inhibit expression of adhesion

molecules in endothelial cells, such as VCAM-1, ICAM-1 and E-selectin. Down-regulation of these molecules causes reduced recruitment of monocytes to the arterial wall [34, 68]. Few HDL functions are responsible for their anti-thrombotic properties. HDL upregulate the expression of NO synthase (eNOS), which reduces vasorelaxation. HDL also activate prostacyclin (PGI2). PGI2 is known to inhibit platelet activation and proliferation of smooth muscle cells [34].

1.4 CETP inhibitors

CETP inhibitors belong to the group of drugs developed to increase HDL-C levels in the blood. The interest in CETP inhibition as a therapeutic approach began with observation that patients with mutations in the CETP gene possess high levels of HDL-C and often a substantial reduction in LDL-cholesterol and apoB levels. According to many epidemiological studies, elevated concentration of HDL-C and decrease in LDL-C concentration reduce the cardiovascular risk. Nevertheless, the exact association between inhibition of CETP function in patients and cardiovascular outcomes remains complex and not fully understood [69].

1.4.1 Mechanism of action of CETP inhibitors

CETP is a plasma protein, mainly secreted by the liver. It binds to HDL in circulation and takes part in transfer of cholesteryl esters (CE) from HDL to apoB-containing lipoproteins LDL and VLDL (heterotypic transfer). In exchange for cholesteryl esters, apoB-containing lipoproteins transfer triglycerides (TGs) to HDL particles. Inhibition of CE transfer results in increased concentration of HDL-C and lower levels of LDL-C, indicating that CETP has proatherogenic properties. Contrary, CETP also facilitates the transfer of CE among HDL subtypes (homotypic transfer) and the conversion of apoA-I-containing α-HDL (HDL2 and HDL3) to small lipid-poor pre-β-HDL particles (remodeling of HDL). Homotypic transfer with remodeling of HDL particles play an important role in RCT thus is antiatherogenic. Generated pre-β-HDL particles are involved in ABCA1-dependent cholesterol efflux [70, 71].

1.4.2 Different CETP inhibitors

There are two types of drugs targeting cholesteryl ester transfer protein, potent CETP inhibitors (torcetrapib, anacetrapib, evacetrapib) and CETP modulators (dalcetrapib). They belong to two different chemical classes, thus their mechanism of action and inhibition of CETP activity presumably differs. Potent CETP inhibitors are 3,5-bis-trifluoromethyl-benzene derivatives and they increase plasma HDL-C levels in humans up to 130%. Their high-affinity binding to CETP

results in formation of CETP-lipoprotein complex. Created complex is inactive and cannot facilitate the transfer of lipids between different lipoproteins, thus blocks both heterotypic and homotypic transfer [71]. CETP modulator dalcetrapib is a benzenethiol derivative and it raises HDL-C levels in humans up to 36%. Dalcetrapib was shown to form disulfide bond with Cys13 of CETP and to impair its conformational change required for its proper function. Such a modulation of CETP activity limits heterotypic CE transfer from HDL to LDL but does not influence CETP activity among HDL subtypes and sustained formation of pre-β-HDL particles [71]. These findings suggest that in comparison to potent CETP inhibitors, dalcetrapib maintains or even potentially enhances the reverse cholesterol efflux.

1.4.3 Evacetrapib and anacetrapib clinical trials

The efficacy of the CETP inhibitor evacetrapib was assessed in placebo-controlled, phase III clinical trial, called ACCELERATE. Study enrolled 12092 patients with high cardiovascular risk. The end point was the occurrence of cardiovascular event, such as death from cardiovascular causes, myocardial infraction, stroke, coronary revascularization or hospitalization from unstable angina. The results showed improvement of patients lipid profile after 3 months of treatment, with 37% absolute decrease in the mean concentration of LDL-C and 131.6% absolute increase in the mean levels of HDL-C. Despite this improved lipid profile in treated patients, evacetrapib did not reduce the number of cardiovascular events among patients with high-risk vascular disease [72, 73]. More favorable results were obtained from phase III clinical trial REVEAL, which aimed to evaluate safety and efficacy of another CETP-inhibitor, anacetrapib. Study enrolled 30449 adults with atherosclerotic vascular disease and demonstrated that treatment with anacetrapib for around 4 years reduced the occurrence of major coronary events by 9% [74]. The lack of observed benefits in clinical trial testing efficacy of evacetrapib may result from insufficient number of study group and too short length of treatment.

1.4.4 Dalcetrapib clinical trials

Dalcetrapib entered phase III clinical trials. A first study named **dal-ACUTE** was a double-blind, placebo-controlled study which aimed to evaluate efficacy and safety of dalcetrapib in 300 patients hospitalized for an acute coronary syndrome. The results showed 33.7% increase of HDL-C after 4 weeks of treatment and 11.8% raise of apolipoprotein A1. Despite one-third increase in HDL-C the total cholesterol efflux raised only by 9,5%, suggesting that increase of HDL-C levels may not correlate with improved HDL function [75].

Next, a significantly bigger phase III clinical trial called **dal-OUTCOMES** intended to test dalcetrapib influence on cardiovascular risk in patients with a recent acute coronary syndrome (ACS). The 15,871 clinically stable patients after a recent ACS were randomly assigned to 600 mg daily treatment with dalcetrapib or placebo on top of basic treatment. The primary endpoint of the study was the time to occurrence of first cardiovascular event, such as coronary heart disease death, nonfatal myocardial infraction, unstable angina, cardiac arrest with resustitation, or ischemic stroke. The results of the study demonstrated that dalcetrapib increased HDL-C levels by around 30% and had minimal effect on LDL-C concentration. Study was terminated sooner than expected due to futility. Dalcetrapib did not influence the risk of cardiovascular events comparing to placebo without significant effect on any component of the primary end point. Patients treated with dalcetrapib had higher levels of inflammatory marker, called C-reactive protein, than placebo group [76, 77].

The **dal-PLAQUE-2** is a phase III substudy designed to evaluate the effect of dalcetrapib on progresion of atherosclerosis in 988 patients with evidence of coronary artery disease. Patients were randomized to receive treatment with 600 mg of dalcetrapib or placebo. Planned time of trial was 24 months, however study was terminated earlier, after 12 months, due to discontinuation of dal-OUTCOMES clinical trial [77]. Dalcetrapib and its increase of HDL-C in patients from dal-PLAQUE-2 study had no effect on carotid intima media thickness (IMT) [78].

Generally all clinical trials confirmed safety of dalcetrapib and lack of off-target effects caused by the other CETP-inhibitor torcetrapib.

1.5 Genome-wide association study

In the past few years big projects such as sequencing of whole human DNA (Human Genome Project) and development of a haplotype map of human genome (HapMap) contributed to significant progress in understanding of genetic causes of human disease. Collected data on human genome and common patterns of human genetic variation enabled identification of numerous loci associated with diseases and drug response. Genome wide association study (GWAS) allows such a discoveries by using genotyping platforms with a large number of markers spread across the human genome, called single nucleotide polymorphisms (SNP). This hypothesis free approach tests if there is a significant association between SNP(s) and phenotype of interest, frequently disease. Due to high number of performed tests, analysis often result in a

high false positive rate, therefore the genomewide statistical significance is usually set at P values of 5×10^{-8} or lower. To be able to reach statistical significance very large sample sizes are needed. Results obtained by GWAS are usually repeated in an independent sample set focusing only on identified SNPs or loci, to validate genuineness of association. Single nucleotide polymorphisms that lay in close proximity tend to be in linkage disequilibrium (LD), what means that their specific alleles are correlated in population. Frequently, GWAS genotyping arrays contain only few variants per LD what limits power of detection. Thus, other SNPs from the identified loci may be additionally imputed in order to discover other potential associations [79].

1.5.1 GWAS Dalcetrapib

After obtaining results showing the lack of benefit from dalcetrapib dal-OUTCOMES trial, the group of researchers under lead of Dr Jean-Claude Tardif and Dr Marie-Pierre Dubé decided to conduct pharmacogenomics analysis on dal-OUTCOMES study using genome-wide association study (GWAS) approach [77]. The goal of this GWAS analysis was to determine if there is an association between genetic variants with cardiovascular events in patients treated with dalcetrapib.

The DNA samples from 5749 white participants of dal-OUTCOMES study that gave informed consent for genomic research were tested in GWAS using Illumina array with 2567845 genetic variants. The genome-wide significant threshold was set at P value $< 5 \times 10^{-8}$ to detect genetic variant(s) associated with cardiovascular events in dalcetrapib treatment arm. Next, the significant variants were tested in the placebo arm to verify if detected association was only observed with dalcetrapib treatment. Tested phenotype included all cardiovascular events from dal-OUTCOME primary outcome with the addition of unanticipated coronary revascularization.

A single region located in adenylate cyclase 9 gene was found to be associated with cardiovascular events in the dalcetrapib arm (Figure 3). In the identified region, genetic variant rs1967309 passed the significance threshold with P value = 2.41×10^{-8}. In the same region, 7 other single nucleotide polymorphisms were identified with P value $< 1 \times 10^{-6}$ (Table 1). In patients with AA genotype (minor allele) for SNP rs1967309, 39% reduction in the cardiovascular risk was observed when treated with dalcetrapib compared to placebo. Contrary, for patients with GG genotype (major allele), 27% increase in cardiovascular events was detected

in the dalcetrapib arm versus placebo. For heterozygote patients AG and placebo arm alone, there was no significant effect observed [77].

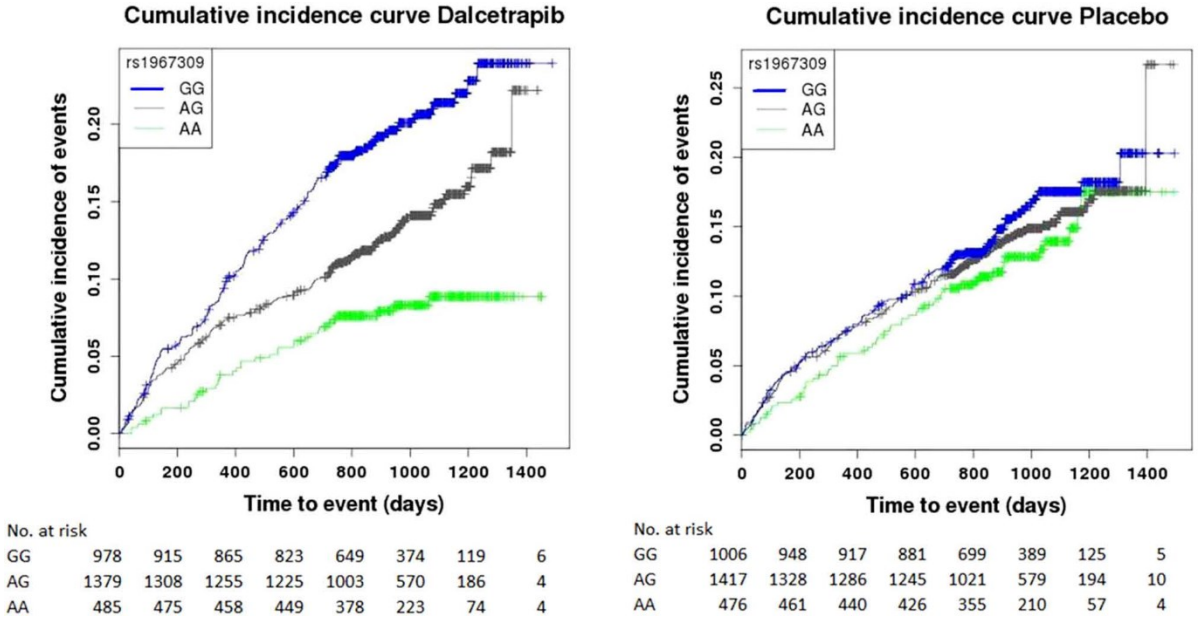

Figure 3 : Cumulative incidence of cardiovascular events from dal-OUTCOMES study for the dalcetrapib and the placebo arm separately. Data stratified by the three genotypes of variant rs1967309 located within ADCY9 gene. Figure from Tardif JC et al., 2015 [77].

To validate the results obtained, 27 SNPs located in the *ADCY9* gene were selected to genotype the samples of participants of dal-PLAQUE-2 study, which gave informed consent for genomic research (n=386). Tested endpoint for dal-PLAQUE-2 studies was the mean intima media thickness of common carotid arteries measured after 6 and 12 months of dalcetrapib treatment. Due to high correlation of tested genetic variants there was no need for adjustment of significance threshold for multiple testing, thus associations were considered as significant reaching P value < 0.05.

Ten single nucleotide polymorphisms displayed association with IMT measures and reached significance threshold (Table 1). Interestingly, significant association was observed for SNP rs2238448 (P value = 0.009), which is in high linkage disequilibrium with rs1967309 ($R^2 = 0.80$) and was also found to be associated with cardiovascular events in dal-OUTCOMES (P value = 8.88×10^{-8}). In patients with TT genotype (minor allele) at variant rs2238448 changes in IMT measure were -0.027 ± 0.079 mm, in heterozygote carriers 0.000 ± 0.048 mm

and in participants with CC genotype (common allele) 0.009 ± 0.038. Again, there was no observable effect in dalcetrapib arm alone. SNP rs1967309, which provided significant association with cardiovascular outcomes in dal-OUTCOMES studies, did not reach significance threshold for association with change in IMT (P=0.114). However, observed changes in intima media thickness measure were consistent with findings in dal-OUTCOMES [77].

Table I : Genetic variants identified by genome-wide association study in dal-OUTCOMES (P value $< 1 \times 10^{-6}$) and dal-PLAQUE-2 (P value < 0.05).

SNP ID	P value dal-OUTCOMES	P value dal-PLAQUE-2
rs1967309	2.41×10^{-8}	0.114
rs2531971	7.74×10^{-8}	0.1415
rs2238448	8.88×10^{-8}	0.009
rs11647778	1.72×10^{-7}	0.0087
rs1259911	1.72×10^{-7}	0.0205
rs12595857	2.02×10^{-7}	0.0193
rs12920508	3.18×10^{-7}	0.114
rs2239310	9.58×10^{-7}	0.6456
rs11647828	1.32×10^{-6}	0.0051
rs8049452	1.93×10^{-6}	0.0126
rs12935810	1.03×10^{-5}	0.0186
rs8061182	4.72×10^{-5}	0.0182
rs4786454	1.91×10^{-4}	0.0222
rs2531967	6.67×10^{-4}	0.0184

In patients treated with dalcetrapib, genotypes of the SNP rs1967309 were also found to be associated with change in total cholesterol after 1 month. The homozygote carriers of common allele (GG) had a smaller total cholesterol increase compared with AG and AA carriers [$10.0 \pm 23.3\, mg/dL$ $(0.2586 + 0.6025\ mmol/L)$, $12.9 \pm 30.3\, mg/dL$ $(0.3336 + 0.7836\ mmol/L)$ and $13.8 \pm 23.3\, mg/dL$ $(0.3569 + 0.6025\ mmol/L)$, respectively]. For the changes in LDL-C after 1 month, similar genotype-dependent pattern was observed. Additionally, rs1967309 also

influenced the changes in body mass index, weight, plasma LDL/HDL cholesterol ratio and triglycerides. Surprisingly, the changes were in opposite direction than could have been anticipated [77].

The association of genetic variant rs1967309, reported to interact with dalcetrapib, was examined in DNA samples obtained from patients participating in ACCELERATE trial. Analysis revealed similar pattern of genetic association as observed with dalcetrapib, but did not reach statistical significance for evacetrapib-treated arm [80]. Meta-analysis of dal-OUTCOMES and ACCELERATE studies aimed to further investigate the evidence that *ADCY9* genotype affects the treatment response to CETP-inhibitors on incidence of cardiovascular events. Analysis demonstrated the evidence of heterogeneity across genetic groups (P=0.004) with regard to relative risk, thus lending support to a potential interaction between CETP inhibition and *ADCY9* genotype for cardiovascular events [81].

1.5.2 Association of dalcetrapib treatment and C-reactive protein levels

HDL are commonly considered to have atheroprotective properties, partly due to their anti-inflammatory functions. One of the standard markers of inflammation is high-sensitivity C-reactive protein (hs-CRP). Studies measuring hs-CRP levels in patients from dal-OUTCOMES trial showed unexpected results. After 3 months of treatment, patients taking dalcetrapib had $+13.7\,mg/dL$ $(+0.3543\ mmol/L)$ increase of HDL-C from baseline and $+0.21\,mg/dL$ $(+0.005431\ mmol/L)$ raise in hs-CRP. Patients from placebo group had slight increase in mean change in HDL-C $+1.7\,mg/dL$ $(+0.04396\ mmol/L)$ and reduction in hs-CRP $-0.49\,mg/dL$ $(-0.012671\ mmol/L)$ [82]. Considering different genotype of patients for variant rs1967309 an association between genotype groups and treatment was observed. Placebo-adjusted geometric mean percent changes in hs-CRP from baseline (12 months) to the end of trial were 18.1% for patients with common genotype GG, 18.7% for participants with AG and -1.0% for patients with minor genotype AA. The increase in hs-CRP levels after treatment with dalcetrapib, which increases HDL-C thought to possess anti-inflammatory properties is difficult to understand but has also been observed for other CETP inhibitors such as anacetrapib and evacetrapib [83]. The fact that only patients with protective genotype AA showed no increase in C-reactive protein levels supports the results obtained from GWAS and suggests that

inflammatory responses may be differently regulated depending on the genotype of the patient [84].

1.5.3 Association of dalcetrapib treatment and cholesterol efflux

Cholesterol efflux is an important step of RCT. In this process HDL particles remove excess cholesterol from cells of peripheral tissues and deliver it to the liver for its excretion [85]. The mean change in cholesterol efflux capacity of serum HDL in patients from dal-PLAQUE-2 trial as measured by the standard J774 macrophage assay was similar between study arms. However, the increase in cholesterol efflux capacity differed between participant groups treated with dalcetrapib. Patients with GG genotype had $7.8 \pm 18.0\%$ increase in cholesterol efflux capacity, participants with AG $12.9 \pm 16.9\%$ and patients with AA $22.3 \pm 22.3\%$. The highest increase in cholesterol efflux capacity for patients with protective genotype is another evidence supporting the genotype-dependent effects of dalcetrapib [84].

1.5.4 Linkage disequilibrium block located at *ADCY9* gene

Single nucleotide polymorphisms associated with responses of patients treated with dalcetrapib are located within the *ADCY9* gene on chromosome 16. Majority of identified SNPs are nonrandomly associated and create a linkage disequilibrium block (Figure 4). This LD block occupies a 27-kb region overlapping part of intron 2 and 3 of the *ADCY9* gene.

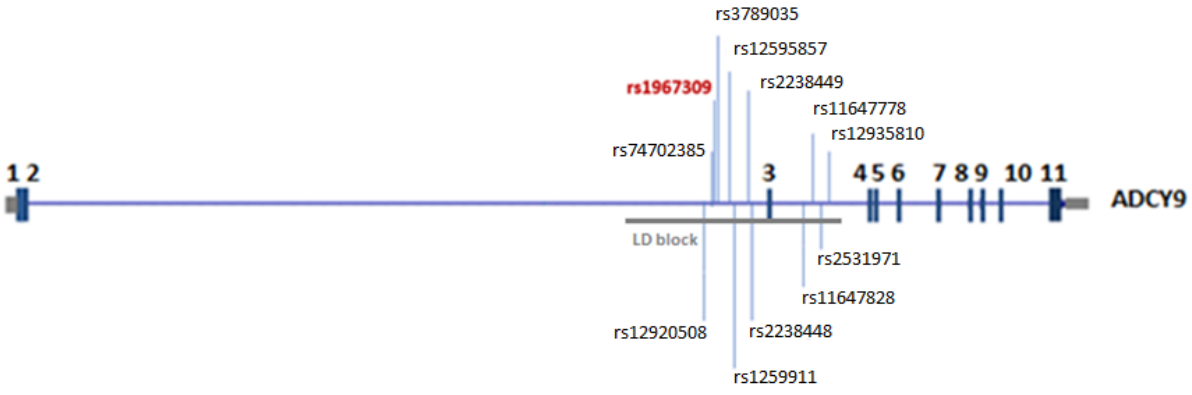

Figure 4 : Graphical representation of adenylyl cyclase 9 gene. Linkage disequilibrium (LD) block lies within second and third intron of *ADCY9* gene. Polymorphism shown in bold red (rs1967309) was found to be significantly associated with cardiovascular response of patients treated with dalcetrapib. The other SNPs presented in black are polymorphisms selected for functional analysis, majority of them is in high linkage disequilibrium with rs1967309.

1.6 Adenylyl cyclases

The adenylyl cyclases (AC) are membrane-associated enzymes that catalyse the conversion of adenosine triphosphate (ATP) to 3',5'-cyclic AMP (cAMP). They play an essential role in signal transduction following stimulation of G-protein coupled receptors, as they are the sole source of cAMP. Cyclic AMP is a second messanger involved in many biological processes, such as hormone responses. Nine membrane-bound adenylyl cyclase isoforms have been identified with a tenth distinct isoform that lacks membrane spans. The primary structure of the first nine isoforms consist of twelve transmembrane (TM) regions featuring cytosolic N and C termini. The first six TM domains (TM1-TM6) are followed by a cytosolic region with catalytic domain (C1a) and C1b domain. The second TM region (TM7-TM12) finishes with a second catalytic domain C2a and C-terminal C2b [86-88]. The tenth, soluble isoform has different regulatory and catalytic properties, and resembles more cyanobacteria enzymes [89].

1.6.1 Adenylyl cyclase 9

In comparison to the previous eight adenylyl cyclases, the ninth isoform (*ADCY9*) is the least characterized and the most divergent one. Human *ADCY9* gene is located at chromosome 16 and its cDNA contains a 4059 bp open reading frame, which encodes a 1353 amino acid protein. Human *ADCY9* cDNA is 85% identical to the mouse *Adcy9* cDNA. On the protein level, ADCY9 is 90% identical to the mouse protein and all the structural motifs predicted in mouse Adcy9, such as 12 transmembrane domains, Asn-linked glycosylation sites and cAMP-dependent protein kinase phosphorylation sites, are conserved in the human form [90].

1.6.2 Expression of ADCY9

Each isoform of adenylyl cyclases has unique tissue distribution. *ADCY9* expresses high levels of mRNA and protein in a wide variety of tissues. The mRNA levels of *ADCY9* are present in most human tissues and the highest levels occur in brain, skeletal muscle, heart and pancreas [90-92]. According to the EMBL-EBI expression atlas, *ADCY9* is also expressed in lower level in CD14-positive, CD16-negative classical monocytes.

1.6.3 Regulation of ADCY9 activity

Like all adenylyl cyclases, ADCY9 main function is catalysis of second messenger cyclic AMP formation. Despite the same basic function, each of the isoforms possess unique regulatory properties. Depending on these properties and expression levels of the isoforms in a tissue or cell

type and cell compartments, extracellular signals received by G-protein coupled receptors can be differently integrated [91]. All the isoforms are known to be upregulated by G-protein ($G_{\alpha s}$), which is involved in well known signaling pathway mediated by one of the G-protein coupled receptors, called β-adrenergic receptor (β2AR). The agonists of β2AR induce its interaction with α subunit of heterotrimeric G-protein ($G_{\alpha s}$). Next, $G_{\alpha s}$ binds GTP and in turn dissociates from G-protein γ subunit. The GTP-bound $G_{\alpha s}$ then binds and activates adenylyl cyclase, which produces cAMP. Increased levels of cAMP may lead to activation of protein kinase A (PKA) and other important signalling cascades. G proteins were found to bind directly to the catalytic core of AC, which includes two homologous cytoplasmic domains (C1a and C2a). However, each of the AC isoforms consists of additional structural elements that are crucial for the correct assembly of G-proteins and proper enzymatic function of ACs. A team of researchers, which determined the 3D structure of the bovine ADCY9 in complex with $G_{\alpha s}$, reported that the C2b region plays an important role in the G protein-bound state of ADCY9. The occlusion of ADCY9 reaction center by the C2b region reduces the affinity for the substrate (ATP) and as a result decreases production of cAMP. The 3D structure of bovine ADCY9 also revealed that transmembrane helices TM6 and TM12 are laterally exposed to the lipid bilayer and they extend into the cytosol becoming helices h1.1 and h2.1. Therefore, it is likely that direct interactions of TM6 and TM12 with lipids and other molecules could directly influence the catalytic domain by altering orientation of helices h1.1 and h2.1 [88]. Additionally, Pálvölgyi et al. reported that the short motif in the C2b isoform-specific domain of ADCY9 is responsible for supression of ADCY9 activation by $G_{\alpha s}$ protein. The authors also suggested that the autoinhibitory effect of C2b domain on ADCY9 activation in the heart may be released by proteolytic cleavage of this domain [93].

The other agents known to regulate the activity of all the isoforms of AC are Ca^{2+} and Mg^{2+}. The rest of regulatory agents are isoform-specific. Since *ADCY9* is the least characterized isoform, its regulation is poorly characterized. Unlike other adenylyl cyclase isoforms, the activity of *ADCY9* is unaffected by forskolin and calcineurin inhibitors. Also, *ADCY9* is stimulated by $G_{\alpha s}$-coupled receptors activation but is not weaken by G_i-coupled receptors [90]. *ADCY9* was reported to be phosphorylated by four different types of protein kinases, protein kinase C (PKC), protein kinase A (PKA), casein kinase 1 (CK1) and cyclin-dependent protein kinase 5/p35 complex (cd5/p35). Additionally, *ADCY9* contains two N-glycosylation sites

located on the sixth extracellular loop and the studies showed that blocking of *ADCY9* glycosylation impair $G_{\alpha s}$ stimulation [94].

1.6.4 ADCY9 in cardiovascular disease

GWAS studies identified single nucleotide polymorphisms located in ADCY9 gene to be associated with asthma, malaria and obesity [95-97]. Patients from dal-OUTCOMES trial showed different responses to HDL-modulating medicine dalcetrapib, depending on the genotype at SNP rs1967309 located in ADCY9 gene.

