Université de Montréal

Genetic association study of plasma creatine kinase levels in the Montreal Heart Institute Hospital Cohort

Par

Rosa Zea Zetler

Unité académique Sciences Biomédicales, Université de Montréal

Faculté Médicine

Mémoire présenté à la Faculté Médicine
en vue de l'obtention du grade de M.Sc
en Sciences Biomédicales
option générale

Juin, 2014

1 RÉSUMÉ

Il a déjà été démontré que les statines (ou inhibiteurs de la HMG-CoA réductase) sont efficaces pour réduire le LDL-cholestérol et elles se sont depuis établies comme étant le pilier dans le traitement de la dyslipidémie. Toutefois, environ 10 pourcent des utilisateurs de statines souffrent d'effets indésirables, généralement sous forme de myopathie qui est souvent accompagnée d'un taux élevé de la créatine kinase (CK) plasmatique. Il est fréquent que les patients doivent arrêter les statines à cause d'un taux de CK dépassant un seuil de référence. Nous avons examiné le taux de CK de près de 6000 participants de la biobanque de l'ICM, qui ont récemment été génotypés à l'aide de la micropuce d'ADN ExomChip d'Illumina. Des études antérieures ont démontré une association significative entre le taux de CK plasmatique et des polymorphismes génétiques et nous avons cherché à répliquer ces résultats par association génétique et à l'aide du test SKAT pour les polymorphismes rares. Nous avons répliqué les résultats dans le gène *CKM* (rs11559024, $p=1.59x10^{-23}$) et le gène *LILRB5* (rs12975366, $p=1.44x10^{-26}$) dans le chromosome 19. Nous espérons que ces résultats seront éventuellement utilisés en clinique pour la prédiction des taux de référence de CK personnalisés selon le profil génétique des patients utilisateurs de statines.

Mots-clés: statine, créatine kinase, myopathie

2 ABSTRACT

Statins (HMG-CoA reductase inhibitors) have been shown to reduce LDL-cholesterol and are undoubtedly the mainstay in the treatment of hyperlipidemia. Approximately 10 percent of statin users suffer from adverse side effects, the most common being muscle myopathy. Muscle myopathy is often accompanied by elevated levels of plasma creatine kinase (CK). Oftentimes, patients are taken off statins after their CK levels surpass a reference threshold. We looked at CK levels in the MHI Biobank, which have recently been genotyped in over 6000 participants with the Illumina ExomChip. Prior studies have found significant association between plasma CK levels and genetic variants and we aimed to replicate these findings using a genome wide association and a SKAT burden test for rare variants. We were able to replicate findings in the *CKM* gene (rs11559024, $p=1.59x10^{-23}$) and *LILRB5* gene (rs12975366, $p=1.44x10^{-23}$) in chromosome 19. We hope that these results will eventually be utilized in clinical statin care by aiding in the prediction of personalized reference CK levels based on genetic information for patients using statins.

Key words: statin, creatine kinase, myopathy

TABLE OF CONTENTS

3 LIST OF TABLES

4 LIST OF FIGURES

5 LIST OF ABBREVIATIONS

ABCB1 – ATP-binding cassette, subfamily B (MDR/TAP), member 1 Transporter
ABCG2 – ATP-binding cassette, subfamily G (WHITE), member 2 Transporter
ADP – adenosine diphosphate
AGTR1 – angiotensin II receptor, type 1
ApoB – apolipoprotein B
ATP – adenosime tripohsphate
ATP2B1 – ATPase, Ca++ transporting, plasma membrane 1
CK – Creatine kinase
CKM – creatine kinase muscle gene
COQ2 – Coenzyme Q2 4-hydroxybenzoate polyprenyltransferase
CYP2C9 – Cytochrome P450, family 2, subfamily C, polypeptide 9
CYP3A4 – Cytochrome P450, family 3, subfamily A, polypeptide 4
CYP3A5 – cytochrome P450, family 3, subfamily A, polypeptide 5
CYP2D6 – cytochrome P450, family 2, subfamily D, polypeptide 6
ER – Exertional rhabdomyolysis
GATM – Glycine amidinotransferase
GLM – general linear model
GWAS – Genome wide association study
HMG-CoA – Hydroxy-methylglutaryl-coenzyme A
HTR3B – 5-hydroxytryptamine (serotonin) receptor 3B
HTR7 – 5-hydroxytryptamine (serotonin) receptor 7,adenylate cyclase-coupled
LDL – Low density lipoprotein
MAF – minor allele frequency
NOS3 – nitric oxide synthase 3, endothelial cell
OATP1B1 – Organic anion-transporting polypeptide, member 1B1
PCA – principal component analysis
RYR2 – ryanodine receptor 2
SKAT – Sequence kernel association test
SLCO1B1 – Solute carrier organic anion transporter family, member 1B1
SNP – single nucleotide polymorphism
SNV – single nucleotide variant
ULN – upper limit of normal

I dedicate this work to Copernicus,
for never being afraid to question science.

6 ACKNOWLEDGEMENTS

I would like to thank everyone at the Pharmacogenomics lab; your patience and guidance made this paper possible. A special thanks to my supervisor Marie-Pierre Dubé – your support was unwavering.

To my family and friends, thank you for making me laugh. Through this experience, I have learnt that it is not how we deal with our success, but with our adversity that defines us. In the words of the late Pete Seeger, "we're waist deep in the Big Muddy and the big fool says to push on!"

7 INTRODUCTION

7.1 ***Background on statins***

Statins (HMG-CoA (hydroxy-methylglutaryl-coenzyme A) reductase inhibitors) have been shown to reduce LDL-cholesterol and are undoubtedly the mainstay in the treatment of hyperlipidemia. Statins have revolutionized primary and secondary prevention of coronary atherosclerotic disease due to their lipid-lowering properties and other pleiotropic effects that affect atherosclerotic plaque stability[2]. Statins work as an inhibitor as they share an HMG-CoA like moiety that acts as a HMG-CoA reductase by competing with the HMG-CoA at the binding site.

Among persons with coronary artery disease, there is an evident decrease in mortality in statin users which persists in all age groups. One study found that when non-statin users were compared to statin users, mortality decreased from 29.5% to 8.5% in participants greater than or equal to 80 years of age, 18.7% to 6.9% in participants 65 to 79 years old and 8.9% to 3.1% in participants younger than 65[3]. The efficacy of statins as primary prevention on patients with high cholesterol but no history of heart disease is unclear; one meta-analysis found no significant mortality benefits[4]. When used in secondary prevention, statins have been found to decrease cardiovascular endpoints in patients with pre-existing cardiovascular disease and lower LDL-cholesterol by an average of 1.8 mmol (70 mg/dl), preventing approximately 60% of cardiac events[5].

The synthesis of a cholesterol lowering drug began in 1971 with Akira Endo, a Japanese biochemist. She speculated that fungi such as molds and mushrooms would produce antibiotics

that inhibited HMG-CoA reductase. Endo's team identified mevastatin, a molecule produced by the fungus *Penicillium citrinum*. Although mevastatin is thought to be the first known statin, it was never deemed safe for human use[6] as adverse effects included tumors, muscle deterioration and death. In 1982, lovastatin was identified in *Pleurotus ostreatus*, and by 1987 it was approved by the FDA.

There are currently six statins available for patients in Canada: atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin and simvastatin. Lovastatin, simvastatin and pravastatin are derived from fungi while the other three are synthetic. Their relative potency, from highest to lowest, is: rosuvastatin, atorvastatin, simvastatin, lovastatin, pravastatin, and fluvastatin[7]. Relative risk rates when comparing statins are considered to be highest in rosuvastatin, then atorvastatin and simvastatin and lastly pravastatin and lovastatin[8]. Relative potency of statins seems to correlate with relative risk, with the exception of fluvastatin[8].

Mechanism of statins. HMG-CoA reductase is a catalyst in the early biosynthesis of cholesterol which derives mevalonate by a 4-electron reduction of HMG-CoA. All statins work by effectually inhibiting this step by competing for the catalytic binding domain for the HMG-CoA in the HMG-CoA reductase molecule. Statins are competitive inhibitors of HMG-CoA reductase and, when competing with HMG-CoA, have a significantly higher binding affinity with the reductase molecule. The HMG-CoA-like moiety of a statin monopolizes the HMG-CoA binding pocket in the HMG-CoA reductase molecule and links via its O5 hydroxy group[9]. When the HMG-CoA moiety is in the binding pocket, the remainder of the statin structure undergoes a conformational change in order to maximize contact with, and binding to, amino acid residues of

the reductase. The number and strength of the bonds vary between statins, dictating the length and degree of inhibition and ultimately resulting in the level of LDL cholesterol-lowering efficacy. Rosuvastatin forms nine bonding interactions with the reductase, making it the most potent statin, followed by atorvastatin, which forms eight bonds. The fungal statins, lovastatin, simvastatin and pravastatin, each have six bonding interactions[9].

Through the inhibition of HMG-CoA reductase, there is a reduction of cholesterol synthesis in the hepatocyte, a reduction in the cholesterol pool and subsequent up-regulation of the nuclear transcription factors sterol regulatory element-binding proteins, resulting in an increase in the transcription of LDL receptors on the hepatic cell surface. The LDL receptors link with apolipoprotein (apo) B and apoE on the surface of circulating LDL and very-low density lipoprotein particles are integrated, along with their lipid content, into the hepatocyte[10]. In addition, when hepatic cholesterol is converted to bile acids the LDL receptors are up-regulated to help with food absorption in the gastrointestinal tract.

Metabolism of statins. Lipid-soluble statins are converted to water-soluble salts and glucuronide conjugates for elimination from the body while simvastatin, lovastatin and, to a lesser extent, atorvastatin are metabolized by cytochrome P450 3A4 (CYP3A4) enzymes. Numerous drugs are either inhibitors or substrates for this enzyme system and can result in drug-drug interactions, leading to increased plasma concentrations of the statin. Approximately 10% of fluvastatin and rosuvastatin are metabolized primarily by CYP2C9[11], the rest remaining fairly unchanged until excretion. These two statins are excreted from the body largely unchanged as parent compounds by transporter-mediated excretion mechanisms in the liver into the feces via bile and into the

urine. Pravastatin is not metabolized by the P450 enzyme system. The open acid forms of several statins often undergo glucuronidation followed by lactonization[12]. In their lactone form, statins may be converted to their open acid form by an esterase enzyme and metabolized, excreted by bile or urine, or directly metabolized by the cytochrome P450 system. Excretion of statins through bile or feces is mediated by transport proteins such as OATP1B1 and multidrug resistance-associated protein-2.

Drug interactions in statin users must be closely monitored, as drugs often affect metabolism time, which in turn affects the systematic concentration of statin. Drugs inhibiting the CYP3A4 enzyme can interfere with the metabolism of atorvastatin, simvastatin, and lovastatin causing an increase in the area under the statin concentration-time curve. Drugs that inhibit CYP2C9 should theoretically affect the metabolism of fluvastatin, although in practice there are very few clinical observations of this. OATP1B1 is an inhibitor that, alongside other inhibitors of CYP3A4, inhibits a vital transporter by which hydrophilic statins are transported into the liver.

Certain fibrates, a class of drugs that lowers blood triglyceride levels, and Warfarin, an anticoagulant mainly used in the prevention of thrombosis and thromboembolism, have the potential to interact with statin therapy. The fibrate gemfibrozil can interact with the OATP1B1-mediated transport resulting in an increase of plasma statin concentration[13]. Additionally, gemfibrozil interferes with the glucuronidation and lactonization of statins, promoting a higher concentration of open acid statins. To avoid these drug interactions, it is recommended that the fibrate fenofibrate is used instead to help avoid adverse effects, or the statins atorvastatin or

fluvastatin should be used. Warfarin-statin interactions may result in a displacement from plasma protein binding sites by the statin, resulting in a minor increase of prothrombin times and bleeding. While adverse effects are uncommon, close monitoring of the patient is recommended for a patient using Warfarin when lovastatin, rosuvastatin or simvastatin is added or removed from a regiment.

7.1 ***Creatine kinase levels as a biomarker***

Creatine kinase is the most commonly used enzyme utilized in the diagnosis of muscle disease. Serum levels of CK are the most sensitive known biomarker and the best known indicator of the course of muscle injury[14]. In extreme cases, this muscle myopathy can manifest into rhabdomyolysis, showcasing a correlated increase in plasma CK levels. Using CK as a biomarker is an effective and non-invasive sensitivity test for various levels of muscle myopathy as well as myocardial infractions.

CK is an enzyme found predominantly in skeletal muscle, the myocardium and the brain. CK catalyzes the reversible reaction of creatine, consuming ATP resulting in the production of phosphocreatine and ADP. CK is expressed as any of the following three isoenzymes: CK-MM, CK-BB, CK-MB where B represents the brain type subunit and M the muscle type subunit.

Disruption of cell membranes can release CK from the cellular cytosol into the systemic circulation. This elevation of plasma CK levels can be used as a biomarker for myocardial infarction, rhabdomyolysis, muscle dystrophy, the autoimmune myositides and in acute renal failure.

Serum CK is a commonly used biomarker for myocardial infarction. 40% of CK in cardiac muscles is CK-MB. Levels of CK-MB tend to increase within 3 to 4 hours of myocardial necrosis, peak within a day and return to normal within 36 hours. Subsequent reinfarctions can be diagnosed if CK-MB levels do not begin to decline after approximately 24 hours[15]. Although highly effective in diagnosing acute MI, it is important not to rely solely on CK-MB as a biomarker, as an electrocardiogram should ultimately be used as an assessment of the patients risk and stability[16].

7.2 ***Patient management through evaluation of CK levels***

Clinicians often measure serum CK levels as a rough proxy for severity of statin-induced myotoxicity, but the correlation between symptoms and CK level remains incomplete. The clinical interpretation of CK level is complex and there is not yet a consensus on the definition of statin myopathy. The American College of Cardiology[17], American Heart Association (AHA), National Heart, Lung and Blood Institute (NHLBI)[18], the FDA[19] and National Lipid Association (NLA)[20] have each proposed different definitions for statin-related muscle effects (**Table II**).

Table I. Definitions of statin-induced myalgia

Clinical entity	ACC/AHA/NHLBI 2002	NLA 2006	FDA
Myopathy	General term referring to any disease of muscles	Complaints of myalgia (muscle pain or soreness), weakness, and/or cramps, plus elevation in serum CK > 10 × upper limit of normal (ULN)	CK ≥ 10 × ULN
Myalgia	Muscle ache or weakness without CK elevation	NA	NA
Myositis	Muscle symptoms with increased CK	NA	NA
Rhabdomyolysis	Muscle symptoms associated with marked CK elevations, typically substantially > 10 × ULN and with creatinine elevation	CK > 10,000 IU/l or CK > 10 × ULN plus an elevation in serum creatinine or medical intervention with i.v. hydration	CK > 50 × ULN and evidence of organ damage, such as renal compromise

The lack of consensus on the definition of statin myopathy hinders estimation of its true incidence. Elevations less than threefold the upper limit of normal (ULN) are typically considered of little consequence. Conversely, clinicians often intervene in statin therapy (change dose or change drug) when CK levels exceed threefold the ULN. At present, the literature supports three diagnostic strata: (i) incipient myopathy (CK above 3-fold the ULN, less than 10-fold the ULN), (ii) myopathy (CK above 10-fold the ULN, less than 50-fold the ULN), and (iii) rhabdomyolysis (CK above 50-fold the ULN)[21]. CK levels are not routinely measured before statin therapy begins. Patients are encouraged to report myalgia by a physician warning or the product information sheet. When CK levels are elevated, the statin is usually withdrawn,

although, largely due to the lack of baseline CK measurement, it is difficult to determine whether statin therapy or another cause is to blame.

Disagreement in the definition of statin-induced muscle events impedes the precise estimation of its true prevalence. Because patients with a considered high susceptibility to statin toxicity are generally excluded from clinical trials of statins, reported adverse event rates from controlled trials may underestimate the true rate of these adverse effects in an unselected patient population. As genetic variants may play a role in elevated levels of plasma CK, it is valuable to analyze the SNPs in the hopes of finding a new biomarker to better diagnose CK levels in patients. There has been a preliminary study done at the Pharmacogenomics center, and strong association was found between variants in the muscle CK gene (CKM) and elevated plasma levels of CK.

A meta-analysis of 21 clinical trials providing 180,000 person-years of follow-up found that myopathy, defined by muscle symptoms and CK levels above a 10-fold ULN, occurs in five patients per 100,000 person-years[20]. Rhabdomyolysis, defined by either CK levels above 10,000 IU/L, CK levels above a 10-fold ULN with an elevation in serum creatinine or requirement for hydration therapy, occurs in 1.6 patients per 100,000 person-years[22]. Less severe manifestations are much more common. Although CK has long been used as a diagnostic biomarker, CK measurement is not always accurate for the clinical management of statin-induced myotoxicity[23], and therefore more accurate tools and/or novel sensitive biomarkers are required.

7.3 ***Predictors of serum creatine kinase levels***

Although elevated levels of CK are often used as a biomarker for muscle myopathy, the expected level of CK in a patient is often unclear. Elevated CK levels tend to be higher in certain populations, and as baseline CK levels are not routinely measured due to cost and time limitations, it is hard to gauge what the actual effects of statins are on CK levels in a patient. Covariates such as gender, ethnicity, age and physical activity levels have been found to be highly correlated with plasma CK levels.

High intensity exercise will spike CK levels for a limited amount of time, usually for a few days following exercise[24]. Females tend to have lower baseline levels of CK, although it has been suggested that females might undergo a greater spike in CK levels with physical activity[25,26]. It has been refuted that this increased spike in CK levels may be attributed to poor study design, and not actually account for the differences in genders exercising[27,28]. Although there is a clear correlation between serum CK levels and recent physical activity, it is unclear whether this spike is a good predictor of muscle damage[29]. When completing a routine measure of a patient's serum CK levels, it is hard to account for the effect of physical activity in the measurement and whether this increase in serum CK levels is leading to muscle myopathy.

Exertional rhabdomyolysis (ER) is the degradation of skeletal muscle cells caused by strenuous exercise. Common indicators of ER include muscle pain, myoglobinuria and an elevated level of CK after exercise. It is unclear why certain individuals experience ER while others, who participate in comparable amounts of physical activity, do not. Confounders include, but are not limited to, fasting, hypokalemia, certain dietary supplements, dehydration, concurrent

illnesses, low baseline fitness levels and extreme or repetitive exercise[30]. Although CK is often used to monitor ER, correlation is unclear. Some patients with CK levels greater than 10,000 U/L have no overt clinical symptoms, while others with a CK level less than 5,000 U/L exhibit symptoms of ER[31]. One study that looked at genetic polymorphisms associated with ER found three noteworthy SNPs in the CKMM, ACTN3 and MYLK2 genes[32]. Another study that looked at association between serum CK and ER in military recruits found the SNPs IL-6 and MLCK 37885 to be positively associated, and furthermore these SNPs were more common in participants of African-American descent[33]. It was noted that this association may have been due to genetic differences in race as opposed to differences in CK levels.

Age and gender have an effect on expected serum CK levels. The average male's baseline CK level is higher than the average females. Additionally, there is an age-dependant decrease of CK levels, seen predominantly in males[34].

Individuals of African American ancestry have been shown to have higher levels of creatine kinase when compared to Caucasian, Hispanic and Asian people[34]. Current standards for measuring muscle myopathy through CK levels disregard biological and environmental differences in patients. Serum CK levels are reported to be approximately 70% higher in a healthy Black person when compared to a healthy Caucasian[35]. A young African American male can be expected to experience higher than normal levels of CK levels without muscle damage, which is not accounted for in the above definitions of myopathy and rhabdomyolysis.

7.4 ***Adverse effects of statins***

Most patients tolerate statins well; however, 10% of patients develop muscle-related adverse effects[36]. The most commonly reported adverse effect of statins is muscle myotoxicity, which can, in the worst cases, progress to rhabdomyolysis. Other adverse effects include toxicity in the kidneys and liver, and may be heightened by age or concomitant medications. Accurate diagnosis of statin-induced myopathy is important given the large and expanding numbers of patients eligible for statin therapy[37] and the fact that myalgia, or muscle pain, is one of the most frequent causes of discontinuation of therapy. Statins can cause a wide range of muscular adverse effects with no specific clinical characteristics, from non-specific myalgias to rhabdomyolysis[38], with symptoms usually developing within four weeks but can be delayed up to four years after statin initiation.

Advocates of statin therapy believe benefits greatly outweigh risks. Myopathy and rhabdomyolysis are estimated to affect 1/100,000 patient-years and by monitoring altered CK levels these adverse effects can be minimized[39]. Monitoring and measuring statin myopathy includes watching for elevated levels of CK, statin dosage reduction, discontinuation, or switching of statin and alternative treatments, such as low-dose or alternative day (as opposed to daily) rosuvastatin[21].

Hepatotoxicity, or chemically driven liver damage, manifesting in statin users as asymptomatic elevation of serum transaminases, hepatits, cholestasis and acute liver failure, has been observed in a small percentage of patients[39]. Serious liver damage is extremely rare in statin

users, one study finding that low to moderate dosages of pravastatin, lovastatin and simvastatin are not associated with liver function test abnormalities[40].

Nephrotoxocity, toxic chemicals or medications to the kidneys, is a rare adverse effect of statins. Studies have reported moderate proteinuria, seen in 1-2% of potent statins given at recommended dosages[41]. Patients with end stage renal disease are more susceptible to nephrotoxicity. Patients with renal dialysis see an increase in cardiovascular events but statins have little or no observed beneficial effects on mortality and cardiovascular events, despite a reduction in serum cholesterol levels[42]. The Study of Heart and Renal Protection (SHARP) found success in reducing major atherosclerotic events when ezetimibe 10 mg was used in addition to simvastatin 20 mg daily in patients with chronic kidney disease[43].