The association of *ADCY9* with cardiovascular diseases is not fully understood. A few fundamental studies showed that heterogeneous distribution of the cholesterol on the cell-membrane surface influences cell structure and function. Cholesterol colocalizes in caveolae, a cell-surface membrane invaginations. Due to the evidence that the majority of caveolae-bound proteins are signal transduction molecules, the hypothesis was proposed that the compartmentalization of signaling molecules is responsible for regulation of signaling-events and cross-talk between different signaling pathways. β2AR can localize in caveolae and exit after adrenergic stimulation [98]. Hence, cholesterol depletion from the plasma membrane induces redistribution of signaling compartments and leads to increased β2AR-stimulated cAMP production. Elevated cholesterol levels had an opposite effect [99]. Study investigating loss of *Adcy9* in mice showed that *Adcy9* inactivation results in no obvious abnormalities in size or structure of an animal. However, in the heart, lack of *Adcy9* significantly decreased basal phosphorylation of heat shock protein 20 (Hsp20), which is responsible for cardioprotective effects of Hsp20 [100]. Other studies demonstrated that protein kinase A is responsible for phosphorylation of ATP-binding cassette A (ABCA1). Since PKA activity is highly influenced by production of cAMP by *ADCY9* and ABCA1 plays major role in cholesterol efflux to apolipoprotein particles, it can be another potential connection of adenylyl cyclase 9 to cardiovascular responses [101].

The activity of *ADCY9* was also found to mediate different functions of immune cells and to regulate various inflammatory responses. Activation of *ADCY9* by rapamycin complex 2 (mTOR2) was found to be crucial for chemotaxis process of white blood cells, specifically neutrophils. Active *ADCY9* produces cAMP from ATP and regulates back contraction via phosphorylation of myosin II (MyoII). It was shown that upon chemoattractant stimulation,

phosphorylated protein kinase CβII (PKCβII) is translocated to the plasma membrane, where it activates *ADCY9* by phosphorylation. Membrane translocation of PKCβII requires its phosphorylation by mTOR2 [102, 103]. cAMP signaling is known to strongly diminish monocyte and macrophage inflammatory responses, with notably strong effect on tumor necrosis factor α (TNFα). Consequently, down-regulation of adenylyl cyclases (isoform 7 and 9) in monocytes was found to significantly promote the TNFα production [104]. In $CD4^+\,CD25^-$ T cells and $CD4^+\,CD25^+$ regulatory T cells, the activity of *ADCY9* is controlled by microRNA miR-142-3p. The levels of miR-142-3p expression regulate the formation of cAMP in those cells, which is the crucial factor for intercellular communication among these two cell subclasses [105]. Also, *ADCY9* expression in mice macrophages determines the production of interleukin 6 and TNFα. Inhibition of *ADCY9* resultes in reduced IL-6 expression [106].

1.6.5 ADCY9 inactivation protects from atherosclerosis only in the absence of CETP

Recently published data showed that inactivation of *Adcy9* in mice has atheroprotective effects in the absence of CETP [107]. This study aimed to evaluate atherosclerosis, vasorelaxation, heart rate and adipose tissue content. Animals used in this study were *Adcy9*-inactivated ($Adcy9^{GT/GT}$) and wild-type (WT) mice, which were or not transgenic for the CETP gene ($CETPtgAdcy9^{GT/GT}\ and\ CETPtgAdcy9^{WT}$). All animals were submitted to an atherogenic protocol that include injection of an AAV8 (adeno-associated virus serotype 8) expressing a PCSK9 gain of function variant, which induced a 90% reduction in hepatic LDL receptor expression, and 0.75% cholesterol diet during 16 weeks. The *Adcy9*-inactivation in $Adcy9^{GT/GT}$ mice was confirmed by reverse transcription-qPCR, which showed at least 90% decrease of *Adcy9* mRNA expression in the heart.

Study demonstrated that the atherosclerotic lesions, covering the aortas surface were reduced by 65% in *Adcy9*-inactivated mice, compared to wild-type ($P<0.01$). Also, a numeric reduction of plaque area all along aortic root was observed in $Adcy9^{GT/GT}$ mice ($P=0.07$). In contrast to *Adcy9*-inactivated mice, WT animals had plaques in the brachiocephalic arteries that are more complex and prone to rupture ($P<0.05$). In addition, plaques found in WT mice had more often fibrin deposits on their surface, compared to $Adcy9^{GT/GT}$ mice ($P=0.06$).

Further, *Adcy9* inactivation resulted in diminished accumulation and proliferation of CD68-positive foam cells, which are main components of atherosclerotic plaques. CD68-positive foam

cells represented $19.0 \pm 1.9\%$ of the lesion area in WT animals, and $11.4 \pm 2.1\%$ in $Adcy9^{GT/GT}$mice (P<0.05). Decrease of CD68-positive foam cells proliferation was observed in *Adcy9*-inactivated mice compared to WT (P<0.05). The blood concentration of total monocytes and their main subpopulations was unchanged by inactivation of *Adcy9*. Also, the migration capacity of bone marrow-derived monocytes toward C-C motif chemokine ligand 2 (CCL2) was unaffected.

The absence of *Adcy9* increased both endothelial-dependent and endothelial-independent vasodilatation in femoral arteries of atherosclerotic mices ($P < 0.05$). To understand the mechanism involved in reduced foam cell accumulation in $Adcy9^{GT/GT}$mice, study assessed the adhesion of splenocytes to the native endothelium. Results showed that basal adhesion of splenocytes to the endothelium was decreased in *Adcy9*-inactivated mice compared to wild-type animals (P=0.07). Only wild-type mice showed significantly increased adhesion of splenocytes after activation of endothelium by histamine. The expression of cell adhesion molecules, such as CD18, CD162, and CD62L on the splenocytes surface was unaffected by inactivation of *Adcy9*. Next, to establish which mechanism may explain the increased endothelial-dependent vasodilatation in *Adcy9*-inactivated mice, researchers performed selective pharmacological blockage of concerned signaling pathways. The use of nitric oxide synthase blocker L-NNA resulted in significant inhibition of endothelial-dependent vasodilatation in both animals, WT and $Adcy9^{GT/GT}$ ($P < 0.01$). Same compound reduced the flow-mediated vasodilatation only in $Adcy9^{GT/GT}$ mice. In turn, the cyclooxygenase inhibited exclusively endothelial-dependent vasodilatation in *Adcy9*-inactivated animals (P<0.01).

It was observed that mice with inactivated *Adcy9* possessed increased body weight (P<0.01). It should be noted that, contrary to patients with deleterious *ADCY9* rs1967309 genotype, dalcetrapib-treated patients with the protective AA genotype did not lose weight [77]. Magnetic resonance imaging demonstrated that $Adcy9^{GT/GT}$ mice had doubled total body adipose tissue volume compared to WT (P<0.01). Atherogenic diet and increased fat deposits did not influence the blood glucose and insulin concentration. The *Adcy9*-inactivated mice demonstrated significantly increased feed efficiency, defined as the ratio of weight gain over energy intake (P<0.01). These data suggest the involvement of autonomic nervous system (ANS), which is known to regulate energy balance. Further evaluation of ANS activity demonstrated that RR

interval was increased in $Adcy9^{GT/GT}$ mice in comparison to WT ($P<0.05$), and the nocturnal heart rate was lower in *Adcy9*-inactivated animals.

Observed atheroprotective properties of *Adcy9* inactivation were lost when human CETP was introduced to *Adcy9*-inactivated mice.

Based on the ADCY9 tissue expression data, showing hight levels of ADCY9 expression in the different brain regions and results obtained from described above study, apart monocytes, brain cells could be other interesting cells to investigate the association of patients genotype with cardiovascular responses induced by dalcetrapib [91, 108].

1.7 From association to function

Although genome wide association studies have identified numerous loci associated with complex diseases and clinical drug response, the majority of associations still lack mechanistical understanding. Since the results of GWAS do not specify which is the causal variant of the identified locus neither the affected target gene, the functional follow up on a GWAS data is highly challenging.

Identified loci generally contain many co-inherited variants in strong linkage disequilibrium with the most significant disease-associated variant (leading SNP). First and major challenge is to determine which of the linked polymorphism(s) is actually causing the association [109, 110]. Multiple causal variants are increasingly identified at the risk locus [111]. After successful prioritization of putative causal SNP(s) and altered mechanism, the next step is to identify target gene(s) and exact impact of genetic variant(s) on phenotype.

1.7.1 Prioritization of functional SNP

A causal SNP is a genetic variant that affects molecular process to alter phenotype. Identification of causal SNP(s) among loci determined by GWAS is complicated by the fact that more than 90% of disease-associated variants are located in the gene non-coding regions [109]. Combination of bioinformatic analysis together with experimental approaches can somewhat facilitate this task.

1.7.2 How intronic SNP(s) can influence gene expression?

Disease-associated variants located in intergenic or intronic regions of the genome often function through modulation of genetic expression of one or more target genes. Gene expression is a process by which gene products, usually proteins, are synthesized from information contained in genes. In eukaryotes it is a complex multistep process regulated by various factors on many levels. Therefore, noncoding variants may influence the expression of a target gene regulating many different steps, including gene transcription, RNA splicing, non-coding RNA function and epigenetic regulation [112].

Regulation of transcription in eukaryotes

All cell types in mammalian body have the same genotype, yet their characteristics and functions differ significantly. The transcriptome is the set of transcribed genes that defines a cell type. The transcriptome of a cell includes RNA species from commonly transcribed genes for all cell types (housekeeping genes) and ones that are transcribed only for one or few specific cell types (cell-type-specific genes). Therefore, the expression of different cell-type-specific genes and their different levels of transcription are responsible for different characteristics of the cells [113].

The transcription process can be viewed as the first step of gene expression and its aim is to synthesize mRNA from DNA template. Transcription in eukaryota is controlled mainly by transcription factors (TFs) that bind to specific DNA sequence motifs in regulatory regions of a gene (Figure 5). The main regulatory region of a gene is generally located upstream of the transcription start site (TSS) and is called the promoter. Commonly, minimal promoter is made up of a TATA box and initiator sequences. Control elements may also be found within introns or downstream of the coding regions of a gene [114]. In eukaryota, RNA polymerase II (pol II) is responsible for transcription of genes that encode proteins. The promoter region is necessary for proper positioning and orientation of pol II at the transcription start site. Although, polymerase itself is enough to catalyse RNA synthesis, it does not recognize the DNA promoter region alone. Therefore, recruitment of RNA polymerase II to TSS is facilitated by the transcription preinitiation complex (PIC). PIC consists of five general transcription factors (TFIIB, TFIID, TFIIE, TFIIF and TFIIH) and its recruitment to DNA promoter is initiated by a subunit of general transcription factor TFIID, called TATA-binding protein (TBP), which binds to TATA element [115].

Except for general transcription factors, gene expression is also regulated by specific transcription factors. These specific TFs can bind to promoter regions of a gene or more distant regulatory regions, called enhancers, and regulate transcription of neighbouring or distant genes. Binding of these specific regulatory proteins provide a unique expression patterns of genes specific for individual cell types and dependent on developmental state of a cell or impact of external stimuli. Transcription factors which bind to specific DNA motifs within enhancer regions typically recruit cofactors, which either induce transcription (coactivators) or inhibit transcription (corepressors). Cofactors then act on transcription by inducing physical interaction between enhancer and core promoter, enabled by looping of DNA. Examples of well characterized cofactors are Mediator complex and cohesin [113, 114].

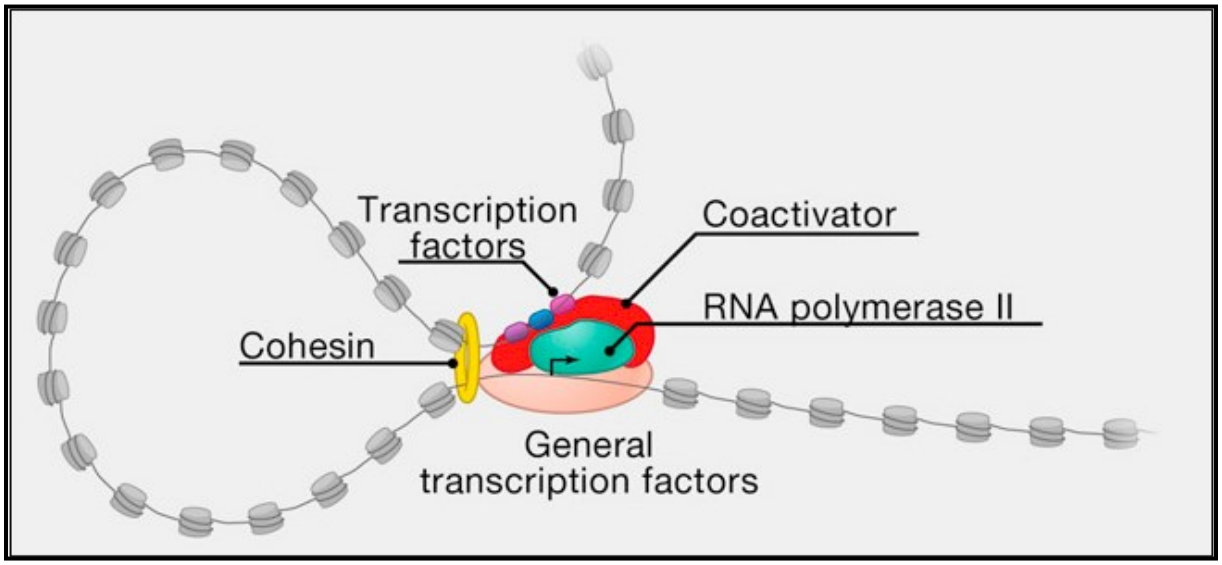

Figure 5 : Assembly of preinitiation complex. Specific transcription factors bind to DNA motifs within enhancer regions and recruit cofactors. Cofactors bind to RNA polymerase II, which with the help of general transcription factors orientate on promoter region of a target gene and binds to transcription start site. Figure presented with permission (Lee TI., 2013) [113].

Beside the control of transcription initiation, transcription factors may also control elongation step. Recruited RNA polymerase II initiate transcription, synthesize short RNA fragment and typically pause. Pol II molecule may finish transcription at this stage and release small RNA species or continue the elongation process by release of the pause. Continuation of elongation is caused by phosphorylation of paused polymerase and associated pause control factors by positive transcription elongation factor b (P-TEFb). This factor is recruited to transcription sites in form of super elongation complex (SEC). Also, other complexes and individual transcription factors were

found to regulate the elongation process. For example, c-Myc stimulate release of pol II from pause sites [113, 115].

SNPs in regulatory DNA binding sites

DNase I hypersensitivity sites (DHSs) are known predictors of accesible chromatin state and thus potential presence of DNA binding proteins, likely transcription factors, at certain DNA regions. Genome-wide data on chromatin accessibility showed that almost 60% of intronic polymorphisms identified by GWAS lie within DHSs. These data suggest that genetic variants in non-coding regions of human genome disrupt DNA binding motifs of transcription factors. More and more functional analysis of GWAS data identify variable TF-DNA interactions as leading cause of phenotypic variation [116].

Intronic polymorphisms can be present within DNA regions where specific regulatory proteins bind. Change of allele may disrupt DNA binding motif and change binding affinity of certain protein, resulting in different regulation of gene transcription. SNP may disrupt already exisiting TF-DNA binding site of coactivator causing transcription of lower mRNA levels or corepressor resulting in higher mRNA levels (Figure 6A). Intronic polymorphism can also create new binding sites for regulatory proteins. Expression quantitative trait loci (eQTL) analysis help to determine the effect of polymorphisms on gene expression levels in specific tissues or cell lines. An eQTL is a genomic locus, which influences the level of mRNA from a given gene. eQTLs may either influence expression of a gene in close proximity (cis-eQTLs) or a gene located further away from regulatory region (trans-eQTLs). The annotation of SNPs to eQTL regions in tissue or cell type of interest facilitates identification of causal SNP(s) [117].

However, the molecular mechanisms mediating TF-binding variation appear to be less obvious than originally anticipated. Only the minority of TF-DNA events were found to be triggered by the nucleotide substitution in the consensus motifs of considered TFs. The other mechanisms, by which SNPs may influence the binding affinity of transcription factors depend on the alterations of proximal motifs. Binding of some TFs depends on the presence of other DNA binding proteins. Since, genomic regulatory regions often comprise multiple TF binding sites, nucleotide changes in one motif can modulate binding of other proximal or farther protein. Transcription factors can bind to the regulatory sites cooperatively (Figure 6B). In this case, TFs physically interact to increase the affinity of complex to the specific region of DNA. Disruption of one of

the two TF binding motifs can affect binding of the entire complex. Regulatory proteins can also bind to particular sites in the genome in collaborative manner (Figure 6C). DNA binding motifs of TFs binding collaboratively may be located in far distances from one another. Since affinity of nucleosome to DNA is generally higher than affinity of single TF, two or more proteins may interact together to displace nucleosome in order to get access to DNA. The influence of distal genetic variants on protein binding to DNA can be explained by long distance TF-TF interactions enabled by chromatin looping (Figure 6D). Polymorphism can alter state and conformation of chromatin and in turn disturb the interactions among TF and DNA, and between TFs [116].

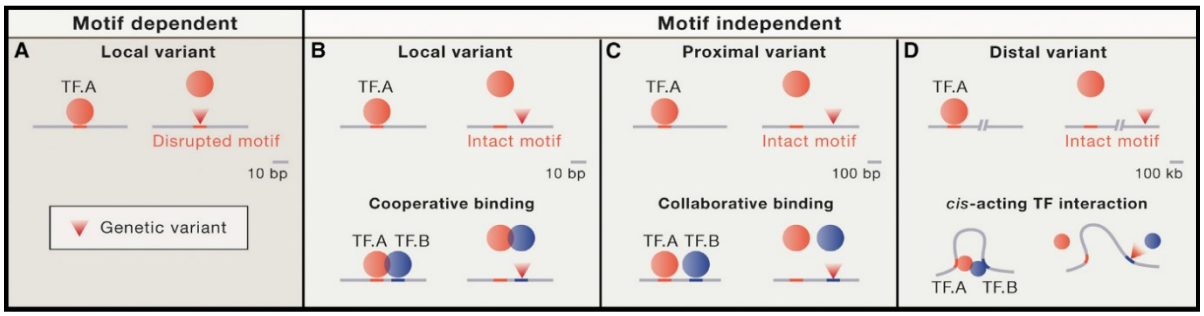

Figure 6 : Different mechanisms of TF-DNA binding variation by genetic variants. (A) Nucleotide substitution disrupts transcription factor binding motif. (B) Polymorphism influences cooperative binding of regulatory proteins. (C) Polymorphism influences collaborative binding of regulatory proteins. (D) Distal genetic variant influences the conformation of chromatin and affect TF-DNA binding. Figure presented with permission (Deplancke B., 2016) [116].

Genetic variants affecting alternative splicing

Another process which affects the levels of gene expression is called alternative splicing. RNA pol II transcribes DNA region of a gene into precursor mRNA (pre-mRNA) that contains coding (exons) and non-coding (introns) parts of a gene. The pre-mRNA further undergoes splicing process in which introns are removed and the final mRNA product contains only exons. Since the majority of genes comprise multiple exons and different splice sites, individual gene can produce various mRNA isoforms with different number of exons in the process known as alternative splicing. Alternative splicing of a gene may result in mRNA isoforms with diverse regulated functions, which can be often translated into different protein isoforms exerting various biological properties. Therefore, this process contributes to the increased diversity of transcriptome and proteome of a human [118].

A large number of human genes undergo alternative splicing and a variety of events may lead to it. The main patterns involve exon skipping, alternative 5' and 3' splice sites, intron retention and mutually exclusive exons. However, alternative splicing may also be the result of other complex patterns that are often cell-type specific. The formation of alternative mRNA isoform is regulated by protein-RNA interactions. Splicing cis-elements in 5' and 3' splice sites, branch point sequences and polypyrimidine tract of pre-mRNA recruit spliceosome complex crucial for the splicing process. Supplementary cis-elements in exon regions or surrounding introns modulate exon splicing through interactions with trans-acting factors, mainly RNA-binding proteins (RBPs). These auxiliary elements can either promote or repress exon splicing [119].

Genetic variants can affect alternative splicing process and lead to changed phenotype. Polymorphisms located in exonic or intronic regions of a gene may disrupt specific motifs in splice sites, enhancer or silencer cis-elements, resulting in formation of abnormal mRNA and protein isoforms [120, 121]. Also, genetic variants may affect trans-acting factors and cause dysregulation of splicing of all downstream target genes.

SNP(s) and non-coding RNA

Beside protein-coding transcripts, human genome transcribes a large number of non-coding RNA (ncRNA). Firstly considered as "genomic noise", ncRNAs have been later found to play important role in gene expression regulation. According to their size, we can distinguish two main classes of ncRNA. Small non-coding RNA with transcript length less than 200 nucleotides and long ncRNA with more than 200 nucleotides in length.

microRNAs

One of the most abundant subclass of small non-coding RNAs are microRNAs (miRNAs). These short (around 19-24 nucleotides long) single-stranded RNAs are known to participate in post-transcriptional gene regulation. MicroRNAs are part of an evolutionary conserved family of ncRNAs, which undergo characteristic process of biogenesis. Initially longer primary miRNA (pri-miRNA) transcript is processed in the nucleus by the enzymes Drosha and DGCR8 into shorter (around 70 nucleotides long) precursor miRNA (pre-miRNA). Next, pre-miRNA is transported into the cytoplasm and shortened by the enzyme Dicer into approximately 20 bp miRNA molecule. RNA-induced silencing complex (RISC) incorporates single strand of miRNA

molecule and targets messenger RNA transcript. Recognition sequence at the 5' end of miRNA, called seed sequence, binds in a complementary manner to nucleotide motif typically present in the 3'-UTR of the target mRNA. Binding of RISC complex to mRNA transcript induces its degradation or inhibit its translation.

Genetic variations located in the sequence of microRNA or their target messenger RNA can influence the expression and functionality of miRNA thus causing downregulation or upregulation of their target genes or affect regulation of the target gene. Such variation in turn influences pathways in which those target genes are involved. There are various ways by which SNPs could influence the levels of miRNA. Polymorphisms present in miRNA regulatory sequences, such as promoters or enhancers, may modulate transcription process and in consequence the expression levels of miRNA. Also, SNPs located in pri-miRNA or pre-miRNA can influence the maturation efficiency of miRNA. Further, genetic variation in miRNA transcript or target mRNA sequence may disrupt miRNA-target interaction. SNP in microRNA can affect loading of miRNA molecule into RISC complex, resulting in failure to recognize target mRNA. Finally, polymorphism in the sequence of target mRNA may destabilize miRNA recognition motifs or create new ones [122, 123].

Long non-coding RNAs

Long non-coding RNAs are a heterogeneous group of transcripts including four subclasses. The biggest group is represented by long intergenic non-coding RNAs (lincRNAs). LincRNA genes are known to not overlap, neither locate close to protein-coding genes. The second subclass of lncRNA are antisense lncRNAs, which are transcribed from the opposite strand of the protein coding genes. The third group are sense lncRNA transcripts. These lncRNAs lie on the same strand than protein-coding gene and are transcribed in the same direction. The fourth subclass is the bi-directional group. Transcripts of these lncRNAs are located on antisense strand and have transcription start site next to TSS of protein-coding gene. However, they are transcribed in opposite direction to protein-coding gene.

The precise mechanisms of action of long non-coding RNA molecules are still not fully understood, however lncRNAs are known to have ability to interact with DNA, RNA and various proteins. They control post-transcriptional processes, such as maturation and transport of RNA and protein synthesis. Additionally, lncRNA can alter transcriptional gene silencing by epigenetic

mechanisms. Common mechanisms of lncRNAs include the signaling of the biological state of a cell to induce or repress genes. Second mechanism called "decoy" rely on the ability of some lncRNA species to compete with another RNA or proteins (e.g. transcription factors) for binding. In the other "guide" mechanism of action, lncRNAs bind to particular regulatory proteins and transport them to specific target regions. Long non-coding RNA molecules may also facilitate interaction of multiple proteins by binding to them and bringing them closer to each other [123].

Still the understanding of how single nucleotide polymorphism could influence lncRNA expression or function is rather limited. SNP in promoter sequence of lncRNA could influence its expression level, likely repress it. Also, polymorphism can disturb binding sites of inhibitory complexes, resulting in gain of function for lncRNAs, not expressed under normal conditions. Additionally, genetic variant in lncRNA genes may regulate its function. SNP may alter the secondary structure of lncRNA or lead to alternative splicing of the transcript [123].

Polymorphisms in epigenetic pathways

Epigenetics involves the heritable changes in the phenotype driven by factors other than DNA sequence. Epigenetic mechanisms lead to the selection of which sets of proteins are transcribed through activation or repression of gene expression [124]. An important process responsible for gene regulation is the organization of chromatin structure. In eukaryotic nucleus, DNA is packed into chromatin in form of basic units called nucleosomes. Single nucleosome contains histone octamer with 146 bp DNA region. The localization of nucleosomes, commonly called "beads-on-a-string" serve as first level of chromatin structure organization. Higher structural level is obtained by linker histones, which compact nucleosomal DNA into higher-order structure, called fiber. Fiber structure can be further compacted into higher levels of chromatin organization but mechanisms driving higher folding are mainly unclear. However, it is known that the dynamic of higher-order chromatin compaction is responsible for transcriptional regulation. For processes, such as DNA replication and transcription initiation, DNA binding factors need to access specific DNA regions. These interactions are prevented by packaging of DNA into chromatin. Dense structure of chromatin (heterochromatin) is generally associated with repression of gene transcription, whereas the open structure of chromatin (euchromatin) promotes active transcription. The basic epigenetic processes regulating the structure of chromatin are DNA methylation and histone modifications [125].

DNA methylation

DNA methylation is mainly associated with cytosine residues in CpG-rich regions of DNA. These regions are primarily present in transcription start sites of genes, but may also be located in the intergenic regions and gene bodies. DNA methylation is commonly considered as silencing epigenetic mark, however, evidence suggests that gene body methylation rather stimulate transcriptional elongation than repress it. Also, methylation in the gene body may affect splicing process [126]. There are two proposed mechanisms explaining how methylation of DNA may affect gene transcription. First, DNA methylation can disrupt binding of regulatory proteins (e.g. transcription factors) to the gene. Second, methylated DNA can recruit chromatin remodeling proteins that control histone modifications, inducing formation of dense, inactive heterochromatin [127, 128].

Histone modifications

Histone proteins possess tails that bulge from the nucleosomes. Residues in these tails can undergo post-translational modifications, which may affect interactions among nucleosomes, further altering overall chromatin structure. Histone tail modifications may also induce binding of remodelling enzymes, which are capable of repositioning nucleosomes using the energy produced by ATP hydrolysis [129, 130]. There are five known types of histone modifications: acetylation, methylation, phosphorylation, ubiquitination and sumoylation. The most established histone acetylation is associated with promotion of gene expression. Acetylation of lysine side chain neutralizes its positive charge and in turn attenuates DNA-histone interactions, resulting in more accesible DNA regions. Similarly, phosphorylation of amino-acid chains at histone tails modulate chromatin structure adding negative charge to histone protein [129]. The regulation of gene expression by histone methylation depends on which amino acid in histone protein undergoes modification as well as the number of attached methyl groups. In some cases histone methylation can promote gene transcription whereas in others it can lead to repression [131].

Genetic variants and epigenetic control

There is more and more evidences suggesting that genetic variants in the genome may influence epigenetic control of gene expression. Few studies identified cis-acting genetic variations, causing epimutations [112]. Among others, in the patients with hereditary nonpolyposis

colorectal cancer, the hypermethylation of MLH1 gene resulted in transcriptional repression, which was associated with polymorphism located at MLH1 5' UTR [132]. Beside cis-regulating variants, polymorphisms may alter DNA methylation of enhancer regions located far from target genes [133]. Further, genetic variants were also found to affect the allele-dependent binding of transcription factors and chromatin structure. For example, the risk of asthma and autoimmune disease was associated with the transcriptional repressor CTCF. This transcription factor is a known chromatin structure modulator, which binds to DNA in allele-specific manner and alters expression of several genes [134].

1.7.3 Bioinformatic analysis to prioritize putative causal SNP(s)

Recent evolution of publicly available data sets from large-scale genomic and epigenomic experiments shared by projects ENCODE, the National Institutes of Health (NIH) Roadmap Epigenomic project and Nuclear Receptor Cistrome (NRCistrome) facilitates functional evaluation of intronic regulatory regions overlapping with genetic variants. Prediction of genomic features, such as TF binding, chromatin state and epigenetic marks help narrowing down the number of putative functional SNPs. Also, bioinformatics predictions can simplify planification of experimental techniques to elucidate regulatory function of variants.

RegulomeDB, HaploReg and FunciSNP are available online tools, which test the potential influence of genetic variants on different genomic features. These databases include information about: expression quantitative-trait loci (eQTL), DNase I hypersensitive sites (DHS), chromatin immunoprecipitation (ChIP), TF-binding motifs, chromatin interactions and evolutionary conservation of sequence [112]. Information about DNase I hypersensitive sites (DHS) and histone modifications help to predict if genetic variant impacts the open state of proximal chromatin region and hence, alter the accessibility of DNA to transcription factors and RNA polymerase II. Prediction of TF-binding motifs and ChIP-seq data facilitate identification of regulatory proteins, which may bind in allele-specific manner to the region of SNP of interest.

Although described databases help predicting function of noncoding genetic variants, it should be taken into account that majority of these tools rely on ENCODE data, which do not include data from all TFs and cell types. Additionally, one should focus on data obtained in relevant cell and tissue types since these tools do not consider tissue specificity [112].

1.7.4 Experimental approaches to identify causal SNP(s)

Even though bioinformatic analysis greatly facilitate prediction of noncoding variant functions, in silico findings need to be confirmed and further investigated by experimental techniques *in vitro* and *in vivo*.

Despite the complexity of GWAS follow-up functional studies, numerous publications can be found that aim to identify causal SNP(s) and molecular mechanism responsible for association. The majority of these studies fail to definitely prove the association of SNP(s) with phenotype but suggest possible mechanisms involved and give the foundation for further studies. Also, exemplary GWAS follow-up studies demonstrate recent approaches worth to employ in functional characterization of risk loci. Some of these studies which intended to identify causal SNP(s) concern associations with complex diseases such as estrogen-receptor positive breast cancer [135], schizophrenia [136] and osteoarthritis [137]. Most often, these functional follow up studies start with examination of associations of lead SNP(s) with expression levels of genes located in close proximity. Edwards et al found that risk-associated G allele of SNP rs10941679 was associated with increased mRNA expression of *FGF10* and *MRPS30* genes [135]. SNP rs1006737, associated with schizophrenia, was identified as cis-eQTL for gene *CACNA1C[136]*. Next, the regulatory capability of DNA regions containing candidate SNP(s) is often evaluated using reporter assays. Regions containing allele of SNP of interest can be subcloned into luciferase reporter construct, either with promoter of particular gene or SV40, and transfected into appropriate cell line. In breast cancer risk loci, putative regulatory element, containing SNP rs10941679 displayed enhancer activity, leading to increase in *FGF10* and *MRPS30* promoter activity. However, the replacement of an allele of the SNP did not change the activity of regulatory element. The authors explains it by the fact that SNP may affect the recruitment of regulatory proteins required for the epigenetic modification of enhancer region, which would not be observed in reporter assay [135]. Interestingly, Loughlin et al observed that the regulatory activity of osteoarthritis risk region may be caused by the combined effect of two SNPs: rs835487 and rs835488 [137]. Further, to seek the evidence that putative causal SNP(s) lies in enhancer element, allele-specific protein binding to a region of interest may be investigated. The most popular technique used for this purpose is called electrophoretic mobility shift assay (EMSA). In the above-mentioned study of the locus associated with breast cancer, EMSA revealed allele-specific binding to SNP rs10941679. Four proteins (FOXA1, FOXA2, CEBPB

and OCT1) were identified by a variation of EMSA, named supershift assay. The authors tried to confirm binding of these proteins *in vivo* by chromatin immunoprecipitation (ChIP). They observed the enrichment with FOXA1 and OCT1 binding to DNA region of SNP rs10941679, but without difference in binding between alleles of the SNP. In the end, authors concluded that the strongest signal of association is mediated through coordinated activation of FGF10 and MRPS30 [135]. The group working on locus associated with schizophrenia found that 13 SNPs from identified locus displayed allele-specific differences in nuclear protein binding affinity. Hence, they suspect that multiple SNPs participate in regulation of *CACNA1C* gene expression [136]. In the osteoarthritis locus, supershift EMSAs identified few TFs (SP1, SP3, YY1 and SUB1) binding stronger to the risk alleles of the two SNPs, rs835487 and rs835488. Researchers concluded that the association of this region is the result of different levels of expression of target gene caused by allele-dependent protein binding to both SNPs [137].

Determining the allele-specific binding of regulatory proteins to DNA

The majority of processes that influence gene expression are mediated through regulatory proteins, mainly transcription factors. Hence, the causal SNPs may affect the binding affinity of transcription factors to DNA regulatory regions and further modulate expression of a target gene [109].

One of the methods investigating the ability of nuclear proteins to bind to DNA region bearing SNP of interest in vitro is called electrophoretic mobility shift assay (EMSA). EMSA is a simple, robust and extremely sensitive technique that allows detection of sequence-specific protein-DNA interactions. This method consists of five basic steps. First, the preparation of protein sample. If protein of interest is unknown, nuclear or cytoplasmic extract are used. Second, the construction and labeling of probes. Probes are DNA or RNA fragments consisting of a particular protein binding site. Labeled DNA fragments can be from 20 up to 300 bp long. Interpretation of results may become complicated for longer probes, since they may contain binding motifs for numerous proteins. Third, the preparation of binding reactions. In the binding reactions, the protein mixture is bound to the probe. The conditions of binding reactions need to be precisely optimized to mimic the binding conditions specific for the studied protein-DNA interaction. Fourth step includes preparation of non-denaturing gel and running of binding reactions on the gel. Common gels for EMSA are 4% to 5% polyacrylamide gels. Very large protein-DNA complexes may be

run on agarose gels. The choice of appropriate running buffer is important to obtain the stability of protein-DNA complexes. Typical buffers used are Tris-borate (TBE), Tris-acetate (TAE) and Tris-glycine electrophoresis buffers. The last step, consists of detection of DNA-protein complexes using chemiluminescence, fluorescence, radioactivity or immunohistochemical approaches. For identification of allele-specific binding proteins, two labeled probes can be designed, containing region of SNP, one with reference and other with variant allele. Next, probes are incubated separately in binding reactions with nuclear extract and affinity of protein binding to the probes can be compared for each allele of genetic variant [138, 139].

Detection of protein binding to the DNA region of interest (labeled probe) by EMSA is based on the principle that in a non-denaturing gel, negatively charged nucleic acid migrates towards anode when subjected to electric current (Figure 7B; lane 1). Longer sequences, having higher molecular weight migrate slower than shorter and lighter probes. The conformation and circularity of nucleic acid can also alter its gel migration. Higher structures could be eliminated by denaturation, however this step cannot be applied in EMSA, because it would disrupt DNA-protein interactions. If protein interacts with studied probe, the protein-nucleic acid complex will have higher molecular weight than probe alone and adjusted charge. Formed complex will alter migration of free probe and result in shift of observed band (Figure 7B; lane 2) [138]. The specificity of formed complex can be assessed by competition binding assay. It is important to verify specificity of binding, especially when using crude nuclear or cytoplasmic protein extracts, which contain specific and non-specific DNA binding proteins. In the competition binding assay, the specific competitor, of the same DNA sequence than probe but not labeled, is added to the protein mix in the binding reaction before addition of labeled probe. If specific competitor competes binding and results in weaker intensity of observed band, binding can be defined as specific (Figure 7B; lane 3). Additionally, non-specific competitor may be used as well, which can be any unrelated sequence but generally it is the same unlabeled sequence like specific competitor but with mutation in the binding site. Addition of non-specific competitor should result in weaker competition than with specific competitor or no competition at all.

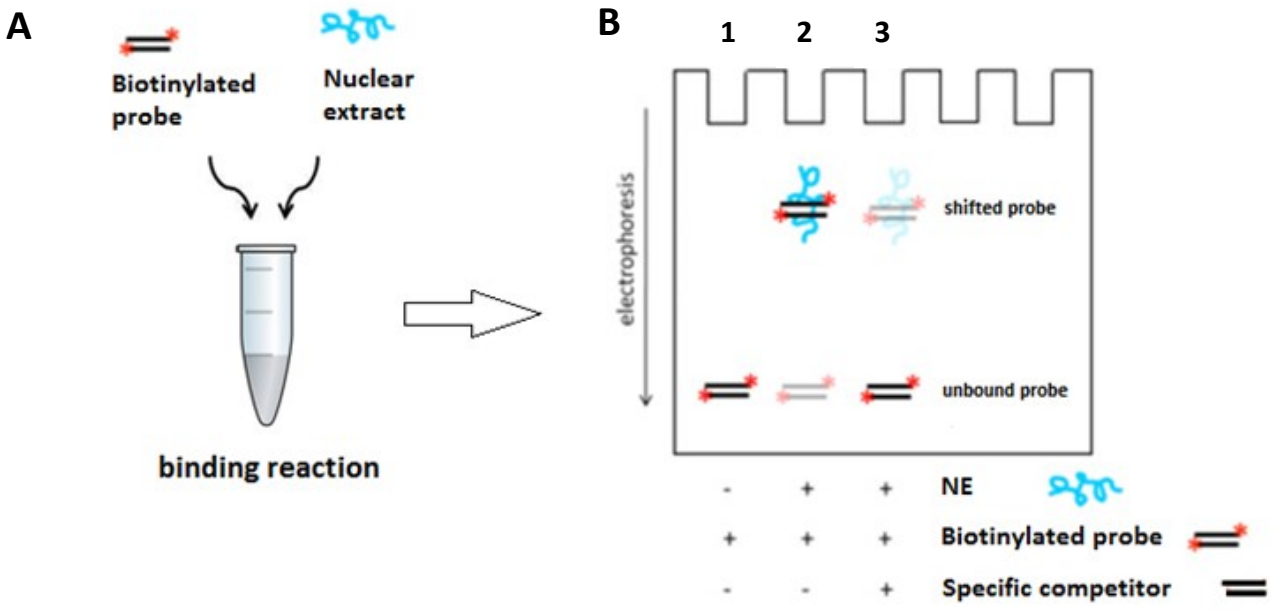

Figure 7 : Schematic representation of electrophoretic mobility shift (EMSA) assay. (A) Labeled probe is incubated together with nuclear extract (NE) in the binding reaction. (B) Different mobility of binding EMSA reactions through non-denaturing gel. Lane 1 – free labeled probe; Lane 2 – Shift of band representing protein-DNA complex; Lane 3 – competitor assay, band intensity decreases after addition of specific competitor to the binding reaction; Retrieved from dx.doi.org/10.17504/protocols.io.k3ccyiw [140].

Although basic EMSA method allows detection of protein-DNA interactions, binding specificity and quantitative determination of binding affinity, it does not provide the answer about molecular weight or identity of protein present in the complex. Variations of EMSA, such as antibody supershift assay and competitor assay with consensus TF-binding sites can help with identification of proteins present in the complex. In the supershift assay, specific antibody is added to the binding reaction. If the protein recognized by selected antibody is present in DNA-protein complex, the addition of antibody should result in formation of antibody-protein-DNA complex and further shift in gel migration (supershift). Second method, competitor assay with consensus TF-binding sites is based on the same principle as competitor assay performed to verify specificity of binding, but uses unlabeled consensus TF-binding sites instead of specific competitor. If addition of consensus binding site for specific transcription factor compete the protein-probe interactions, the related TF may be present in the complex [139].

Described variations of EMSA can facilitate the identification of protein present in the complex, but they require prior knowledge of putative proteins that may bind to studied region. The list of

putative proteins can be created using bioinformatic predictions and be tested either by supershift or competitor assay. However, it is a difficult task to predict which regulatory proteins can bind DNA region of interest. The method enabling indentification of DNA-binding proteins without prior knowledge of putative proteins is DNA-affinity chromatography followed by mass spectrometry. Similarly like in the EMSA technique labeled probes containing DNA region of interest are incubated with the protein extract. Then DNA-protein complexes are purified, proteins are eluted and analysed by mass spectrometry. This method is unbiased and allowes detection of allele-specific protein binding [141].

Since, EMSA and DNA-affinity chromatography followed by mass spectrometry are both *in vitro* methods, allele-dependent protein binding should be confirmed *in vivo* by chromatin immunoprecipitation (ChIP). In the ChIP assay DNA-protein complexes are crosslinked *in vivo* and after shearing of chromatin into fragments, DNA regions bound to protein of interest are immunoprecipitated by specific antibody. Enriched DNA fragments can be analysed by qPCR with allele-specific primers to detect difference in protein binding between alleles of SNP in a heterozygous cell line [109, 142].

Testing transcriptional regulation of genetic variant

The effect of allele-specific binding proteins on transcription can be further tested by cell-culture based reporter assays. Cassettes containing the region of SNP can be cloned into plasmid with reporter gene and transfected into relevant cell type. Depending on which regulatory activity of genetic variant is analysed, cassettes may be cloned into different positions with respect to the reporter gene. Depending on activity of cloned element, different levels of mRNA are transcribed and then translated into active protein. Activity of protein can be measured and compared in between alleles of a SNP. The effect of interactions between multiple variants on transcription can be tested by subcloning to the reporter vector an insert bearing more than one SNP.

In general, reporter assays are very useful in determining potential function of genetic variants, however these assays have few limitations. High variability in results is often observed for reporter assays, due to significant amount of transcriptional noise. Also, slight differences in reporter activity can be caused by small differences in amount of transfected plasmid DNA or transfection efficiency. And most importantly, reporter assays do not test the regulatory activity of genetic variants in the native genomic context. The method overcoming these issues is

genome-editing. The most recent clustered regularly interspaced short palindromic repeat (CRISPR) systems allow either deletion of whole or parts of the linkage disequilibrium block detected by GWAS to analyze the regulatory function of this region, or mutation of single allele of SNP to determine allele-specific regulatory functions [109].

After successful identification of causal variant and its regulatory function next steps are to determine affected target gene(s) and to understand how the altered regulation or function of causal genes are associated with disease risk or certain phenotype.

1.8 Justification of the cell type choice for identification of causal SNP associated with cardiovascular responses of patients treated with dalcetrapib

The results from many GWAS follow up studies suggest that majority of causal variants influence the phenotype by modulating the function of regulatory elements, what leads to altered expression of target genes in cell-type specific manner. That is why selection of cell-type for follow up studies is very important. Very often, eQTL effect is seen in cell types known to be relevant in affected disease [109]. Therefore, since monocytes play important role in atherosclerotic plaque formation and reverse cholesterol transport the monocytic cell lines are good candidates to study association of genetic variants and cardiovascular response of patients treated with dalcetrapib. Moreover, SNP rs1967309 and many other variants in high LD with it, were found to be eQTLs for *ADCY9* gene expression in blood and peripheral blood $CD14^{+}$ $CD16^{-}$ monocytes [143, 144]. Since, the activity of ADCY9 mediate different functions of immune cells and regulate various inflammatory responses [102-106], the fact that genotype of patients at SNP rs1967309 was found to be associated with changes of C-reactive protein levels suggests possible effects of genetic variants on inflammatory responses of patients [84]. Further, according to the data collected in FANTOM5 database, transcription start sites (indicating presence of promoters) are located at DNA regions overlapping SNPs rs1967309 and rs12920508, the last showing signal mainly in cells of monocytic origin. Additionally ENCODE data suggest open chromatin state in these regions in primary monocytes.

Therefore, two cell line models for human monocytes were selected for these functional studies. In the majority of experiments, undifferentiated THP-1 monocytes were used, which are immortalized cells isolated from the pheripheral blood of a 1-year old male patient with acute monocytic leukemia [145]. A second cell line, U937, was used in minority of experiments. The

U937 cells were isolated from the histiocytic lymphoma of a 37-year-old male patient [146]. Both cell lines have many common features. They are used often as models to examine monocyte and macrophage-realted physiological processes and they are both immortalized cell lines what brings them many advantages over human primary monocytes. The most obvious advantages are their easy cultivation and safety of use, high growing rate that enables collection of large amount of cells for downstream analysis and their ability to grow up to high passage without significant changes of cell sensitivity and activity. The other advantage of THP-1 and U937 cells over human primary monocytes is their homogeneous genetic background, which minimizes the degree of variability in the cell phenotype and results in higher reproducibility of findings [147]. The main disadvantage of THP-1 and U937 cell lines is their cultivation under controlled conditions, what may result in different sensitivities and responses compared to somatic cells in their natural environment [148]. The other drawback is lack of possibly relevant interactions between the target cells and surrounding cells, as in natural tissues. This issue may be important for our studies since the association of patient genotype was visible only in patients treated with dalcetrapib, therefore the effect may not be observed in isolated cells without interactions with surrrounding [149].

Despite high similarity of THP-1 and U937, few differences exist between those cell lines. Main differences are their origin and maturation stage [147]. THP-1 are isolated from peripheral blood when U937 originate from tissue, hence more mature stage. Other differences significant for our studies include differences in their genotype. THP-1 cells are homozygotes GG for SNP rs1697309. THP-1 cells are homozygotes for allele CC for SNP rs12920508, which allele, due to high linkage disequilibrium between SNPs rs1967309 and 12920508 ($R^2 = 0.98$ in dal − OUTCOMES population) is associated to the deleterious allele. Wild-type U-937 cells are heterozygotes for both SNPs. Our laboratory designed CRISPR-Cas9 clones of THP-1 cells for SNP rs1937309 what gives advantage for use of this cell line. However, wild-type U937 cells are heterozygotes for our both SNPs of interest rs1967309 and rs12920508, what enables us to perform analysis of allele-specific protein binding in vivo with more physiologically natural state. Therefore the main reason for using two different cell lines in this work is their genotype.

2 HYPOTHESIS AND OBJECTIVES

Personalized medicine is a promising approach to improve patient care. It has potential to adjust therapies depending on patient's genetic profile, to result in best response and the highest safety of the treatment [150]. The response of patients treated with dalcetrapib is associated with different genotype of patients. Understanding the molecular mechanisms driving these associations is important for further development of dalcetrapib-based personalized treatment, for patients having suitable profile. Genetic variants associated with responses of patients to dalcetrapib are located within intronic regions of *ADCY9* gene. Many of them are in high LD, and it is unknown which SNP causes the association. Since the majority of causal intronic SNPs function through regulation of target gene expression by allele-specific recruitment of transcription factors, we hypothesize that SNP rs1967309 or one of the other identified SNPs in high LD with it affects *ADCY9* expression. Given the importance of monocytes in atherosclerosis and inflammatory responses, and the association between rs1967309 and monocytic counts [80], monocytes represent cells of particular interest. The possible mechanism would imply altered binding affinity of regulatory protein(s) to alleles of the causal SNP, what may further lead to altered target gene expression, and modulate response of patients under dalcetrapib treatment.

The general objectives of this project are to identify the causal single nucleotide polymorphisms (SNPs) and to understand the molecular mechanism connecting causal SNPs and dalcetrapib effects on cardiovascular outcomes. To answer general objectives, I set three specific aims:

1. Identification of specific DNA-protein complexes

DNA-binding regulatory proteins mediate most processes that modulate gene expression. Therefore, I will investigate the allele-specific DNA-binding of proteins to selected SNPs in order to prioritize functional variants.

2. Determination of transcriptional activity of selected genetic variants

The effects of SNP on gene expression levels in specific cell lines are great predictors of functionality of genetic variant. I will analyze the transcriptional and enhancer activities of DNA regions containing SNPs of interest to narrow down list of putative causal variants.

3. Identification of causal SNP(s) and elucidation of molecular mechanism altering *ADCY9* gene expression

Based on the results obtained from specific DNA-protein binding and transcriptional activity assays I will restrict the list of putative causal SNPs and I set out to identify possible mechanisms linking putative causal SNPs and *ADCY9* gene expression.

3 ARTICLE

Student contribution

Over the time of my master project, I performed the majority of presented results. I optimized and carried out the approaches, such as electrophoretic mobility shift assay to detect protein binding to the SNPs, dual luciferase reporter assay to assess the effect of selected polymorphisms on gene transcription, multiplexed competitor EMSA, DNA-affinity chromatography followed by mass spectrometry and chromatin immunoprecipitation to identify proteins binding to chosen SNP. Furthermore, I mainly contributed to writing of presented article. Co-author Rocio Sanchez prepared plasmids for dual luciferase reporter assay and performed transfections (n=4) for cassettes 3 and 9, containing SNPs rs1967309 and rs12920508, respectively. Co-author Gabriel Théberge-Julien contributed to the presented article mainly with his knowledge about PCRs. Co-author Marie-Pierre Dubé together with my director Jean-Claude Tardif and co-director Eric Rhéaume are the researchers who discovered the *ADCY9* risk loci and supervised my work.

Toward identification of the causal SNP determining dalcetrapib responses

Magdalena Burchert[1], Rocio Sanchez[1], Gabriel Théberge-Julien[1], Marie-Pierre Dubé[1,2], Eric Rhéaume[1,2], Jean-Claude Tardif[1,2*]

[1] Montreal Heart Institute, Montreal, Quebec, Canada

[2] Faculty of medicine, Université de Montréal, Quebec, Canada

---

*To whom correspondence should be addressed:

Jean-Claude Tardif, MD

Montreal Heart Institute Research Center,

5000 Belanger Street, Montreal, Canada, H1T 1C8

Fax: 1-514-593-2500

Phone: 1-514-376-3330 ext. 3604

e-mail: jean-claude.tardif@icm-mhi.org

Abstract

The cardiovascular responses of patients with coronary artery disease receiving dalcetrapib treatment differ depending on their genetic profile. Genome-wide association study identified an associated locus within intronic regions of adenylate cyclase 9 (*ADCY9*) gene, but the detailed molecular mechanism underlying this association remains unknown. Our main objective was to identify putative causal genetic variant(s) from associated locus to gain knowledge into the mechanism involved.

Using electrophoretic mobility shift assay (EMSA), we demonstrated that seven single nucleotide polymorphisms (SNPs) of *ADCY9* linkage disequilibrium block show differential allelic binding to nuclear proteins from THP-1 monocytic cell line, however, binding was not influenced by direct exposure of cells to dalcetrapib. Among these variants, the regions bearing SNPs rs1967309 and rs12920508 demonstrated increased promoter activity in dual luciferase reporter assays. Mass spectrometry identified 11 proteins binding significantly stronger to allele C (deleterious) of rs12920508 and we confirmed binding of TAR DNA-binding protein 43 by chromatin immunoprecipitation. We discovered that DNA construct containing both SNPs rs1967309 and rs12920508 showed elevated promoter activity with significantly higher activity for deleterious haplotype, what corresponded the tendency of SNP rs1967309.

Taken together, our data support hypothesis that polymorphisms located in *ADCY9* gene are associated with cardiovascular responses of patients induced by dalcetrapib. Additionally, although SNP rs12920508 appears in a region that may possess transcriptional activity, SNP rs1967309 may overshadow its effect and seems to be responsible for the observed effect of this region in THP-1 monocytic cell line. Still, further work is required to understand the link between potentially causal SNPs and cardiovascular responses induced by dalcetrapib.

Keywords: Dalcetrapib, *ADCY9*, SNP, EMSA, cardiovascular disease, HDL-C

Introduction

According to data of the World Health Organization from May 2017, cardiovascular diseases (CVDs) are the number one cause of death globally [1]. Reduced levels of high-density lipoprotein cholesterol (HDL-C) are commonly recognized as a major CVD risk factor. One of the main antiatherogenic function of HDL may involve its role in reverse cholesterol transport (RCT). In this process high-density lipoprotein particles allow the excretion of cholesterol accumulated in macrophage foam cells, what leads to plaque regression. Other atheroprotective properties of HDL include its antioxidant, anti-inflammatory and anti-thrombotic functions [2].

Dalcetrapib is a cholesteryl ester transfer protein (CETP) inhibitor, which aims to increase levels of HDL-C in patients. Its mechanism of action lies in inhibition of transfer of cholesteryl esters from HDL to apolipoprotein B-containing particles, such as low-density lipoproteins (LDLs). Although clinical trials showed the safety of dalcetrapib and its efficacy to raise HDL-C levels by around 30 %, its development was suspended due to the lack of clinically significant benefits [3]. Genome-wide association study (GWAS) of the dalcetrapib clinical trial dal-OUTCOMES identified a single region located in adenylate cyclase 9 (*ADCY9*) gene that is associated with cardiovascular events in dalcetrapib-treated patients. In the identified region, genetic variant rs1967309 passed the significance threshold with P value = 2.41×10^{-8}. In the same region, 7 other single nucleotide polymorphisms (SNPs) were identified with P value $< 1 \times 10^{-6}$. In patients bearing the AA genotype (minor protective allele) for SNP rs1967309 and treated with dalcetrapib, there was a 39% reduction in the cardiovascular risk compared to placebo. Contrary, for patients with GG genotype (major deleterious allele), a 27% increase in cardiovascular events was detected in dalcetrapib arm versus placebo. For heterozygote patients AG and placebo arm alone, there was no significant effect observed. A panel of 27 SNPs located in the *ADCY9* gene was selected for targeted genotyping and analysis of samples from dal-PLAQUE-2 study participants to confirm association of identified locus and cardiovascular responses of patients to dalcetrapib. Selected panel included 5 imputed SNPs with P value $= 10^{-6}$ and 15 genotyped SNPs with P value < 0.05 identified in the discovery GWAS to the dal-OUTCOMES study, together with 7 SNPs chosen based on literature review and functional prediction. Ten single nucleotide polymorphisms displayed association with measure of intima-media thickness and reached the significance threshold (P value≤ 0.05) [4]. Additionally, the genotype of patients

from dal-OUTCOMES at SNP rs1967309 was associated with changes of C-reactive protein (CRP) levels under treatment with dalcetrapib. Only patients with protective haplotype AA did not show an increase in CRP levels [5].