Table II. other rare adverse effects of statins are demonstrated in the following table[39]

Organ/systems	Statin-induced adverse effects
Nervous	Hemorrhagic stroke: ↓ LDL-C level → ↑ hemorrhagic stroke; strong data unavailable; risk benefit ratio of statins completely outweighs Peripheral neuropathy: Generally appear after 1-2 months → resolves on discontinuation → low attributable risk Cognitive impairment: Occasionally reported in statin-treated patients → large controlled trials do not confirm Sleep: Sleep disturbances and nightmares
Cardio-vascular	Vascular reactivity: Statins stimulate vascular CYP2C-derived ROS → inactivation of NO; farnesyl the product of mevalonate cascade deficiency LDL oxidizability: Ubiquinone deficiency → ↑ LDL oxidizability
Immune	Induction of apoptosis and release of intracellular antigens (i.e., histones or nucleic acids) → triggers immune response ↑ Auto-antibodies formation → immune response shifts from Th1-mediated (cellular) to Th2-mediated (humoral)
Endocrine	Insulin sensitivity: Stimulation of farnesyl and geranylgeranyl transferases both *in vitro* and *in vivo* ↓ Insulin sensitivity and ↑ of plasma insulin concentration after statin therapy Statins impairs insulin signaling and insulin secretion
Cancer	Cancer risk: Low coenzyme Q or low serum cholesterol →↑ breast cancer rates
Gastrointestinal tract	Rare side-effects: Nausea, dyspepsia, abdominal pain, diarrhea or constipation, gastric ulcer, gallstones
Skin	Rare side-effects: Alopecia, rashes, lichenoid eruption, dermographism, chronic urticaria, and toxic epidermal necrolysis
Eye	Rarely cause cataract and ocular hemorrhage
Reproductive	Rarely cause erectile dysfunction, decrease libido, and gynecomastia
Blood	↓ Blood clotting → thrombocytopenia and thrombotic thrombocytopenic purpura
Respiratory	Rarely cause interstitial lung diseases

LDL: Low-density lipoprotein

7.5 ***Compliance with statins***

Despite reducing clinical cardiovascular end points by 30%[44] and having a positive benefit to risk ratio, statins are underutilized. A recent European study estimated that 20% of patients with coronary heart disease do not use statins[45]. The primary reason for discontinuation of statins is due to muscle pain, followed by cost and then by perceived lack of efficacy[46]. If physician-patient communication were to increase, there may be an improvement in patients adherence to a statin[47]. Through personalized medication and pharmacogenomics, we aim to eventually increase patient-doctor insight in statin therapy, henceforth increasing patient wellbeing and adherence while minimizing muscle adverse effects.

Patient adherence to statins is low, and even more so in geriatric patients. In elderly patients, most of the statin discontinuation occurs within six months of commencement and long term use remains low[48]. As a result, elderly patients who discontinue may have little to no benefit from receiving statin therapy[49]. In order to gain insight on discontinuation, studies have compared two specific statins, in order to better gauge each of their adherence rates. Discontinuation was found to be much lower in atorvastatin than simvastatin, although it is unclear whether this was due to effectiveness, cost, adverse effects or another confounder[50]. It is difficult to gauge adherence of prescription medicine, as WHO estimates that 50% of patients do not take their medication as prescribed[51].

7.6 ***Loci associated with statin-induced myotoxicity***

Patient-related risk factors for statin-related myotoxicity include female gender, low body mass index, concomitant treatment with certain cytochrome P450 inhibitors, a decline in renal and

hepatic function, and changes in albumin and α-1 glycoprotein levels with subsequent changes in free concentration levels of statins[21]. Statin myopathy is dose-related. An increase in statin dose and statin systemic exposure magnifies the risk of muscle toxicity[21]. It is speculated that there may be a genetic link to a patient's susceptibility to statin adverse effects. Below is a review of genetic variants found in prior studies that are associated with an increase in adverse muscle related effects in statin users.

SLCO1B1. A genome-wide study has identified common genetic variants in *SLCO1B1* that are associated with substantial alterations in the risk of simvastatin-induced myopathy[52]. The finding of an association between the variant **rs4149056** in the *SLCO1B1* gene and statin-induced myotoxicity has since been replicated in both an independent trial and a practice-based longitudinal cohort[53,54]. Recently, the Clinical Pharmacogenomics Implementation Consortium published a guideline paper that discusses the relationship between rs4149056 and the clinical outcome for simvastatin[55]. The mechanism of statin-induced muscle injury is not completely understood, although several mechanisms have been suggested including isoprenoid depletion, decreased sarcolemmal membrane cholesterol, inhibition of ubiquinone or coenzyme Q10 (CoQ10) synthesis, disturbed calcium metabolism or an autoimmune occurrence[21].

Other variants, such as **rs4363657** located in *SLOC1B1*, have been found to be associated with a higher risk of statin induced myopathy[56]. This variant is located within intron 11 of *SLCO1B1* on chromosome 12. These two SNPs identified to be associated with statin metabolism, rs4363657 and rs4149056, were found to be in linkage disequilibrium (r^2=0.97)[52].

rs2306283 has been found to be borderline associated with risk of myopathy in statin users, validated in the SEARCH trial among the HPS subjects[52].

COQ2. 2 SNPs in the COQ2 gene that are in linkage disequilibrium were identified in a study that is associated with adverse muscle symptoms in statin users[57]. Statins reduce CoQ10 levels by inhibiting the HMG-CoA reductase and diminishing the CoQ10 transport capacity by decreasing LDL levels[58] which may hinder statin mechanisms and somehow increase the frequency of muscle adverse affects in statin users.

Another variant associated with statin induced myopathy, **rs4693570,** was identified in the COQ2 gene[59].

GATM. GATM is an enzyme required for the synthesis of creatine that encodes glycine aminotransferase. Since the phosphorylation of creatine is the primary downstream product of *GATM* activity, it was hypothesized that variants in *GATM* may affect muscle adverse effects in statin users. In a primary case control study, and then again in a secondary replication study, it was found that statin users with the minor allele at the *GATM* differential eQTL locus had a reduced incidence of myopathy[60]. **rs9806699** was identified, along with 5 other loci, a *cis*-eQTL for the gene glycine amidinotransferase (*GATM*) that encodes the rate limiting enzyme in creatine synthesis[60]. Although the mechanism behind this association is unclear, a possible explanation could be that diminished capacity for phosphocreatine storage modifies cellular energy storage and adenosine monophosphate-activated protein kinase signalling[61,62] in a manner that is protective against cellular stress as induced by glucose deprivation[62] or, potentially, by

cholesterol depletion. It is possible that statin users may be at high risk for muscle toxicity partially through metabolic effects in the liver, the primary site of statin's pharmacologic actions.

ATP2B1. *ATP2B1*, located in chromosome 12, aids in intracellular calcium regulation and may play a role in statin induced myalgia. Variant **rs17381194** was identified in the Statin Induction and Neuro-Myopathy study as being associated with intracellular calcium regulation, and thereby may be associated with myalgia in statin users[59].

DMPK. Statins have been found to expose mild forms of muscle dystrophy, caused by certain genes. *DMPK* is thought to be related to muscle dystrophy, and variant **rs672348** in the *DMPK* gene, located in chromosome 12, may play a part in muscle related adverse effects in statin users[59].

ABCB1. The SNP **rs2235046** (C1236T) located in the *ABCB1* gene, in chromosome 7, has been found to be associated with elevated plasma CK levels in statin users[63,64].

ABCG2. The *ABCG2* gene, located in chromosome 4, helps to limit the absorption of statins in the gut. The SNP **rs2231142** (genotype *ABCG2* c.421C>A) SNP was found to significantly affect the risk of atorvastatin-induced myopathy[65].

CYP2D6. The SNP **rs3892097** (CYP2D6*4), located in chromosome 22, has been found to be associated with muscle adverse events caused by at least two structurally dissimilar HMG-CoA reductase inhibitors, and may be related to statin induced myopathy[66].

CYP3A5. The *CYP3A5* gene, located in chromosome 7, is involved in drug metabolism and synthesis of cholesterol. One study found the variant **rs776746** (CYP3A5*3) to be associated with atorvastatin-induced myalgia[67]. It was found that persons who were homozygous for CYP3A5*3 had a greater risk for increased plasma CK levels than those that were heterozygous for the SNP[67].

RYR2. Mutations in *RYR2*, an intronic gene located in chromosome 1, may indicate an increased risk for muscle myopathy[68]. One study noted an increased risk for cerviastatin-induced rhabdomyolysis which correlated with a mutation in the SNP **rs2819742**.

HTR3B and HTR7. People who have a lower tolerance for pain can be expected to experience muscle pain at a higher frequency. Two SNPs, **rs2276307** SNP in the *HTR3B* gene, located in chromosome 11, and **rs1935349** SNP in the *HTR7* gene, located in chromosome 10, were found to be associated with the myalgia score in statin users[68,70]. This is thought to be due to these mutations which affect an individual's pain tolerance.

NOS3 and AGTR1. SNPs associated with endothelial homeostasis may be correlated to plasma CK levels in statin users. One study identified two variants, r**s1799983** in the *NOS3* gene, located in chromosome 7, and **rs12695902** in the *AGTR1* gene, located in chromosome 3[68,71].

7.7 ***Myotoxicity reports in clinical trials***

Data from systematic reviews, meta-analyses, clinical trials, and post marketing surveillance indicate that statin-associated myalgia typically affects around 5.0% of patients, as myopathy in 0.1% and as rhabdomyolysis in 0.01%. The frequency of muscle symptoms associated with statin therapy was evaluated in the Prediction of Muscular Risk in Observational Conditions study, an observational study conducted in an unselected population of 7,924 hyperlipidemic patients receiving high-dose statins in an outpatient setting in France[36]. Muscle-related symptoms were reported by 832 patients (10.5%), which is a rate at least 2 times higher than that observed in clinical trials involving statins (1–5%)[72]. The number of patients reporting muscle-related symptoms was the highest in those receiving simvastatin (18.2%), followed by atorvastatin (14.9%) and pravastatin (10.9%). Myalgia is among the leading reasons patients discontinue statins. It is thought that hyroliphic statins such as fluvastatin and pravastatin are less likely to result in muscle adverse events. The number of muscle complaint incidences varies between studies is mainly due to the contradictory definitions of myotoxicity.

One study that assessed the benefits and harms of early administered statins to patients with acute coronary syndrome found that there were no short term benefits (less than four months) although there were long term favourable benefits. Therefore, statins are more useful as a long term preventative measure against cardiovascular incidences. It was found that severe muscle toxicity (found in 0.13% of participants) was mostly limited to those taking 80 mg of simvastatin[73]. When compared to atorvastatin, cerivastatin had a higher rate of muscle myopathy and rhabdomyolysis[74], which was at times fatal, and was taken off the market in 2001. Since statin induced myotoxocity is dose dependant, there may be some success with intermittent statin

dosing regimens in patients with previous intolerance due to myopathy, particularly with atorvastatin and rosuvastatin[75]. The efficacy of this method is questionable, as the benefits of statin may go down as dosage is lowered.

A study from the University of Toronto noted that there has been a shift towards intensive versus moderate statin therapy[76]. This shift has primarily taken place within atorvastatin users, and is not generalizable to other statins. This shift has led some to believe that statins are oftentimes excessively administered. These studies call for a more personalized approach to statin care[77].

Because hypercholesterolemia is largely asymptomatic, any unpleasant effects of pharmacologic agents used to manage it can undermine adherence. However, patients often underestimate the degree to which elevated cholesterol increases their risk of coronary events. Given these perceptions, adverse effects, such as myalgia, can assume an important role in a patient's decision to discontinue a much-needed lipid-modifying medication. Indeed, several studies document a significant decrease in adherence to lipid-lowering treatments over time among outpatients, and some indicate that patient's perception of adverse effects is a primary reason for discontinuation[48,49]. Statin intolerance may substantially reduce treatment compliance and statin efficacy in preventing cardiovascular endpoints. Physicians, pharmacists and patients are often hasty to assume that a variety of common musculo-skeletal aches and pains are related to statin use.

7.8 ***Summary of prior project***

The Pharmacogenomics Center at the MHI conducted a case-control study on statin myotoxicity evaluating genetic determinants of statin-induced muscle pain. 4391 participants were recruited, all taking statins and half reporting muscle adverse events. 979 participants were removed for the following reasons: they were recruited into the study because of a history of high CK levels, they had a missing CK measure or they were not currently on a statin. An additional 3 outliers were removed.

The objective of this project was to perform a genome wide association study evaluating the association between CK measures and SNPs.

It was found that patients with ongoing muscle pain had, on average, slightly higher plasma CK levels than those without a prior history of statin-induced muscle pain who generally tolerated high-dose statins well. In a multivariate analysis, when adjusting for factors known to influence CK levels including age, sex, physical activity, statin and statin dose, a significant difference in plasma CK levels between cases and controls (p-value = 9.90×10^{-5}) was found, despite the fact that the majority of CK values were within the "normal" clinical range.

To follow up on this observation, a genome-wide search for genetic determinants of CK levels in all statin users in the study population was performed. The 610quad and iselect were used to identify 584509 SNPs for the project. A strongly significant association signal between the muscle CK gene (*CKM*) variants and plasma levels of CK ($p = 10^{-16}$; **Figure 1**) was found.

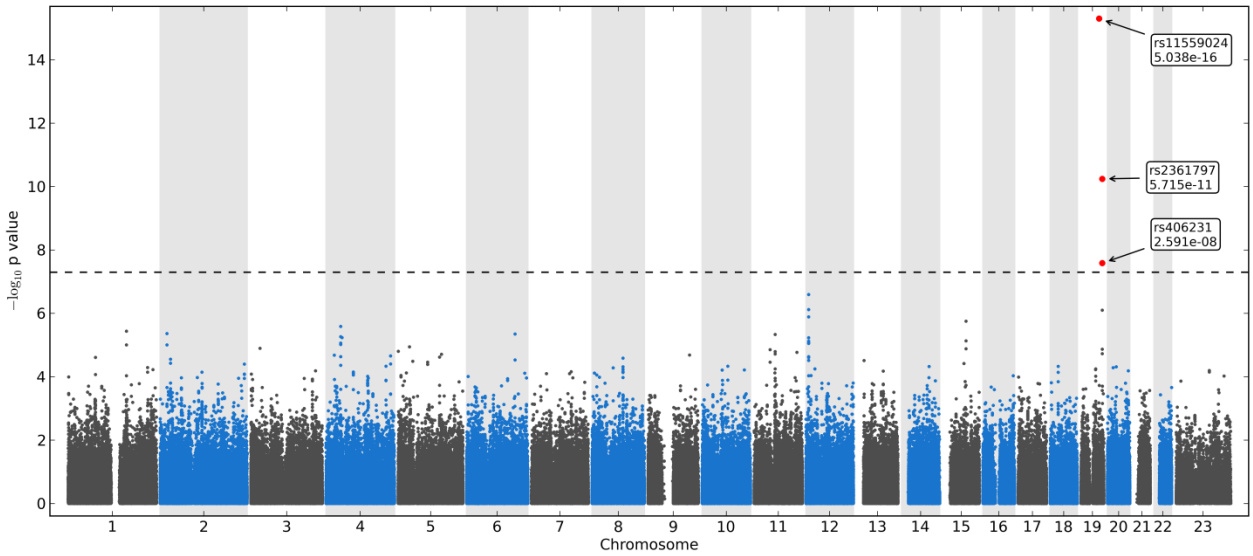

Figure 1**.** Manhattan plot showing the results of a genome-wide association study for genetic determinants of CK levels measured in statin users in the MHI statin study. **A GLM regression with the natural logarithm of CK was used, with adjustment for the case-control status, the lab where the CK measures were taken, gender, physical activity level, age, diabetes and BMI.** There were 3388 participants in the statin myotoxicity case-control study that were using statins at the time of CK measurement. **Genetic variants in the *CKM* gene ($p=5.03x10^{-16}$) and the *LILRB5* gene ($p=5.71x10^{-11}$) were identified.**

The *CKM* variant rs11559024 encodes a missense variant (421A/G) in the 3rd exon of the gene causing an amino acid change at GLU83GLY, a polymorphism with a population frequency of 2%. This gene encodes creatinine-kinase-M which reversibly catalyzes the transfer of phosphate between ATP and creatine phosphate and functions as a homodimer in striated muscle. The two other associated variants (rs2361797 and rs406231) are located in the *LILRB5* gene, over 5M base pairs away from the *CKM* gene. The *LILRB5* gene encodes a leukocyte immunoglobulin-like receptor which binds to MHC class I molecules and is involved in immune response. Genetic association of the *LILRB5* variants by conditioning on the *CKM* variant rs11559024 provides a strongly significant signal ($p=2.55x10^{-11}$) and is therefore expected to function as an independent factor from *CKM*. As the findings in this study are issued from a genome-wide investigation, it is likely that the associations found are indirect associations and

that the identified variants are in linkage disequilibrium with other neighbouring and potentially causal variants.

7.9 ***Reason for replication in a larger cohort***

The prior project warrants a follow up study that will more finely characterize the genetic factors in the *CKM* and *LILRB5* genes. Pharmacogenetics has the potential to optimize patient benefice from drugs by removing the 'fixed dosage' method and instead looking at variants in the DNA in order to gauge the patient's potential benefits and adverse effects of a drug[78]. Replication on a larger cohort will help to ensure that prior finding were credible associations and not a chance finding or due to uncontrolled biases[79]. The current replication study on 5091 patients will validate prior association found with variants in the *CKM* and *LILRB5* genes, and possibly lead to further discoveries, helping to pave the way towards better statin care.

Bias in a GWAS can occur for a number of reasons. One major source of bias when conducting genetic association tests in a population is heterogeneity, especially if the population has a similar ethnic background. A GWAS done for a continuous trait using a population of a certain ethnic background can not necessarily be applied to a population of a different ethnic background. Bias in a study can produce an outcome that differs from the truth, resulting in false negatives, false positives or inaccurate effect sizes. Hardy-Weinberg disequilibrium, genotyping quality, how you arrive at a population, incorrect confounders, and Winner's curse can all cause bias in a GWAS. Replication studies and careful choosing of a population, as well as being aware of all limitations can help to reduce bias.

Conclusive evidence should not be drawn from one study, but rather from a study and the replication studies it spawns. When a regression model shows an association with a p-value below 0.05, this is an indicator that there may be significant correlation, but by no means a complete argument[80]. The probability of a study yielding true results is based on the probability of truth before the study was done, the statistical power of the study and the statistical significance of the study[81,82]. Therefore it is best if replication is shown before publishing findings, in order to minimize the number of contradictory publications and better further studies in the area.

In a preliminary genome wide association study, a phenomenon called the *winner's curse* can arise. This refers to the regression toward the mean. Since numerous tests are done at once, a preliminary GWAS will report the most significant findings, or the *winners*, which tend to be inflated. When association of a rare variant is discovered, the power can be less than 1%, grossly inflating the effects and creating a need for further replication before confirmation of association[83].

Preliminary, hypothesis testing studies are a first step, and are not sufficient on their own. Carefully conducted replication studies are necessary, followed by a sensible meta-analysis that incorporates not only statistical analysis but also provides a cautious interpretation of the results[84]. The prior study done at the Pharmacogenomics Center of the MHI warrants a replication project in a larger cohort. Before confirmation of the genetic variants previously identified, it is necessary to test for replication in order to increase confidence that identified variants are not false negatives.

7.10 ***Discovery Study with the MHI Biobank***

The MHI Biobank study uses the ExomeChip from Illumina to provide genetic data of the coding regions of genes in participants, which allows for the testing of rare genetic variants in a large cohort. We propose to use this wealth of genetic data to examine these rare SNPs and their relation to plasma CK levels in order to potentially further validate or discover novel variants that were not necessarily covered in the prior study. The discovery study will ultimately pinpoint any genetic variants in this population that are associated with plasma CK levels, providing a deeper comprehension of genetic variants related to statin adverse muscle effects.

The MHI Cohort is diverse in terms of age and sex and is thorough in its questionnaire regarding alcohol intake, dietary habits, tobacco, medication and depression. Vital signs (heart rate, blood pressure, weight, height, waist circumference) are obtained by a nurse, and blood, DNA, and plasma are collected. The self-reported physical activity is lacking in comparison to the WHO physical activity standards, but this pitfall has been amended for any new participants in the continuation of the study. Participants were recruited from MHI and at its affiliated EPIC centre. Over 11,000 participants have been genotyped with the Illumina ExomeChip. The MHI Biobank is a robust cohort that can be used for the analysis of a multitude of studies. As in any study, it must be noted that the cohort population was located in Montreal, so this population can be expected to differ in genetic makeup than a population residing elsewhere.

The ExomeChip from Illumina tests for over 270,000 genome-wide variants located in the exon of genes. The ExomeChip aims to create an intermediate platform which will enable examination of coding variants, down to singletons. Both relatively common variants currently

used in genotyping arrays, and exome sequencing of a very large numbers of samples are looked at using the ExomeChip.

8 RATIONAL AND HYPOTHESIS

We hypothesize that there exist genetic variants in the coding regions of the exome that would regulate plasma CK levels. We further hypothesize that we would be able to detect those genetic variants by genetic association tests with the Illumina ExomeChip.

Objective 1. We aim to test the association of genetic variants in the *CKM* and the *LILRB5* genes with plasma CK levels in order to replicate the results of a prior study where those genes were identified.

Objective 2. We aim to search for novel genetic variants in an exome-wide discovery study. The study will consist of two major parts:

1. GWAS: This will look at common genetic variants (MAF>5%)
2. SKAT: This will look at rare genetic variants (MAF<5%)

9 METHODS

9.1 *Study Design*

The goal of this study is to provide information that may, in the future, help to better clinical statin care by improving the expected CK level algorithm for a patient. Along with covariates already known to affect CK levels, genetic factors and the role they played were looked at that may help to ultimately create a new method for calculating expected CK levels of a patient. In doing this, one can hope to better future statin care by creating a more realistic threshold for CK levels and decreasing the number of patients taken off statins unnecessarily.

This study is a cohort study for a continuous trait. Data on participants was collected from the MHI Biobank, including both descriptive statistics and genetic data. The study was limited to those who had an available CK measure and participants were stratified by three subgroups: statin users, non statin users and all participants. The study consists of two phases: **a replication study using 17 variants from the *CKM* and *LILRB5* genes,** both areas identified in a previous study conducted at the Pharmacogenomics center followed by a discovery study. **The discovery study was genome wide, genotyping 584,509 SNPs, consisting of both a GWAS for common variants and a SKAT burden test for rare variants (MAF < 5%).**

9.2 *Study population*

The MHI Hospital Cohort participants came from the MHI Biobank. This is a longitudinal cohort study led by Dr. Jean-Claude Tardif which was initiated in 2005 with the aim to recruit 30,000 patients of the MHI for clinical and genetic research. Participants are recruited from different departments within the MHI and its affiliated EPIC centre; the largest fitness centre in Canada

for coronary patients and research in primary and secondary prevention. The MHI Cohort collects data by using a 35-page questionnaire administered by a research nurse at baseline and questions about demographics, personal and family medical history, physical activity, diet, tobacco, medication use, as well as an additional depression and hostility questionnaire. Vital signs (heart rate, blood pressure, weight, height, waist circumference) are obtained by the nurse and blood, DNA, and plasma are collected and stored at the Beaulieu-Saucier Pharmacogenomics Center. Patient's health status is confirmed by the nurse by using the Hospital's health record and with governmental databases (RAMQ, MedEcho) for retrospective and prospective follow up until patient's death. The Cohort's database is updated daily with patient's medical information from the hospital's electronic records and a follow-up study questionnaire is administered every three years. The MHI Biobank comprises of 17,000 participants as of August 2012 with a mean follow up period of 2.8 years including over 5000 participants who are using or have used a statin during the follow-up period.