Single nucleotide polymorphisms associated with responses of patients treated with dalcetrapib are located within *ADCY9* gene. Majority of identified SNPs are nonrandomly associated and create a linkage disequilibrium block that occupies the regions of intron 2 and 3 of *ADCY9* gene [4]. Adenylate cyclase 9 is a membrane-associated protein that catalyzes the conversion of adenosine triphosphate (ATP) to second messenger cyclic AMP (cAMP) [6]. The mechanisms underlying the association of *ADCY9* with cardiovascular responses to dalcetrapib are still not fully understood. However, the activity of ADCY9 was found to mediate different functions of immune cells and to regulate various inflammatory responses [7-11]. Also, recently published data showed that inactivation of *Adcy9* in mice has atheroprotective effects but only in the absence of CETP [12].

Despite the rapid advancement of GWAS, many identified associations still lack mechanistical understanding. Functional follow up on GWAS data includes multiples steps, such as fine mapping of GWAS signal, bioinformatic analysis, identification of causal SNP(s) and its function, and identification of target gene(s) and its link to the phenotype [13]. Identification of causal SNP(s) is a challenging task, since majority of identified loci lie in non-coding parts of a genome [14]. Our task may be even more challenging as the cardioprotective response was only observed in dalcetrapib-treated patients.

Disease-associated variants located in intergenic or intronic regions often function through modulation of genetic expression of one or more target genes. Gene expression is a complex multistep process regulated by various factors on many levels. Therefore, noncoding variants may influence the expression of a target gene by regulating many different steps, including gene transcription, RNA splicing, non-coding RNA function and epigenetic regulation [13]. Genome-wide data on chromatin accessibility showed that almost 60% of intronic polymorphisms identified by GWAS lie within DNase I hypersensitivity sites (DHSs). These data suggest that genetic variants in non-coding regions of human genome disrupt DNA binding motifs of regulatory proteins. Numerous functional analysis of GWAS data identify variable transcription factor (TF)-DNA interactions as leading cause of phenotypic variation [14, 15].

Expression quantitative trait loci (eQTL) analysis allows to determine the effect of polymorphisms on gene expression levels in specific tissues or cell lines. The annotation of SNPs to eQTL regions in tissue or cell type of interest facilitates identification of causal SNP(s) [16]. Polymorphism rs1967309 and/or other genetic variants in high linkage disequilibrium with it, are associated with *ADCY9* gene expression in blood and peripheral blood $CD14^{+}CD16^{-}$ monocytes [17, 18].

Given the above, we hypothesize that SNP rs1967309 or one of the other identified SNPs in high LD with it, regulate expression of *ADCY9* gene in monocytes. The putative mechanism would imply altered binding affinity of regulatory protein(s) to alleles of the causal SNP, what may further lead to altered target gene expression, and modulated response of patients under dalcetrapib treatment.

Here, we characterize genetic variant rs1967309 and 11 other selected SNPs with final goal of identifying the SNP(s) that cause the association (causal SNP), and understanding its link with dalcetrapib effects on cardiovascular outcomes. First, we tested if any proteins bind in allele-dependent manner to regions of selected SNPs. Next, we assessed the influence of chosen genetic variants on gene transcription. Finally, we identify proteins binding to the regions of the putative causal SNP.

Materials and methods

SNP selection

Twelve lead SNPs were selected as putative causal SNPs: rs1967309, rs2531971, rs2238448, rs11647778, rs1259911, rs12595857, rs12920508, rs11647828, rs12935810, rs74702385, rs3789035, rs2238449 (Table I and Figure 1). All SNPs (n=7) with p values < 10^{-6} for association with cardiovascular responses of patients treated with dalcetrapib in dal-OUTCOMES were chosen [4]. Three SNPs with p value between 1.32×10^{-6} and 7.31×10^{-5} were selected because they seemed to occur in regions of interest based on ENCODE data. Variant rs3789035 was selected due to its low P value (9.85×10^{-7}) in high-density imputation. SNP rs2238449 was chosen because it was found to be an eQTL for *ADCY9* in peripheral blood $CD14^{+}CD16^{-}$ monocytes [17] and is in moderate LD with rs1967309.

Cell culture

THP-1 and U-937 cell lines were obtained from ATCC® and both are human monocytic cell lines. THP-1 and U-937 cells were grown in Roswell Park Memorial Institute medium (RPMI-1640, Wisent Bioproducts) supplemented with 10 % fetal bovine serum (FBS, Wisent Bioproducts). Both cell lines were cultured at 37 ℃ and 5 % $\boldsymbol{CO_2}$.

THP-1 cells are homozygotes GG for SNP rs1697309. THP-1 cells are homozygotes for allele CC for SNP rs12920508, which allele, due to high linkage disequilibrium between SNPs rs1967309 and 12920508 ($R^2 = 0.98$ in dal − OUTCOMES population) is associated to the deleterious allele. Wild-type U-937 cells are heterozygotes for both SNPs.

Electrophoretic mobility shift assay (EMSA)

Probe synthesis

For each selected SNP, biotinylated DNA probes of 31 bp long sequences, 15 bp on each side of the SNPs, were designed (Table II). Synthesized probes encompassed the major or minor variant of every polymorphism. For each allele of the SNP, forward and reverse complementary sequences single-strand DNA were first synthesized and were further annealed together to generate double-strand DNA probes. Specific competitor probes encompassed the same sequence and the same allele of variant like biotinylated probes but were not biotinylated.

Nuclear proteins extraction

Nuclear proteins were extracted from THP-1 monocytic cell line. Extractions were performed using NE-PER nuclear and cytoplasmic extraction kit reagents (Thermo-Fisher), according to the manufacturer's protocol. Protein concentration was further assessed by Lowry method.

To assess if DNA-protein binding was influenced by dalcetrapib, THP-1 cells were exposed to 5 μM dalcetrapib for 24 hours and then nuclear proteins were extracted from those cells as above.

Binding reactions preparation

Electrophoretic mobility shift assay was performed according to the manufacturer's protocol (LightShift Chemiluminescent EMSA Kit; Thermo Scientific). In brief, the components of binding reactions: 1X binding buffer, 2.5 % glycerol, 5 mM $MgCl_2$, 0.5 μg Poly(dI-dC), 25 ng calf thymus DNA and 0.05 % NP-40 were incubated with 4 μg of nuclear proteins and 200 fmol of biotinylated probe. In competition assays, 100x non-labeled competitor probes relative to the labeled probe concentration were added to the reactions before addition of biotinylated probe. Reactions were incubated at room temperature for 20 min, followed by addition of 5X loading buffer.

Gel electrophoresis and exposure

Samples were electrophoresed on a 4 % polyacrylamide gel for 40 min at 100 V and then transferred to nylon membrane. Membranes were UV crosslinked and visualized by Chemiluminescent nucleic acid detection module kit (Thermo Scientific). The images were taken using ChemiDoc MP (Bio-Rad) and bands were quantified using Image Lab 5.2.1.

Dual luciferase reporter assay

Dual luciferase reporter assays were used to measure the effects of selected polymorphisms on gene transcription activity. To assess enhancer activity of genetic variants, DNA fragments with sequences flanking either the deleterious or protective alleles of selected SNPs were individually amplified by PCR from genomic DNA and then inserted into the enhancer position of the pGL3SV40 promoter vector. To assess promoter activity of genetic variants, DNA fragments containing either the deleterious or protective alleles of selected SNPs were individually inserted into the promoter site of pGL4.10[luc2] vector, lacking promoter. Isolated genomic DNA from

five individuals with minor haplotype and five individuals with major haplotype were obtained from patients from dal-OUTCOMES study [4]. Next, one genomic DNA was selected for PCR amplification. The list of primers used is presented in Table III. Amplified DNA fragments were cleaved by BamHI restriction enzymes, and then subcloned into BamHI restriction site of pGL3SV40 promoter vector, or in the BamHI-compatible BglII restriction site of pGL4.10[luc2] vector. Cloned amplified fragments were sequenced to validate the absence of unwanted mutation. DNA fragments, commonly called cassettes, were $\sim 500-1000$ bp long DNA fragments obtained from individual with either the deleterious or protective haplotype. See Table IV for details about size and location of the tested DNA fragments.

Exceptionally, longer DNA fragment (4185 bp), called Insert A was prepared differently than shorter cassettes. The synthetic gene HapP/D-Insert A was assembled from synthetic oligonucleotides and PCR products (Life technologies). Next, assembled fragment was cloned into plasmid pMA (ampR) using PacI and AscI cloning sites. Plasmids were then purified and inserts were verified by sequencing. Inserts were cloned into enhancer position of pGL3SV40 vector, using XhoI-SacI restriction sites. Since only deleterious insert could successfully be subcloned, the corresponding protective insert was created by site-directed mutagenesis. DNA fragment of insert A contains both SNPs rs1967309 and 12920508. Hence, to generate insert with protective haplotype (PA(7309-0508)-pGL4Basic) these two SNPs were mutated in deleterious insert (DA-pGL4Basic). Site-directed mutagenesis was carried out using QuickChange Lighting Multi Site-Directed Mutagenesis Kit (Stratagene), according to the manufacturer's protocol. The resulting insert sequence was verified by sequencing.

Luciferase reporter constructs were transfected into THP-1 cells along with the renilla control reporter vector (pRL-CMV) (Promega). Transfection was carried out in 96-well plates with white bottom (BrandTech Scientific). In each well, 200 000 THP-1 cells were plated and transfected with 9 µl of reagent-DNA complex. Reagent-DNA complex contained 0.4 µl of TransIT-2020 (Mirus), OPTI-MEM reduced-serum media (Gibco), 90 ng of luciferase reporter vector and 10 ng of renilla control vector, per well. After 48 hours of transfection, luciferase and renilla activities were measured by Synergy 2 plate reader (BioTek), using Dual-Glo Luciferase Assay System (Promega) according to manufacturer's protocol. Normalized firefly luciferase activity was

calculated by dividing measured luciferase activity by the renilla luciferase activity from the same transfected sample.

Multiplexed competitor EMSA (MC-EMSA)

We used two different approaches to identify the proteins binding to SNP rs12920508 location. First, multiplexed competitor EMSA uses cocktails of consensus DNA-binding sequence oligonucleotides of well characterized transcription factors to compete the specific band shift from conventional EMSA [19]. Seven cocktails were used, each containing unlabeled consensus short sequence double-strand DNA for ten TFs (Table V). Competition cocktails were incubated with nuclear extracts in EMSA binding reactions for 30 minutes on ice, before addition of labeled probe.

Consensus binding sequences DNA for transcription factors from the cocktail competing specific band shift from EMSA were further added separately to the binding reactions to determine which TF-binding DNA was responsible for competition.

DNA-affinity chromatography followed by mass spectrometry

The second method we used to identify proteins binding to the site of polymorphism rs12920508 is called DNA-affinity chromatography followed by mass spectrometry. Firstly, 17 pmol of biotinylated probes bearing either deleterious (C) or protective (G) allele of variant rs12920508 were incubated with 340 µg of nuclear extracts derived from THP-1 monocytes for 25 minutes at room temperature. The conditions of binding reactions were the same as for the EMSA experiments. Next, 10 µl of streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen by Thermo Scientific) were incubated overnight with reactions at 4℃ on a rotating platform. After magnetic separation, the beads were washed three times with 400 µl 50 mM ammonium bicarbonate solution. Afterwards, beads were suspended in 20 µl of the washing buffer and sent for mass spectrometry analyses to the Proteomics technology platform located at the Research Institute - McGill University Health Centre (RI-MUHC). Beads with bound proteins from samples, were boiled in Laemmli buffer and the resulting beads & buffer were loaded onto a single stacking gel band to remove lipids, detergents and salts. A single gel band containing all proteins was reduced with DTT, alkylated with iodoacetic acid and digested with trypsin. Extracted peptides were re-solubilized in 0.1% aqueous formic acid and loaded onto a Thermo Acclaim Pepmap (Thermo, 75 µm ID X 2 cm C18 3 µm beads) precolumn and then onto an

Acclaim Pepmap Easyspray (Thermo, 75 μm X 15 cm with 2 μm C18 beads) analytical column separation using a Dionex Ultimate 3000 uHPLC at 220 nl/min with a gradient of 2-35% organic solvent (0.1% formic acid in acetonitrile) over 2 hours. Peptides were analyzed using a Thermo Orbitrap Fusion mass spectrometer operating at 120,000 resolution (FWHM in MS1) with HCD sequencing at top speed (15,000 FWHM) of all peptides with a charge of 2+ or greater. The raw data were converted into *.mgf format (Mascot generic format) for searching using the Mascot 2.5.1 search engine (Matrix Science) against human protein sequences (Uniprot 2018). The database search results were loaded onto Scaffold Q+ Scaffold_4.4.8 (Proteome Sciences) for statistical treatment and data visualization. Significance was determined by Fisher's exact test.

To control for nonspecific binding, two controls were designed. The first control contained biotinylated scrambled probe in order to exclude nonspecific binding to the probes. Second control was lacking biotinylated probe, what enabled us to detect nonspecific binding of proteins directly to the magnetic beads.

Chromatin immunoprecipitation (ChIP)

We performed chromatin immunoprecipitation to assess if proteins identified by mass spectrometry are binding to region of SNP rs12920508 *in vivo*. The assay was carried out according to the manufacturer's protocol (ChIP-IT Express, Active Motif). Approximately 15 million of U-937 cells were fixed with 0.54 ml of 37 % formaldehyde for 10 minutes at room temperature on a shaking platform. Next, glycine-stop solution was added to the cells and incubated at room temperature for 5 minutes. Supernatant was removed after centrifugation for 10 minutes at 2500 rpm in 4 ℃ and cells were washed with 1X PBS. Cells were centrifuged one more time in same conditions and resuspended in 1X PBS supplemented with PMSF. Crosslinked cells were next lysed and homogenized by 10 strokes in Dounce homogenizer to help nuclei release. Chromatin was sheared by 2 pulses of sonication, each lasting 15 seconds with 30 second pause on ice in between.

The concentration of DNA was measured by NanoDrop. The size of chromatin fragments was verified by migration of sheared chromatin sample on a 1 % agarose gel in TAE buffer. The size of fragments obtained after shearing was 200-500 base pairs.

ChIP reactions contained 25 µl of protein G magnetic beads, 20 µl ChIP buffer 1, 60 µg of sheared chromatin DNA, 2 µl of protease inhibitor cocktail and 4 µg of antibody (Table VI). After addition of all components, reactions were incubated overnight at 4 ℃ on a rotating platform. Magnetic beads were next washed once with 800 µl of ChIP Buffer 1 and two times with 800 µl of ChIP Buffer 2. Bound fragments of chromatin were eluted and analyzed by quantitative PCR and ddPCR (supplementary methods). Primers used for amplification of region containing SNP rs12920508 are presented in Table VII.

Statistical analysis

Statistical analyzes were carried out using GraphPad Prism 7 software or SAS 9.4 for multilevel model.

For all collected data, Shapiro-Wilk normality test was performed in order to check distribution of data. Further, outliers located below Q1-1.5 IQR or above Q3+1.5 IQR were calculated and data were analyzed with and without them to identify any possible false results.

For dual luciferase results, significance was determined by the multilevel model specified in term of fixed effects for vector construct (deleterious, protective, control) as repeated factors. The random effects within the model was the intercept and vector, to take into account correlation within cells and within vector x cells.

For ChIP data analyzed by qPCR, results were presented after log_2 transformation and significance was determined by one sample t-test.

Results

Protein binding to regions of selected SNPs

Seven SNPs show allele-specific protein binding

Since all 12 selected SNPs are located in intronic regions of *ADCY9*, the most likely way they could influence gene expression is by allele-specific transcription factor binding. To test if any nuclear proteins from THP-1 monocytic cell line bind to the regions of the selected SNPs and if allele alteration affects binding, we performed electrophoretic mobility shift assay (EMSA).

First, we assessed if nuclear proteins derived from THP-1 bind to 31-bp-DNA probes centered around each of the 12 selected SNPs, and if the binding is allele-dependent. An example of such results is shown for rs1967309 (Figure 2). We set a minimum ratio of quantified bands between alleles of the same SNP higher than 2 to assign the protein binding to be allele-dependent.

Next, we checked protein binding specificity by competition assays. According to this principle, protein binding is specific if band intensity decreases after addition of specific competitor to the binding reaction. To conclude that binding was specific to variants of interest, we set the cut off to be at least 2 times decrease after addition of specific competitor for the stronger binding allele.

EMSA experiments revealed that binding of nuclear proteins derived from THP-1 can be detected to all tested probes for the genetic variants. Eight probes displayed allele-dependent protein binding which was also specific, according to competition experiments (Table VIII). Probes corresponding to the regions of polymorphisms rs1967309, rs12920508 and rs2238449 displayed the strongest protein binding, while protein binding to variant rs11647778 was very weak, thus we did not select it for further functional analysis. Four SNPs, rs2238448, rs12595857, rs12920508 and rs12935810 demonstrated increased binding to their haplotype-associated deleterious allele, whereas three SNPs rs1967309, rs2531971 and rs2238449 showed stronger binding to protective allele probes.

Exposure of THP-1 cells to dalcetrapib did not influence nuclear protein binding

After discovery of allele-dependent protein binding to seven variants of interest, we decided to verify if protein binding could be directly influenced by dalcetrapib. To do this, we performed EMSA experiments with nuclear extract derived from THP-1 monocytes exposed to 5 μM of dalcetrapib for 24 hours.

We calculated the ratio of binding intensity between alleles of SNPs before and after exposure to dalcetrapib (Figure 3). The ratio of binding intensity between alleles of variants did not change significantly after exposure to dalcetrapib. We conclude that direct exposure of THP-1 cells to dalcetrapib did not influence significantly nuclear protein binding for all seven tested SNP-containing probes.

Allele-specific regulatory activity of selected genetic variants

DNA regions containing SNPs rs1967309 and rs12920508 show transcriptional activities

In order to assess regulatory capacity of genetic variants that demonstrated allele-specific protein binding in EMSA experiments, we carried out dual luciferase reporter assays. For each selected genetic variant, two DNA fragments, thereafter called cassettes (see Table IV), were generated by PCR amplification from patients' DNA bearing protective and deleterious haplotype. To compare enhancer activity of selected SNPs' allele, cassettes were cloned into downstream of the luciferase cDNA in reporter vector. Rs2238449 and rs2238448 were included in one construct because of their physical proximity. Additionally, for SNPs rs1967309, rs12920508, 12595857 and rs12935810, luciferase reporter vectors with cassettes cloned into the promoter position, upstream of luciferase cDNA, were also generated to assess promoter activity.

For each SNP, four independent experiments were performed, and four replicate samples were assayed in each experiment. Relative luciferase activity was obtained by normalization to the luciferase activity of the control vectors pGLSV40 or pGL4Basic. Two cassettes, those bearing SNPs rs1967309 and rs12920508, initially showed increased transcriptional activity. SNP rs12920508 showed a tendency towards higher activity for protective haplotype. Exceptionally, for these two cassettes, assays with cassettes in promoter position were repeated ten times (Figure 4). After increasing repeats to n=10, P value illustrating difference in between alleles in promoter position was 0.0674 for SNP rs1967309 and 0.0389 for rs12920508. SNPs rs12595857 and rs12935810 did not display change in promoter activity in comparison to control vector (data not shown).

None of the 7 selected variants showed enhancer activity. Interestingly, constructs containing alleles of SNPs rs12920508 and rs1967309 reduced expression of reporter gene suggesting potential repressor activities, nevertheless without significant differences in between the alleles (Figure 4). Additionally, protective haplotype of cassette 4 (rs2531971) displayed lower enhancer

activity comparing to control plasmid, but without significant difference in activity with the deleterious haplotype (data not shown). In conclusion, of the 7 SNPs showing allele-specific protein binding in EMSA experiments, only cassettes with rs1967309 and rs12920508 displayed significantly increased transcriptional activity but without significant difference between alleles.

Identification of proteins binding to the DNA probe containing SNP rs12920508

Eleven proteins were identified binding significantly more to the probe bearing the deleterious allele of rs12920508

Considering that the DNA cassette region of variant rs12920508 displayed a tendency for higher activity of the protective allele in luciferase reporter assays, we decided to identify proteins binding to this region. We used an approach consisting of DNA-affinity chromatography followed by protein mass spectrometry. DNA-protein complexes binding specifically to EMSA probes with rs12920508 presenting each allele separately were pulled down by streptavidin coated magnetic beads. Afterwards, proteins were eluted and detected by mass spectrometry. Obtained data were analyzed using Scaffold 4 proteome software. P values representing enrichment of proteins binding to deleterious C allele over protective G, or over the control scramble sequence probe or magnetic beads without probe were calculated by Fisher's exact test.

In the EMSA experiments, the probe with variant rs12920508 showed three bands with higher protein binding to the C allele (data not shown). Ratios of band intensity between alleles (C/G) of samples sent to mass spectrometry analysis for higher, middle and lower bands were 1.7, 3.1 and 3.6, respectively.

Mass spectrometry identified 11 proteins, which were significantly enriched in the sample with probe C compared to probe G and both controls (Table IX). The majority of detected proteins are DNA or RNA-binding proteins. Some of them were found to be involved in splicing (HNRNPM, TARDBP), regulation of transcription (TARDBP, FUBP1 and FUBP3) or RNA degradation (MTREX, EXOSC10 and ZCCHC8).

TAR DNA-binding protein 43 binds to the chromatin region of variant rs12920508

To validate *in vivo* the *in vitro* results obtained by mass spectrometry, we performed chromatin immunoprecipitation (ChIP) in U-937 cells.

From 11 identified proteins, we selected four which showed the highest number of peptides detected for prepared sample with probe containing allele C (deleterious allele). Although previous analysis was done with nuclear extract derived from THP-1 monocytes, we decided to use U-937 monocytes for ChIP assays. Wild type THP-1 cells are homozygotes CC for SNP rs12920508, while U-937 cells are heterozygotes for this variant. Thus, U-937 cells could potentially enable determination of allele-specific protein binding. EMSA reactions with nuclear extract derived from U-937 demonstrated similar binding pattern to those with nuclear extract derived from THP-1 cells (data not shown).

DNA fragments enriched by ChIP were quantified in two ways, by classical approach quantitative PCR (qPCR) and droplet digital PCR (ddPCR). Sample with RNA polymerase II antibody served as positive control.

For qPCR analysis we used normalization based on the fold enrichment method. The results obtained represent the ChIP signal as the fold increase in signal relative to the background signal with non-specific IgG, calculated by the formula: $2^{-(Ct\ IP - Ct\ IgG)}$ (Figure 5A). The TAR DNA-binding protein 43 (TARDBP) demonstrated the highest fold enrichment at the region of SNP rs12920508 (≈ 26). Other tested proteins hnRNPM, FUBP3 and FUBP1 showed much lower fold enrichment 3.7, 0.47 and 2.1, respectively. Since fold enrichment (FE) over IgG data are not normally distributed, we represented TARDBP data as $log_2(FE)$ and we applied one-sample t-test to assess if enrichment with TARDBP antibody was significant (Figure 5B). Obtained P value did not reach significance with value of 0.0553.

Results from ChIP assay analyzed by ddPCR are presented in the supplementary results section but presented high variability between replicates and high background signal from the IgG control sample and thus did not allow firm conclusions to support the qPCR data or to quantify separately immunoprecipitation of allele-specific chromatin (supplementary Figure 13).

MC-EMSA for detection of potential TFs binding to variant rs12920508

We used multiplexed competitor EMSA (MC-EMSA) as a second method to identify proteins binding to DNA probe with SNP rs12920508. This high-throughput technique uses cocktails of small DNA probes bearing consensus binding sites of well characterized transcription factors, together with a conventional EMSA approach. We tested seven competitor cocktails. Each cocktail contained 10 DNA probes with consensus binding sites for TF, giving in total the

potential to alter binding of 70 TFs to the probe. The ratios of band intensity, representing DNA-protein complex, before and after addition of competitors were calculated for each cocktail and the set of TFs showing the highest competition was selected for further analysis. Next, DNA probes with consensus binding sites from the most competing cocktail were tested separately to detect which of the ten TF was responsible for observed competition.

The specific band shift (middle band) observed in EMSA for the probe bearing C allele of rs12920508 and nuclear extract derived from THP-1 monocytes was significantly decreased in three independent experiments by cocktail #3, with ratio of band intensity before and after addition of this mix of competitors equals 10.8 (Figure 6A). Subsequent experiments using the individual competitor probes from cocktail 3 identified interferon-stimulated response element (ISRE) as the element responsible for competition in binding (ratio 3.1 for middle band) (Figure 6B).

Region containing both SNPs rs1967309 and rs12920508 shows increased promoter activity with significantly higher activity for deleterious haplotype

Since both regions bearing genetic variants rs1967309 and rs12920508 showed increased promoter activity in reporter assays but resulting in opposite direction with regard to transcriptional activity of the protective allele compared to the deleterious allele, we decided to assess the activity of a longer DNA fragment containing both SNPs. An insert encompassing a deleterious haplotype background was cloned into promoter position of the luciferase reporter vector. The protective alleles of SNPs rs1967309 and rs12920508 were then generated by site-directed mutagenesis. The two vectors were transfected into THP-1 cells. The DNA region containing both rs1967309 and rs12920508 SNPs displayed increased promoter activity with significant difference between deleterious and protective alleles (n=10, P value=0.0004) (Figure 7). Higher activity was observed for deleterious alleles, what corresponds with the pattern obtained with the shorter region cassette containing SNP rs1967309 but with transcriptional activity that is intermediate that obtained with the small cassette bearing rs12920508 (Figure 4).

Discussion

Genome-wide association study of the dal-OUTCOMES trial revealed the association of a locus in the *ADCY9* gene and cardiovascular responses to dalcetrapib of patients with coronary artery disease [4]. Molecular mechanisms linking identified locus with dalcetrapib effects on cardiovascular outcomes are not fully understood. Identification of the causal variant from the locus and its function could facilitate understanding of the pathways involved and mechanisms. Here we report a functional study aimed to narrow down the list of putative causal SNPs and to investigate their possible transcriptional modulation effect in monocytic cell line, THP-1. Functional analysis included identification of allele-specific protein binding by EMSA, assessment of transcriptional activity of DNA regions bearing selected variants by reporter assays and identification of proteins binding to prioritized polymorphism using two different approaches.

Intronic SNPs predominantly function through regulation of target gene expression [13]. They often modulate expression of target gene(s) by allele-specific binding of regulatory proteins, such as transcription factors. Non-coding genetic variants may create new DNA binding site or disrupt existing one, what results in altered gene expression [20]. Our results from EMSA experiments indicate that 7 out of 12 selected SNPs demonstrate allele-specific binding with nuclear proteins derived from THP-1 monocytic cells. Among them, three variants: rs1967309, rs12920508 and rs2238449 show the strongest binding. These SNPs may change binding affinity of transcription factors and in turn regulate the expression of the target gene(s). EMSA also revealed that direct exposure of cells to dalcetrapib did not affect protein binding affinity to SNPs that previously showed allele-specific protein binding. However, we need to consider the fact that EMSA is an *in vitro* method that does not reflect conditions of living system, such as chromatin organization and long-range interactions. Chromatin structure is an important factor in DNA-protein binding, since open and close chromatin state affects proteins ability to bind DNA [21]. Also the assembly of DNA-protein complex may require special binding conditions that often cannot be reached in *in vitro* assay. Also, and maybe more importantly, the effect of dalcetrapib that is modulated by the genome may be indirect, for instance, it may be the result of increased HDL possibly combined to altered HDL composition. Therefore, we do not exclude the posibility that dalcetrapib may influence allele-specific protein binding in living system and that the studied SNPs or even other SNPs could also demonstrate allele-specific protein binding in different conditions of the assay.

We understand this limitation and therefore cannot eliminate completely the remaining SNPs as causal SNPs. These limitations aside, we focused our attention on three SNPs showing the strongest evidence to bind regulatory proteins in allele-dependent manner.

Polymorphism rs1967309 and other variants in high LD with it are eQTLs for *ADCY9* gene in monocytes and blood [17, 18]. This suggest that at least one of these SNPs are located in a DNA regulatory region modulating *ADCY9* transcription. Our data from reporter assay support this evidence showing that regions encompassing SNPs rs1967309 and rs12920508 displayed increased promoter activity in THP-1 cells. Longer DNA fragment containing both SNP rs1967309 and rs12920508 also demonstrated increased promoter activity with significantly higher activity for deleterious haplotype. Since a similar tendency was observed for the short cassette with SNP rs1967309, we suspect that this variant is responsible for majority of observed effect. However, the overall level of transcriptional activity was higher for the small cassette with rs12920508 and the longer fragment with both SNP than for the cassette with SNP rs1967309 alone, what suggest the influence of variant rs12920508 or its region on the effect of rs1967309. Moreover, increasing number of studies identifies cooperation of mutliple variants that drives the associations [22, 23]. Therefore, we should keep in mind the possibility of interaction between SNPs at *ADCY9* loci.

Commonly, the associated variants influence the expression of genes located nearby [20]. Thus we believe that DNA region containing SNPs rs1967309 and rs12920508 regulates the expression of *ADCY9* gene. Moreover, the observations that our group made related to *Adcy9* inactivation in the mouse, the observation that despite increased weight, the *Adcy9*-inactivated mouse are protected from atherosclerosis but only in absence of CETP, support *ADCY9* as a target gene. The exact mechanisms linking putative causal SNPs with *ADCY9* activity and cardiovascular responses to dalcetrapib are still unknown but we speculate that different levels of *ADCY9* gene expression dependent on genotype of patients in rs1967309 could result in modulated inflammatory pathways in monocytes and lead to changed cardiovascular responses of patients treated with dalcetrapib.

Despite this promissing evidence, reporter assay should be repeated in order to overcome its limitations. One of the limitations we faced is the presence of multiple variants present in some cassettes, including the one with SNP rs1967309. Hence, we cannot exclude the possibility that

different haplotypes of the same DNA region could yield different results. However, for cassette containing SNP rs1967309, two other variants in this DNA fragment (rs74702385 and rs3789035) did not demonstrate allele-specific protein binding. Also, site-directed mutagenesis of rs74702385 and rs3789035 did not result in significant alteration of transcriptional activity (Sanchez et al, data not shown). What suggest that observed effect of the cassette is drived mostly by SNP rs1967309.

Before we tested the transcriptional activity of the longer DNA fragment, that now suggest a major role of SNP rs1967309, variant rs12920508 seemed more interesting due to the higher transcriptional activity of the cassette bearing it. Therefore, we were interested to identify proteins binding in allele-specific manner to region of SNP rs12920508. We identified 11 proteins binding significantly more to allele C (deleterious) than allele G (protective) by DNA-affinity chromatography followed by mass spectrometry. Among the four most enriched proteins, we confirmed *in vivo* binding of TAR DNA-binding protein 43 (TARDBP). We also observed high binding of RNA polymerase II to the region of SNP rs12920508, what supports evidence that this region has indeed transcriptional activity or that it may be a region of RNA polymerase II pausing [24]. TARDBP is a DNA and RNA binding protein involved in multiple processes, such as transcription, mRNA splicing and translation. TARDBP was reported to behave as a transcriptional repressor [25], what correlates with our observation that the protein had higher binding affinity to the allele C of variant rs12920508, which showed lower transcriptional activity. Therefore, allele-dependent binding of TARDBP could repress the transcription of target gene, possibly *ADCY9*.

Since DNA-protein binding is very sensitive to changes in binding conditions, we used another assay, called multiplexed competitor EMSA. This method pointed to interferon-stimulated response element (ISRE) as an element competing DNA-protein binding to SNP rs12910508. Taking into account wide involvement of interferon proteins in signaling pathways that trigger immune responses of immune cells, such as monocytes and macrophages [26], we found this lead very interesting and worth following. Unfortunatelly the identified competing consensus binding sequence (ISRE) is an element with ability to bind multiple proteins involved in modulation of IFN-regulated genes. Hence, further tests are required to identify binding protein.

In conclusion, presented evidence reinforces our hypothesis that tested SNPs lie in regulatory elements and function through modulation of gene expression. Additional work is still required to understand the link between causal SNP function, inflammation and cardiovascular response induced by dalcetrapib**.** Presented results allow prioritization of potential causal SNP(s) involved in response of patients to dalcetrapib and bring new evidence to help understanding the mechanisms involved. Identification of causal SNP(s) associated with response of patients to dalcetrapib treatment may allow proper selection of patients, using this SNP(s) as biological marker. Such accurate selection of patients would ensure safety and best response to the treatment and avoid unnecessary use of medicine in patients with unsuitable genetic profile. Also, understanding the mechanism of dalcetrapib action could allow further development of new drugs and personalized therapeutic approaches targeting atherosclerosis.

Acknowledgments

We would like to express our special thanks to Sylvie Levesque from Montreal Health Innovations Coordinating Center for her help with multilevel model statistical analysis of dual luciferase reporter assays.

1. Organization, W.H.
2. Woudberg, N.J., et al., *Pharmacological Intervention to Modulate HDL: What Do We Target?* Front Pharmacol, 2017. **8**: p. 989.
3. Rhainds, D., et al., *An update on the clinical development of dalcetrapib (RO4607381), a cholesteryl ester transfer protein modulator that increases HDL cholesterol levels.* Future Cardiol, 2012. **8**(4): p. 513-31.
4. Tardif, J.C., et al., *Pharmacogenomic determinants of the cardiovascular effects of dalcetrapib.* Circ Cardiovasc Genet, 2015. **8**(2): p. 372-82.
5. Tardif, J.C., et al., *Genotype-Dependent Effects of Dalcetrapib on Cholesterol Efflux and Inflammation: Concordance With Clinical Outcomes.* Circ Cardiovasc Genet, 2016. **9**(4): p. 340-8.
6. Hacker, B.M., et al., *Cloning, chromosomal mapping, and regulatory properties of the human type 9 adenylyl cyclase (ADCY9).* Genomics, 1998. **50**(1): p. 97-104.
7. Alper, S., et al., *Identification of innate immunity genes and pathways using a comparative genomics approach.* Proceedings of the National Academy of Sciences, 2008. **105**(19): p. 7016-7021.
8. Huang, B., et al., *miR-142-3p restricts cAMP production in CD4+CD25- T cells and CD4+CD25+ TREG cells by targeting AC9 mRNA.* EMBO Rep, 2009. **10**(2): p. 180-5.
9. Liu, L., et al., *mTORC2 regulates neutrophil chemotaxis in a cAMP- and RhoA-dependent fashion.* Dev Cell, 2010. **19**(6): p. 845-57.
10. Liu, L., D. Gritz, and C.A. Parent, *PKCβII acts downstream of chemoattractant receptors and mTORC2 to regulate cAMP production and myosin II activity in neutrophils.* Molecular biology of the cell, 2014. **25**(9): p. 1446-1457.
11. Petter K. Risøe, A.R., Joanna Ågren, Ingrid B.M. Kolseth, Signe Flood Kjeldsen, and J.V. Guro Valen, Maria K. Dahle, *Higher TNFa responses in young males compared to females are associated with attenuation of monocyte adenylyl cyclase expression.* Hum Immunol, 2015. **76:427–43**.
12. Rautureau, Y., et al., *ADCY9 (Adenylate Cyclase Type 9) Inactivation Protects From Atherosclerosis Only in the Absence of CETP (Cholesteryl Ester Transfer Protein).* Circulation, 2018. **138**(16): p. 1677-1692.
13. Edwards, S.L., et al., *Beyond GWASs: illuminating the dark road from association to function.* Am J Hum Genet, 2013. **93**(5): p. 779-97.
14. Gallagher, M.D. and A.S. Chen-Plotkin, *The Post-GWAS Era: From Association to Function.* Am J Hum Genet, 2018. **102**(5): p. 717-730.
15. Deplancke, B., D. Alpern, and V. Gardeux, *The Genetics of Transcription Factor DNA Binding Variation.* Cell, 2016. **166**(3): p. 538-554.
16. Westra, H.J. and L. Franke, *From genome to function by studying eQTLs.* Biochim Biophys Acta, 2014. **1842**(10): p. 1896-1902.
17. Raj, T., et al., *Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes.* Science (New York, N.Y.), 2014. **344**(6183): p. 519-523.
18. Westra, H.J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations.* Nat Genet, 2013. **45**(10): p. 1238-1243.
19. Smith, A.J. and S.E. Humphries, *Characterization of DNA-binding proteins using multiplexed competitor EMSA.* J Mol Biol, 2009. **385**(3): p. 714-7.
20. Cavalli, M., et al., *Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals.* Genomics, 2016. **107**(6): p. 248-254.
21. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation.* Current Opinion in Genetics & Development, 2011. **21**(2): p. 175-186.

22. Cannon, M.E. and K.L. Mohlke, *Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci.* Am J Hum Genet, 2018. **103**(5): p. 637-653.
23. Hormozdiari, F., et al., *Identifying causal variants at loci with multiple signals of association.* Genetics, 2014. **198**(2): p. 497-508.
24. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease.* Cell, 2013. **152**(6): p. 1237-51.
25. Ou, S.H., et al., *Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs.* J Virol, 1995. **69**(6): p. 3584-96.
26. Delneste, Y., et al., *Interferon-gamma switches monocyte differentiation from dendritic cells to macrophages.* Blood, 2003. **101**(1): p. 143-50.

Article Table I

Table II : Polymorphisms selected for functional analysis. P values obtained from Genome Wide Association in dal-OUTCOMES. R^2 values represent linkage disequilibrium in dal-OUTCOMES study. MAF represents minor allele frequency from population of dal-plaque 2 study. Adapted from (Tardif et al., 2015) [4].

SNP	P_{value}	Linkage disequilibrium with rs1967309 (R^2)	Deleterious allele (major)	Protective allele (minor)	MAF (minor allele frequency)
rs1967309	P=2.4×10^{-8}	1	G	A	0.411
rs2531971	P=7.74×10^{-8}	0.84	T	G	0.434
rs2238448	P= 8.88×10^{-8}	0.81	C	T	0.459
rs11647778	P= 1.72×10^{-7}	0.77	G	C	0.459
rs1259911	P= 1.72×10^{-7}	0.86	T	T	0.444
rs12595857	P= 2.02×10^{-7}	0.86	A	G	0.447
rs12920508	P= 3.18×10^{-7}	0.98	C	G	0.420
rs11647828	P= 1.32×10^{-6}	0.69	T	C	0.438
rs12935810	P= 1.03×10^{-5}	0.48	A	G	0.430
rs74702385	P= 7.31×10^{-5}	0.2	A	A	0.125
rs3789035	P= 9.85×10^{-7} *	0.485	T	C	0.276***
rs2238449	P= 5.70×10^{-9} **	0.482	G	C	0.416***

* in high-density imputation

** eQTL in peripheral blood $CD14^{+}CD16^{-}$ monocytes

*** MAF retrieved from NCBI public database, 1000 Genomes phase 3, European population

Article Table II

Table III : Sequences of 5'-biotinylated oligonucleotides used in EMSA. Complementary 5'-biotinylated oligonucleotides were also synthesized and annealed to prepare double-stranded DNA probes.

	SNP	Deleterious allele	Protective allele
1	rs1967309	ATTTCTTTCAACCCTCAGCCCAGATCCTAAC	[T]
2	rs2531971	ACCCCCATAGGCTGGTGGTGAGCAGGGGGCA	[G]
3	rs2238448	TTCTTTCTTTCTTTTCTGAGATGGGGGTCTC	[T]
4	rs11647778	GGTGCTTTCTCAGAGCAGACTGAGGTTTGGG	[G]
5	rs1259911	CACATGGACCCTGGGTTCCAAGTTCATTAGA	[G]
6	rs12595857	AACCTCAACAACAGCAATGTCTTTTATCAGC	[G]
7	rs12920508	GAAAATGTAAAATTACGTTGTGGTGATGGTT	[G]
8	rs11647828	AAGGTGGCATCTGCCTGGTTTGTCCACTGTG	[C]
9	rs12935810	CGACTGTATGATCTCATCCTTTGCAGCCACA	[G]
10	rs74702385	CCCACCAGGTGCTATCGCTGTCATTCATTTG	[T]
11	rs3789035	CCTGCCTCAGCCTCCTGAGCAGCTGGGACTA	[C]
12	rs2238449	GACGCAGGAGTGAAGGCATCTTTATACTAAT	[C]

Article Table III

Table IV : List of primers used for PCR amplification of cassettes**.**

Cassette	***Primers***
Cass 1 *forward*	5′-GGGCTTTGTCAGTCACTCTT-3′
reverse	5′-GCTCCTGATTTGTTTCTGACTTG-3′
Cass 3 *forward*	5′-TGATCTCGGCTATCTGCAAC-3′
reverse	5′-AGTACAGTGGCTCCCTCA-3′
Cass 4 *forward*	5′-ACCTCTGCGGCATTCTTT-3′
reverse	5′-CTAACTCTATGTCCCTGGCATT-3′
Cass 6 *forward*	5′-TGAAGGCGGGTATAAGCC-3′
reverse	5′-GGTCATAGCTGAGAATCA-3′
Cass 9 *forward*	5′-CAAGCAGTCCTCCTGCCTC-3′
reverse	5′-TAACAGCTCATCCGTAGGTTTCT-3′
Cass 10 *forward*	5′-GATGCCTCTGTGAAATGTTTC-3′
reverse	5′-ATGACTATTCCTGCTCACCATG-3′

Article table IV

Table V : Location and size of cassettes used for luciferase reporter assay.

Cassette	Size	Location (hg19)	SNPs present
1	760 bp	chr16: 4062068-4062827	rs12595857, rs12599911
3	932 bp	chr16: 4064993-4065924	rs1967309, rs74702385, rs3789035
4	1006 bp	chr16: 4050416-4051421	rs2531971, rs183973203, rs3222326, rs11647778, rs2238446
6	668 bp	ch16: 4059187-4059854	rs22238448, rs2238449, rs73490533, rs113201607
9	562 bp	ch16: 4066523-4067084	rs12920508
10	1104 bp	ch16: 3999038-4000141	rs12935810, rs12917711, rs10852639

Article table V

Table VI : Transcription factors covered in seven competitor probe cocktails used in multiplexed competitor EMSA [19].

1	2	3	4	5	6	7
AP1	E2F1	HIF1a	NFATc	RAR	Stat4	VDR
AP2	Egr	ISRE	NFkB	RXR	Stat5	YY1
AR	ER	HNF-4	NR5A2	SIE	Stat56	ZEB
Brn3	Ets	IRF1	Oct-01	Smad	Tbet	HNF-1
CBP	Ets1	MEF1	p53	Smad34	TFE3	ARP1
CDP	FAST1	MEF2	PAX5	Smuc	TFEB	NFY
CEBP	GAS	MIBP1	Pbx1	Sp1	TFIID	HNF-3
cMyb	GATA	MycMax	Pit1	SRE	TGIF	BARP
CREB	G61	NF1	PPAR	Stat1	TR	SREBP1
CTCF	GR	NFE2	PR	Stat3	USF1	HSF1

Article Table VI

Table VII : Antibodies used in chromatin immunoprecipitation assays.

Name	Reference no	Company	Type of antibody
RNA polymerase II	39097	Active Motif	mouse monoclonal IgG1
IgG	ab18413	Abcam	mouse control IgG2a
hnRNP M1-M4	NB200-314SS	Novus Biologicals	mouse monoclonal IgG1
TDP-43	12892-1-AP	Proteintech	rabbit polyclonal
FUBP1	GTX104579	GeneTex	rabbit polyclonal
FUBP3	10623-1-AP	Proteintech	rabbit polyclonal

Article Table VII

Table VIII : Primers used for amplification of DNA regions containing SNP rs12920508, enriched by chromatin immunoprecipitation.

rs12920508 forward	5′-CTAACGGATGCAGGGTTACTT-3′
rs12920508 reverse	5′-TCACCCTCTCAAGTGTATGATTT-3′

Article table VIII

Table IX : Summary of EMSA results for 12 tested SNPs.

SNP	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Protein binding	+++	++	++	+	+	+	+++	+	++	+	+	+++
Ratio between alleles	+	+	+	+	-	+	+	-	+	-	+	+
Competition assay	+	+	+	+	+	+	+	-	+	+	-	+

EMSA assay was repeated 3 times for each SNP. For all 12 SNPs, band shift was observed, suggesting possible protein binding to all of these regions. SNPs are presented as a number from 1 to 12 labelled accordingly: 1-rs1967309, 2-rs2531971, 3-rs2238448, 4-rs11647778, 5-rs1259911, 6-rs12595857, 7-rs12920508, 8-rs11647828, 9-rs12935810, 10-rs74702385, 11-rs3789035, 12-rs2238449. For protein binding, one plus "+" indicates very weak binding, three pluses "+++" strong binding. Nine variants showed allele-specific protein binding (indicated by a "+"), which was assessed by the ratio between alleles ≥ 2. The specificity of the binding was determined by competition assays. For binding to be specific as indicated by "+" sign, ratio between band quantification of protein and probe complex, before and after addition of specific competitor ≥ 2 for stronger binding allele.

Article Table IX

Table X : Eleven proteins identified by mass spectrometry binding significantly more to the C allele of SNP rs12920508.

Protein ID	*C allele*	*G allele*	*Fold change by sample allele C/G*	*Allele C vs Allele G (pValue)*	*Allele C vs Control 1 (pValue)*	*Allele C vs Control 2 (pValue)*
HNRNPM	98	11	8.9	<0.00010	<0.00010	<0.00010
TARDBP	58	4	15	<0.00010	<0.00010	-
FUBP3	35	-	35	<0.0001	<0.0001	-
FUBP1	30	9	3.3	<0.00017	<0.015	<0.00025
KCNAB2	25	2	13	<0.0001	<0.0006	-
PDCD11	24	6	4	<0.00028	<0.0001	-
MTREX	23	-	23	<0.0001	<0.0012	-
EXOSC10	14	3	4.7	<0.0036	<0.034	-
RPA1	13	4	3.2	<0.015	<0.0026	-
ZCCHC8	11	-	-	<0.009	<0.0065	-
DAZAP1	8	-	8	<0.0027	<0.026	-

First column represents identified protein ID, second and third columns number of protein peptides detected by mass spectrometry. Fourth column represents fold change by sample allele C/G. P values calculated by Fisher's exact test representing enrichment of sample with probe C over probe G, Control 1 and Control 2. Control 1 represents reaction with scrambled probe, whereas control 2 represents reaction without probe.

Article Figure 1

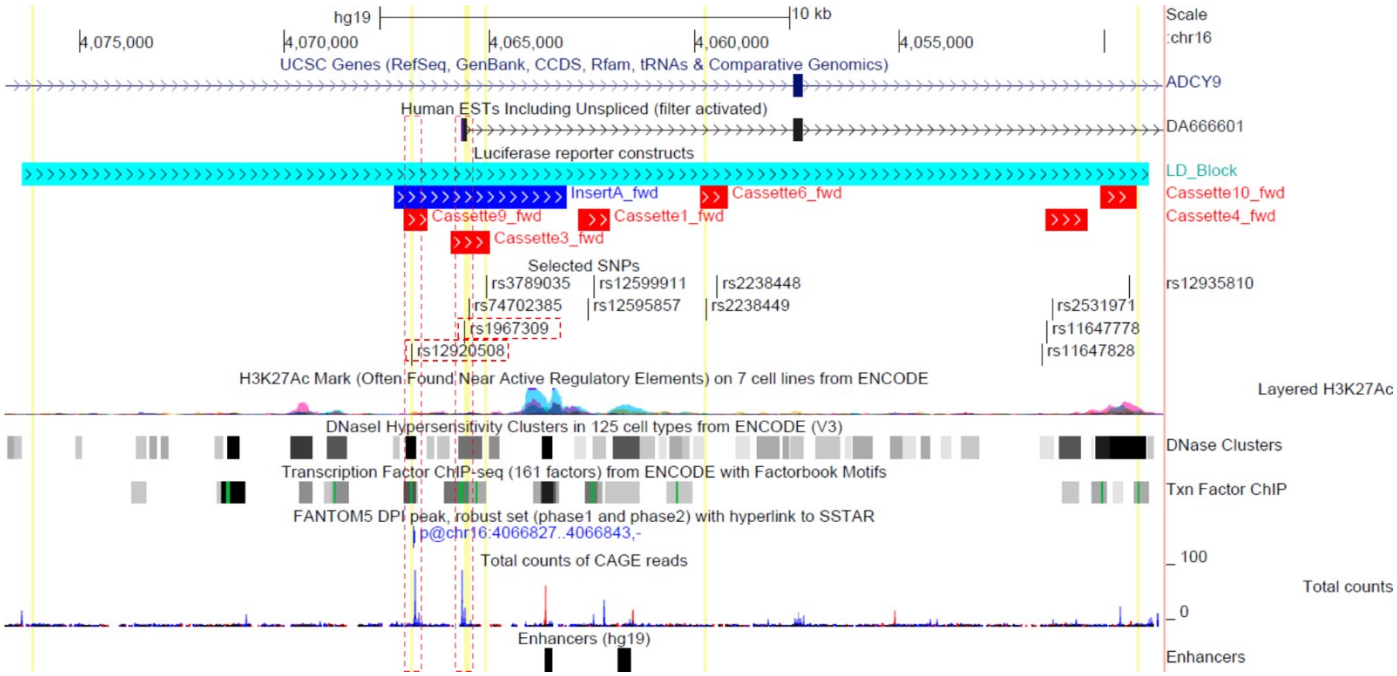

Figure 8 : SNPs selected for functional analysis with supplementary ENCODE and FANTOM5 data.

Linkage disequilibrium (LD) block is presented in cyan. Transcription start sites (TSS) overlapping SNPs rs1967309 and rs12920508 are indicated by increased counts of CAGE reads in FANTOM5 database.

Article figure 2

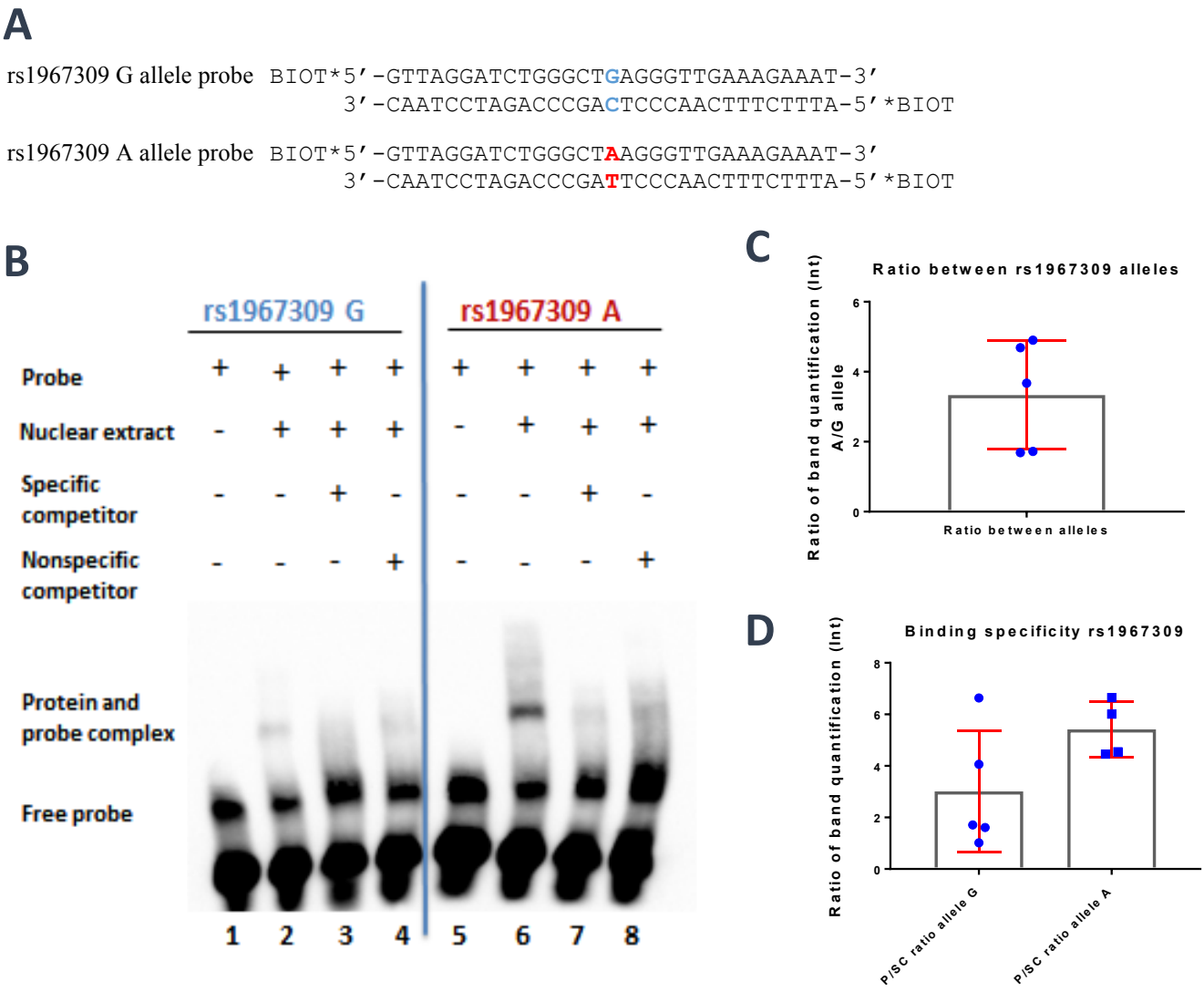

Figure 9 : SNP rs1967309 shows allele-specific protein binding.