The association of genetic factors with plasma CK levels in both the replication (objective 1) and discovery studies (objective 2) was investigated in a subset of the MHI Biobank. The MHI Biobank is a large hospital-based longitudinal population of over 11,000 patients with a wide spectrum of cardiovascular diseases who have recently had their DNA genotyped using the ExomeChip from Illumina. The MHI Biobank was initiated in 2005 with the aim of clinical and genetic research.

In the present study, a sub-cohort of the original Biobank consisting of 7821 participants with available genotyping and CK levels was chosen of which the data of 5875 participants had

usable CK levels. All participants that had a creatinine level exceeding 200 mmol/L (n=85) taken within a month of the CK measurement used were excluded from the study. The reasoning behind this is that patients with severe renal impairment or muscle disease not related to statins should be removed from the study. To control for this, patients with a creatinine level greater than 200 umol/L were removed. Additional outliers were also removed using jack-knife residuals. Therefore, discovery study was performed on 5875 available participants including 3677 on a statin and 2198 not on a statin.

9.3 ***Creatine kinase***

Creatine kinase data collection. From the cohort, we used a subset of participants with available plasma CK levels and genetic data. CK is an enzyme and can be assayed from the blood serum with a detection kit. This was done in the MHI laboratory with a blood sample taken from each participant.

Creatine kinase measurements. The most recent available CK measurement was used, using a selection of total CK measures (variables CKi1 and CK1 from softlab, n=43518). We have opted to exclude all measurements that were requested for hospitalized patients and emergency patients in order to exclude CK measurements. This is because measurements would be inflated by elevation of CK-MB, which is typically the case in patients who have suffered a recent myocardial infarction. Specifically, measurements requested from the following clinics were excluded:

Code	Clinic
3EST	3ieme est

4EST	4ieme est
5EST	5ieme est
3CEN	3ieme centre
PSERP	projet serp-1
PBARI	projet barath
PPRIM	projet primo
ANDG	introuvable
PPGXS	génomie
PREPA	projet repare
URG	urgence
CLDIA	cl.dialyse
4CEN	4ieme centre
URGAM	urgence ambulatoire
5CEN	5ieme centre
CLPGR	clinique de pré greffe

9.4 ***Statins***

For patients taking a statin, their dosage is related to the type of statin used. The equivalent dosage to atorvastatin was calculated for each participant in order to allow for comparisons between patients on applicable statins. The statin dosage was transformed to the equivalent dosage as shown in table III.

Different statins use varying pathways, and therefore different statins affect statin induced muscle adverse effects differently. **We used one all-inclusive category of statin, instead of stratifying by each statin type, which could be a limitation of our study.** The majority of participants in the MHI cohort with available CK levels used atorvastatin. It has been found that the most hydrophilic statins, such as pravastatin and fluvastatin, are less likely to cause myalgia, while lipophilicone statins such as simvastatin are more likely to induce muscle adverse effects[85]. We did not stratify the users by statin type because most of the subsets were too small to preserve power.

Table III. Table of equivalent dosage of statins

Atorvastatin equivalent	**Atorva-statin**	**Fluva-statin**	**Lova-statin**	**Prava-statin**	**Rosuva-statin**	**Simva-statin**
5 mg	--	40 mg	20 mg	20 mg	--	10 mg
10 mg	10 mg	80 mg	40 mg	40 mg	--	20 mg
20 mg	20 mg	--	80 mg	80 mg	5 mg	40 mg
40 mg	40 mg	--	--	--	10 mg	80 mg
80 mg	80 mg	--	--	--	20 mg	--
160 mg	--	--	--	--	40 mg	--

9.5 ***Physical activity***

Muscle damage, often accompanied with a spike in CK levels, is commonplace after intense physical activity. The increase in CK levels that occurs after physical activity varies in relation to the type (low impact verse high impact) and the intensity of exercise. Participants in the MHI cohort completed a survey of their physical activity which was based off the World Health Organization (WHO) guidelines. The two surveys differed only in that **the MHI Cohort survey did not take into account unpaid labour, such as gardening and household chores, that are commonly accepted as physical activity.** This is an expected pitfall of the survey, as we have no way of adjusting for this and may lose some accuracy in the analysis. The MHI Cohort survey is being amended to adjust for this in future studies.

In order to clean up data from the survey and split the participants into three classes of physical activity, guidelines from the WHO survey were followed. The procedure included calculating the metabolic minutes (a measure to combine the counts of physical and moderate activity) of physical activity. Using the metabolic minutes and the days of activity per week, the participants were split into three classes of physical activity: elevated, moderate or limited.

9.6 ***Ethnicity and principal components***

In the MHI Hospital Cohort survey, participants were asked to choose between Caucasian, Hispanic, Black, Native American, Asian or other as their principal ethnicity. The vast majority of participants identified as Caucasian. Although self-reported ancestry is sometimes a viable way to stratify data in a large cohort[86], it can have limited reliability, especially when an individual has more than one country of origin[87]. As the vast majority of participants identified as Caucasian, principal component analysis (PCA) was used. In a GWAS, a small percentage of

the SNPs can provide useful information on a participant's ancestry. These ancestry informative markers can be used in a PCA analysis to help correct for any ancestry discrepancies[88].

9.7 ***Hardy-Weinberg equilibrium and linkage disequilibrium***

If a population is in Hardy-Weinberg equilibrium, allele and phenotype frequency will remain stable throughout generations, assuming there is no outside evolutionary influence. It is important to look at Hardy-Weinberg equilibrium when genotyping, as deviation from equilibrium may indicate genotyping errors, population stratification or, in rare cases, true association[89]. A goodness-of-fit test was performed on SAS 9.3 to verify that **all variants were in Hardy-Weinberg equilibrium**.

Linkage disequilibrium is the non-random association of alleles at two or more loci. When two alleles are located near each other and are inherited together more often than expected, this association is called linkage disequilibrium[90]. In a traditional GWAS such as this, the one-at-a-time SNP method increases the chances of randomness seen in associated SNPs, and can potentially increase the number of false positives in the study. GWAS studies can help to identify loci where causal relationships are suspected, but associated SNPs should be cautiously verified due to this increase in false positives. **Linkage disequilibrium was looked at using a HapMap from haploview.**

9.8 ***Genetic data and cleanup***

All patients were genotyped using the IlluminaExomeChip array (version InfiniumHumanExome v1.0 DNA Analysis BeadChip), that tests for over 270,000 genome-wide variants located in the

exome of genes. Whole genome analysis was performed with 620,901 SNPs using the IlluminaInfinium HD Assay and the Human 610quad BeadChips (Illumina, San Diego, CA).

Complementing this, a Multi-Sample iSelect Custom BeadChip was used, containing 11,568 SNPs from candidate genes involved in lipid homeostasis, hypertension and drug metabolism. Control samples provided 100% consistency. Following genotype quality procedures and removal of 65 non-Caucasian samples, there remained 584,509 SNPs for analysis. All genotyping was performed at the Beaulieu-Saucier Pharmacogenomics Centre.

The cleanup of genetic data was done at the Beaulieu-Saucier Pharmacogenomics Centre at the MHI, using the following guidelines:

- Checked for duplicate samples, then for duplicated SNPs.
- Completely failed SNPs were removed.
- Poorly performing DNA samples with a genome-wide genotyping success rate < 90% were removed, followed by SNPs with a genotyping success rate < 98%, and finally by removing IDs with less than 98% successful call rates.
- All individuals with discrepancies between phenotype and genotype data for sex were removed.
- SNPs with genotyping plate-bias were filtered out.
- 610quad and iselect data sets were merged.
- Pairwise IBD was used to conduct close familial relationship checks, and all but one pair-member of sibling pairs and sample duplicated (IBS > .86 or IBS > 98%) were removed. This was based on a selection of uncorrelated SNPs ($r^2 < 0.1$).

- The pairwise IBS matrix was used as a distance metric to identify cryptic relatedness among samples and sample outliers by multi-dimensional scaling (MDS). The first two MDS components of each subject have been plotted. In order to scale the outlying distance, the genotypes were merged with those of HapMap CEU data (keeping only founder individuals) prior to calculating the IBS matrix and the MDS.
- For each SNP, Hardy-Weinberg equilibrium was tested for, and SNPs with p-value < $8x10^{-8}$ were removed. SNPs with p-value < 10^{-4} were flagged.
- All individuals with no sex in the phenotype file or heterozygous haploid genotypes were removed.

9.9 ***Summary of available variants for replication study***

In the prior project to be replicated, there were significantly associated variants found in the *CKM* and *LILRB5* genes. As a result, all available variants from the *CKM* and *LILRB5* genes in the ExomeChip were used. The build number for all genes in the ExomeChip is 37. There were 12 available variants in the CKM gene and 20 available in the LILRB5 gene. In the CKM gene, rs11559024 was found to be significantly associated to plasma CK levels in the prior project.

Table IV. Table of variants in CKM and LILRB5 genes available for replication study

Chr	Position (build 37)	Gene	SNP	Intron/ exon	Minor Allele Frequency (MAF)*	minor allele *	major allele *
19	45810010	CKM	rs142092440	exonic	0.001143	G	A

19	45810035	CKM	rs4884	exonic	0.2892	A	G
19	45810108	CKM	rs199913939	exonic	0	0	A
19	45810862	CKM	rs200633759	exonic	0	0	A
19	45815163	CKM	rs17357122	exonic	0.008429	A	G
19	45818628	CKM	rs377993	intronic	0.4742	A	G
19	45818750	CKM	rs149872015	exonic	0.000286	A	G
19	45818809	CKM	rs147574145	exonic	0.000286	A	G
19	45818825	CKM	rs17875653	exonic	0.002999	C	G
19	45821183	CKM	rs11559024	exonic	0.01095	G	A
19	45822809	CKM	rs201048164	exonic	0.00019	A	G
19	45822882	CKM	rs145633772	exonic	0.000333	A	C
19	54754667	LILRB5	rs150133483	exonic	0.000143	A	C
19	54755937	LILRB5	rs144185169	exonic	0	0	G
19	54755955	LILRB5	rs3848614	exonic	0.000907	G	C
19	54756246	LILRB5	rs117421142	exonic	0.02247	G	A
19	54756401	LILRB5	rs139593424	exonic	0.000476	A	G
19	54756735	LILRB5	rs143517976	exonic	0.000048	A	G
19	54756767	LILRB5	rs201004751	exonic	0.000095	A	G
19	54758762	LILRB5	rs149797743	exonic	0	0	G
19	54758810	LILRB5	rs201168501	exonic	0.000048	G	A
19	54759306	LILRB5	rs139291477	exonic	0.000238	C	G
19	54759349	LILRB5	rs199626075	exonic	0	0	A
19	54759361	LILRB5	rs12975366	exonic	0.4355	G	A

19	54759395	LILRB5	rs145597773	exonic	0.000381	A	G
19	54759931	LILRB5	rs199539096	exonic	0.000381	C	A
19	54760105	LILRB5	rs200687007	exonic	0.00019	A	G
19	54760189	LILRB5	rs150247609	exonic	0.00019	A	G
19	54760376	LILRB5	rs144666082	exonic	0.000238	A	C
19	54760381	LILRB5	rs149664511	exonic	0.002713	A	G
19	54760453	LILRB5	rs191146326	exonic	0	0	A
19	54761023	LILRB5	rs150778096	exonic	0.001	G	C

* MAF of MHI Hospital Cohort, calculated in PLINK

9.10 ***Statistical analysis***

Statistical analyses were performed using SAS 9.3 (SAS Institute Inc., Cary, NC) and PLINK v1.07 (http://pngu.mgh.harvard.edu/~plink). The Shapiro Wilk test was used to check the normality of CK measures. Additionally, a goodness-of-fit test was performed to verify that **all variants were in Hardy-Weinberg equilibrium** and linkage disequilibrium (LD) was looked at using a HapMap from haploview.

9.11 ***Regression model for replication study***

The regression model used in this analysis was based on that of the prior project. The replication project used the same covariates as the prior project done at the Pharmacaogenomcis lab while the discovery study left room for exploration of covariates. A general linear model (GLM) was constructed between the CK level and the additive effect x (Equation 1). Furthermore, a GLM with adjustment for sex, statin, age and activity levels (Equation 2) was constructed. The analysis was done on SAS 9.3.

$$\ln(CK\ level) = \beta_0 + \beta_1 x \quad \text{[Equation 1]}$$

$$\ln(CK\ level) = \beta_0 + \beta_1 x + \beta_2 \text{gender} + \beta_3 \text{statin} + \beta_4 \text{age} + \beta_5 \text{activity levels} \quad \text{[Equation 2]}$$

Where x is the additive effect for each individual, β_0 corresponds to the mean effect and β_1, β_2, β_3, $\beta_4$and β_5 are the coefficients of x, gender, statin, age and activity levels respectively.

Note that the variables x is defined for the additive effect as seen below:

$$x = \begin{cases} 1 & \text{if the SNP has two major alleles,} \\ 0 & \text{if the SNP has one major allele and one minor allele,} \\ -1 & \text{if the SNP has 2 minor alleles.} \end{cases}$$

The data was stratified by three subgroups: all participants, statin users and non-statin users. Linear regression was run on each of these subgroups twice, once with only the additive model as an independent variable and then a multivariate analysis with gender, age, activity levels and statin dose (where applicable) as covariates.

The frequencies of rare alleles among the significantly associated variants were looked at.

9.12 ***Statistical analysis of multiple variants at a time for replication study***

This analysis looked at each associated variant, two or three at a time, to check for linkage equilibrium. Linkage equilibrium occurs when the genotype present at one locus is independent of the genotype at a second locus. A regression model was constructed, and if the two or three variants together were relatively similar as the model where only one variant was looked at, the variants were said to not be in linkage equilibrium, and we could consider them to be independent for this analysis. The model was run using SAS 9.3. The following equations were used:

$$ln(CK\ level) = \beta_0 + \beta_1 x\_rs11559024$$

$$+ \beta_2 x\_rs12975366$$

[Equation 3]

$$ln(CK\ level) = \beta_0 + \beta_1 x\_rs11559024$$

$$+ \beta_2 x\_rs12975366$$

$$+\beta_3 gender + \beta_4 statin + \beta_5 age + \beta_6 activity\ levels$$

[Equation 4]

$$ln(CK\ level) = \beta_0 + \beta_1 x\_rs11559024$$

$$+ \beta_2 x\_rs142092440$$

[Equation 5]

$$ln(CK\ level) = \beta_0 + \beta_1 x\_rs11559024$$

$$+ \beta_2 x\_rs142092440 + \beta_3 gender + \beta_4 statin + \beta_5 age + \beta_6 activity\ levels$$

[Equation 6]

$$ln(CK\ level) = \beta_0 + \beta_1 x\_rs11559024$$

$$+ \beta_2 x\_rs142092440$$

$$+ \beta_3 x\_rs12975366$$

[Equation 7]

$$ln(CK\ level) = \beta_0 + \beta_1 x\_rs11559024$$

$$+ \beta_2 x\_rs142092440$$

$$+ \beta_3 x\_rs12975366$$

$$+ \beta_4 gender + \beta_5 statin + \beta_6 age + \beta_7 activity\ levels$$

[Equation 8]

9.13 ***Imputation analysis for replication study***

An imputation analysis allows for the accurate evaluation of association of genetic markers that are not directly genotyped[91]. **Since variant rs2361797 (BP 54751064 build37) from the *LILRB5* gene in chromosome 19 was identified in a previous study but not directly genotyped in this one, we aimed to impute data for this variant in the hopes of validating prior findings.** Imputation of common variants is extremely accurate, and allows for the combination of two or more data sets, which often analyze different genetic variants. According to protocol, we only used SNPs with a completion rate greater than 98%. The imputation analysis was done using PLINK v1.07.

9.14 ***GWAS for discovery study***

When used correctly, GWAS is a powerful tool used in identifying variants associated with disease in humans. In a typical GWAS, thousands of samples and their hundreds of thousands of SNP markers throughout the human genome are scanned. Since a GWAS scans the entire genome, there is the possibility of finding novel susceptibility factors, as opposed to candidate gene studies which only look at variants either taken at random or with a suspected association to the disease.

An analysis for principal components one and two was performed to try and account for ethnicity in the plasma CK levels. Ethnicity is known to be a strong confounder in genetic association studies. As there is a well-known difference in CK levels in some ethnic populations, this is relevant in the present study. This said, one must be careful as it has been observed that when studying rare SNPs, covariates can sometimes decrease power of the analysis[92]. Certain SNPs already account for factors such as gender and ethnicity, so when a covariate is added, they may be double counted for, resulting in a less powerful test overall. We performed the GWAS with principal components one and two as covariates in the hopes of increasing power for rare variants and minimizing false positives.

When using all available variants, the QQ from the GWAS were not within the expected range, insinuating inflated results. This may be because the rare variants were not being properly accounted for in the algorithm. To adjust for this we compared the QQ plots of all available SNPs with QQ plots using just common SNPs that had an allele frequency greater than 5%. Through a comparison of variants with different MAFs, we found that the QQ plots were within an expected range when only looking at common variants, increasing our confidence in the Manhattan plots shown in the results section. Thus we used all variants with an allele frequency greater than 5% for the GWAS and opted to use a SKAT burden test for the rare variants. This method allowed for better management of inflation, as demonstrated in the QQ plots seen in the results section.

The GWAS was done using PLINK v1.07.

9.15 ***SKAT for discovery study***

In order to find association of rare variants (defined in this study as less than 5% allele frequency) in a GWAS, the sample size must often be very large. Burden tests can be used to summarize the rare variants in a certain area and find association within a locus. These tests have been shown to be more powerful than the univariate test used in GWAS when looking at rare variants[93]. In a burden test, all rare variants in a given area are summarized by a single value, and association for this is tested. This must be implemented cautiously, as it can be expected that summarizing variants will increase noise in an analysis.

SKAT calculates a p-value for association of rare variants in a region of a phenotype, while adjusting for any covariates specified. SKAT has been found to be more powerful than many other burden tests when analyzing rare variants genome wide[94].

The SKAT analysis was done using PLINK v1.07.

10 RESULTS

10.1 ***Participants***

A sub-cohort was chosen from the MHI Biobank consisting of participants with available genotyping and CK levels. After exclusions, there were 5875 available participants, 3677 on a statin and 2198 not on a statin. Participants were excluded if they had a creatinine level exceeding 200 mmol/L and then further exclusions were made based on an analysis of studentized residuals.

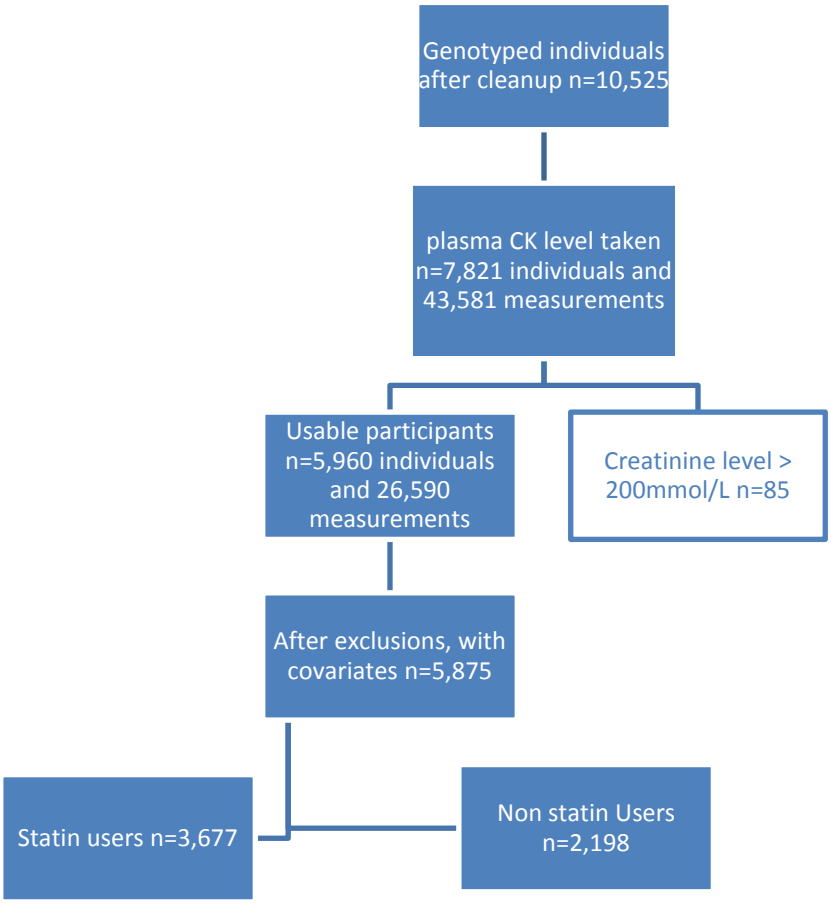

Figure 2. Flowchart of participants from MHI Biobank that tracks where participants were excluded

10.2 ***Descriptive statistics from MHI Biobank***

Frequency and percentages were calculated for categorical variables while mean, median, standard deviation and ranges were looked at for continuous variables. Only participants with an available CK level were used.