Typical analysis of nuclear protein binding to region of SNP, based on variant rs1967309. (A) Sequences of the rs1967309 double-stranded, biotinylated probes with alleles of SNP marked in color. (B) EMSA for oligonucleotides containing SNP rs1967309 with the G (deleterious allele) and A (protective allele) assayed using THP-1 nuclear extracts. (C) Analysis of allele-dependent protein binding. Bands were quantified and ratio between alleles of SNP was calculated. Error bars denote standard deviation from at least 3 independent experiments. (D) Analysis of binding specificity. Ratio between band quantification of protein and probe complex, before and after addition of specific competitor. Error bars denote standard deviation from at least 3 independent experiments. P/SC – probe/ specific competitor.

Article Figure 3

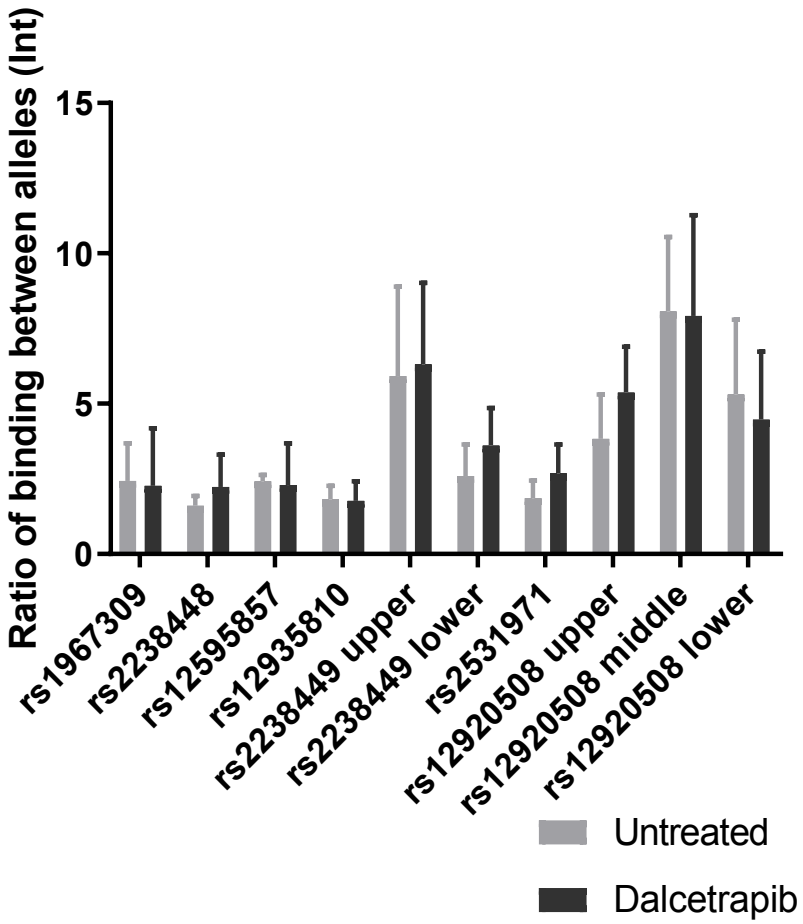

Figure 10 : Exposure of cells to dalcetrapib did not influence protein binding.

Summary of EMSA results for seven selected SNPs, assayed using nuclear proteins from THP-1 monocytes exposed or not to dalcetrapib. Ratio of binding intensity between alleles of SNPs before and after 24 h exposure to dalcetrapib (presented as Untreated and Dalcetrapib, respectively). For SNPs rs2238449 and rs12920508, multiple bands representing protein-probe complexes were observed, which are described on the figure as “upper”, “middle” and “lower”. Error bars denote standard deviation from at least 3 independent experiments.

Article Figure 4

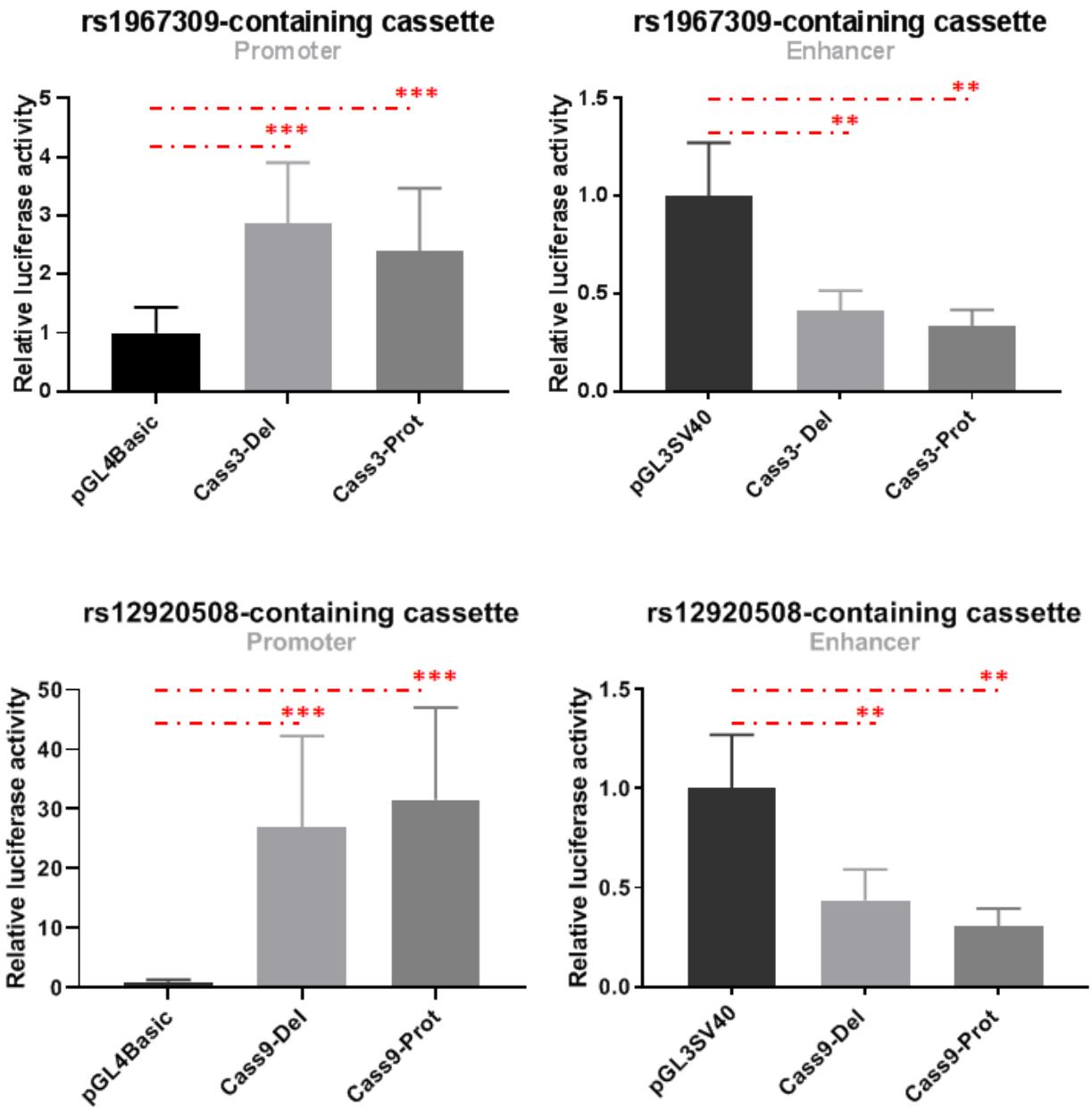

Figure 11 : Luciferase reporter assays for SNPs rs1967309 and rs12920508 after transfection of THP-1 cell lines.

Average relative luciferase activity for cassettes containing selected genetic variants bearing deleterious (Del) and protective (Prot) alleles. First bar represents control vector, pGL4Basic control vector without promoter for assessment of promoter activity or pGL3SV40 control vector for assessment of enhancer activity containing SV40 promoter. Error bars denote standard

deviation from 4 independent experiments for plasmids inserted into enhancer position and from 10 independent experiments for plasmids inserted into promoter position. P values were calculated by multilevel model (*≤ 0.05; **≤ 0.01; ***≤ 0.001).

Article Figure 5

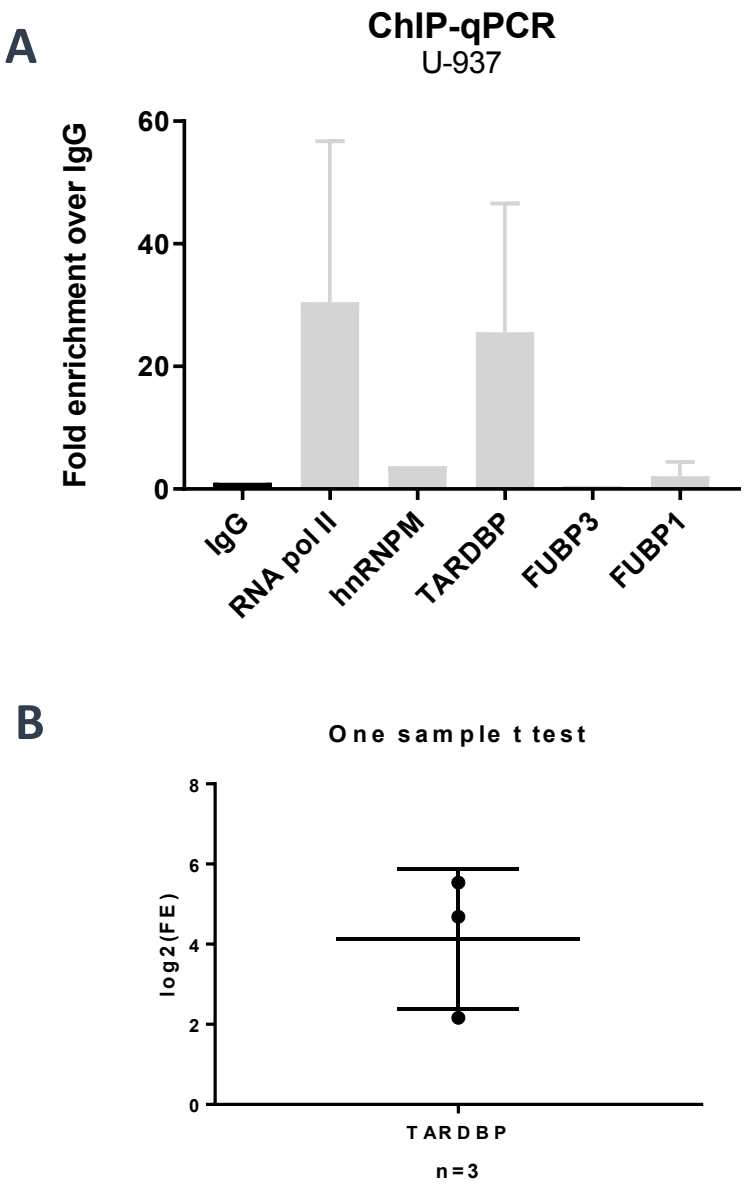

Figure 12 : TARDBP binds to the region of SNP rs12920508 *in vivo*.

Summary of the results from chromatin immunoprecipitation. (A) qPCR analysis results representing average fold enrichment over IgG for region of SNP rs12920508. Error bars denote standard deviation from replicas of independent experiments (n=3 for IgG, RNA pol II, TARDBP; n=2 for FUBP1 and n=1 for hnRNPM, FUBP3). (B) Statistical analysis of enrichment with TARDBP antibody. Data of fold enrichment over IgG (FE) were converted into $log_2(FE)$ and analyzed by one sample t-test. Obtained P value=0.0553.

Article Figure 6

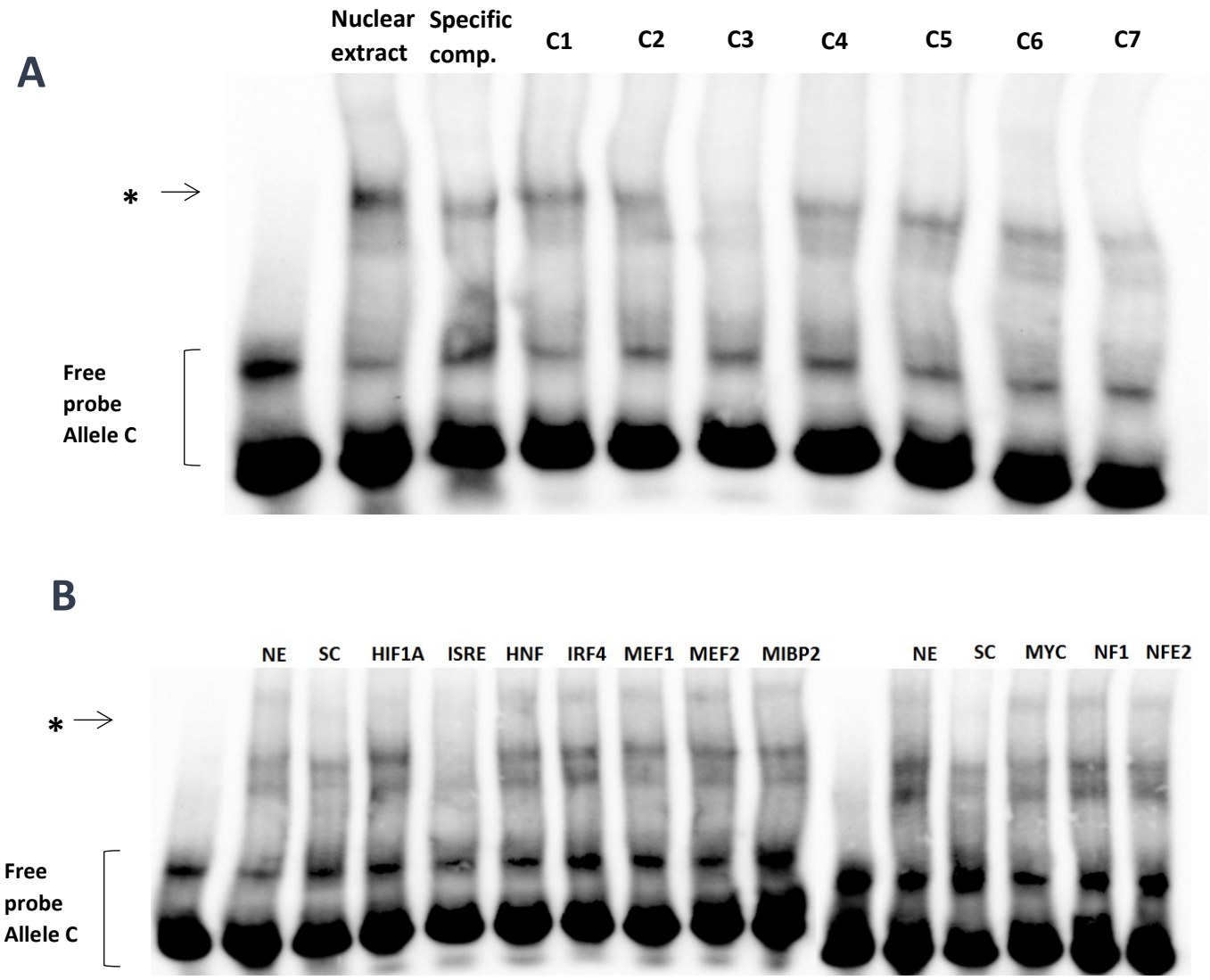

Figure 13 : Multiplexed competitor EMSA with cocktails against the probe bearing C allele of variant rs12920508 assayed using nuclear extract derived from THP-1 monocytes.

(A) MC-EMSA with 7 cocktails of unlabeled DNA competitors. First lane represents free probe, second lane (nuclear extracts) the band shift representing DNA-protein complex created after addition of nuclear extract, third lane (specific competitor) reaction with addition of specific competitor, and subsequently reactions with addition of seven competitor cocktails. * shows competed specific band. (B) EMSA with individual competitors from cocktail 3.

Article Figure 7

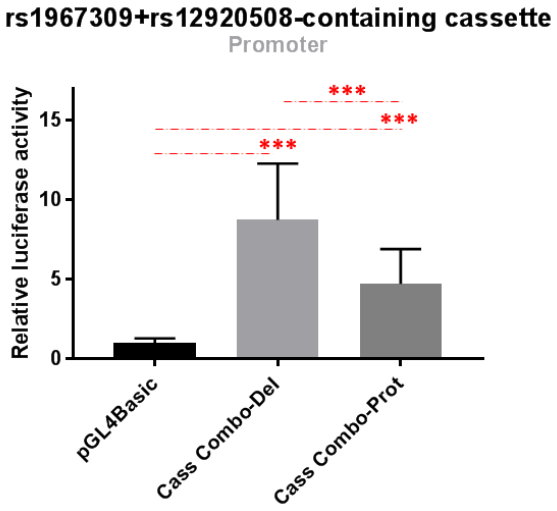

Figure 14 : Luciferase reporter assay for DNA region containing SNPs rs1967309 and rs12920508 after transfection of THP-1 cell lines.

Average relative luciferase activity for cassette containing polymorphisms rs1967309 and rs12920508, showing deleterious (Del) and protective (Prot) alleles. First bar represents control vector pGL4Basic, lacking promoter. Error bars denote standard deviation from 10 independent experiments. P values were calculated by multilevel model (*≤ 0.05; **≤ 0.01; ***≤ 0.001).

4 SUPPLEMENTARY METHODS

Digital droplet PCR (ddPCR) analysis of DNA regions enriched in ChIP

We used digital droplet PCR (ddPCR) aiming to assess allele-specific binding of TARDBP *in vivo*. In the ddPCR method using Bio-Rad (QX200 system), PCR solution is separated into 20000 partitions by a water oil emulsion technique, which allows an absolute counts of target DNA copies. Moreover, we combined this method with a TaqMan assay using activatable probes enabling the differentiation between alleles of SNP. The distinction between alleles of SNP is possible thanks to specific quantification by use of probes with fluorophore attached at the 5'-end and quencher at the 3'-end. We used two probes, one recognizing G allele of SNP rs12920508 with HEX fluorophore, and second one recognizing C allele with FAM fluorophore. Probes were designed and synthesized by Integrated DNA Technologies. In each probe, four different nucleic acids were "locked" by inserting a new covalent bound in the ribose of those nucleic acids to increase the affinity of the probes. This creates a significant difference in melting temperature (Tm) between the probes binding their correct target and one with a single mismatch. Consequently, using the proper annealing temperature (60℃), only a probe specific to a SNP can bind to the region with that allele of the SNP. The specific probe bound to nucleic acid sequence does not emit fluorescence signals as long that the quencher molecule is proximal to the fluorophore and inhibits its fluorescence. When the Taq polymerase reaches the probe during the extension, the polymerase cleaves the hybridized probe and causes the separation of the quencher and of the fluorophore, increasing the distance between both molecules. Lack of quenching effect results in fluorophore emitting fluorescence, which can be detected by the PCR thermal cycler. Detected fluorescence is directly proportional to the released fluorophore and also DNA template amount present in assessed reaction. Depending on allele of SNP in the sequence, different probe will bind, and different fluorophore will be released, allowing separate quantification of DNA regions with either G or C allele.

Impact of interferons treatment on transcriptional activity of region containing SNP rs12920508

To assess the influence of exposure of cells to interferons on promoter activity, IFNγ, IFNα-2a and IFNβ were added separately to the cells in dual luciferase reporter assay. Luciferase reporter assays were performed according to earlier described protocol and IFNs in concentration 20 ng/ml were added 24, 6 and 2 hours before luciferase activity read-out.

5 COMPLEMENTARY RESULTS

This chapter represents preliminary data which follow results presented in the article. Here we present data on allele-dependent protein binding of TARDBP analyzed by droplet digital PCR and follow up on interferon-stimulated response element identified by multiplexed competitor EMSA.

Probable allele-dependent binding of TAR-DNA binding protein 43 to region of SNP rs12920508

To detect allele-specific binding of proteins identified by mass spectrometry to the region of SNP rs12920508 in the natural chromatin context, DNA fragments enriched by chromatin immunoprecipitation were analyzed by digital droplet PCR. This assay could allow to distinguish between alleles of variant and presents results in the form of absolute copies detected per µl (Figure 15).

TARDBP was enriched the most among all tested proteins, with slightly higher absolute copies number for C allele (30.9 average absolute copies/µl for C allele and 23.6 for G allele). However, high variability was observed in between 3 independent repeats of the experiment. RNA polymerase II displayed relatively low copy number with no significant difference in between alleles of genetic variant (6.2 average absolute copies/µl for C allele and 7 for G allele).

IFN-gamma-activated site (GAS) also competes protein binding to SNP rs12920508

Sequence of interferon-stimulated response element (ISRE), which competed protein binding to SNP rs12920508 turned out to be the same as sequence of IFN-gamma-activated site (GAS) from cocktail 2. Since cocktail 2 did not result in visible competition of protein binding, we decided to confirm if this sequence decrease specific band shift observed in EMSA for the probe bearing C allele of rs12920508 and nuclear extract derived from THP-1 monocytes. When ran separately, GAS competitor probe competed protein binding, with ratio of band intensity before and after addition of competitor probe equals 3.90 (Figure 16).

IFNα-2a tend to decrease promoter activity of protective allele of SNP rs12920508

Since interferon-stimulated response element competed DNA-specific binding to variant rs12920508, we decided to check if treatment of cells with IFNs could modulate promoter

activity of mentioned SNP. We repeated luciferase reporter assay for constructs containing variant rs12920508 with treatment of cells with IFNγ, IFNα-2a and IFNβ for 2, 6 and 24 hours. Next, luciferase and renilla activities were measured and normalized luciferase activity was calculated.

Treatment of cells with IFNα-2a for 24 hours resulted in decrease in promoter activity of protective allele (Figure 17). Treatments with IFNγ and IFNα-2a for 6 and 2 hours, as well as treatment with IFNβ did not influence promoter activity of DNA region containing SNP rs12920508.

Additional results

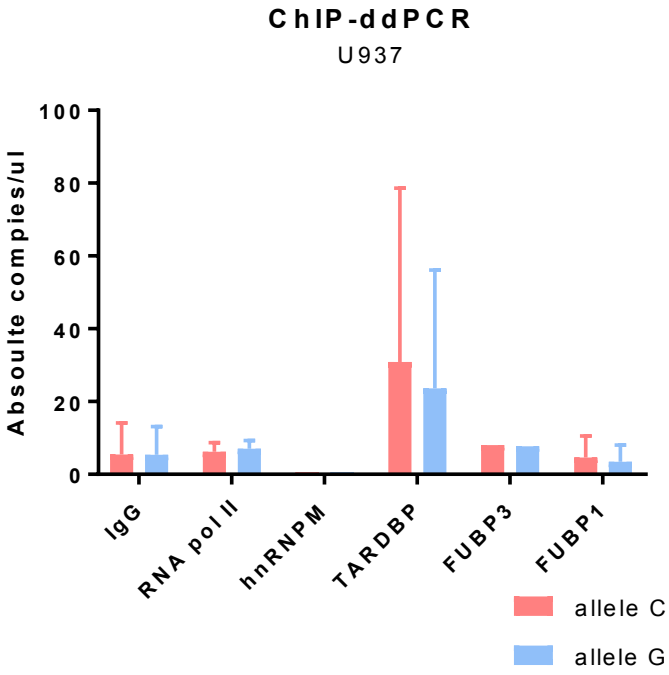

Figure 15 : Chromatin immunoprecipitation results analyzed by droplet digital PCR.

Results represent absolute copies detected per µl for variant rs12920508, separately for deleterious allele C shown in red and protective allele G in blue. Error bars denote standard deviation from replicates of independent experiment (n=3 for IgG, RNA pol II, TARDBP; n=2 for FUBP1 and n=1 for hnRNPM, FUBP3).

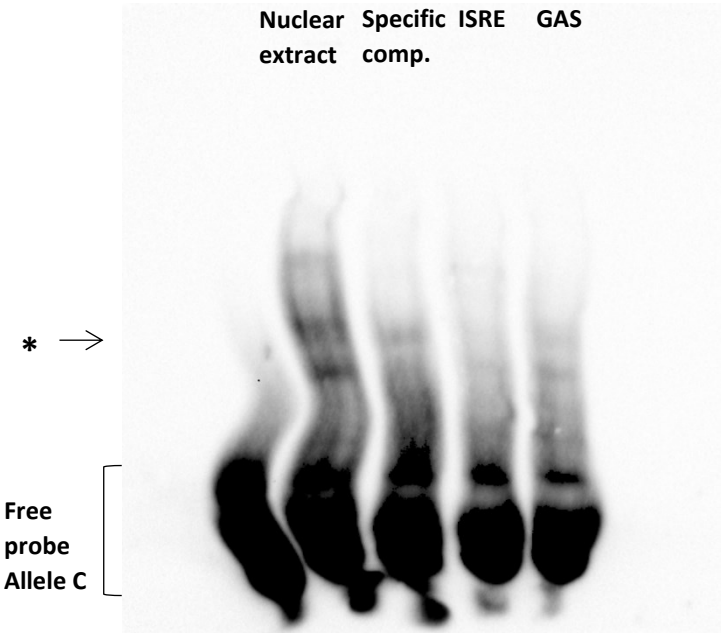

Figure 16 : EMSA with ISRE and GAS competitor probes.

Competitor EMSA with individual competitors for interferon stimulated response element (ISRE) and IFN-gamma-activated site (GAS) against the probe bearing C allele of variant rs12920508 assayed using nuclear extract derived from THP-1 monocytes.