Table V**.** Descriptive statistics of participants from the MHI Biobank after exclusions

		All	**Statin Users**	**Non-Statin Users**
		n=5875	**n=3677**	**n=2198**
Age (years)	Mean (Std)	68.38 (±10.59)	70.55 (±8.81)	64.76 (±12.21)
Men	n (%)	3959 (67%)	2718 (74%)	1241 (56%)
Race	n (%)			
Caucasian		5773 (98.4%)	3624 (98.7%)	2149 (97.9%)
Hispanic		30 (0.5%)	15 (0.4%)	15 (0.7%)
Black		36 (0.6%)	18 (0.5%)	18 (0.8%)
Asian		16 (0.3%)	8 (0.2%)	8 (0.4%)
Native American		0 (0%)	0 (0%)	0 (0%)
Other		10 (0.2%)	5 (0.1%)	5 (0.2%)
Creatine kinase (U/L)	Mean (Std)	106.38 (±89.37)	106.77 (±80.70)	105.74 (±102.25)
	Min/Max	7/2120	7/2120	8/2120
ln(Creatine kinase)	Mean (Std)	4.47 (±.60)	4.49 (±.59)	4.44 (±.61)
	Min/Max	1.95/7.66	1.95/7.47	2.08/7.66
Physical activity	n (%)			
Low		1734 (41.3%)	1133 (40.1%)	601 (43.8%)
Moderate		1699 (40.4%)	1178 (41.7%)	521 (37.9%)
Elevated		768 (18.3%)	517 (18.3%)	251 (18.3%)

*13 participants identified with two ethnicities, and were categorized as other

Table VI**.** Overview of the 3516 patients currently taking statins, with statin dosage available, and with a creatinine level under 200 mmol/L

		Dosage (mg)	**Atorvastatin Equivalent (mg)**
	n (%)	Mean	Mean
Current statin			
Atorvastatin	2113 (57%)	34.50	6.46
Fluvastatin	36 (1%)	51.67	5.02
Lovastatin	9 (0%)	20.06	7.67
Pravastatin	274 (7%)	30.67	72.40
Rosuvastatin	892 (24%)	18.10	15.12
Simvastatin	353 (10%)	30.23	6.46

Ethnicity and principal components. Principal components were used to control for ethnic discrepancies between individuals. As seen in the graphs below, by using just the first two principal components one can reduce discrepancies between self-reported questionnaires and genetic ethnicity.

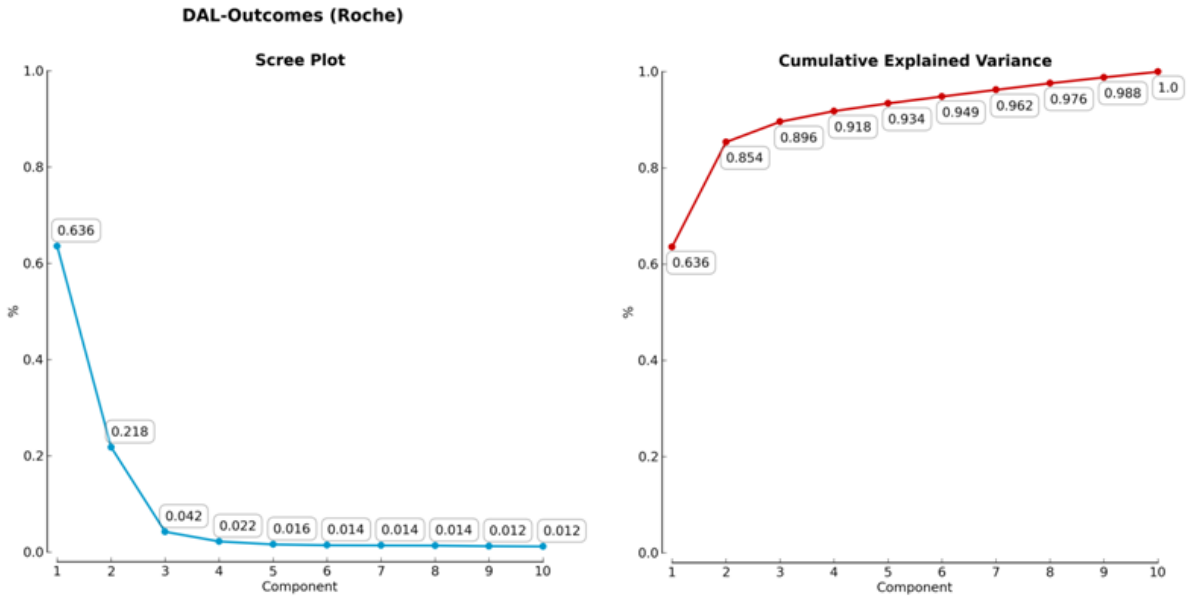

Figure 3. The principal components from the MHI cohort data using PCA as a proxy for ethnicity. Charts, from left to right, represent DAL-Outcomes (Roche) and proportional cumulative explained variance of the first 10 components. The third principal component seems to be an inflection point on the screen plot, so in further analysis only the first two principal components were used.

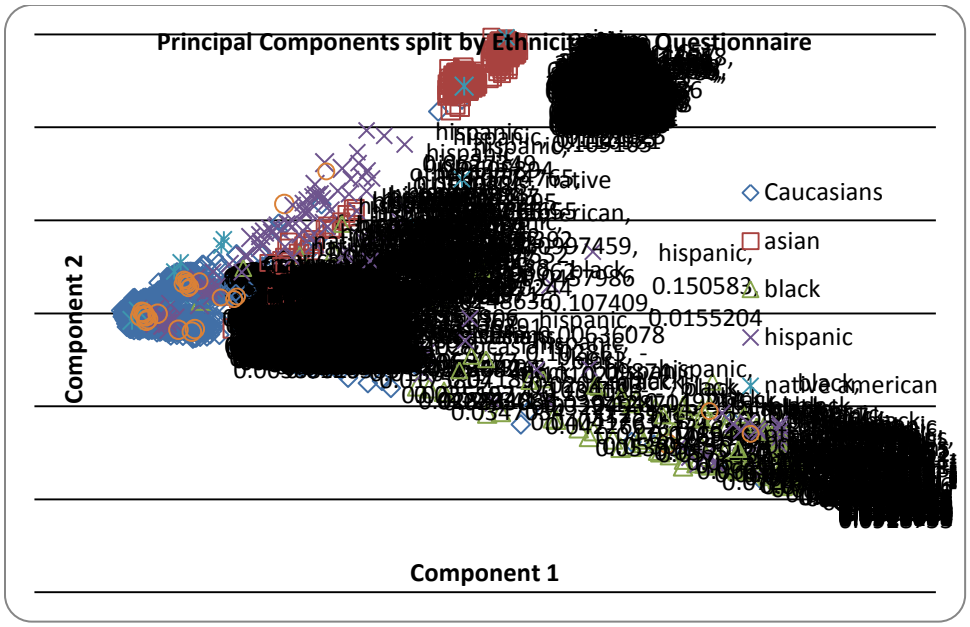

Figure 4. Principal components 1 and 2, marked by self reported ethnicity

11 RESULTS FOR REPLICATION STUDY - Objective 1

11.1 *Results of replication model*

There are 17 SNPs available for testing, so the **Bonferroni-adjusted significance threshold was set to 0.05/17=0.002941**.

There are 5091 genotyped participants available for analysis after exclusions. In the discovery study there were 5875 participants. The replication study was done before all the participants were genotyped, explaining the smaller population size.

Table VIIa. Association analysis between log(CK levels) and SNPs for **all participants** in the MHI Hospital Cohort. The multivariate analysis with a general linear model of CK is performed first with **additive effect only**. P-values surpassing the Bonferoni threshold are highlighted.

	Position			**No ajustement**		
Chr	**(build 37)**	**Gene**	**SNP**	**N**	**p-value**	**R-squared***
19	45810010	CKM	rs142092440	5091	0.000156	0.002805
19	45810035	CKM	rs4884	5091	0.73213	0.000023
19	45815163	CKM	rs17357122	5091	0.008329	0.001367
19	45818628	CKM	rs377993	5091	0.446943	0.000114
19	45818750	CKM	rs149872015	5091	0.167151	0.000375
19	45818809	CKM	rs147574145	5091	0.656476	0.000039
19	45818825	CKM	rs17875653	5091	0.573712	0.000062
19	45821183	CKM	rs11559024	5091	8.05×10^{-19}	0.015306
19	54755955	LILRB5	rs3848614	5091	0.71948	0.000025
19	54756246	LILRB5	rs117421142	5091	0.425831	0.000125
19	54756401	LILRB5	rs139593424	5091	0.178514	0.000356
19	54759306	LILRB5	rs139291477	5091	0.810846	0.000011
19	54759361	LILRB5	rs12975366	5091	4.28×10^{-21}	0.017307
19	54759395	LILRB5	rs145597773	5091	0.531165	0.000077
19	54759931	LILRB5	rs199539096	5091	0.983371	0.000085
19	54760381	LILRB5	rs149664511	5091	0.883995	0.0000042
19	54761023	LILRB5	rs150778096	5091	0.324039	0.000191

*R-squared is the coefficient of determination for the regression model including the genetic term only

VIIb. Association analysis between log(CK levels) and SNPs for **all participants** in the MHI Hospital Cohort. The multivariate analysis with a general linear model of CK is performed first with **additive effect, gender, age and physical activity** (two adjusted versions, one with statin yes/no and one with equivalent dosage of statins). P-values surpassing the Bonferoni threshold are highlighted.

		Adjusted GLM (statin yes/no)				Adjusted GLM (statin equivdos)			
Gene	**SNP**	**N**	**Global p-value**	**p-value**	**R-squared**	**N**	**Global p-value**	**p-value**	**R-Squared**
CKM	rs142092440	3960	$1.87x10^{-80}$	$5.06824x10^{-06}$	0.093182	3960	$4.05x10^{-80}$	$6.04301x10^{-06}$	0.092825
CKM	rs4884	3960	$5.33x10^{-76}$	0.760151119	0.088416	3960	$9.62x10^{-76}$	0.730986936	0.088141
CKM	rs17357122	3960	$1.00x10^{-76}$	0.062563017	0.089194	3960	$1.88x10^{-76}$	0.064443715	0.088902
CKM	rs377993	3960	$5.40x10^{-76}$	0.792186263	0.088411	3960	$9.78x10^{-76}$	0.770811158	0.088133
CKM	rs149872015	3960	$1.22x10^{-76}$	0.079296937	0.089105	3960	$2.28x10^{-76}$	0.081909345	0.088812
CKM	rs147574145	3960	$5.34x10^{-76}$	0.765684026	0.088415	3960	$9.82x10^{-76}$	0.780829741	0.088132
CKM	rs17875653	3960	$2.91x10^{-76}$	0.251484862	0.088698	3960	$5.52x10^{-76}$	0.265340131	0.0884
CKM	rs11559024	3960	$8.10x10^{-91}$	$9.95869x10^{-17}$	0.104166	3960	$1.29x10^{-90}$	$8.64888x10^{-17}$	0.103953
LILRB5	rs3848614	3960	$4.44x10^{-76}$	0.495481621	0.088502	3960	$8.22x10^{-76}$	0.508639223	0.088215
LILRB5	rs117421142	3960	$3.98x10^{-76}$	0.408770468	0.088552	3960	$7.23x10^{-76}$	0.404392051	0.088274
LILRB5	rs139593424	3960	$2.83x10^{-76}$	0.241041739	0.088712	3960	$5.17x10^{-76}$	0.241457728	0.08843
LILRB5	rs139291477	3960	$5.24x10^{-76}$	0.719233949	0.088425	3960	$9.72x10^{-76}$	0.754801237	0.088136
LILRB5	rs12975366	3960	$3.50x10^{-92}$	$4.00975x10^{-18}$	0.105601	3960	$4.36x10^{-92}$	$2.71181x10^{-18}$	0.1055
LILRB5	rs145597773	3960	$3.83x10^{-76}$	0.381955307	0.088571	3960	$7.11x10^{-76}$	0.392817672	0.088282
LILRB5	rs199539096	3960	$5.55x10^{-76}$	0.916017295	0.088397	3960	$1.01x10^{-75}$	0.907607254	0.088117
LILRB5	rs149664511	3960	$2.52x10^{-76}$	0.204241401	0.088766	3960	4.63x10-$^{-76}$	0.20644526	0.088482
LILRB5	rs150778096	3960	$4.46x10^{-76}$	0.501270815	0.088499	3960	$8.18x10^{-76}$	0.504128214	0.088217

*R-squared is the coefficient of determination for the regression model including the genetic term, sex, physical activity, age and statin dose (either yes/no or atorvastatin equvilant dosage)

Table VIII**.** Association analysis between log(CK levels) and SNPs for participants in the MHI Hospital Cohort **taking statins**. The multivariate analysis with a general linear model of CK is performed first with **additive effect only** and second with **additive effect, gender, statin dose, age and physical activity**. P-values surpassing the Bonferoni threshold are highlighted.

				No adjustment			Adjusted GLM			
Chr	**Position (build 37)**	**Gene**	**SNP**	**N**	**p-value**	**R-squared**	**N**	**Global p-value**	**p-value**	**R-Squared**
19	45810010	CKM	rs142092440	3446	0.001217	0.003034	2673	6.43×10^{-47}	0.000472	0.082719
19	45810035	CKM	rs4884	3446	0.828302	0.000014	2673	2.58×10^{-44}	0.865465	0.078512
19	45815163	CKM	rs17357122	3446	0.009412	0.001956	2673	2.36×10^{-45}	0.026989	0.080192
19	45818628	CKM	rs377993	3446	0.129494	0.000667	2673	1.62×10^{-44}	0.322431	0.078841
19	45818750	CKM	rs149872015	3446	0.149883	0.000602	2673	5.70×10^{-45}	0.078273	0.079574
19	45818809	CKM	rs147574145	3446	0.615888	0.000073	2673	2.49×10^{-44}	0.755146	0.078536
19	45818825	CKM	rs17875653	3446	0.469796	0.000152	2673	1.19×10^{-44}	0.206306	0.079055
19	45821183	CKM	rs11559024	3446	4.19×10^{-15}	0.017733	2673	2.40×10^{-54}	7.25×10^{-12}	0.094594
19	54755955	LILRB5	rs3848614	3446	0.944248	0.0000014	2673	2.06×10^{-44}	0.486798	0.07867
19	54756246	LILRB5	rs117421142	3446	0.962001	0.00000066	2673	1.89×10^{-44}	0.41625	0.078731
19	54756401	LILRB5	rs139593424	3446	0.218915	0.000439	2673	1.66×10^{-44}	0.336723	0.078822
19	54759306	LILRB5	rs139291477	3446	0.769741	0.000025	2673	2.22×10^{-44}	0.564758	0.078617
19	54759361	LILRB5	rs12975366	3446	9.39×10^{-14}	0.015986	2673	5.97×10^{-56}	1.6×10^{-13}	0.097135
19	54759395	LILRB5	rs145597773	3446	0.479128	0.000145	2673	1.95×10^{-44}	0.439394	0.078709
19	54759931	LILRB5	rs199539096	3446	0.932539	0.0000021	2673	2.57×10^{-44}	0.852214	0.078515
19	54760381	LILRB5	rs149664511	3446	0.524918	0.000117	2673	2.02×10^{-44}	0.469468	0.078683
19	54761023	LILRB5	rs150778096	3446	0.961878	0.00000066	2673	2.60×10^{-44}	0.899056	0.078508

*The first R-squared is the coefficient of determination for the regression model including the genetic term, and the second is including the genetic term, sex, physical activity, age and statin atorvastatin equvilant dosage

Table IX. Association analysis between log(CK levels) and SNPs for participants in the MHI Hospital Cohort **not taking statins**. The multivariate analysis with a general linear model of CK is performed first **with additive effect only** and second with **additive effect, gender, age and physical activity**. P-values surpassing the Bonferoni threshold are highlighted.

				No adjustment			Adjusted GLM			
Chr	**Position (build 37)**	**Gene**	**SNP**	**N**	**p-value**	**R-squared***	**N**	**Global p-value**	**p-value**	**R-Squared***
19	45810010	CKM	rs142092440	1642	0.035658	0.002688	1287	8.75×10^{-35}	0.003283	0.123893
19	45810035	CKM	rs4884	1642	0.882256	0.000013	1287	5.94×10^{-33}	0.794537	0.118006
19	45815163	CKM	rs17357122	1642	0.352934	0.000526	1287	5.95×10^{-33}	0.799412	0.118004
19	45818628	CKM	rs377993	1642	0.368015	0.000494	1287	3.99×10^{-33}	0.349906	0.118561
19	45818825	CKM	rs17875653	1642	0.997037	0.0000000084	1287	5.98×10^{-33}	0.817759	0.117996
19	45821183	CKM	rs11559024	1642	1.48×10^{-05}	0.011375	1287	8.47×10^{-38}	1.86×10^{-06}	0.133484
19	54755955	LILRB5	rs3848614	1642	0.587221	0.00018	1287	6.04×10^{-33}	0.856411	0.117982
19	54756246	LILRB5	rs117421142	1642	0.187937	0.001057	1287	6.06×10^{-33}	0.86995	0.117978
19	54756401	LILRB5	rs139593424	1642	0.519905	0.000253	1287	3.68×10^{-33}	0.307514	0.118676
19	54759361	LILRB5	rs12975366	1642	7.65×10^{-09}	0.020143	1287	4.20×10^{-37}	1.01×10^{-05}	0.131282
19	54759931	LILRB5	rs199539096	1642	0.899592	0.0000097	1287	8.76×10^{-34}		0.117959
19	54760381	LILRB5	rs149664511	1642	0.220883	0.000914	1287	3.05×10^{-33}	0.232979	0.118939
19	54761023	LILRB5	rs150778096	1642	0.134263	0.001367	1287	2.19×10^{-33}	0.147969	0.1194

rs149872015, exm1480276, rs147574145 and rs145597773 were removed since no minor alleles were represented in the data when only patients not on statin were used. rs199539096 does not have a p-value for adjusted model because no minor alleles were represented in the data.

*The first R-squared is the coefficient of determination for the regression model including the genetic term, and the second is including the genetic term, sex, physical activity and age

Table X. Mean serum CK values by genotype of the associated variants in the *CKM* and *LILRB5* genes in MHI Hospital Cohort

Gene	Variant	Genotype	Statin users (n=3448)		Statin non-users (n=1465)		All participants (n=5093)	
			N	Mean CK (U/L)	N	Mean CK (U/L)	N	Mean CK (U/L)
CKM	rs142092440	GG	0	N/A	0	N/A	0	N/A
		AG	10	52.70±4.38	3	40.33±1.20	13	49.85±19.51
		AA	3438	106.09±81.31	1642	103.36±103.01	5080	105.21±88.91
CKM	rs11559024	GG	0	N/A	0	N/A	0	N/A
		AG	66	59.24±13.47	42	66.86±8.86	108	62.20±40.70
		AA	3382	106.84±81.62	1603	104.20±103.85	4985	105.99±89.37
LILRB5	rs12975366	GG	661	93.44±65.03	315	84.17±66.88	976	90.45±0.57
		AG	1714	103.40±74.76	811	103.08±58.13	2525	103.30±0.59
		AA	1073	117.67±96.15	519	115.10±127.50	1592	116.83±0.61

Mean ± standard deviation; N/A: not available.

The general trend seen in table X is that **as the number of rare alleles increase, the mean CK level decreases**. This holds true for all four variants. However **for SNP rs12975366 in the *LILRB5* gene, the populations mean CK is best represented by the heterozygous genotype**, with homozygotes showing either an increase or a decrease in CK. In this SNP, occurrences of the minor allele are common. Over 50% of the population has at least one minor allele. The participants who do not have either of these minor alleles tend to have above average plasma CK levels. The overall average level of CK in the 5093 patients is 105.07 U/L.

11.2 ***Linkage disequilibrium analysis***

Haploview was used to look at linkage disequilibrium for the variants used in the replication study.

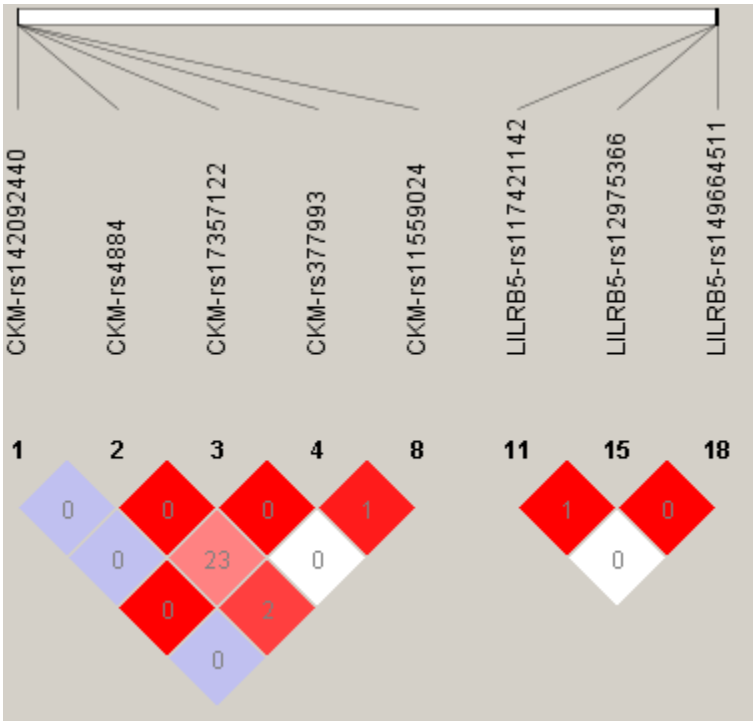

Figure 5. Linkage disequilibrium of available SNPs in the *CKM* and *LILRB5* gene obtained from Haploview. The above chart looks at the r^2 between variants of the 5093 participants from the MHI Hospital Cohort. **All SNPs are in Hardy-Weinberg equilibrium. Each two different SNPs are in linkage disequilibrium.**

11.3 ***Statistical analysis of multiple variants***

The following analysis aims to find whether two SNPs are independent. Since there are SNPs from both the *LILRB5* gene and *CKM* gene, it is hypothesized that these two SNPs will exhibit independence. All equations used are shown in section 9.12.

As seen in tables XI and XII, the most significant association between CK measures and the two SNPs is in all participants, then statin users, and lastly non-statin users. The R-squared is slightly higher in non statin users than statin users. This follows the same general trend as in tables VII – IX. These two SNPs appear to be independent. The greatest amount of variability in plasma CK that could be explained was 15%, by using a model that includes the genetic variants in *CKM* and *LILRB5*, gender, age, and physical activity in non-statin users.

Similar conclusions can be drawn using tables XIII to XVI.

This further supports the haplomap that the two SNPs are independent.

1. **rs11559024 in CKM and rs12975366 in LILRB5** (equations 3 and 4)

Table XI. **Association analysis between ln(CK levels) and SNPs rs11559024 and rs12975366** genotypes stratified by statin/no statin/all participants in the MHI Hospital Cohort. The multivariate analysis with a general linear model of CK is performed first with genotypes only and second with additive effect, gender, statin dose, age and physical activity.

		SNPs			
		rs11559024		**rs12975366**	
		Estimate	**Standard error**	**Estimate**	**Standard error**
Statin user	additive effect	0.59	0.07	0.11	0.01
	adjusted additive effect	0.55	0.08	0.12	0.02
Non statin user	additive effect	0.42	0.09	0.13	0.02
	adjusted additive effect	0.44	0.09	0.10	0.02
All participants	additive effect	0.53	0.06	0.12	0.01
	adjusted additive effect	0.51	0.06	0.11	0.01

Table XII. Variance explained for each of the stratified models in Table XII summarized by R-squared.

Dependent	R-squared	Coefficient of variation	Root of mean squared error	Mean	Category
ln(CK levels)	0.04	12.83	0.57	4.48	additive effect in statin users
ln(CK levels)	0.11	12.04	0.54	4.49	adjusted additive effect in statin users
ln(CK levels)	0.03	13.57	0.60	4.42	additive effect in non statin users
ln(CK levels)	0.15	12.54	0.55	4.40	adjusted additive effect in non statin users
ln(CK levels)	0.03	13.08	0.58	4.46	additive effect in all patients
ln(CK levels)	0.12	12.24	0.55	4.46	adjusted additive effect in all patients

*The R-squared is the coefficient of determination for the regression model including the genetic term in the unadjusted model, and the genetic term, sex, physical activity, age and statin atorvastatin equivilant dosage in the adjusted model.

2. **rs11559024 and rs142092440 in CKM** (equations 5 and 6)

Table XIII. **Association analysis between log(CK levels) and SNPs rs11559024 and rs12975366** genotypes stratified by statin/no statin/all participants in the MHI Hospital Cohort. The multivariate analysis with a general linear model of CK is performed first with genotypes only and second with additive effect, gender, statin dose, age and physical activity.

		SNPs			
		rs11559024		**rs142092440**	
		Estimate	**Standard error**	**Estimate**	**Standard error**
Statin user	additive effect	0.57	0.07	0.61	0.18
	adjusted additive effect	0.53	0.08	0.62	0.17
Non statin user	additive effect	0.41	0.09	0.74	0.35
	adjusted additive effect	0.44	0.09	0.97	0.32
All participants	additive effect	0.57	0.07	0.61	0.18
	adjusted additive effect	0.53	0.08	0.62	0.17

Table XIV. Variance explained for each of the stratified models in Table XIV summarized by R-squared.

Dependent	R-squared	Coefficient of variation	Root of mean squared error	Mean	Category
ln(CK levels)	0.02	12.93	0.58	4.48	additive effect in statin users
ln(CK levels)	0.10	12.14	0.54	4.49	adjusted additive effect in statin users
ln(CK levels)	0.01	13.70	0.60	4.42	additive effect in non statin users
ln(CK levels)	0.14	12.60	0.55	4.40	adjusted additive effect in non statin users
ln(CK levels)	0.02	13.19	0.59	4.46	additive effect in all patients
ln(CK levels)	0.11	12.34	0.55	4.46	adjusted additive effect in all patients

*The R-squared is the coefficient of determination for the regression model including the genetic term in the unadjusted model, and the genetic term, sex, physical activity, age and statin atorvastatin equivilant dosage in the adjusted model.

3. rs12975366 in LILRB5, rs11559024 and rs142092440 in CKM (equations 7 and 8)

Table XV. **Association analysis between log(CK levels) and SNPs rs11559024, rs12975366 and exm 1480239** genotypes stratified by statin/no statin/all participants in the MHI Hospital Cohort. The multivariate analysis with a general linear model of CK is performed first with genotypes only and second with additive effect, gender, statin dose, age and physical activity.

		SNPs					
		rs11559024		**rs142092440**		**rs12975366**	
		Estimate	**Standard error**	**Estimate**	**Standard error**	**Estimate**	**Standard error**
Statin user	additive effect	0.59	0.07	0.58	0.18	0.11	0.01
	adjusted additive effect	0.55	0.08	0.60	0.17	0.11	0.01
Non statin user	additive effect	0.42	0.09	0.69	0.35	0.12	0.02
	adjusted additive effect	0.44	0.09	0.93	0.32	0.10	0.02
All participants	additive effect	0.53	0.06	0.60	0.16	0.11	0.01
	adjusted additive effect	0.51	0.06	0.68	0.15	0.11	0.01

Table XVI. Variance explained for each of the stratified models in Table XVI summarized by R-squared.

Dependent	R-squared	Coefficient of variation	Root of mean squared error	Mean	Category
ln(CK levels)	0.04	12.82	0.57	4.48	additive effect in statin users
ln(CK levels)	0.12	12.02	0.54	4.49	adjusted additive effect in statin users
ln(CK levels)	0.03	13.56	0.60	4.42	additive effect in non statin users
ln(CK levels)	0.15	12.50	0.55	4.40	adjusted additive effect in non statin users
ln(CK levels)	0.04	13.07	0.58	4.46	additive effect in all cases
ln(CK levels)	0.13	12.21	0.54	4.46	adjusted additive effect in all cases

*The R-squared is the coefficient of determination for the regression model including the genetic term in the unadjusted model, and the genetic term, sex, physical activity, age and statin atorvastatin equivilant dosage in the adjusted model.

As expected, the significant SNPs in the *CKM* and *LILRB5* genes exhibit independence. The two SNPs in the *CKM* gene seem to be somewhat dependant.

11.4 ***Results from imputation analysis***

The imputation analysis was done but unfortunately the SNP of interest (rs2361797) had an extremely low completion rate of 4.18%, as seen in table XVII. Since protocol calls only for SNPs with a completion rate over 98%, this value was much too low to force into the rest of the analysis, and therefore **the SNP could not be tested for replication using this method.**

We opted to continue to conduct the association test between CK levels in the SNPs where completion rate was above 98% in the off chance that any of these imputations may surpass the threshold of significance and potentially find novel variants. The imputation analysis used 53,301 variants between position 52272131 and 57182204 on chromosome 19, build 37. 52,535 of these variants were imputed, 766 were genotyped.

The imputation analysis was done using an earlier version of the MHI Biobank data. The new data set has an additional 718 new participants which was not in the earlier version. There were 6009 participant with usable CK levels used in this analysis, but hospitalized and emergency participants were not excluded. This is a potential pitfall of the Manhattan plot seen in figure 7. These participants may have had CK measurements that were inflated by elevation of CK-MB. To control for this, jackknife residuals were removed, excluding a further 19 participants.

Table XVII. Information on imputation of SNP rs2361797

SNP	position	Null	Geno	Completion (%)
rs2361797	54751064	9777	426	4.18

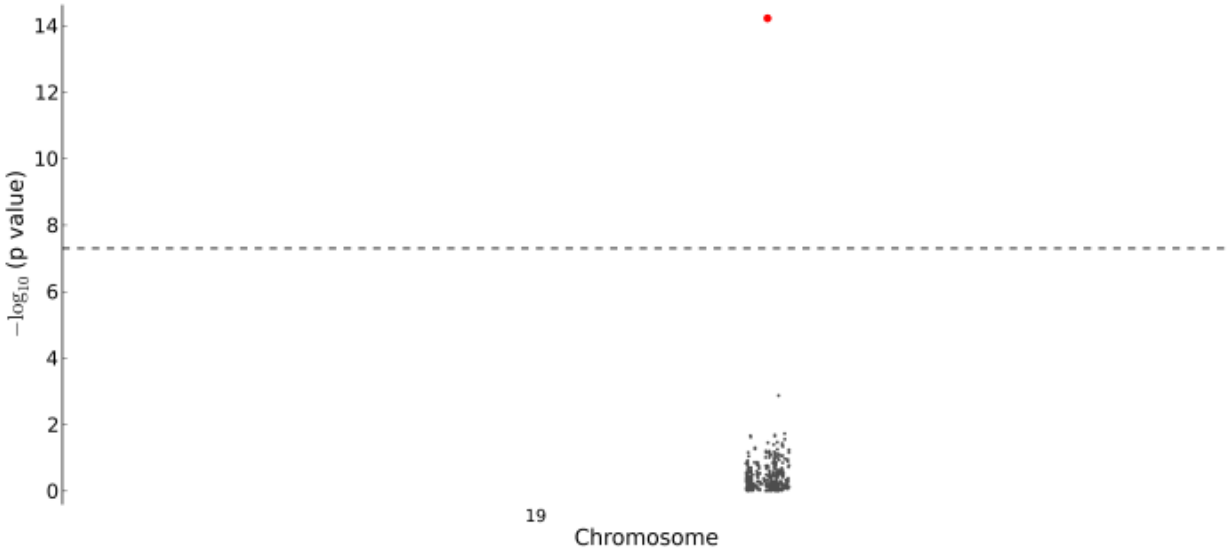

Figure 6. Manhattan plot showing the results for genetic determinants of CK levels measured in **all participants** in the MHI cohort study on chromosome 19 between position 52272131 and 57182204 (build 37) including variants from the imputation analysis. A GLM regression with the natural logarithm of CK was used, with adjustment for **components 1 and 2**. **A genetic variants in the *LILRB5* gene ($p= 6.01x10^{-15}$) was identified on rs12975366**.

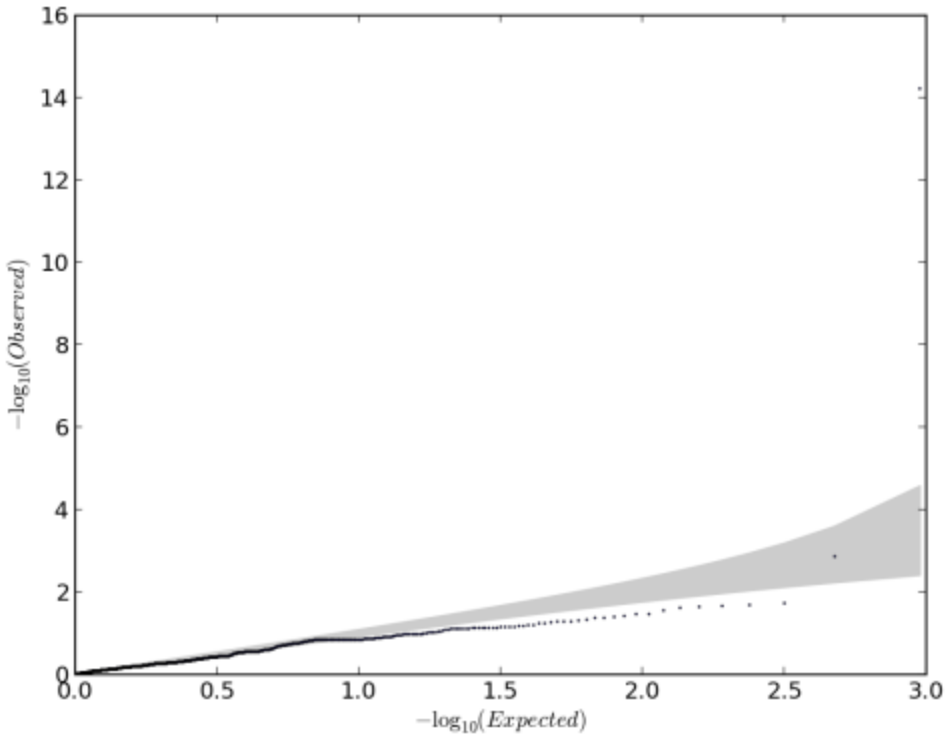

Figure 7. QQ plot corresponding to figure 1 for **all participants** in the MHI cohort study. A GLM regression with the natural logarithm of CK was used with adjustment for **principal components 1 and 2**.

The goal of this imputation was to look at position 54751064 (build 37) located in chromosome 19, but unfortunately this exact location could not be analyzed. The closest imputations were at 54751043 and 54751166. There were no new findings from this imputation analysis.

12 RESULTS FOR THE DISCOVERY STUDY - Objective 2

12.1 *GWAS Results*

For the GWAS, the analysis was limited to variants with a MAF > 5% in order to limit bias due to unequal variance in small genotype categories which can cause inflation of the statistics. Instead, we used a SKAT burden test to test the rare variants for association with plasma CK levels. The GWAS was conducted on three separate subgroups: all participants, statin users and non statin users. Participants removed in genetic data cleanup were not used. Additional phenotypic outliers were removed immediately before the analysis using jack-knife for residuals (n=4 for all participants, n=2 for statin users, n=4 for non statin users).

Figures 9, 11 and 13 are Manhattan plots showing results of the genome-wide association study for serum CK levels in statin users showing significant (in red) association signals in the *LILRB5* gene region. Each dot represents the $-\log_{10}$ P value for the genetic association using a multiple regression model adjusted for 2 principal components for genetic ancestry. Only variants with a MAF>5% were used. The dotted line shows the significance threshold.

Figures 10, 12 and 14 are the corresponding QQ-plots to the Manhattan plots. The two QQ-plots that were overlaid are: all variants and variants with a MAF>95%. Since rare variants can have been shown to cause inflation, and there was inflation when all variants were used, we used the variants with a MAF>95% in the analysis and used SKAT to test the rarer variants.

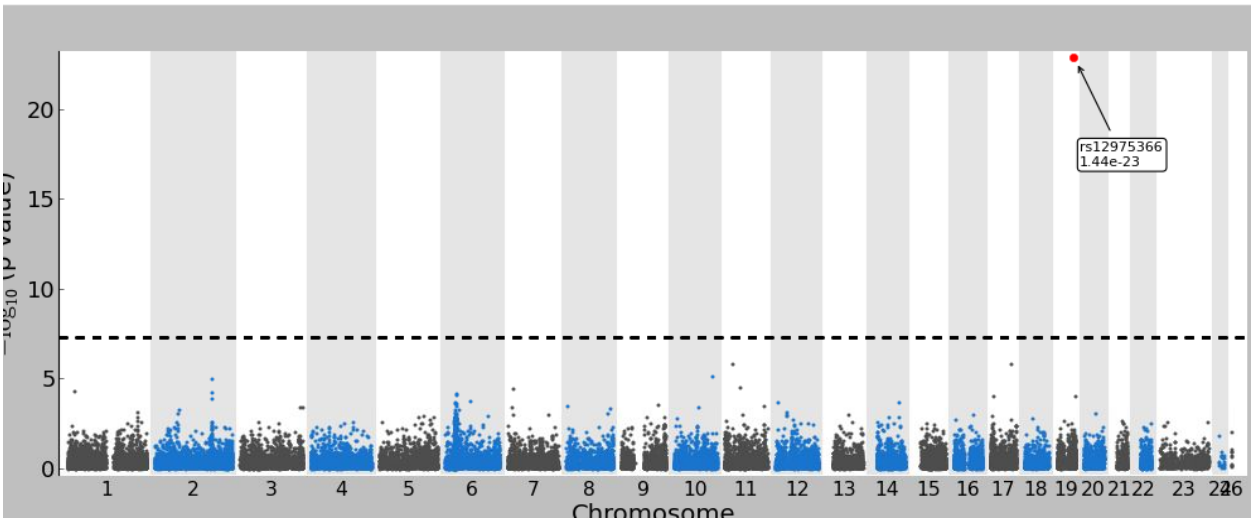

Figure 8. Manhattan plot showing the results of a genome-wide association study for genetic determinants of CK levels measured in **all participants** (n=5809) in the MHI cohort study. Only variants with MAF > 5% were used. A GLM regression with the natural logarithm of CK was used, with adjustment for **components 1 and 2. The genetic variant rs12975366 in the *LILRB5* gene ($p=1.44x10^{-23}$) was identified.**

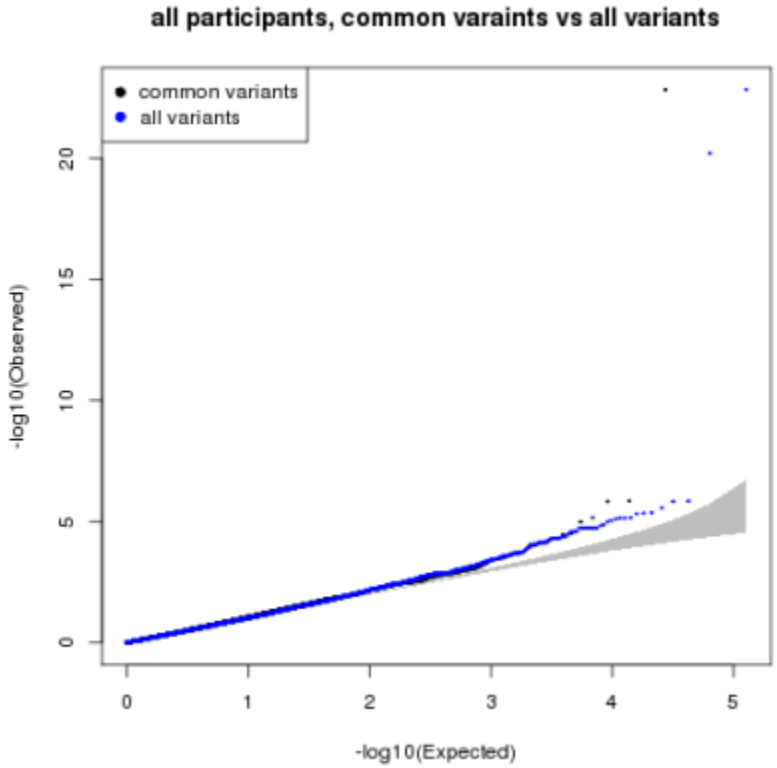

Figure 9. QQ plot corresponding to figure 8 and compared with a QQ plot of all variants, common SNPs with greater than 5% allele frequency for **all participants** in the MHI cohort study. A GLM regression with the natural logarithm of CK was used with adjustment for **principal components 1 and 2**.

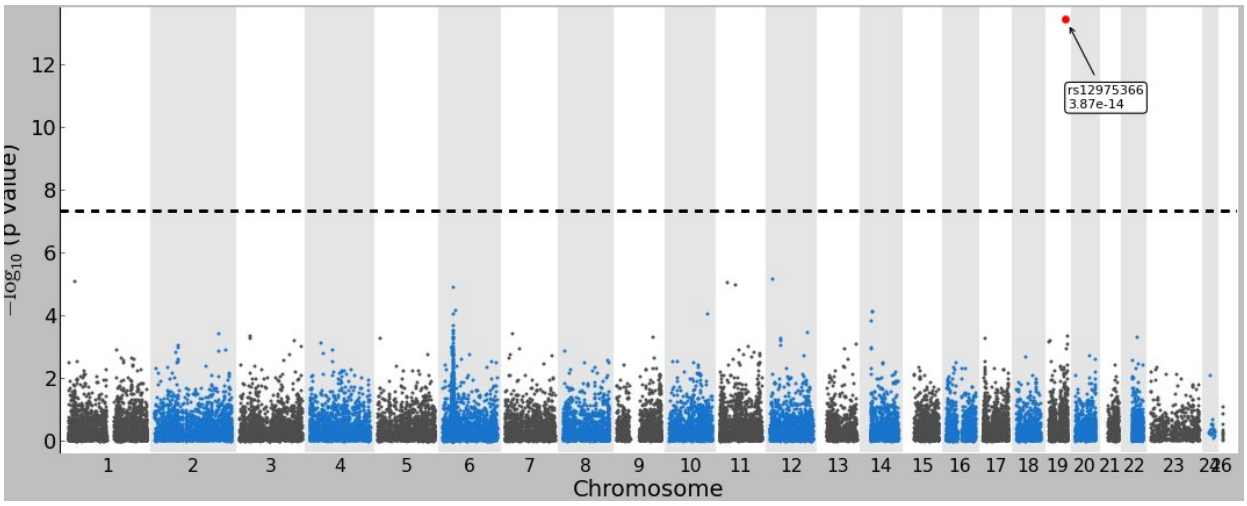

Figure 10. Manhattan plot showing the results of a genome-wide association study for genetic determinants of CK levels measured in **statin users** (n=3673) in the MHI cohort study. Only variants with MAF > 5% were used. A GLM regression with the natural logarithm of CK was used, with adjustment for **components 1 and 2**. **The genetic variant rs12975366 in the *LILRB5* gene ($p=3.87x10^{-14}$) was identified.**

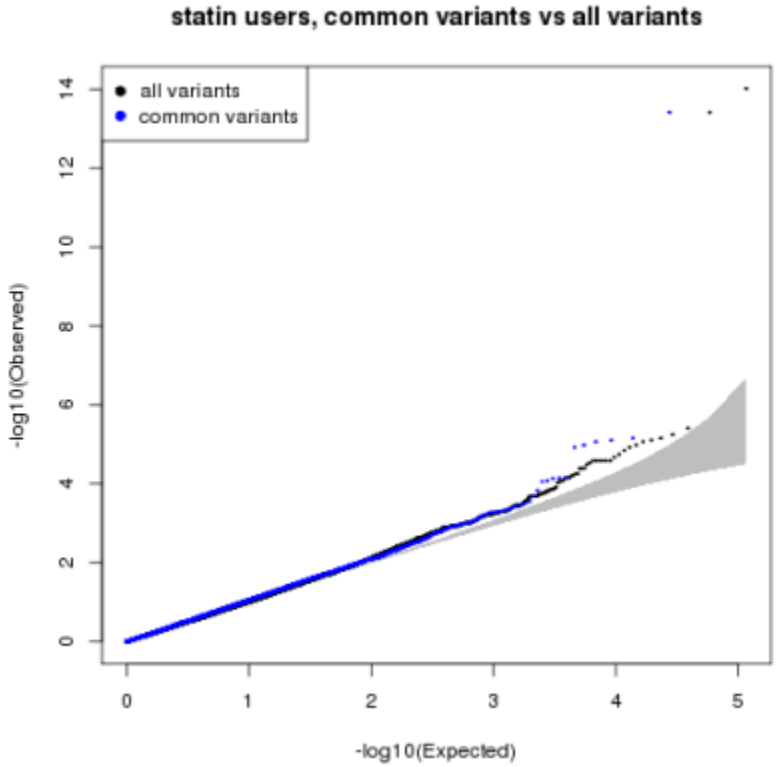

Figure 11. QQ plot corresponding to figure 10 and compared with a QQ plot of all variants, common SNPs with greater than 5% allele frequency for **statin users** in the MHI cohort study. A GLM regression with the natural logarithm of CK was used with adjustment for **principal components 1 and 2**.