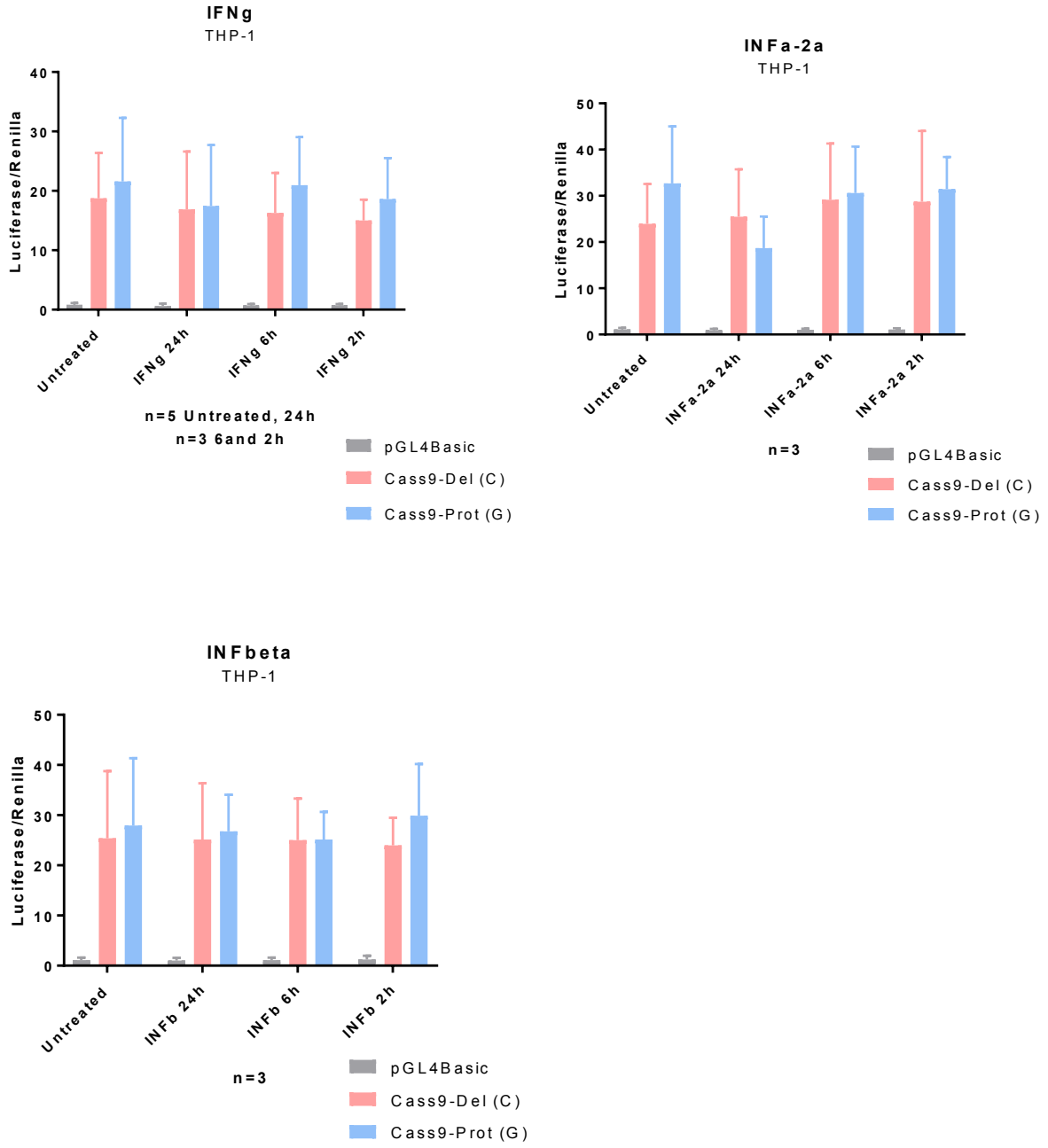

Figure 17 : Dual luciferase reporter assay with IFN-γ, INFα-2a and INF-β treatment in THP-1 monocytic cell line.

Average luciferase/renilla ratio for alleles of SNP rs12920508 with IFNγ, INFα-2a and INF-β treatments for 24, 6 and 2 hours. Grey bar represents pGL4Basic control plasmid, red bar cassette with deleterious alleles of rs12920508 and blue bar cassette with protective allele of rs12920508. Error bars denote standard deviation from at least 3 independent experiments.

6 DISCUSSION

In the last few years genome-wide association studies identified numerous DNA loci that influence the risk for complex diseases and response of patients to various treatments. Yet most of these associations remain poorly understood. The main goal of GWAS functional follow up studies is to identify causal genetic variant within associated loci and its function [109]. Identification of causal SNP is challenging because very often a lead variant detected by GWAS is in linkage disequilibrium with many variants on the same haplotype. Hence, observed effect may be mediated by any of the SNPs in LD block [151]. GWAS of the dal-OUTCOMES study identified a region within the *ADCY9* gene that is associated with cardiovascular responses of patients induced by dalcetrapib [77]. Molecular pathways underlying this association remain unknown. Identification of variant causing the association and its function could contribute significantly to provide explanation of the mechanisms involved. The majority of intronic SNPs were found to function through modulation of genetic expression of one or more target genes [112]. In this study, we investigated putative causal SNP candidates and their possible effect on *ADCY9* gene expression in THP-1 monocytic cell line. Functional analysis included identification of allele-specific protein binding by electrophoretic mobility shift assay, assessment of transcriptional activity of region of selected variants and possible allelic modulation by reporter assays and identifiaction of proteins binding to prioritized polymorphism using two different approches.

Intronic SNPs often regulate expression of target genes through allele-specific protein binding. Genetic variant can disrupt or create new DNA binding motif of regulatory proteins, such as transcription factors and alter gene expression [151]. Bioinformatic analysis of selected polymorphisms by RegulomeDB and MatInspector revealed possibility of protein binding to their position. Therefore we assessed the ability of selected SNPs to bind TFs in allele-specific manner by EMSA. We demonstrated that 7 out of 12 selected SNPs showed allele-specific binding with nuclear proteins derived from THP-1 monocytic cells. The strongest binding was observed to variants rs1967309, rs12920508 and rs2238449, implicating that these SNPs may change binding affinity of regulatory proteins and in turn regulate the expression of target gene. Exposure of cells to dalcetrapib did not affect protein binding to regions of selected SNPs assesed by EMSA. Lack of observed effect may be explained by the fact that this technique evaluates only if treatment with dalcetrapib changes the amount of proteins in the nucleus, when it may be other mechanism

involved. Moreover dalcetrapib increases the amount of circulating HDL-C what may lead to different levels of cholesterol in the plasma membrane and influence other mechanisms not directly affecting protein binding to DNA region [99]. Additionally EMSA is an *in vitro* assay, so it does not reflect the same conditions as living system. *In vitro* assay does not take into account the structure and organization of chromatin, which is important since open and close state of chromatin greatly affects the ability of proteins to bind to DNA [125]. Also *in vitro* assay does not reflect DNA looping which may be crucial for binding of protein complexes. Nevertheless we demonstrated that seven SNPs from the LD block have ability to bind nuclear proteins in allele-specific manner, what may lead to further influence of target gene expression. Since DNA-protein binding is very sensitive for changes in binding conditions, we do not exclude possibility that other SNPs could also show allele-dependent protein binding with different conditions of the assay. Therefore, we do not consider the exclusion of remaining SNPs for any further study but rather a reason to focus on three SNPs showing the strongest allele-dependent protein binding.

Predictions of eQTLs, genomic locus influencing level of mRNA expression, are often indicators of SNP functionality. Few among selected SNPs are known to be eQTLs for *ADCY9* gene in monocytes and blood [143, 144]. To follow up on this evidence, we investigated the enhancer and promoter activities of regions containing SNPs showing allele-specific protein binding by reporter assays. SNPs rs1967309 and rs12920508 demonstrated increased promoter activity. Interestingly DNA region containing both mentioned SNPs together also showed increased promoter activity with significantly higher activity for deleterious haplotype. These results resemble more the pattern of the results obtained with SNP rs1967309, what suggests that SNP rs1967309 is responsible for the majority of observed effect. However, difference in activity between alleles and activity in general was much higher for longer DNA region than that of the smaller DNA region with SNP rs1967309 alone, what let us believe in possible interactions between both SNPs. With increasing number of GWAS follow up studies, it is starting to be more common that observed phenotype is caused by more than one genetic variant. Multiple functional SNPs can act cooperatively to regulate the same target gene or they may modulate the expression of different genes [152]. Since seven of SNPs in *ADCY9* LD block showed allele-specific protein binding and transcriptional activity of SNPs rs1967309 seems to be modulated by variant rs12920508, we should not exclude posibility that function of multiple SNPs in our loci of interest leads to observed phenotype.

The exact mechanism by which DNA regions containing SNPs rs1967309 and rs12920508 could induce *ADCY9* promoter activity is challenging, since these SNPs are located in the intronic region of *ADCY9* gene, and not in its known promoter region. However, some evidence from FANTOM5 database identifies some transcription start sites located at DNA regions overlapping SNPs rs1967309 and rs12920508 which supports our data that these regions behave like promoters. Intron DNA sequences were already shown to have the potential to affect transcription initiation and to have important role in determining the site of transcript initiation [153]. An interesting example of promoter region located in intronic region is illustrated by a study which investigated a risk loci associated with piglet splay leg (PSL) syndrome [154]. This study identified a promoter region in intron 4 of homer scaffolding protein 1 (HOMER1) gene. By dual luciferase reporter assays, researchers detected allele-dependent promoter activity driven by SNP rs235197091. Therefore, they proposed that G allele of rs235197091 may create binding site for aryl hydrocarbon receptor nuclear translocator (ARNT) protein, which could regulate the intronic promoter activity and in turn affect the expression of HOMER1 gene. Another study qualified fragment of the first intron of the c-fos gene as genuine promoter, because this fragment induced luciferase expression in the absence of any promoter [155]. Authors confirmed promoter activity of intron 1 *in vivo* in embryonic and adult tissues, proposing novel, tissue-specific promoter function for gene c-fos. They suggested that an enhancer-like effect may result from the sum of canonical and intronic promoter activities. Another group of researchers showed that higher-order chromosomal organization regulates transcription and that promoter-promoter interactions may play important role in transcription regulation of housekeeping and tissue-specific genes [156]. Taken together, we suspect that intronic promoter containing SNPs rs1967309 and rs12920508 may interact with the canonical promoter of *ADCY9* and contribute to modulation of *ADCY9* transcription. Direct interaction of these two loci should be confirmed by chromosome conformation capture (3C) analysis. Another explanation could be that this region lie in an alternative promoter for initiation of transcription of an alternative RNA, possibly a long non-coding RNA, which could further influence the expression of *ADCY9* and/or other uncharacterized target gene(s).

Since common assumption consider that most associated variants lie in regulatory element that affects the activity of neighboring gene [151], we believe that DNA region containing rs1967309 and rs12920508 regulates the transcriptional activity of *ADCY9* gene. Our speculations are

further supported by the evidence that silencing of *ADCY9* in mices in the absence of CETP has atheroprotective effect [107]. In reporter assays, deleterious haplotype of DNA region containing two putative functional SNPs (rs1967309 and rs12920508) showed increased transcriptional activity. Further mechanisms underlying the association of *ADCY9* activity with cardiovascular responses to dalcetrapib are hard to predict with our present knowledge, but *ADCY9* was reported to mediate different functions of immune cells and to regulate various inflammatory responses [102-106]. Therefore, we believe it is possible that different levels of *ADCY9* gene expression dependent on genotype of patients in rs1967309 could modulate inflammatory pathways in monocytes and lead to changed cardiovascular responses of patients treated with dalcetrapib. One possible hypothesis that links *ADCY9* activity with responses to dalcetrapib assumes that *ADCY9* dependent cAMP production could modulate activity of protein kinase A that is responsible for phosporylation of ATP-binding cassette A (ABCA1), widely involved in cholesterol efflux [101, 157]. However, many other mechanisms could be responsible for this association. Finding causal SNP and its functional pathway will greatly help understanding of involved mechanisms.

Also we need to keep in mind that regulatory elements such as enhancers and silencers can regulate expression of genes over a considerable distance. Thus we cannot exclude possibility that our loci of interest regulates further located gene than *ADCY9*, or may modulate multiple target genes. Once causal variant is found, affected gene can be identified experimentally. The most common assays used for this purpose are: changing sequence of SNP using CRISPR, overexpressing or silencing TFs binding to causal SNP or by testing long-range chromatin interactions [151].

Despite solid evidence pinpointing variant rs1967309 to influence gene expression of target gene, we need to keep in mind limitations of our assay. Some cassettes, including the one with SNP rs1967309, contained more that one genetic variant. Since our transfection repeats did not reflect independent DNA preparations of each construct, from different donors, we cannot exclude that possible changes could appear if we used different haplotypes of the same DNA fragment. Also unknown subtle variables could affect the regulatory activity of DNA fragments. What reasures us in the functional role of SNP rs1967309 is the fact that cassette containing this SNP DNA includes two more polymorphisms (rs74702385 and rs3789035), which did not show allele-specific protein binding in EMSA assays. Therefore, we think that SNP rs1967309, which

demonstrated very strong allele-dependent binding to its region is more likely to cause the effect. Additionally, the construction of longer cassette containing both SNPs, was different. To create protective haplotype in DNA fragment with deleterious haplotype, the alleles of SNPs rs2967309 and 12920508 were changed by site-directed mutagenesis. Thus, in this case the effect could be associated only with these two changed alleles.

Before we obtained the results suggesting that longer DNA region containing both SNPs rs1967309 and rs12920508 reflects more the effect pattern of SNP rs1967309, we were more focused on SNP rs12920508. Hence the rest of our experiments and discussed results concerns this variant. In light of these more recent results it would be interesting to perform similar analysis to characterize protein binding to SNP rs1967309.

To understand the mechanism how rs12920508 could influence the transcripion of target gene, we decided to identify proteins binding to its region using our 31-bp DNA probe used previously in EMSA experiments. Mass spectrometry identified 11 proteins that were binding significantly stronger to deleterious allele (allele C). From the four most enriched proteins in the assay, we identified TAR DNA-binding protein 43 (TARDBP) to bind *in vivo* to region of SNP rs12920508. Interestingly, antibodies recognizing RNA polymerase II also resulted in high enrichment of regions containing SNP rs12920508, what seems to support evidence that region with this genetic variant has increased promoter activity. TARDBP demonstrates the ability to bind both DNA and RNA and to affect multiple processes, such as transcription, mRNA splicing and translation. Originally, TARDBP was found to repress the transcription of HIV-1 by binding to chromosomally integrated trans-activation response element (TAR) DNA [158]. Further, TARDBP was reported to be a splicing factor binding to intron/exon junction of cystic fibrosis transmembrane conductance regulator (CFTR) and apolipoprotein A-II (apoAII) genes, and to regulate alternative splicing of these genes [159, 160]. TARDBP was also shown to be a neuronal activity response factor in the dendrites of hippocampal neurons. What suggests its probable function in regulation of mRNA transport, stability and local translation in neuronal cells [161]. The function of TARDBP as a repressor of transcriptional activity correlates with observation that the protein had higher binding affinity to the allele C of variant rs12920508, which in reporter assays showed lower transcriptional activity. Thus, allele-specific binding of TARDBP could repress the transcription of target gene, possibly *ADCY9*. The results from reporter assay

for longer DNA fragment, encompassing both SNPs rs1967309 and rs12920508, showed lower transcriptional activity for protective haplotype. However, the protective allele of SNP rs1967309 which showed higher affinity to proteins at EMSA assays, also demostrated lower transcriptional activity in reporter assays. What could indicate that TARDBP may bind also to region of SNP rs1967309 or there may be a repressor protein complex binding to both SNPs. Detection of high levels of RNA polymerase II in this region may also suggest that allele-specific binding of proteins could induce RNA polymerase II pausing or release and in turn affect the level of transcriptional activity. For example, in *Drosophila* cells protein Argonaute 2 (Ago2) was found to localize to promoters with paused RNA polymerase II and modulate its activity. Studies in *Drosophila* also showed that probability of RNA polymerase II pausing depends on the nucleotide sequence of initialy transcribed region on a given gene [162, 163].

Unfortunatelly we failed to confirm allele-specific binding of TARDBP *in vivo* to SNP rs12920508. Analysis of DNA fragments enriched from chromatin immunoprecipitation assay by droplet digital PCR could have given the possibility to distinguish between alleles of this variant. However we faced difficulties in optimizing the assay to obtain enough input material to reach the power to compare the affinity of binding between alleles. Still, it is important to validate allele-specific binding of TARDBP to rs12920508 *in vivo*, hence it would be necessary to test alternative methods. One method that could answer our question is based on same protocol like chromatin immunoprecipitation followed by droplet digital analysis, but the input cells would be transfected with reporter vectors containing DNA regions of SNP of interest. This would increase the amount of DNA regions enriched in ChIP with antibody recognising TARDBP. Once reaching enough signal at ddPCR to compare the affinity of binding between alleles of variant, it would be interesting to analyze if RNA polymerase II (run as positive control) demonstrates different affinity to alleles of SNPs showing increased promoter activity (rs1967309 and rs12920508). However, the drawback of this technique is that assessed protein binding is not to the chromatin DNA, thus it does not takes into account epigenetic factors and chromatin conformation.

Succesfull identification of DNA-protein interactions depends on the binding conditions of the assay. To increase the power of detection we performed one more assay to identify proteins binding to region of SNP rs12920508. MC-EMSA identified interferon stimulated response

element (ISRE) as an element competing one of the DNA-protein complexes. Moreover the sequence of ISRE turned out to be the same like sequence of IFN-gamma-activated site (GAS), which also competed the same binding when run separately. Since interferons are widely implicated in signaling pathways which trigger immune responses of immune cells, such as monocytes and macrophages [164], we found this result very intriguing. Identification of the exact protein binding to this region is not evident, however, since both ISRE and GAS are elements with ability to bind multiple regulatory proteins. In response to the external stimuli (pathogens), cell release interferons, which transfer signal to modulate immune system. Type I interferons (α or β) induce expression of IFN-stimulated genes (ISGs). Their transcription is regulated by binding of transcription factors family called IFN regulatory factors (IRFs) to the ISRE located in promoter regions of these genes. There are nine cellular IRFs which bind directly to ISRE element [165]. Type II interferons (γ) induce transcription of variety of genes implicated in inflammatory responses by binding of gamma interferon activation factors (GAFs) to GAS elements. GAF complex is a dimer of transcription factor Stat. Family of Stat proteins contains at least six Stats able to recognise GAS element [166, 167]. Taking into account that protein binding to SNP rs12920508 could be either IRF or one of the Stat protein and that these TFs often bind in form of heterodimers, identification of exact protein still requires further work.

We tried to narrow down the list of possible candidate proteins that could be the protein competed by ISRE and GAS unlabeled probe, and to understand more the mechanisms involved. We presumed that if protein binding to region of SNP rs12920508 would belong to family of IRFs or GAF complex and would be responsible for promoter activity of this region, then treatment with IFN type I or type II should have an effect on observed promoter activity. Indeed, after 24 hours of treatment with IFNα-2a, we noticed a tendency of the protective allele to show decreased promoter activity. Interestingly, difference in activity between alleles of rs12920508 took the opposite direction. Referring to our assumption that protein binding to variant rs12920508 has a repressor activity (stronger binding for deleterious allele with lower promoter activity), we suspect that treatment with IFNα-2a promotes binding of protein to protective allele, which in turn decrease its transcriptional activity. Since the effect was observed with IFNα-2a which is a type I interferon, this indicates that the binding protein belongs to the family of IRFs. Some of the proteins from IRFs family were found to exert repressor activities. IRF-2 could be a promising candidate since it was reported to negatively regulate many IFN-responsive genes,

such as iNOS, MHC class I and 2'-5'-oligoadenylate synthetase [168]. Additionally IRF2 can create heterocomplex with IRF8/ICSBP which was also found to possess repressor function in immune cells [169, 170]. Binding of IRF-2 or other protein from IRF family exerting repressor activity could be tested by chromatin immunoprecipitation assay. However, we need to keep in mind that these are preliminary results that should be repeated more times and verified testing additional controls, for example reporter vectors with elements known to be regulated by IFNα. Additionally, mass spectrometry analysis of proteins binding to this region did not identify proteins from IRFs family.

7 CONCLUSION AND PERSPECTIVES

In this master thesis we present functional analysis of SNPs associated with cardiovascular responses induced by dalcetrapib. We present strong evidence about functionality of variant rs1967309 in THP-1 monocytic cell line with possibility of its interaction with variant rs12920508.

Throughout our experiments, we showed that seven SNPs located in intronic LD block of *ADCY9* gene have ability to bind nuclear proteins derived from THP-1 monocytes, in allele-specific manner. Among these SNPs, two genetic variants rs1967309 and rs12920508 demonstrated increased promoter activity. Using DNA-affinity chromatography followed by mass spectrometry, we further identified TAR-DNA binding protein as putative protein binding to region of rs12920508. Binding of TARDBP *in vivo* was confirmed by chromatin immunoprecipitation. Subsequent results showed that DNA region containing both rs1967309 and rs12920508 together had higher promoter activity for the deleterious haplotype, what corresponds with tendency observed for SNP rs1967309 alone. Therefore, we prioritize SNP rs1967309 as putative causal SNP, but we do not exclude the possibility that other SNPs in LD block, such as rs12920508, may cooperate with it. In addition, lower promoter activity for the protective haplotype supports our beliefs that this region modulates expression of *ADCY9* gene, since silencing of *Adcy9* gene in mouse was shown to exert atheroprotective properties.

Moving forward, it will be important to identify proteins binding in allele-specific manner to region of rs1967309 in order to understand molecular mechanism involved. Also, it would be interesting to perform similar analysis in different cell lines, since this study was focused exclusively on monocytic cells which may not be the sole target cell involved.

In conclusion, presented findings support the evidence that SNP(s) located within *ADCY9* intronic LD block exert functional effect, and may regulate expression of target gene, most likely *ADCY9* itself. Additionally, these results narrow down the list of putative causal SNP(s) associated with patient's response to dalcetrapib treatment. Still, further work is required to identify causal SNP(s) and to understand the link between its function, inflammation and cardiovascular response induced by dalcetrapib. Understanding the mechanism of dalcetrapib action will allow further development of new drugs and personalized therapeutic approaches targeting atherosclerosis.

REFERENCES

1. Gistera, A. and G.K. Hansson, *The immunology of atherosclerosis.* Nat Rev Nephrol, 2017. **13**(6): p. 368-380.
2. Lusis, A.J., *Atherosclerosis.* Nature, 2000. **407**(6801): p. 233-41.
3. Wu, M.-Y., et al., *New Insights into the Role of Inflammation in the Pathogenesis of Atherosclerosis.* International Journal of Molecular Sciences, 2017. **18**(10): p. 2034.
4. Rafieian-Kopaei, M., et al., *Atherosclerosis: process, indicators, risk factors and new hopes.* Int J Prev Med, 2014. **5**(8): p. 927-46.
5. Fruchart, J.-C., et al., *New Risk Factors for Atherosclerosis and Patient Risk Assessment.* Circulation, 2004. **109**(23 suppl 1): p. III-15-III-19.
6. Roberts, W.C., *Atherosclerotic risk factors — are there ten, or is there only one?* Atherosclerosis. **97**: p. S5-S9.
7. Tegos, T.J., et al., *The genesis of atherosclerosis and risk factors: a review.* Angiology, 2001. **52**(2): p. 89-98.
8. Khera, A.V. and S. Kathiresan, *Genetics of coronary artery disease: discovery, biology and clinical translation.* Nat Rev Genet, 2017. **18**(6): p. 331-344.
9. Abifadel, M., et al., *Mutations in PCSK9 cause autosomal dominant hypercholesterolemia.* Nat Genet, 2003. **34**(2): p. 154-6.
10. Soria, L.F., et al., *Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100.* Proc Natl Acad Sci U S A, 1989. **86**(2): p. 587-91.
11. Davignon, J. and P. Ganz, *Role of Endothelial Dysfunction in Atherosclerosis.* Circulation, 2004. **109**(23_suppl_1): p. III-27-III-32.
12. Deanfield, J.E., J.P. Halcox, and T.J. Rabelink, *Endothelial function and dysfunction: testing and clinical relevance.* Circulation, 2007. **115**(10): p. 1285-95.
13. Hunt, B.J. and K.M. Jurd, *Endothelial cell activation. A central pathophysiological process.* BMJ (Clinical research ed.), 1998. **316**(7141): p. 1328-1329.
14. Liao, J.K., *Linking endothelial dysfunction with endothelial cell activation.* The Journal of clinical investigation, 2013. **123**(2): p. 540-541.
15. Falk, E., *Pathogenesis of Atherosclerosis.* Journal of the American College of Cardiology, 2006. **47**(8, Supplement): p. C7-C12.
16. Weber, C. and H. Noels, *Atherosclerosis: current pathogenesis and therapeutic options.* Nat Med, 2011. **17**(11): p. 1410-22.
17. Gieroba, B., *Miażdżyca - czynniki ryzyka i patogeneza.* 2016: p. 35-45.
18. Singh, S.K., et al., *The connection between C-reactive protein and atherosclerosis.* Ann Med, 2008. **40**(2): p. 110-20.
19. Libby, P., *Inflammation in atherosclerosis.* Nature, 2002. **420**(6917): p. 868-74.
20. Bergheanu, S.C., M.C. Bodde, and J.W. Jukema, *Pathophysiology and treatment of atherosclerosis : Current view and future perspective on lipoprotein modification treatment.* Neth Heart J, 2017. **25**(4): p. 231-242.
21. S El Agamy, D. and E. Said, *Diagnosis and Treatment of Atherosclerosis: A Review Article*. Vol. 101. 2015.
22. Newman, C.B., et al., *Statin Safety and Associated Adverse Events: A Scientific Statement From the American Heart Association.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2019. **39**(2): p. e38-e81.
23. Maron, D.J., S. Fazio, and M.F. Linton, *Current perspectives on statins.* Circulation, 2000. **101**(2): p. 207-13.