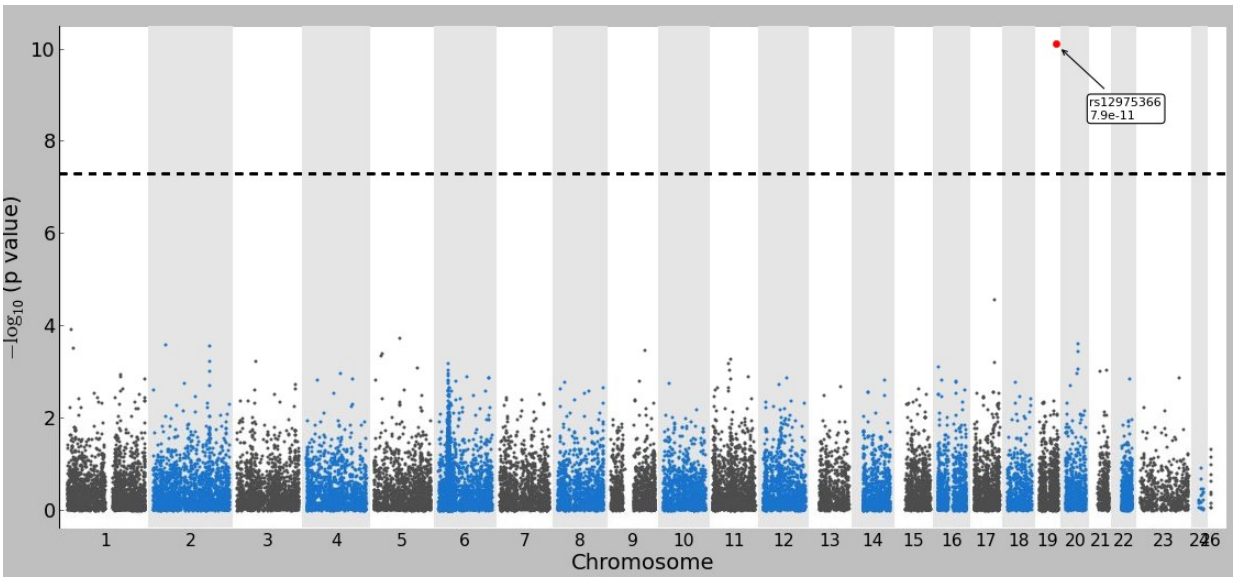

Figure 12. Manhattan plot showing the results of a genome-wide association study for genetic determinants of CK levels measured in **non-statin users** (n=2134) in the MHI cohort study. Only variants with MAF > 5% were used. A GLM regression with the natural logarithm of CK was used, with adjustment for **components 1 and 2**. **The genetic variant rs12975366 in the *LILRB5* gene ($p=7.9x10^{-11}$) was identified.**

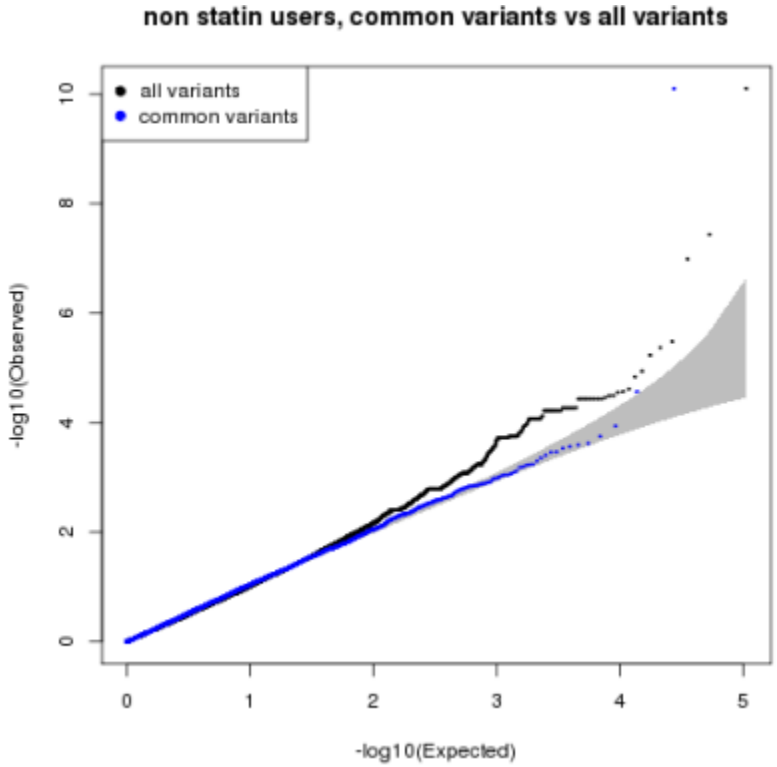

Figure 13. QQ plot corresponding to figure 12 and compared with a QQ plot of all variants, common SNPs with greater than 5% allele frequency for **non-statin users** in the MHI cohort study. A GLM regression with the natural logarithm of CK was used with adjustment for **principal components 1 and 2**.

12.2 ***Results of SKAT***

For the SKAT analysis, only rare variants (MAF < 5%) were used. The beta weighted graph was looked at, and components 1 and 2 were used as covariates. The same subgroups were used as in the GWAS: all participants, statin users and non statin users.

Figures 15, 17 and 19 are Manhattan plots showing results of the sequence kernel association test for serum CK levels in statin users showing significant (in red) association signals in the *CKM* gene. Each dot represents the $-\log_{10}$ P value for the genetic association using a multiple regression model adjusted for 2 principal components for genetic ancestry. The dotted line shows the significance threshold. Participants removed in genetic data cleanup were not used. Additional outliers were removed immediately before the analysis using jack-knife residuals (n=4 for all participants, n=2 for statin users, n=4 for non statin users).

Figures 16, 18 and 20 are the corresponding QQ-plots to the Manhattan plots.

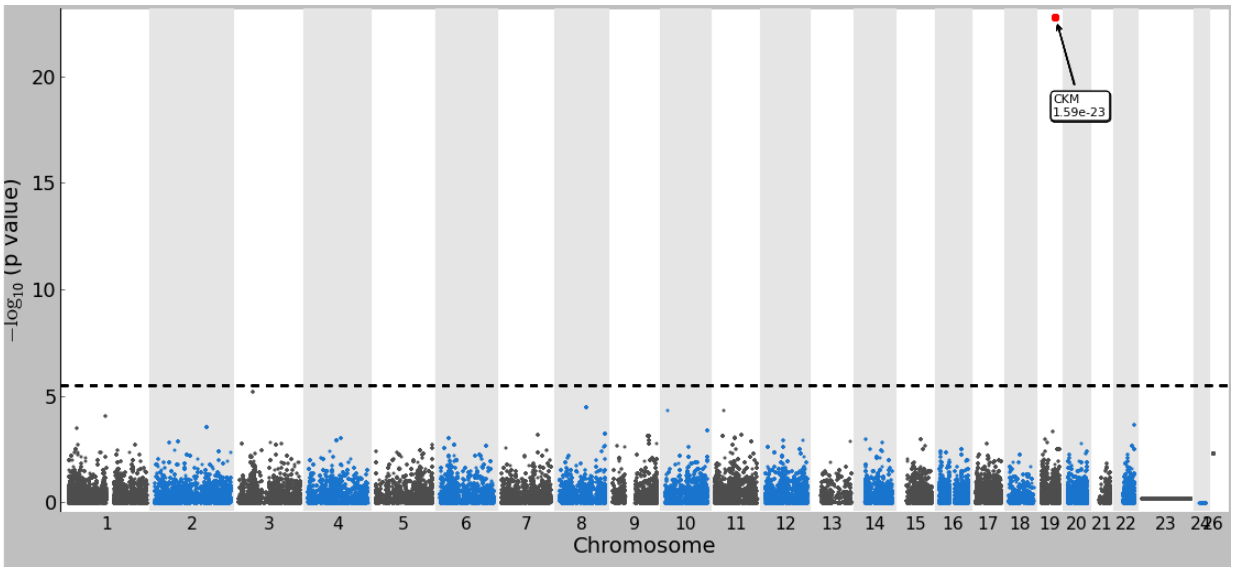

Figure 14. Manhattan graph showing the results for a **beta weighted** SKAT analysis for **all participants** (n=5349) from the MHI hospital Cohort, rare SNPs only, with adjustment for **C1 and C2**, with intergenic genes**. A genetic variant in the *CKM* gene (p=1.59x10^{-23}) was found**.

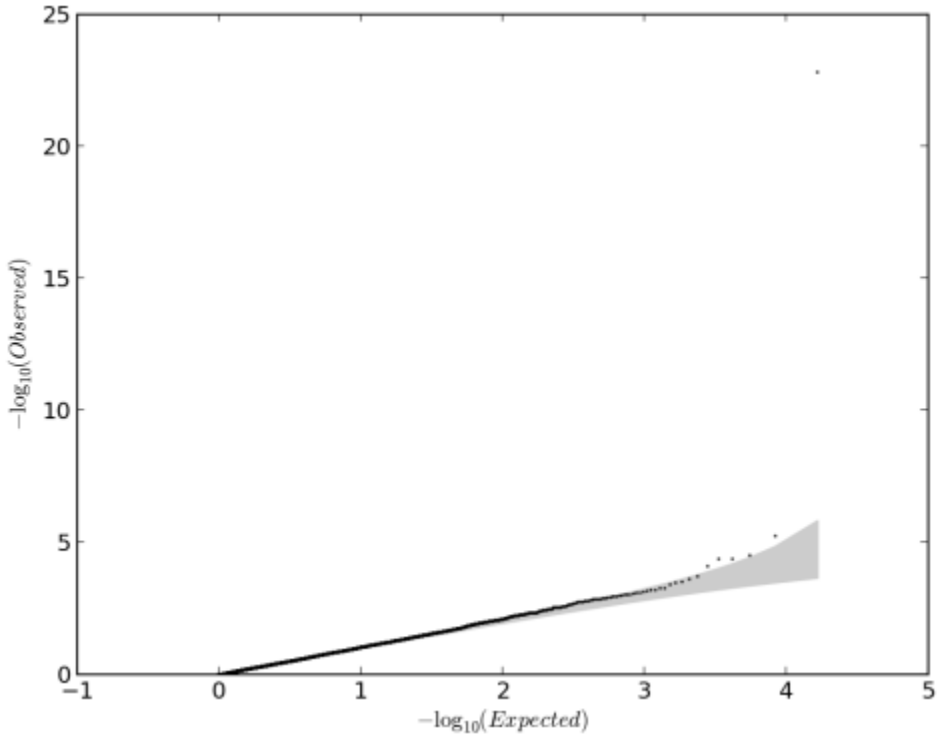

Figure 15. QQ plot corresponding figure 14 for the beta weighted SKAT analysis for **all participants** from the MHI hospital Cohort, rare SNPs only, with adjustment for **C1 and C2**, with intergenic genes.

Table XVIII. This table expands upon regions that surpassed the significant threshold level in the SKAT analysis where all participants were looked at, adjusted for genetic components 1 and 2

Gene	chr	Position (build=37)	SNP	Alleles (minor/major)	MAF	Homozygote rare/ heterozygote (count)	Type of mutation*	Mutation*	P-value from SKAT
CKM	19	45810010	rs142092440	G/A	0.00128	0/13	stop-lost	Synonymous SNV	$1.6x10^{-23}$
CKM	19	45815163	rs17357122	A/G	0.00846	1/89	missense	Synonymous SNV	$1.6x10^{-23}$
CKM	19	45818750	rs149872015	A/G	0.00019	0/2	missense	Synonymous SNV	$1.6x10^{-23}$
CKM	19	45818809	rs147574145	A/G	0.000095	0/2	missense	Nonsynonymous SNV	$1.6x10^{-23}$
CKM	19	45818825	rs17875653	C/G	0.00314	0/31	missense	Synonymous SNV	$1.6x10^{-23}$
CKM	19	45821183	rs11559024	G/A	0.0114	0/118	missense	Nonsynonymous SNV	$1.6x10^{-23}$

*from ANNOVAR (openbioinformatics.org/**annovar**, build hg19)
SNV stands for single nucleotide variant

The variant rs11559024 in the *CKM* gene, located in chromosome 19, was identified in the replication study done at the Pharmacogenomics lab, and therefore it is likely that this is in fact an actual association, and not just a false positive.

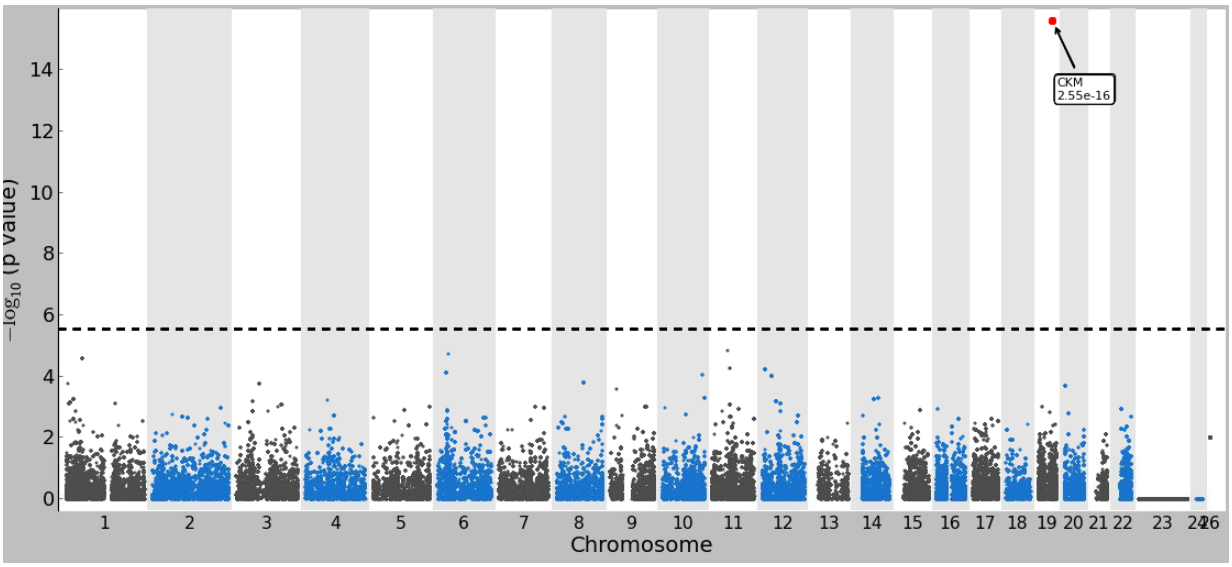

Figure 16. Manhattan graph showing the results for a **beta weighted** SKAT analysis for **statin users** (n=3403) from the MHI hospital Cohort, rare SNPs only, with adjustment for **C1 and C2**, with intergenic genes. **A genetic variant in the *CKM* gene ($p=2.55x10^{-16}$) was found.**

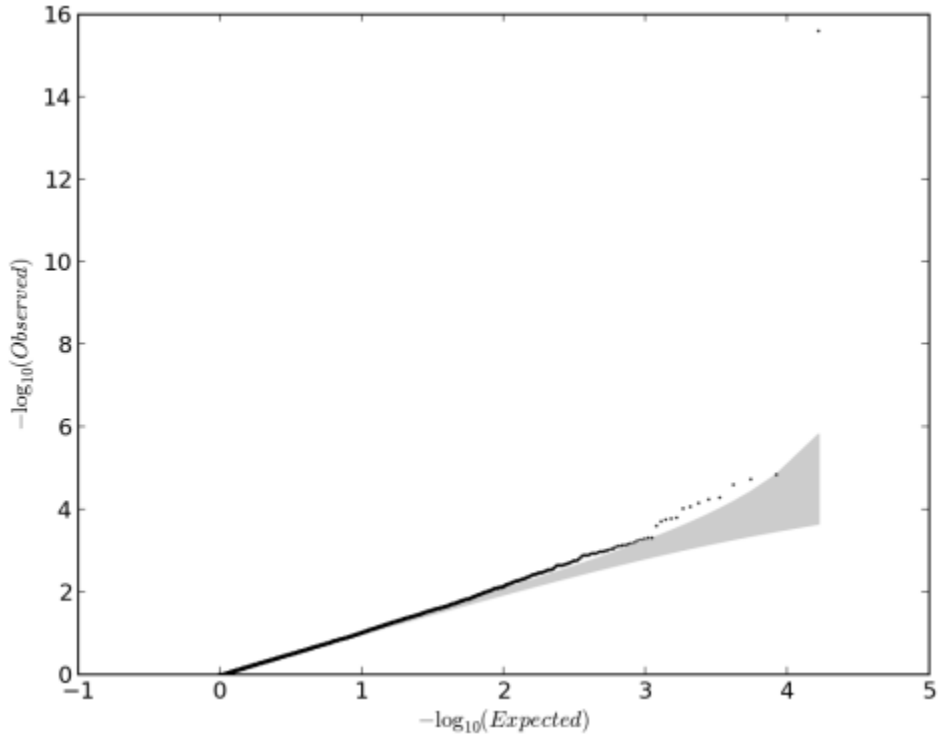

Figure 17. QQ plot corresponding to figure 16 for the beta weighted SKAT analysis for **statin users** from the MHI hospital Cohort, rare SNPs only, with adjustment for **C1 and C2**, with intergenic genes.

Table XIX. This table expands upon regions that surpassed the significant threshold level in the SKAT analysis where statin users were looked at, adjusted for genetic components 1 and 2

Gene	chr	Position (build=37)	SNP	Alleles (minor/major)	MAF	Homozygote rare/ heterozygote (count)	Type of mutation*	Mutation*	P-value from SKAT
CKM	19	45810010	rs142092440	G/A	0.00128	0/10	stop-lost	Synonymous SNV	$2.6x10^{-16}$
CKM	19	45815163	rs17357122	A/G	0.00846	1/58	missense	Synonymous SNV	$2.6x10^{-16}$
CKM	19	45818750	rs149872015	A/G	0.00019	0/1	missense	Synonymous SNV	$2.6x10^{-16}$
CKM	19	45818809	rs147574145	A/G	0.000095	0/2	missense	Nonsynonymous SNV	$2.6x10^{-16}$
CKM	19	45818825	rs17875653	C/G	0.00314	0/24	missense	Synonymous SNV	$2.6x10^{-16}$
CKM	19	45821183	rs11559024	G/A	0.0114	0/70	missense	Nonsynonymous SNV	$2.6x10^{-16}$

*from ANNOVAR (openbioinformatics.org/**annovar**, build hg19)
SNV stands for single nucleotide variant

The variant rs11559024 in the *CKM* gene, located in chromosome 19, was identified in the replication study done at the Pharmacogenomics lab, and therefore it is likely that this is in fact an actual association, and not just a false positive.

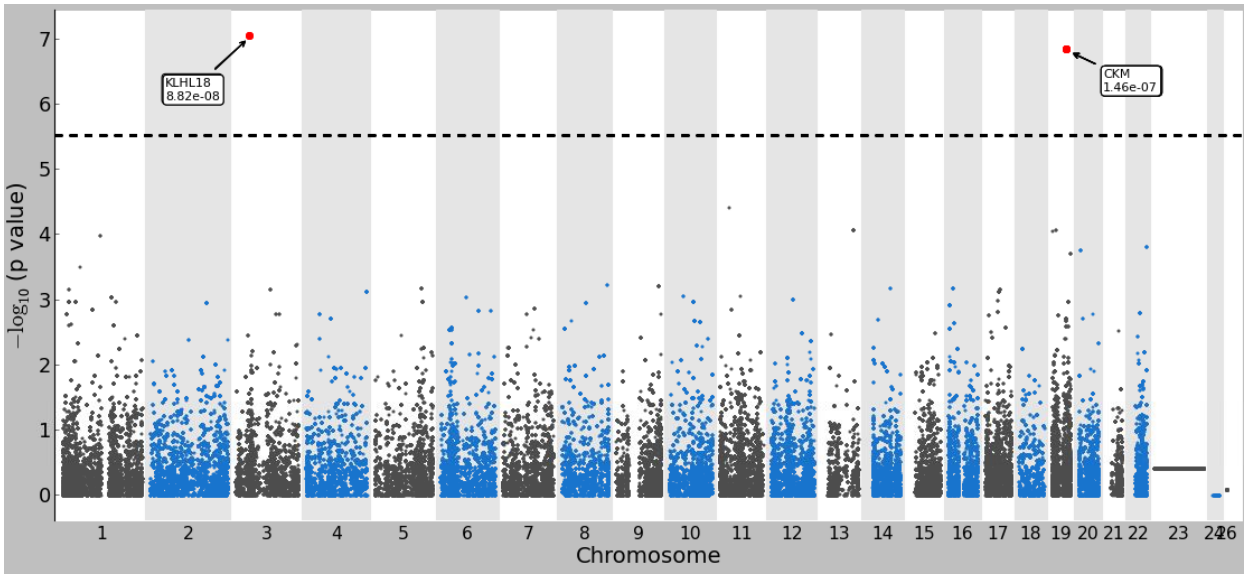

Figure 18. Manhattan graph showing the results for a **beta weighted** SKAT analysis for **non statin users** (n=1944) from the MHI hospital Cohort, rare SNPs only, with adjustment for **C1 and C2**, with intergenic genes. **Genetic variants in the *CKM* gene ($p=1.46=x10^{-7}$) and the *KLHL18* gene ($p=8.82x10^{-8}$) were found.**

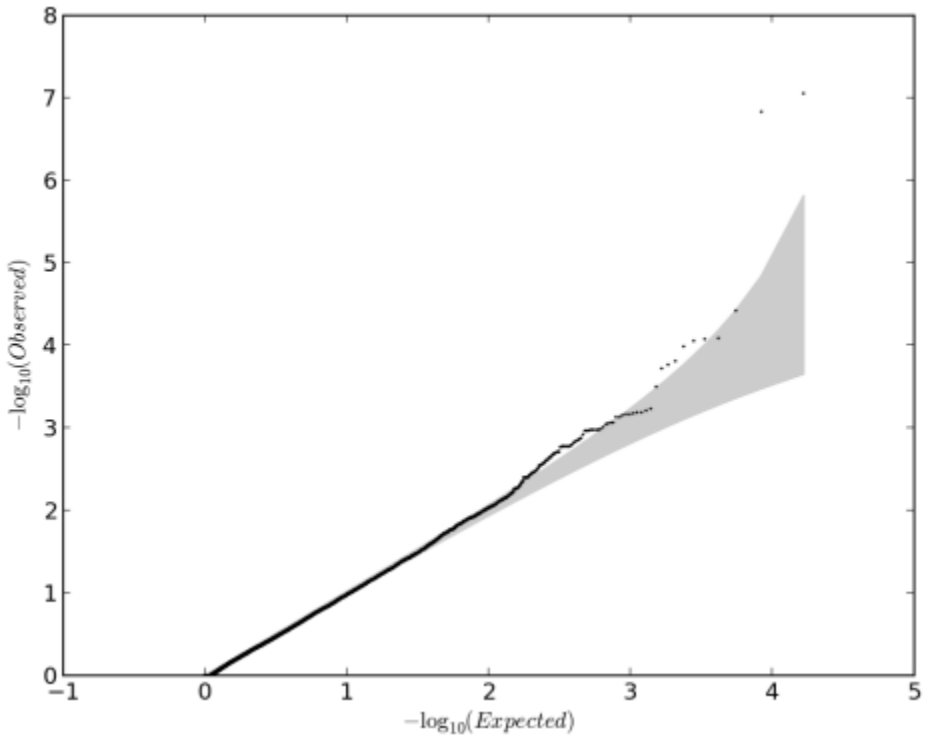

Figure 19. QQ plot for beta weighted SKAT analysis for **non statin users** from the MHI hospital Cohort, rare SNPs only, with adjustment for **C1 and C2**, with intergenic genes

Table XX. This table expands upon regions that surpassed the significant threshold level in the SKAT analysis where non-statin users were looked at, adjusted for genetic components 1 and 2.

Gene	chr	Position (build=37)	SNP	Alleles (minor /major)	MAF	Homozygote rare/ heterozygote (count)	Type of mutation*	Mutation*	P-value from SKAT
KLHL18	3	47374680	rs147834934	A/G	0.00019	0/2	missense	Synonymous SNV	$8.8x10^{-08}$
KLHL18	3	47374744	rs143338649	A/G	0.000048	0/0	missense	Synonymous SNV	$8.8x10^{-08}$
KLHL18	3	47378058	rs139767015	A/G	0.000095	0/1	missense	Nonsynonymous SNV	$8.8x10^{-08}$
CKM	19	45810010	rs142092440	G/A	0.00128	0/3	stop-lost	Synonymous SNV	$1.5x10^{-07}$
CKM	19	45815163	rs17357122	A/G	0.00846	0/31	missense	Synonymous SNV	$1.5x10^{-07}$
CKM	19	45818750	rs149872015	A/G	0.00019	0/1	missense	Synonymous SNV	$1.5x10^{-07}$
CKM	19	45818809	rs147574145	A/G	0.000095	0/0	missense	Nonsynonymous SNV	$1.5x10^{-07}$
CKM	19	45818825	rs17875653	C/G	0.00314	0/7	missense	Synonymous SNV	$1.5x10^{-07}$
CKM	19	45821183	rs11559024	G/A	0.0114	0/48	missense	Nonsynonymous SNV	$1.5x10^{-07}$

*from ANNOVAR (openbioinformatics.org/**annovar**, build hg19)
SNV stands for single nucleotide variant

It is unlikely the *KLHL18* gene is in fact a true positive, given the low representation of genetic variants in the gene in this population.

12.3 ***Candidate genes of statin-induced myotoxocity***

A literature review of genetic variants associated with statin-induced muscle myotoxicity (presented in introduction) has identified numerous variants. Although our study was not designed to assess statin-induced myotoxicity, we find it of relevance to report on the association results with plasma CK levels, as CK is a common biomarker of myotoxicity. The table below is an overview of these variants identified in prior studies and their affiliated p-values for association with plasma CK levels from the GWAS and SKAT analyses.

Table XXI. Table of candidate genes associated with statin-induced myotoxicity found in a literature review. The table includes p-values from the GWAS; p-values below 0.05 are highlighted. Variants were only included in the chart if they had a MAF > 0.05 in the MHI Hospital Cohort.

Gene	chr	Position (build 37)	SNP	Alleles (minor /major)	MAF*	Type of mutation**	Mutation**	P-value from GWAS		
								all participants	statin users	non statin users
RYR2	1	237841390	rs34967813	G/A	0.3054	missense	synonymous SNV	0.7551	0.2658	0.1151
RYR2	1	237266603	rs7554607	G/A	0.4543	intron	NULL	0.06218	0.305	0.131
RYR2	1	237271632	rs2485579	A/G	0.1344	intron	NULL	0.1428	0.3805	0.304
RYR2	1	237911004	rs2819770	A/G	0.2685	intron	NULL	0.2015	0.09982	0.9426
RYR2	1	237921596	rs1464461	G/A	0.3914	intron	NULL	0.03751	0.01456	0.8229
RYR2	1	237990122	rs2819742	A/G	0.4012	intron	NULL	0.6775	0.8432	0.3319
AGTR1	3	148460700	rs380400	G/A	0.133	untranslated-3	NULL	0.8201	0.7638	0.8063
ABCG2	4	89026109	rs2725264	G/A	0.07074	intron	NULL	0.4821	0.3886	0.9723
NOS3	7	150704843	rs891511	A/G	0.3167	intron	NULL	0.3161	0.205	0.9571
CYP3A5	7	99266318	rs4646450	A/G	0.1565	intron	NULL	0.4847	0.4524	0.898
ABCB1	7	87160618	rs2032582	A/C	0.457	missense	nonsynonymous SNV	0.9922	0.8936	0.8844
ABCB1	7	87229440	rs9282564	G/A	0.09943	missense	nonsynonymous SNV	0.4458	0.9008	0.2867
ABCB1	7	87220886	rs3789243	A/G	0.493	intron	NULL	0.5207	0.9571	0.2676
HTR3B	11	113803028	rs1176744	C/A	0.3338	missense	synonymous SNV	0.3121	0.8777	0.04759
ATP2B1	12	90008959	rs2681472	G/A	0.1639	intron	NULL	0.4618	0.8405	0.2853
ATP2B1	12	90013089	rs2681492	G/A	0.164	intron	NULL	0.5035	0.919	0.2748
SLCO1B1	12	21329738	rs2306283	G/A	0.4327	missense	synonymous SNV	0.6255	0.7039	0.7843
SLCO1B1	12	21329813	rs11045819	A/C	0.1687	missense	synonymous SNV	0.3113	0.547	0.4613

SLCO1B1	12	21331549	rs4149056	G/A	0.1621	missense	nonsynonymous SNV	0.5747	0.3202	0.5106
SLCO1B1	12	21331625	rs2291075	A/G	0.4117	coding-synon	nonsynonymous SNV	0.8761	0.6642	0.6645
SLCO1B1	12	21391976	rs34671512	C/A	0.05982	missense	synonymous SNV	0.4207	0.4394	0.7993
SLCO1B1	12	21317791	rs4149032	A/G	0.3631	intron	NULL	0.985	0.9359	0.8539
SLCO1B1	12	21368722	rs4363657	G/A	0.1691	intron	NULL	0.4515	0.2868	0.6749
GATM	15	45661678	rs1288775	T/A	0.2513	missense	nonsynonymous SNV	0.2520	0.05954	0.5143
DMPK	19	46275976	rs527221	G/C	0.108	missense	nonsynonymous SNV	0.2061	0.8083	0.1072

*of MHI Hospital Cohort, calculated in PLINK

from ANNOVAR (openbioinformatics.org/annovar**, build hg19)

Table XXII. Table of candidate genes associated with statin-induced myotoxicity found in a literature review. The table includes gene-based p-values from the SKAT analysis; p-values below 0.05 are highlighted. Variants were only included in the chart if they had a MAF > 0.0001 and <0.05 in the MHI Hospital Cohort.

Gene	chr	P-value from SKAT		
		all participants	statin users	non statin users
RYR2	1	0.886491	0.842326	0.768509
AGTR1	3	0.446288	0.946719	0.42878
COQ2	4	0.696129	1	0.584439
ABCG2	4	0.801945	0.711438	0.838153
NOS3	7	0.13698	0.156989	0.653147
CYP3A5	7	0.990837	0.987342	0.972358
ABCB1	7	0.535148	0.279578	0.575275
HTR7	10	0.059676	0.187646	0.297454
HTR3B	11	0.00116	0.026969	0.078215
ATP2B1	12	1	1	0.700123
SLCO1B1	12	0.543457	0.191031	0.943387
GATM	15	0.777022	0.687016	0.293927
DMPK	19	0.767762	0.440695	0.640323
CYP2D6	22	0.414223	0.188027	0.732096

13 CONCLUSION

13.1 ***Key results***

The first objective of this work was to replicate prior findings of association between the *CKM* and *LILRB5* gene and plasma CK levels by using data from over 5000 participants to the MHI Biobank Cohort. We report finding a significant association with the same *CKM* variant (rs11559024) as the prior finding. **This strengthens the argument that the rare allele at rs11559024 in the *CKM* gene is associated with a decrease in plasma CK** (p-value=8.05×10^{-19} without adjustment; and p-value=9.96×10^{-17} in the adjusted model). **An additional SNP, rs142092440 in the *CKM* gene was also identified in the present study** (p-value=0.00016, without adjustment). Previously, two intronic SNPs in the *LILRB5* gene (rs2361797 and rs406231) were found to be associated with plasma CK levels. Although these SNPs are not present on the data from the ExomeChip in our data set, **an additional SNP in the *LILRB5* gene, rs12975366, was found to be associated with plasma CK levels (p-value=4.28×10^{-21}, without adjustment).** Together, those findings add strong evidence that supports the replication of the association of *CKM* and *LILRB5* with plasma CK levels.

The second objective of this work was to conduct a genome wide discovery study with plasma CK levels by using over 5000 participants from the MHI Biobank Cohort. Analyses conducted with adjustment for ethnic ancestry with the inclusion of principal components 1 and 2 also support **association in the *LILRB5* gene, rs1297366 (p-value=1.44×10^{-23})**, which looked at all common variants (MAF > 5%). The SKAT analysis, which looked at rare variants **replicated finding in the *CKM* gene (p-value=1.59×10^{-23}).** We are unsure of which variant(s) are causing this association, although **rs142092440 causes a stop lost mutation, and**

rs147574145 and rs11559024 are nonsynonomous so these variants are likely candidates. rs11559024 was identified in the prior project at the Pharmacogenomics lab, so this may be a causal variant or associated with a causal variant in the locus. **The SKAT analysis identified a novel variant in the *KLHL18* gene among the non statin users (p-value=3.66×10^{-08})**. The MAF of genetic variants in *KLHL18* gene in the MHI Biobank Cohort was very low, and requires replication in a larger cohort where more of the rare alleles are present.

Statin users had, on average, a higher plasma CK level than non statin users. Participants who had a rare allele from the associated variants in the *CKM* and *LILRB5* genes had, on average, lower CK levels. Furthermore, when two rare alleles were present, these average CK levels were even lower. Carriers of these rare variants may have a reduced risk of elevated CK levels while on a statin.

Table XXIII. Summary table of SNP rs12975366 located in the *LILRB5* gene in chromosome 19 identified in the GWAS analysis

	Beta (95% CI)	**Standard error**	**P-value**	**MAF of MHI Cohort**	**Global MAF (1000 Genome)**
All participants	-0.12 (-0.14, -0.09)	0.01	$1.44x10^{-23}$	0.44	0.2663
Statin users	-0.11 (-0.14, -0.08)	0.01	$3.87x10^{-14}$		
Non statin users	-0.12 (-0.16, -0.09)	0.02	$7.90x10^{-11}$		

Table XXIV. Summary table of SNP variants identified in the SKAT analysis variants surpassing the significance threshold

	Chr	**Gene**	**SNP**	**Position (build 37)**	**Type of mutation***	**Mutation***	**P-value**	**MAF of MHI Cohort**	**Global MAF (1000 Genome)**
All participants	19	CKM	rs142092440	45810010	stop-lost	Synonymous SNV	$1.6x10^{-23}$	0.00128	0.0009
			rs17357122	45815163	missense	Synonymous SNV		0.00846	0.0041
			rs149872015	45818750	missense	Synonymous SNV		0.00019	
			rs147574145	45818809	missense	Nonsynonymous SNV		0.000095	0.0009
			rs17875653	45818825	missense	Synonymous SNV		0.00314	0.003
			rs11559024	45821183	missense	Nonsynonymous SNV		0.0114	0.006

Statin users	19	CKM	rs142092440	45810010	stop-lost	Synonymous SNV	$2.6x10^{-16}$	0.00128	0.0009
			rs17357122	45815163	missense	Synonymous SNV		0.00846	0.0041
			rs149872015	45818750	missense	Synonymous SNV		0.00019	
			rs147574145	45818809	missense	Nonsynonymous SNV		0.000095	0.0009
			rs17875653	45818825	missense	Synonymous SNV		0.00314	0.003
			rs11559024	45821183	missense	Nonsynonymous SNV		0.0114	0.006
Non statin users	3	KLHL18	rs147834934	47374680	missense	Synonymous SNV	$8.8x10^{-08}$	0.00019	
			rs143338649	47374744	missense	Synonymous SNV		0.000048	
			rs139767015	47378058	missense	Nonsynonymous SNV		0.000095	
	19	CKM	rs142092440	45810010	stop-lost	Synonymous SNV	$1.5x10^{-07}$	0.00128	0.0009
			rs17357122	45815163	missense	Synonymous SNV		0.00846	0.0041
			rs149872015	45818750	missense	Synonymous SNV		0.00019	
			rs147574145	45818809	missense	Nonsynonymous SNV		0.000095	0.0009
			rs17875653	45818825	missense	Synonymous SNV		0.00314	0.003
			rs11559024	45821183	missense	Nonsynonymous SNV		0.0114	0.006

*from ANNOVAR (openbioinformatics.org/**annovar**, build hg19)
SNV stands for single nucleotide variant

14 DISCUSSION

14.1 ***Limitations***

The MHI Biobank is a longitudinal study with a growing array of data, and although large, has limiting factors. **There is a genetic bias**, as all participants were recruited from within Montreal, and for the most part self-identified as Caucasian. In order to answer the MHI questionnaire, participants had to be proficient in either English or French, resulting in the exclusion of all allophones that could not speak either of these languages. We would expect allophones to be made up predominately of immigrants, so when allophones are excluded this could further homogenize the cohort. To try and account for this, we used genetic variants that can further pinpoint a more exact ethnic makeup of each individual. As in all cohort studies, there is a risk that this cohort has a different genetic makeup than another, and that the genetic variants identified in this study that are associated with plasma CK are different from those found in another cohort. In order to reduce chances of this, we compared findings to prior studies. As this is a fairly large replication, we hope to account for as many ethnic minorities as possible and reduce errors that genetic heterogeneity may incur.

The ExomeChip from Illumina only looks at the exome of the genome. Exome sequencing is cost effective but fails to identify variants found in the coding region of genes which affect protein function. Whole genome sequencing can identify many more variants, but is currently not time or cost effective although it may be a standard approach in the future. This said, the exome still represents a portion of the genome that can be used to identify variants with a large effect size and makes way for a valuable study as long as limitations are realized.

The MHI Biobank includes the prevalence rate of statin users, but not the incidence rate. This may create a bias in the study as it is probable that some participants were put on statins and subsequently removed after the occurrence of adverse muscle effects. Optimally, these participants should have their CK level used from when they were on the statin and be included in the statin user subgroup as it is likely they have genetic variants influencing their CK levels. In addition, many participants in the MHI Biobank are on concomitant medicine that may impact their statin dosage or CK levels. The incidence rate of statin usage among participants would have been a worthwhile marker to know, but only current medication is recorded in the MHI Biobank.

When using a hospital cohort, it is important to be aware of general biases that may arise. **A hospital cohort, such as the MHI Biobank, often has a higher incidence of sickness, is more homogenous and more medicated than a general population.** This may incur biases when studies are used to assess a general population.

We stratified all useable participants as either statin users or non users. **We did not further stratify the statin users by statin type**, which has been found to affect the risk of a statin user developing muscle related adverse effects. We did not find any change in statistical significance when we tried to stratify participants by statin type. Furthermore, the vast majority of participants used atorvastatin while only a very small percentage used fluvastatin or lovastatin. Stratification by statin type is something that could be looked at in future studies.

14.2 ***Study strengths***

A GWAS is a useful research tool which allows a project to be designed without knowing the biological pathway of the trait. A GWAS can potentially find multiple novel candidate genes with one analysis, or further validate genetic variants identified in prior studies. When used correctly and with caution, a GWAS is a powerful statistical method. The discovery project used a large cohort to identify novel variants and further confirm association in previously replicated genetic variants. This project paid close attention to STrengthening the REporting of Genetic Associations (STREGA) guidelines when conducting all aspects of the analysis. STREGA is considered to be the gold standard when it comes to conducting a GWAS. It was written by approximately 30 epidemiologists, geneticists, statisticians and journal editors as well as the general community of researchers, journal editors and stakeholders providing guidelines on all aspects of genetic association tests. We used SREGA alongside replication in order to reduce bias and conduct the most statistically powerful test possible.

14.3 ***Interpretations***

This study replicated prior findings of variants in both the *CKM* and *LILRB5* genes. When present, and even more so in the homozygote rare type, these variants appear to decrease CK levels. The *LILRB5* gene has a minor allele that is extremely common while the variant in the *CKM* gene is considerably rarer. As this is a replication study, the coupling of pre-existing studies with similar findings, a large cohort population and highly significant p-values gives evidence for this not being a false positive, but instead a strong candidate for variants that could be used for CK management in a clinical setting. Using CK as a biomarker for adverse muscle effects, one can hope to use genetic variants in the diagnosis and prognosis of muscle myopathy in statin users.

The influence of each genetic variant is quite small, accounting for only a small percentage of change in predicting plasma CK levels. Although these genetic variants could be a useful tool in a clinical setting, they are not powerful enough on their own and should be coupled with pre-existing methods for predicting plasma CK levels in patients.

The variants in the *LILRB5* gene and *CKM* gene make for strong candidates for true variation in plasma CK levels. It should be noted that these variants do not appear to help predict plasma CK levels in statin users or non-users by more than a few percent.

Variants identified in other studies were not necessarily replicated in this cohort study. Previously, there has been significant association, often replicated in multiple studies, that shows association between CK levels and myotoxocity in statin users in genes including *SLCO1B1, GATM, CYP3A4* and *CYP2D6*. None of these findings were replicated in the GWAS or SKAT analysis presented in this paper. Association may not have been detectable due to the fact that we did not always test for the exact SNP or because SNPs previously found to be associated with muscle myotoxicity are not necessarily expected to be associated with CK levels.

14.4 ***Generalizability***

Personalized pharmacogenomics aims to bring an individualized medical plan for each patient based on their genome. As this is a relatively new area of research, it is necessary to be thorough and sure of the findings before implementing them in a clinical setting. This study is one of many needed in order to bring personalized statin care to an individual in a clinical setting.

Once these variants in the CKM and LILRB5 genes are either refuted or confirmed, a sensitivity analysis can help develop an algorithm that takes into account genetic variants and provides a personalized reference level of plasma CK for each patient to be used in a clincia setting. We are far from this point, as complex gene interactions are hard to interpret, and this report is just one of the many necessary on the road to understanding the effects of genetic variants on CK levels in statin users.

14.5 ***Future Studies***

This replication study confirms genetic variants associated with muscle adverse effects among statin users in the MHI Biobank. A major limitation of this study is the homogeneity of the population. A future study in a different cohort with a different ethnic makeup is a necessary next step. Future studies in different cohorts could further test these variants, as well as look at replicated variants found in the discovery study of this project.

14.6 ***Relevance of the genes***

CKM

Creatine kinase (CK) catalyzes the reversible transfer of high energy phosphates between ATP/ADP and creatine systems. CK is important for normal energy homeostasis and exerts several integrated functions, including temporary energy buffering, metabolic capacity, energy transfer and metabolic control. CK level and activity are clinically important and CK serves as a reliable biomarker for several diseases including myocardial infarction, rhabdomyolysis, muscular dystrophy and autoimmune myositis. There is only limited documentation of polymorphisms in the *CKM* gene in the literature.[95,96] Wu *et al.* have reported on the finding of

the point mutation E79G in *CKM* in two acute myocardial infarction patients with muscle CK activity deficiency.[19] This variant is situated 12 base pairs upstream from variant rs11559024 (Glu83Gly) reported in our study.

LILRB5

LILRB5 is a member of the leukocyte immunoglobulin-like receptor (LIR) family, which is found in a gene cluster at chromosomal region 19q13.4.[97] LIR subfamily B receptors are expressed on immune cells where they bind to MHC class I molecules on antigen-presenting cells and inhibit stimulation of an immune response. The protein is an integral membrane protein with receptor activity and contains four extracellular immunoglobulin-like domains and the rs2361797 variant causes a missense (D -> G) mutation in its third domain. The protein also includes two immunoreceptor tyrosine-based inhibition motifs and three phosphorylation sites in its cytoplasmic part.[98] The *LILRB5* gene presents multiple transcript variants encoding different isoforms and is highly expressed in skeletal muscle, liver and gallbladder.[99] Mass spectrometry has detected the protein in plasma, liver and aorta and this plasmatic protein has been ascertained in the HUPO plasma proteome project.[99,100] Furthermore, we could not find evidence of LILRB5 modulation by statins.[101] There was no report in the literature until our work for possible implication of this gene in human disease, but the implication of immunity in inter-individual variation in CK is not entirely new. Inter-individual variability in CK levels is influenced by the rate of CK leakage from injured muscle fibers into the circulation but it is possible that a portion of the variability is also due to the rate of CK clearance from the circulation.[102-104] CK clearance occurs via the mononuclear phagocytic system in the liver and via Fc receptors that mediate the endocytosis of immune complexes. CK immune complexes are found in the blood and are

commonly referred to as macro CK type 1, which is a complex formed by an immunoglobulin, often IgG, and a CK isoenzyme, often CK-BB[102,105].

KLHL18

KLHL18 is a gene that, in humans, encodes the kelch-like protein 18. *KLHK* genes have been found to be responsible for many Mendelian diseases and are possibly associated with certain cancers[106], but there is currently no literature supporting its association to plasma CK levels. Since this locus was identified only for the statin subgroup, and there were very few minor alleles available, it is not excluded at this point that this could be a false positive. It has been suggested that for the loci in a SKAT anaylsis containing 4 or less allele observations, the results are likely biased[107], as is the case which this the *KLHL18* gene which has only 3 allele observtions. Further investigations as part of a replication study will be necessary to confirm the validity of this finding.

15 REFERENCES

1 Pignone, M., Phillips, C. & Mulrow, C. Use of lipid lowering drugs for primary prevention of coronary heart disease: meta-analysis of randomised trials. *BMJ* **321**, 983-986 (2000).

2 Fletcher, G. F. *et al.* Efficacy of drug therapy in the secondary prevention of cardiovascular disease and stroke. *Am J Cardiol* **99**, 1E-35E, doi:10.1016/j.amjcard.2007.02.001 (2007).

3 Allen Maycock, C. A. *et al.* Statin therapy is associated with reduced mortality across all age groups of individuals with significant coronary disease, including very elderly patients. *Journal of the American College of Cardiology* **40**, 1777-1785 (2002).