24. Spence, J.D. and G.K. Dresser, *Overcoming Challenges With Statin Therapy.* J Am Heart Assoc, 2016. **5**(1).
25. Nussbaumer, B., et al., *Ezetimibe-Statin Combination Therapy.* Dtsch Arztebl Int, 2016. **113**(26): p. 445-53.
26. Chaudhary, R., et al., *PCSK9 inhibitors: A new era of lipid lowering therapy.* World J Cardiol, 2017. **9**(2): p. 76-91.
27. Brahmkshatriya, P.S., M.H. Jani, and M.T. Chhabria, *Recent developments in the treatment of atherosclerosis.* J Enzyme Inhib Med Chem, 2006. **21**(1): p. 1-15.
28. Steiner, G., *Fibrates and coronary risk reduction.* Atherosclerosis, 2005. **182**(2): p. 199-207.
29. *A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Report from the Committee of Principal Investigators.* Br Heart J, 1978. **40**(10): p. 1069-118.
30. Alexander, C., et al., *Third National Health and Nutrition Examination Survey (NHANES III)I; National Cholesterol Education Program (NCEP). NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older*. Vol. 52. 2003. 1210-4.
31. Tenkanen, L., M. Mänttäri, and V. Manninen, *Some Coronary Risk Factors Related to the Insulin Resistance Syndrome and Treatment With Gemfibrozil.* Circulation, 1995. **92**(7): p. 1779-1785.
32. Riaz, H., et al., *Effects of high-density lipoprotein targeting treatments on cardiovascular outcomes: A systematic review and meta-analysis.* European Journal of Preventive Cardiology, 2019. **26**(5): p. 533-543.
33. Staels, B., et al., *Mechanism of Action of Fibrates on Lipid and Lipoprotein Metabolism.* Circulation, 1998. **98**(19): p. 2088-2093.
34. Woudberg, N.J., et al., *Pharmacological Intervention to Modulate HDL: What Do We Target?* Front Pharmacol, 2017. **8**: p. 989.
35. Bruikman, C.S., et al., *New Drugs for Atherosclerosis.* Can J Cardiol, 2017. **33**(3): p. 350-357.
36. Barr, D.P., E.M. Russ, and H.A. Eder, *Protein-lipid relationships in human plasma. II. In atherosclerosis and related conditions.* Am J Med, 1951. **11**(4): p. 480-93.
37. Gordon, T., et al., *High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study.* Am J Med, 1977. **62**(5): p. 707-14.
38. Vergeer, M., et al., *The HDL hypothesis: does high-density lipoprotein protect from atherosclerosis?* Journal of lipid research, 2010. **51**(8): p. 2058-2073.
39. Gordon, D.J., et al., *High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies.* Circulation, 1989. **79**(1): p. 8-15.
40. Hansel, B., et al., *One third of the variability in HDL-cholesterol level in a large dyslipidaemic population is predicted by age, sex and triglyceridaemia: The Paris La Pitie Study.* Curr Med Res Opin, 2006. **22**(6): p. 1149-60.
41. Criqui, M.H., et al., *Cigarette smoking and plasma high-density lipoprotein cholesterol. The Lipid Research Clinics Program Prevalence Study.* Circulation, 1980. **62**(4 Pt 2): p. IV70-6.
42. Kodama, S., et al., *Effect of Aerobic Exercise Training on Serum Levels of High-Density Lipoprotein Cholesterol: A Meta-analysis.* Archives of Internal Medicine, 2007. **167**(10): p. 999-1008.
43. Busetto, L., et al., *Short-term effects of weight loss on the cardiovascular risk factors in morbidly obese patients.* Obes Res, 2004. **12**(8): p. 1256-63.
44. Rashid, S. and J. Genest, *Effect of obesity on high-density lipoprotein metabolism.* Obesity (Silver Spring), 2007. **15**(12): p. 2875-88.
45. van Leuven, S.I., et al., *Systemic inflammation as a risk factor for atherothrombosis.* Rheumatology (Oxford), 2008. **47**(1): p. 3-7.
46. Oldoni, F., R.J. Sinke, and J.A. Kuivenhoven, *Mendelian Disorders of High-Density Lipoprotein Metabolism.* Circulation Research, 2014. **114**(1): p. 124-142.

47. Rader, D.J., *Spotlight on HDL biology: new insights in metabolism, function, and translation.* Cardiovasc Res, 2014. **103**(3): p. 337-40.
48. Canner, P.L., et al., *Fifteen year mortality in Coronary Drug Project patients: long-term benefit with niacin.* J Am Coll Cardiol, 1986. **8**(6): p. 1245-55.
49. Boden, W.E., et al., *Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy.* N Engl J Med, 2011. **365**(24): p. 2255-67.
50. *Effects of Extended-Release Niacin with Laropiprant in High-Risk Patients.* New England Journal of Medicine, 2014. **371**(3): p. 203-212.
51. Barter, P.J., et al., *Effects of torcetrapib in patients at high risk for coronary events.* N Engl J Med, 2007. **357**(21): p. 2109-22.
52. Funder, J.W., *Aldosterone, Sodium, and Hypertension.* Hypertension, 2010. **55**(2): p. 221-223.
53. Forrest, M.J., et al., *Torcetrapib-induced blood pressure elevation is independent of CETP inhibition and is accompanied by increased circulating levels of aldosterone.* British journal of pharmacology, 2008. **154**(7): p. 1465-1473.
54. Rached, F.H., M.J. Chapman, and A. Kontush, *HDL particle subpopulations: Focus on biological function.* Biofactors, 2015. **41**(2): p. 67-77.
55. Davidson, W.S. and T.B. Thompson, *The structure of apolipoprotein A-I in high density lipoproteins.* J Biol Chem, 2007. **282**(31): p. 22249-53.
56. Joshi, P.H., et al., *Association of high-density lipoprotein subclasses and incident coronary heart disease: The Jackson Heart and Framingham Offspring Cohort Studies.* Eur J Prev Cardiol, 2016. **23**(1): p. 41-9.
57. Rothblat, G.H. and M.C. Phillips, *High-density lipoprotein heterogeneity and function in reverse cholesterol transport.* Curr Opin Lipidol, 2010. **21**(3): p. 229-38.
58. Bhatt, A. and A. Rohatgi, *HDL Cholesterol Efflux Capacity: Cardiovascular Risk Factor and Potential Therapeutic Target.* Curr Atheroscler Rep, 2016. **18**(1): p. 2.
59. Curtiss, L.K., et al., *What is so special about apolipoprotein AI in reverse cholesterol transport?* Arterioscler Thromb Vasc Biol, 2006. **26**(1): p. 12-9.
60. Phillips, M.C., W.J. Johnson, and G.H. Rothblat, *Mechanisms and consequences of cellular cholesterol exchange and transfer.* Biochim Biophys Acta, 1987. **906**(2): p. 223-76.
61. Davidson, W.S., et al., *Effects of acceptor particle size on the efflux of cellular free cholesterol.* J Biol Chem, 1995. **270**(29): p. 17106-113.
62. Ji, Y., et al., *Scavenger receptor BI promotes high density lipoprotein-mediated cellular cholesterol efflux.* J Biol Chem, 1997. **272**(34): p. 20982-5.
63. de Beer, M.C., et al., *Apolipoprotein A-I conformation markedly influences HDL interaction with scavenger receptor BI.* J Lipid Res, 2001. **42**(2): p. 309-13.
64. Thuahnai, S.T., et al., *Scavenger receptor class B type I-mediated cholesteryl ester-selective uptake and efflux of unesterified cholesterol. Influence of high density lipoprotein size and structure.* J Biol Chem, 2004. **279**(13): p. 12448-55.
65. Gelissen, I.C., et al., *ABCA1 and ABCG1 synergize to mediate cholesterol export to apoA-I.* Arterioscler Thromb Vasc Biol, 2006. **26**(3): p. 534-40.
66. Vaughan, A.M. and J.F. Oram, *ABCA1 and ABCG1 or ABCG4 act sequentially to remove cellular cholesterol and generate cholesterol-rich HDL.* J Lipid Res, 2006. **47**(11): p. 2433-43.
67. Soran, H., J.D. Schofield, and P.N. Durrington, *Antioxidant properties of HDL.* Frontiers in Pharmacology, 2015. **6**: p. 222.
68. Barter, P.J., et al., *Antiinflammatory properties of HDL.* Circ Res, 2004. **95**(8): p. 764-72.
69. de Grooth, G.J., et al., *A review of CETP and its relation to atherosclerosis.* J Lipid Res, 2004. **45**(11): p. 1967-74.

70. Ferri, N., et al., *Present therapeutic role of cholesteryl ester transfer protein inhibitors.* Pharmacol Res, 2018. **128**: p. 29-41.
71. Niesor, E.J., et al., *Modulating cholesteryl ester transfer protein activity maintains efficient pre-beta-HDL formation and increases reverse cholesterol transport.* J Lipid Res, 2010. **51**(12): p. 3443-54.
72. Filippatos, T.D. and M.S. Elisaf, *Evacetrapib and cardiovascular outcomes: reasons for lack of efficacy.* Journal of thoracic disease, 2017. **9**(8): p. 2308-2310.
73. Lincoff, A.M., et al., *Evacetrapib and Cardiovascular Outcomes in High-Risk Vascular Disease.* New England Journal of Medicine, 2017. **376**(20): p. 1933-1942.
74. Louise Bowman, M.D., F.R.C.P., Jemma, et al., *Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease.* New England Journal of Medicine, 2017. **377**(13): p. 1217-1227.
75. Ray, K.K., et al., *The effect of cholesteryl ester transfer protein inhibition on lipids, lipoproteins, and markers of HDL function after an acute coronary syndrome: the dal-ACUTE randomized trial.* Eur Heart J, 2014. **35**(27): p. 1792-800.
76. Schwartz, G.G., et al., *Effects of Dalcetrapib in Patients with a Recent Acute Coronary Syndrome.* New England Journal of Medicine, 2012. **367**(22): p. 2089-2099.
77. Tardif, J.C., et al., *Pharmacogenomic determinants of the cardiovascular effects of dalcetrapib.* Circ Cardiovasc Genet, 2015. **8**(2): p. 372-82.
78. Rhainds, D., et al., *Abstract 17913: Inhibition of CETP by Dalcetrapib Results in a Modest Increase in Cholesterol Efflux Capacity Associated With an Increase in Large HDL Particles, but Does Not Impact Carotid Intima-Media Thickness in a dal-PLAQUE-2 Substudy.* Circulation, 2014. **130**(suppl_2): p. A17913-A17913.
79. Hardy, J. and A. Singleton, *Genomewide association studies and human disease.* N Engl J Med, 2009. **360**(17): p. 1759-68.
80. Nissen, S.E., et al., *ADCY9 Genetic Variants and Cardiovascular Outcomes With Evacetrapib in Patients With High-Risk Vascular Disease: A Nested Case-Control Study.* JAMA Cardiol, 2018. **3**(5): p. 401-408.
81. Holmes, M.V. and G. Davey Smith, *CETP inhibition and ADCY9 genotype: evidence of a qualitative pharmacogenetic interaction in cardiovascular disease?* bioRxiv, 2018: p. 336875.
82. Pitts, R., et al., *Treatment With Dalcetrapib Modifies the Relationship Between High-Density Lipoprotein Cholesterol and C-Reactive Protein.* J Am Coll Cardiol, 2016. **68**(22): p. 2488-2490.
83. Tardif, J.C., et al., *CETP: Pharmacogenomics-Based Response to the CETP Inhibitor Dalcetrapib.* Arterioscler Thromb Vasc Biol, 2017. **37**(3): p. 396-400.
84. Tardif, J.C., et al., *Genotype-Dependent Effects of Dalcetrapib on Cholesterol Efflux and Inflammation: Concordance With Clinical Outcomes.* Circ Cardiovasc Genet, 2016. **9**(4): p. 340-8.
85. Favari, E., et al., *Cholesterol efflux and reverse cholesterol transport.* Handb Exp Pharmacol, 2015. **224**: p. 181-206.
86. Sadana, R. and C.W. Dessauer, *Physiological roles for G protein-regulated adenylyl cyclase isoforms: insights from knockout and overexpression studies.* Neurosignals, 2009. **17**(1): p. 5-22.
87. Zhang, G., et al., *Structure of the adenylyl cyclase catalytic core.* Nature, 1997. **386**(6622): p. 247-53.
88. Qi, C., et al., *The structure of a membrane adenylyl cyclase bound to an activated stimulatory G protein.* Science, 2019. **364**(6438): p. 389-394.
89. Kamenetsky, M., et al., *Molecular details of cAMP generation in mammalian cells: a tale of two systems.* J Mol Biol, 2006. **362**(4): p. 623-39.
90. Hacker, B.M., et al., *Cloning, chromosomal mapping, and regulatory properties of the human type 9 adenylyl cyclase (ADCY9).* Genomics, 1998. **50**(1): p. 97-104.

91. Defer, N., M. Best-Belpomme, and J. Hanoune, *Tissue specificity and physiological relevance of various isoforms of adenylyl cyclase.* Am J Physiol Renal Physiol, 2000. **279**(3): p. F400-16.
92. Paterson, J.M., et al., *Characterisation of human adenylyl cyclase IX reveals inhibition by Ca(2+)/Calcineurin and differential mRNA plyadenylation.* J Neurochem, 2000. **75**(4): p. 1358-67.
93. Palvolgyi, A., et al., *Auto-inhibition of adenylyl cyclase 9 (AC9) by an isoform-specific motif in the carboxyl-terminal region.* Cell Signal, 2018. **51**: p. 266-275.
94. Cumbay, M.G. and V.J. Watts, *Novel Regulatory Properties of Human Type 9 Adenylate Cyclase (AC9).* Journal of Pharmacology and Experimental Therapeutics, 2004.
95. Apinjoh, T.O., et al., *Association of candidate gene polymorphisms and TGF-beta/IL-10 levels with malaria in three regions of Cameroon: a case–control study.* Malaria Journal, 2014. **13**: p. 236-236.
96. Berndt, S.I., et al., *Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.* Nat Genet, 2013. **45**(5): p. 501-12.
97. Kim, S.H., et al., *Combined pharmacogenetic effect of ADCY9 and ADRB2 gene polymorphisms on the bronchodilator response to inhaled combination therapy.* J Clin Pharm Ther, 2011. **36**(3): p. 399-405.
98. Niesor, E.J. and R. Benghozi, *Potential Signal Transduction Regulation by HDL of the beta2-Adrenergic Receptor Pathway. Implications in Selected Pathological Situations.* Arch Med Res, 2015. **46**(5): p. 361-71.
99. Pontier, S.M., et al., *Cholesterol-dependent Separation of the β2-Adrenergic Receptor from Its Partners Determines Signaling Efficacy: INSIGHT INTO NANOSCALE ORGANIZATION OF SIGNAL TRANSDUCTION.* Journal of Biological Chemistry, 2008. **283**(36): p. 24659-24672.
100. Li, Y., et al., *Loss of type 9 adenylyl cyclase triggers reduced phosphorylation of Hsp20 and diastolic dysfunction.* Scientific reports, 2017. **7**(1): p. 5522-5522.
101. See, R., et al., *Protein Kinase A Site-specific Phosphorylation Regulates ATP-binding Cassette A1 (ABCA1)-mediated Phospholipid Efflux*. Vol. 277. 2002. 41835-42.
102. Liu, L., et al., *mTORC2 regulates neutrophil chemotaxis in a cAMP- and RhoA-dependent fashion.* Dev Cell, 2010. **19**(6): p. 845-57.
103. Liu, L., D. Gritz, and C.A. Parent, *PKCβII acts downstream of chemoattractant receptors and mTORC2 to regulate cAMP production and myosin II activity in neutrophils.* Molecular biology of the cell, 2014. **25**(9): p. 1446-1457.
104. Petter K. Risøe, A.R., Joanna Ågren, Ingrid B.M. Kolseth, Signe Flood Kjeldsen, and J.V. Guro Valen, Maria K. Dahle, *Higher TNFa responses in young males compared to females are associated with attenuation of monocyte adenylyl cyclase expression.* Hum Immunol, 2015. **76:427–43**.
105. Huang, B., et al., *miR-142-3p restricts cAMP production in CD4+CD25- T cells and CD4+CD25+ TREG cells by targeting AC9 mRNA.* EMBO Rep, 2009. **10**(2): p. 180-5.
106. Alper, S., et al., *Identification of innate immunity genes and pathways using a comparative genomics approach.* Proceedings of the National Academy of Sciences, 2008. **105**(19): p. 7016-7021.
107. Rautureau, Y., et al., *ADCY9 (Adenylate Cyclase Type 9) Inactivation Protects From Atherosclerosis Only in the Absence of CETP (Cholesteryl Ester Transfer Protein).* Circulation, 2018. **138**(16): p. 1677-1692.
108. Hanoune, J. and N. Defer, *Regulation and role of adenylyl cyclase isoforms.* Annu Rev Pharmacol Toxicol, 2001. **41**: p. 145-74.
109. Gallagher, M.D. and A.S. Chen-Plotkin, *The Post-GWAS Era: From Association to Function.* Am J Hum Genet, 2018. **102**(5): p. 717-730.

110. Oldoni, F., et al., *Post-GWAS methodologies for localisation of functional non-coding variants: ANGPTL3.* Atherosclerosis, 2016. **246**: p. 193-201.
111. Hormozdiari, F., et al., *Identifying causal variants at loci with multiple signals of association.* Genetics, 2014. **198**(2): p. 497-508.
112. Edwards, S.L., et al., *Beyond GWASs: illuminating the dark road from association to function.* Am J Hum Genet, 2013. **93**(5): p. 779-97.
113. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease.* Cell, 2013. **152**(6): p. 1237-51.
114. Jameson J.L., K.P., *Harrison's Principles of Internal Medicine 16th Edition.* 2005. **Principles of Human Genetics**: p. 359-363.
115. Krishnamurthy, S. and M. Hampsey, *Eukaryotic transcription initiation.* Curr Biol, 2009. **19**(4): p. R153-6.
116. Deplancke, B., D. Alpern, and V. Gardeux, *The Genetics of Transcription Factor DNA Binding Variation.* Cell, 2016. **166**(3): p. 538-554.
117. Westra, H.J. and L. Franke, *From genome to function by studying eQTLs.* Biochim Biophys Acta, 2014. **1842**(10): p. 1896-1902.
118. Lu, Z.-X., P. Jiang, and Y. Xing, *Genetic variation of pre-mRNA alternative splicing in human populations.* Wiley interdisciplinary reviews. RNA, 2012. **3**(4): p. 581-592.
119. Park, E., et al., *The Expanding Landscape of Alternative Splicing Variation in Human Populations.* The American Journal of Human Genetics, 2018. **102**(1): p. 11-26.
120. Bonnevie-Nielsen, V., et al., *Variation in antiviral 2',5'-oligoadenylate synthetase (2'5'AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene.* Am J Hum Genet, 2005. **76**(4): p. 623-33.
121. Lu, Z.X., et al., *Context-dependent robustness to 5' splice site polymorphisms in human populations.* Hum Mol Genet, 2011. **20**(6): p. 1084-96.
122. Calore, M., L.J. De Windt, and A. Rampazzo, *Genetics meets epigenetics: Genetic variants that modulate noncoding RNA in cardiovascular diseases.* J Mol Cell Cardiol, 2015. **89**(Pt A): p. 27-34.
123. Hrdlickova, B., et al., *Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease.* Biochim Biophys Acta, 2014. **1842**(10): p. 1910-1922.
124. Simmons, D., *Epigenetic influences and disease*. Vol. 1. 2008.
125. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation.* Current Opinion in Genetics & Development, 2011. **21**(2): p. 175-186.
126. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
127. Choy, M.-K., et al., *Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated.* BMC Genomics, 2010. **11**: p. 519-519.
128. Clouaire, T. and I. Stancheva, *Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin?* Cellular and molecular life sciences : CMLS, 2008. **65**(10): p. 1509-1522.
129. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
130. Lawrence, M., S. Daujat, and R. Schneider, *Lateral Thinking: How Histone Modifications Regulate Gene Expression.* Trends Genet, 2016. **32**(1): p. 42-56.
131. Gupta, S., et al., *HISTONE METHYLATION REGULATES MEMORY FORMATION.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2010. **30**(10): p. 3589-3599.
132. Hitchins, M.P., et al., *Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR.* Cancer Cell, 2011. **20**(2): p. 200-13.

133. Aran, D., S. Sabato, and A. Hellman, *DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes.* Genome Biol, 2013. **14**(3): p. R21.
134. Verlaan, D.J., et al., *Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease.* Am J Hum Genet, 2009. **85**(3): p. 377-93.
135. Ghoussaini, M., et al., *Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation.* American journal of human genetics, 2016. **99**(4): p. 903-911.
136. Eckart, N., et al., *Functional Characterization of Schizophrenia-Associated Variation in CACNA1C.* PloS one, 2016. **11**(6): p. e0157086-e0157086.
137. Reynard, L.N., et al., *Functional Characterization of the Osteoarthritis Susceptibility Mapping to CHST11-A Bioinformatics and Molecular Study.* PloS one, 2016. **11**(7): p. e0159024-e0159024.
138. Alves C, C.C., *Electrophoretic Mobility Shift Assay: Analyzing Protein - Nucleic Acid Interactions.* 2012.
139. Frederick M.Ausubel, R.B., Robert E.Kingston, David D.Moore, J.G.Seidman, John A.Smith, Kevin Struhl, *Current Protocols in Molecular Biology.* 2003: p. 1809-1819.
140. LI-COR Biosciences, M.D., *Electrophoretic Mobility Shift Assay (EMSA) Using IRDye® Oligonucleotides.*
141. Tacheny, A., et al., *Mass spectrometry-based identification of proteins interacting with nucleic acids.* J Proteomics, 2013. **94**: p. 89-109.
142. Wiehle, L. and A. Breiling, *Chromatin Immunoprecipitation.* Methods Mol Biol, 2016. **1480**: p. 7-21.
143. Raj, T., et al., *Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes.* Science (New York, N.Y.), 2014. **344**(6183): p. 519-523.
144. Westra, H.J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations.* Nat Genet, 2013. **45**(10): p. 1238-1243.
145. Tsuchiya, S., et al., *Establishment and characterization of a human acute monocytic leukemia cell line (THP-1).* Int J Cancer, 1980. **26**(2): p. 171-6.
146. Sundstrom, C. and K. Nilsson, *Establishment and characterization of a human histiocytic lymphoma cell line (U-937).* Int J Cancer, 1976. **17**(5): p. 565-77.
147. Chanput, W., J.J. Mes, and H.J. Wichers, *THP-1 cell line: an in vitro cell model for immune modulation approach.* Int Immunopharmacol, 2014. **23**(1): p. 37-45.
148. Schildberger, A., et al., *Monocytes, peripheral blood mononuclear cells, and THP-1 cells exhibit different cytokine expression patterns following stimulation with lipopolysaccharide.* Mediators Inflamm, 2013. **2013**: p. 697972.
149. Chanput, W., V. Peters, and H. Wichers, *THP-1 and U937 Cells*, in *The Impact of Food Bioactives on Health: in vitro and ex vivo models*, K. Verhoeckx, et al., Editors. 2015, Springer International Publishing: Cham. p. 147-159.
150. Vogenberg, F.R., C. Isaacson Barash, and M. Pursel, *Personalized medicine: part 1: evolution and development into theranostics.* P & T : a peer-reviewed journal for formulary management, 2010. **35**(10): p. 560-576.
151. Cavalli, M., et al., *Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals.* Genomics, 2016. **107**(6): p. 248-254.
152. Cannon, M.E. and K.L. Mohlke, *Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci.* Am J Hum Genet, 2018. **103**(5): p. 637-653.
153. Gallegos, J.E. and A.B. Rose, *Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation.* The Plant cell, 2017. **29**(4): p. 843-853.

154. Xu, S., et al., *Polymorphisms of HOMER1 gene are associated with piglet splay leg syndrome and one significant SNP can affect its intronic promoter activity in vitro.* BMC genetics, 2018. **19**(1): p. 110-110.
155. Coulon, V., et al., *A novel mouse c-fos intronic promoter that responds to CREB and AP-1 is developmentally regulated in vivo.* PloS one, 2010. **5**(6): p. e11235-e11235.
156. Li, G., et al., *Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation.* Cell, 2012. **148**(1-2): p. 84-98.
157. Haidar, B., et al., *cAMP induces ABCA1 phosphorylation activity and promotes cholesterol efflux from fibroblasts.* Journal of Lipid Research, 2002. **43**(12): p. 2087-2094.
158. Ou, S.H., et al., *Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs.* J Virol, 1995. **69**(6): p. 3584-96.
159. Buratti, E., et al., *Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping.* EMBO J, 2001. **20**(7): p. 1774-84.
160. Mercado, P.A., et al., *Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene.* Nucleic acids research, 2005. **33**(18): p. 6000-6010.
161. Wang, I.-F., et al., *TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor.* Journal of Neurochemistry, 2008. **105**(3): p. 797-806.
162. Chen, K., et al., *A global change in RNA polymerase II pausing during the Drosophila midblastula transition.* eLife, 2013. **2**: p. e00861-e00861.
163. Scheidegger, A. and S. Nechaev, *RNA polymerase II pausing as a context-dependent reader of the genome.* Biochemistry and cell biology = Biochimie et biologie cellulaire, 2016. **94**(1): p. 82-92.
164. Delneste, Y., et al., *Interferon-gamma switches monocyte differentiation from dendritic cells to macrophages.* Blood, 2003. **101**(1): p. 143-50.
165. Meraro, D., et al., *IFN-Stimulated Gene 15 Is Synergistically Activated Through Interactions Between the Myelocyte/Lymphocyte-Specific Transcription Factors, PU.1, IFN Regulatory Factor-8/IFN Consensus Sequence Binding Protein, and IFN Regulatory Factor-4: Characterization of a New Subtype of IFN-Stimulated Response Element.* The Journal of Immunology, 2002. **168**(12): p. 6224-6231.
166. Decker, T., P. Kovarik, and A. Meinke, *GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression.* J Interferon Cytokine Res, 1997. **17**(3): p. 121-34.
167. Tessitore, A., et al., *Two gamma-interferon-activation sites (GAS) on the promoter of the human intercellular adhesion molecule (ICAM-1) gene are required for induction of transcription by IFN-gamma.* Eur J Biochem, 1998. **258**(3): p. 968-75.
168. Childs, K.S. and S. Goodbourn, *Identification of novel co-repressor molecules for Interferon Regulatory Factor-2.* Nucleic acids research, 2003. **31**(12): p. 3016-3026.
169. Bovolenta, C., et al., *Molecular interactions between interferon consensus sequence binding protein and members of the interferon regulatory factor family.* Proc Natl Acad Sci U S A, 1994. **91**(11): p. 5046-50.
170. Sharf, R., et al., *Functional domain analysis of interferon consensus sequence binding protein (ICSBP) and its association with interferon regulatory factors.* J Biol Chem, 1995. **270**(22): p. 13063-9.