4 Ray, K. K. *et al.* Statins and all-cause mortality in high-risk primary prevention: a meta-analysis of 11 randomized controlled trials involving 65,229 participants. *Archives of internal medicine* **170**, 1024-1031, doi:10.1001/archinternmed.2010.182 (2010).

5 Law, M. R., Wald, N. J. & Rudnicka, A. R. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *Bmj* **326**, 1423, doi:10.1136/bmj.326.7404.1423 (2003).

6 Endo, A. The origin of the statins. 2004. *Atherosclerosis. Supplements* **5**, 125-130, doi:10.1016/j.atherosclerosissup.2004.08.033 (2004).

7 Shepherd, J., Hunninghake, D. B., Barter, P., McKenney, J. M. & Hutchinson, H. G. Guidelines for lowering lipids to reduce coronary artery disease risk: a comparison of rosuvastatin with atorvastatin, pravastatin, and simvastatin for achieving lipid-lowering goals. *The American journal of cardiology* **91**, 11C-17C; discussion 17C-19C (2003).

8 Hoffman, K. B., Kraus, C., Dimbil, M. & Golomb, B. A. A survey of the FDA's AERS database regarding muscle and tendon adverse events linked to the statin drug class. *PloS one* **7**, e42866, doi:10.1371/journal.pone.0042866 (2012).

9 Istvan, E. S. & Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* **292**, 1160-1164, doi:10.1126/science.1059344 (2001).

10 Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34-47 (1986).

11 McTaggart, F. *et al.* Preclinical and clinical pharmacology of Rosuvastatin, a new 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor. *The American journal of cardiology* **87**, 28B-32B (2001).

12 Prueksaritanont, T. *et al.* Mechanistic studies on metabolic interactions between gemfibrozil and statins. *The Journal of pharmacology and experimental therapeutics* **301**, 1042-1051 (2002).

13 Shitara, Y., Hirano, M., Sato, H. & Sugiyama, Y. Gemfibrozil and its glucuronide inhibit the organic anion transporting polypeptide 2 (OATP2/OATP1B1:SLC21A6)-mediated hepatic uptake and CYP2C8-mediated metabolism of cerivastatin: analysis of the mechanism of the clinically relevant drug-drug interaction between cerivastatin and gemfibrozil. *The Journal of pharmacology and experimental therapeutics* **311**, 228-236, doi:10.1124/jpet.104.068536 (2004).

14 Bohlmeyer, T. J., Wu, A. H. & Perryman, M. B. Evaluation of laboratory tests as a guide to diagnosis and therapy of myositis. *Rheumatic diseases clinics of North America* **20**, 845-856 (1994).

15 Chattington, P., Clarke, D. & Neithercut, W. D. Timed sequential analysis of creatine kinase in the diagnosis of myocardial infarction in patients over 65 years of age. *Journal of clinical pathology* **47**, 995-998 (1994).

16 Karras, D. J. & Kane, D. L. Serum markers in the emergency department diagnosis of acute myocardial infarction. *Emergency medicine clinics of North America* **19**, 321-337 (2001).

17 Brancaccio, P., Maffulli, N. & Limongelli, F. M. Creatine kinase monitoring in sport medicine. *Br Med Bull* **81-82**, 209-230, doi:10.1093/bmb/ldm014 (2007).

18 Pasternak, R. C. *et al.* ACC/AHA/NHLBI Clinical Advisory on the Use and Safety of Statins. *Circulation* **106**, 1024-1028 (2002).

19 Sewright, K. A., Clarkson, P. M. & Thompson, P. D. Statin myopathy: incidence, risk factors, and pathophysiology. *Curr Atheroscler Rep* **9**, 389-396 (2007).

20 McKenney, J. M., Davidson, M. H., Jacobson, T. A., Guyton, J. R. & National Lipid Association Statin Safety Assessment Task, F. Final conclusions and recommendations of the National Lipid Association Statin Safety Assessment Task Force. *Am J Cardiol* **97**, 89C-94C, doi:10.1016/j.amjcard.2006.02.030 (2006).

21 Jacobson, T. A. Toward "pain-free" statin prescribing: clinical algorithm for diagnosis and management of myalgia. *Mayo Clinic proceedings. Mayo Clinic* **83**, 687-700, doi:10.4065/83.6.687 (2008).

22 Law, M. & Rudnicka, A. R. Statin safety: a systematic review. *The American journal of cardiology* **97**, 52C-60C, doi:10.1016/j.amjcard.2005.12.010 (2006).

23 Rallidis, L. S., Fountoulaki, K. & Anastasiou-Nana, M. Managing the underestimated risk of statin-associated myopathy. *International journal of cardiology*, doi:10.1016/j.ijcard.2011.07.048 (2011).

24 Baird, M. F., Graham, S. M., Baker, J. S. & Bickerstaff, G. F. Creatine-kinase- and exercise-related muscle damage implications for muscle performance and recovery. *Journal of nutrition and metabolism* **2012**, 960363, doi:10.1155/2012/960363 (2012).

25 Amelink, G. J., Koot, R. W., Erich, W. B., Van Gijn, J. & Bar, P. R. Sex-linked variation in creatine kinase release, and its dependence on oestradiol, can be demonstrated in an in-vitro rat skeletal muscle preparation. *Acta physiologica Scandinavica* **138**, 115-124 (1990).

26 Komulainen, J., Koskinen, S. O., Kalliokoski, R., Takala, T. E. & Vihko, V. Gender differences in skeletal muscle fibre damage after eccentrically biased downhill running in rats. *Acta physiologica Scandinavica* **165**, 57-63 (1999).

27 Rinard, J., Clarkson, P. M., Smith, L. L. & Grossman, M. Response of males and females to high-force eccentric exercise. *Journal of sports sciences* **18**, 229-236, doi:10.1080/026404100364965 (2000).

28 Clarkson, P. M. & Hubal, M. J. Exercise-induced muscle damage in humans. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists* **81**, S52-69, doi:10.1097/01.PHM.0000029772.45258.43 (2002).

29 Manfredi, T. G. *et al.* Plasma creatine kinase activity and exercise-induced muscle damage in older men. *Medicine and science in sports and exercise* **23**, 1028-1034 (1991).

30 Melli, G., Chaudhry, V. & Cornblath, D. R. Rhabdomyolysis: an evaluation of 475 hospitalized patients. *Medicine* **84**, 377-385 (2005).

31 Chevion, S. *et al.* Plasma antioxidant status and cell injury after severe physical exercise. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5119-5123, doi:10.1073/pnas.0831097100 (2003).

32 Deuster, P. A. *et al.* Genetic polymorphisms associated with exertional rhabdomyolysis. *European journal of applied physiology* **113**, 1997-2004, doi:10.1007/s00421-013-2622-y (2013).

33 Landau, M. E. *et al.* Investigation of the relationship between serum creatine kinase and genetic polymorphisms in military recruits. *Military medicine* **177**, 1359-1365 (2012).

34 Neal, R. C., Ferdinand, K. C., Ycas, J. & Miller, E. Relationship of ethnic origin, gender, and age to blood creatine kinase levels. *Am J Med* **122**, 73-78, doi:10.1016/j.amjmed.2008.08.033 (2009).

35 Brewster, L. M., Coronel, C. M., Sluiter, W., Clark, J. F. & van Montfrans, G. A. Ethnic differences in tissue creatine kinase activity: an observational study. *PloS one* **7**, e32471, doi:10.1371/journal.pone.0032471 (2012).

36 Bruckert, E., Hayem, G., Dejager, S., Yau, C. & Begaud, B. Mild to moderate muscular symptoms with high-dosage statin therapy in hyperlipidemic patients--the PRIMO study. *Cardiovascular drugs and therapy / sponsored by the International Society of Cardiovascular Pharmacotherapy* **19**, 403-414, doi:10.1007/s10557-005-5686-z (2005).

37 Ridker, P. M. *et al.* Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med* **359**, 2195-2207, doi:NEJMoa0807646 [pii] 10.1056/NEJMoa0807646 (2008).

38 Harper, C. R. & Jacobson, T. A. Evidence-based management of statin myopathy. *Curr Atheroscler Rep* **12**, 322-330, doi:10.1007/s11883-010-0120-9 (2010).

39 Maji, D., Shaikh, S., Solanki, D. & Gaurav, K. Safety of statins. *Indian journal of endocrinology and metabolism* **17**, 636-646, doi:10.4103/2230-8210.113754 (2013).

40 de Denus, S., Spinler, S. A., Miller, K. & Peterson, A. M. Statins and liver toxicity: a meta-analysis. *Pharmacotherapy* **24**, 584-591 (2004).

41 Agarwal, R. Effects of statins on renal function. *The American journal of cardiology* **97**, 748-755, doi:10.1016/j.amjcard.2005.09.110 (2006).

42 Palmer, S. C. *et al.* HMG CoA reductase inhibitors (statins) for dialysis patients. *The Cochrane database of systematic reviews* **9**, CD004289, doi:10.1002/14651858.CD004289.pub5 (2013).

43 Baigent, C. *et al.* The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial. *Lancet* **377**, 2181-2192, doi:10.1016/S0140-6736(11)60739-3 (2011).

44 Vaughan, C. J. & Gotto, A. M., Jr. Update on statins: 2003. *Circulation* **110**, 886-892, doi:10.1161/01.CIR.0000139312.10076.BA (2004).

45 Kotseva, K. *et al.* EUROASPIRE III: a survey on the lifestyle, risk factors and use of cardioprotective drug therapies in coronary patients from 22 European countries. *European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology* **16**, 121-137, doi:10.1097/HJR.0b013e3283294b1d (2009).

46 Wei, M. Y., Ito, M. K., Cohen, J. D., Brinton, E. A. & Jacobson, T. A. Predictors of statin adherence, switching, and discontinuation in the USAGE survey: Understanding the use

of statins in America and gaps in patient education. *Journal of clinical lipidology* **7**, 472-483, doi:10.1016/j.jacl.2013.03.001 (2013).

47 Simpson, R. J., Jr. *et al.* Treatment pattern changes in high-risk patients newly initiated on statin monotherapy in a managed care setting. *Journal of clinical lipidology* **7**, 399-407, doi:10.1016/j.jacl.2013.06.002 (2013).

48 Benner, J. S. *et al.* Long-term persistence in use of statin therapy in elderly patients. *JAMA : the journal of the American Medical Association* **288**, 455-461 (2002).

49 Jackevicius, C. A., Mamdani, M. & Tu, J. V. Adherence with statin therapy in elderly patients with and without acute coronary syndromes. *JAMA : the journal of the American Medical Association* **288**, 462-467 (2002).

50 Foody, J. M., Joyce, A. T., Rudolph, A. E., Liu, L. Z. & Benner, J. S. Persistence of atorvastatin and simvastatin among patients with and without prior cardiovascular diseases: a US managed care study. *Current medical research and opinion* **24**, 1987-2000, doi:10.1185/03007990802203279 (2008).

51 Brown, M. T. & Bussell, J. K. Medication adherence: WHO cares? *Mayo Clinic proceedings. Mayo Clinic* **86**, 304-314, doi:10.4065/mcp.2010.0575 (2011).

52 Group, S. C. *et al.* SLCO1B1 variants and statin-induced myopathy--a genomewide study. *The New England journal of medicine* **359**, 789-799, doi:10.1056/NEJMoa0801936 (2008).

53 Voora, D. *et al.* The SLCO1B1*5 genetic variant is associated with statin-induced side effects. *Journal of the American College of Cardiology* **54**, 1609-1616, doi:10.1016/j.jacc.2009.04.053 (2009).

54 Bulbulia, R. *et al.* Effects on 11-year mortality and morbidity of lowering LDL cholesterol with simvastatin for about 5 years in 20,536 high-risk individuals: a randomised controlled trial. *Lancet* **378**, 2013-2020, doi:10.1016/S0140-6736(11)61125-2 (2011).

55 Wilke, R. A. *et al.* The Clinical Pharmacogenomics Implementation Consortium: CPIC Guideline for SLCO1B1 and Simvastatin-Induced Myopathy. *Clin Pharmacol Ther* **92**, 112-117 (2012).

56 Sathasivam, S. Statin induced myotoxicity. *European journal of internal medicine* **23**, 317-324, doi:10.1016/j.ejim.2012.01.004 (2012).

57 Oh, J., Ban, M. R., Miskie, B. A., Pollex, R. L. & Hegele, R. A. Genetic determinants of statin intolerance. *Lipids in health and disease* **6**, 7, doi:10.1186/1476-511X-6-7 (2007).

58 Marcoff, L. & Thompson, P. D. The role of coenzyme Q10 in statin-associated myopathy: a systematic review. *Journal of the American College of Cardiology* **49**, 2231-2237, doi:10.1016/j.jacc.2007.02.049 (2007).

59 Ruano, G. *et al.* Mechanisms of statin-induced myalgia assessed by physiogenomic associations. *Atherosclerosis* **218**, 451-456, doi:10.1016/j.atherosclerosis.2011.07.007 (2011).

60 Mangravite, L. M. *et al.* A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. *Nature* **502**, 377-380, doi:10.1038/nature12508 (2013).

61 Choe, C. U. *et al.* L-arginine:glycine amidinotransferase deficiency protects from metabolic syndrome. *Human molecular genetics* **22**, 110-123, doi:10.1093/hmg/dds407 (2013).

62 Ide, T. *et al.* GAMT, a p53-inducible modulator of apoptosis, is critical for the adaptive response to nutrient stress. *Molecular cell* **36**, 379-392, doi:10.1016/j.molcel.2009.09.031 (2009).

63 Ferrari, M. *et al.* Association between statin-induced creatine kinase elevation and genetic polymorphisms in SLCO1B1, ABCB1 and ABCG2. *European journal of clinical pharmacology* **70**, 539-547, doi:10.1007/s00228-014-1661-6 (2014).

64 Fiegenbaum, M. *et al.* The role of common variants of ABCB1, CYP3A4, and CYP3A5 genes in lipid-lowering efficacy and safety of simvastatin treatment. *Clinical pharmacology and therapeutics* **78**, 551-558, doi:10.1016/j.clpt.2005.08.003 (2005).

65 Keskitalo, J. E. *et al.* ABCG2 polymorphism markedly affects the pharmacokinetics of atorvastatin and rosuvastatin. *Clinical pharmacology and therapeutics* **86**, 197-203, doi:10.1038/clpt.2009.79 (2009).

66 Frudakis, T. N. *et al.* CYP2D6*4 polymorphism is associated with statin-induced muscle effects. *Pharmacogenetics and genomics* **17**, 695-707, doi:10.1097/FPC.0b013e328012d0a9 (2007).

67 Wilke, R. A., Moore, J. H. & Burmester, J. K. Relative impact of CYP3A genotype and concomitant medication on the severity of atorvastatin-induced muscle damage. *Pharmacogenetics and genomics* **15**, 415-421 (2005).

68 Needham, M. & Mastaglia, F. L. Statin myotoxicity: a review of genetic susceptibility factors. *Neuromuscular disorders : NMD* **24**, 4-15, doi:10.1016/j.nmd.2013.09.011 (2014).

69 Marciante, K. D. *et al.* Cerivastatin, genetic variants, and the risk of rhabdomyolysis. *Pharmacogenetics and genomics* **21**, 280-288, doi:10.1097/FPC.0b013e328343dd7d (2011).

70 Ruano, G. *et al.* Physiogenomic association of statin-related myalgia to serotonin receptors. *Muscle & nerve* **36**, 329-335, doi:10.1002/mus.20871 (2007).

71 Ruano, G. *et al.* Physiogenomic analysis links serum creatine kinase activities during statin therapy to vascular smooth muscle homeostasis. *Pharmacogenomics* **6**, 865-872, doi:10.2217/14622416.6.8.865 (2005).

72 Thompson, P. D., Clarkson, P. & Karas, R. H. Statin-associated myopathy. *JAMA* **289**, 1681-1690, doi:10.1001/jama.289.13.1681 (2003).

73 Vale, N. *et al.* Statins for acute coronary syndrome. *The Cochrane database of systematic reviews*, CD006870, doi:10.1002/14651858.CD006870.pub2 (2011).

74 Cziraky, M. J. *et al.* Risk of hospitalized rhabdomyolysis associated with lipid-lowering drugs in a real-world clinical setting. *Journal of clinical lipidology* **7**, 102-108, doi:10.1016/j.jacl.2012.06.006 (2013).

75 Keating, A. J., Campbell, K. B. & Guyton, J. R. Intermittent nondaily dosing strategies in patients with previous statin-induced myopathy. *The Annals of pharmacotherapy* **47**, 398-404, doi:10.1345/aph.1R509 (2013).

76 Austin, P. C. & Mamdani, M. M. Impact of the pravastatin or atorvastatin evaluation and infection therapy-thrombolysis in myocardial infarction 22/Reversal of Atherosclerosis with Aggressive Lipid Lowering trials on trends in intensive versus moderate statin therapy in Ontario, Canada. *Circulation* **112**, 1296-1300, doi:10.1161/CIRCULATIONAHA.104.531582 (2005).

77 Keidar, S. & Gamliel-Lazarovich, A. Viewpoint: personalizing statin therapy. *Rambam Maimonides medical journal* **4**, e0008, doi:10.5041/RMMJ.10108 (2013).

78 Zhou, S. F. *et al.* Clinical pharmacogenetics and potential application in personalized medicine. *Current drug metabolism* **9**, 738-784 (2008).

79 Kraft, P., Zeggini, E. & Ioannidis, J. P. Replication in genome-wide association studies. *Statistical science : a review journal of the Institute of Mathematical Statistics* **24**, 561-573, doi:10.1214/09-STS290 (2009).

80 Ioannidis, J. P. Why most published research findings are false. *PLoS medicine* **2**, e124, doi:10.1371/journal.pmed.0020124 (2005).

81 Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L. & Rothman, N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute* **96**, 434-442 (2004).

82 Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-856, doi:10.1038/35015718 (2000).

83 Ioannidis, J. P., Thomas, G. & Daly, M. J. Validating, augmenting and refining genome-wide association signals. *Nature reviews. Genetics* **10**, 318-329, doi:10.1038/nrg2544 (2009).

84 Trikalinos, T. A., Ntzani, E. E., Contopoulos-Ioannidis, D. G. & Ioannidis, J. P. Establishment of genetic associations for complex diseases is independent of early study findings. *European journal of human genetics : EJHG* **12**, 762-769, doi:10.1038/sj.ejhg.5201227 (2004).

85 Bitzur, R., Cohen, H., Kamari, Y. & Harats, D. Intolerance to statins: mechanisms and management. *Diabetes care* **36 Suppl 2**, S325-330, doi:10.2337/dcS13-2038 (2013).

86 Dumitrescu, L. *et al.* Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. *Genetics in medicine : official journal of the American College of Medical Genetics* **12**, 648-650, doi:10.1097/GIM.0b013e3181efe2df (2010).

87 Burnett, M. S. *et al.* Reliability of self-reported ancestry among siblings: implications for genetic association studies. *American journal of epidemiology* **163**, 486-492, doi:10.1093/aje/kwj057 (2006).

88 Lee, S., Epstein, M. P., Duncan, R. & Lin, X. Sparse principal component analysis for identifying ancestry-informative markers in genome-wide association studies. *Genetic epidemiology* **36**, 293-302, doi:10.1002/gepi.21621 (2012).

89 Turner, S. *et al.* Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet* **Chapter 1**, Unit1 19, doi:10.1002/0471142905.hg0119s68 (2011).

90 Pearson, T. A. & Manolio, T. A. How to interpret a genome-wide association study. *JAMA : the journal of the American Medical Association* **299**, 1335-1344, doi:10.1001/jama.299.11.1335 (2008).

91 Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annual review of genomics and human genetics* **10**, 387-406, doi:10.1146/annurev.genom.9.081307.164242 (2009).

92 Pirinen, M., Donnelly, P. & Spencer, C. C. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nature genetics* **44**, 848-851, doi:10.1038/ng.2346 (2012).

93 Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American journal of human genetics* **83**, 311-321, doi:10.1016/j.ajhg.2008.06.024 (2008).

94 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93, doi:S0002-9297(11)00222-9 [pii] 10.1016/j.ajhg.2011.05.029 (2011).

95 Doring, F. *et al.* Single nucleotide polymorphisms in the myostatin (MSTN) and muscle creatine kinase (CKM) genes are not associated with elite endurance performance. *Scandinavian journal of medicine & science in sports* **21**, 841-845, doi:10.1111/j.1600-0838.2010.01131.x (2011).

96 Wu, Q. Y. *et al.* Disrupting of E79 and K138 interaction is responsible for human muscle creatine kinase deficiency diseases. *International journal of biological macromolecules* **54**, 216-224, doi:10.1016/j.ijbiomac.2012.12.034 (2013).

97 Borges, L., Hsu, M. L., Fanger, N., Kubin, M. & Cosman, D. A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. *J Immunol* **159**, 5192-5196 (1997).

98 Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551-1561, doi:10.1002/pmic.200300772 (2004).

99 Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248-1250, doi:10.1038/nbt1210-1248 (2010).

100 Wang, M. *et al.* PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics* **11**, 492-500, doi:10.1074/mcp.O111.014704 (2012).

101 *http://www.ncbi.nlm.nih.gov/geoprofiles/81912982*, <http://www.ncbi.nlm.nih.gov/geoprofiles/81912982> (

102 Warren, G. L. *et al.* CK-MM autoantibodies: prevalence, immune complexes, and effect on CK clearance. *Muscle Nerve* **34**, 335-346, doi:10.1002/mus.20594 (2006).

103 Ebbeling, C. B. & Clarkson, P. M. Muscle adaptation prior to recovery following eccentric exercise. *European journal of applied physiology and occupational physiology* **60**, 26-31 (1990).

104 Hyatt, J. P. & Clarkson, P. M. Creatine kinase release and clearance using MM variants following repeated bouts of eccentric exercise. *Medicine and science in sports and exercise* **30**, 1059-1065 (1998).

105 Delanghe, J., De Scheerder, I., De Buyzere, M., Algoed, L. & Robbrecht, J. Macro CK type 1 as a marker for autoimmunity in coronary heart disease. *Atherosclerosis* **60**, 215-219 (1986).

106 Dhanoa, B. S., Cogliati, T., Satish, A. G., Bruford, E. A. & Friedman, J. S. Update on the Kelch-like (KLHL) gene family. *Human genomics* **7**, 13, doi:10.1186/1479-7364-7-13 (2013).

107 Liu, L. *et al.* Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS genetics* **9**, e1003443, doi:10.1371/journal.pgen.1003443 (2013).