GENOME-WIDE ASSOCIATION STUDY OF STATIN-INDUCED MYOPATHY IN PATIENTS RECRUITED USING THE UK CLINICAL PRACTICE RESEARCH DATALINK

Daniel F. Carr[1], Ben Francis[2], Andrea L. Jorgensen[2], Eunice Zhang[1], Hector Chinoy[3,4], Susan R. Heckbert[5], Joshua C. Bis[6], Jennifer A. Brody[6], James Floyd[6], Bruce M. Psaty[6], Mariam Molokhia[7], Maryse Lapeyre-Mestre[8], Anita Conforti[9], Ana Alfirevic[1], Tjeerd van Staa[10,11], Munir Pirmohamed[1]

[1]Wolfson Centre for Personalised Medicine, Dept. of Molecular and Clinical Pharmacology,
[2]Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, Liverpool, UK.
[3]Rheumatology Department, Salford Royal NHS Foundation Trust, Manchester Academic Health Science Centre, Salford, UK.
[4]NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK
[5]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA, USA
[6]Cardiovascular Health Research Unit, Department of Medicine, University of Washington; Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA
[7]School of Population Health and Environmental Sciences, King's College London, UK.
[8] Medical and Clinical Pharmacology, CHU Toulouse, Toulouse, France
[9] University Hospital, Verona, Italy
[10]Health e-Research Centre, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.
[11]Faculty of Science, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, the Netherlands.

RUNNING TITLE: **Genome-wide association study of statin myopathy**

CORRESPONDING AUTHOR:
Dr Daniel F. Carr
Wolfson Centre for Personalised Medicine
Department of Molecular and Clinical Pharmacology,
University of Liverpool
Block A: Waterhouse Buildings,
1-5 Brownlow Street
Liverpool, L69 3GL
Phone: 0151 795 5392, Email: d.carr@liv.ac.uk

CONFLICT OF INTEREST

B. Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. No other author declares any conflict of interest.

FUNDING INFORMATION

This work was funded by a grant from the e-Health Initiative funded jointly by the Medical Research Council (reference: MC_qA137929), Wellcome Trust, EPSRC and ESRC, and HL078888. . MP is a NIHR Senior Investigator. MP also thanks the MRC Centre for Drug Safety Science and NIHR CLAHRC North West Coast for infrastructure support. Collaborators of UKMYONET are acknowledged in (1). This report includes independent research supported by the NIHR Biomedical Research Centre Funding Scheme. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

The EUDRAGENE project was supported with a concerted action grant from the European Commission 5th Framework QLRI-CT-2002- 02757, by the Serious Adverse Events Consortium (SAEC), and the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London.

KEYWORDS: statin, myopathy, genome wide association study, adverse dug reaction

ABSTRACT

Statins can be associated with myopathy. We have undertaken a Genome-wide Association Study (GWAS) to discover and validate genetic risk-factors for statin-induced myopathy in a "real-world" setting. 135 statin myopathy patients recruited via the UK Clinical Practice Research Datalink were genotyped using the Illumina OmniExpress Exome v1.0 Bead Chip, and compared to the Wellcome Trust Case-Control Consortium (n=2501). Nominally statistically significant SNP signals in the GWAS ($p<5x10^{-5}$) were further evaluated in several independent cohorts (comprising 332 cases and 449 drug-tolerant controls). Only one (rs4149056/c.521C>T in the *SLCO1B1* gene) SNP was genome-wide significant in the severe myopathy (CK>10xULN or rhabdomyolysis) group ($p=2.55x10^{-9}$; OR 5.15, 95%CI 3.13-8.45). The association with *SLCO1B1* was present for several statins and replicated in the independent validation cohorts. The data highlight the role of *SLCO1B1* c.521C>T SNP as a replicable genetic risk-factor for statin myopathy. No other novel genetic risk-factors with a similar effect size were identified.

INTRODUCTION

HMG-CoA inhibitors, or statins, are a widely prescribed class of drugs for the treatment of hyperlipidaemia. Though generally well-tolerated, a small proportion of patients can develop muscle related adverse-effects (2). These can range from mild muscle pain without creatine phosphokinase CPK elevation, where causality can be difficult to assess, to myopathy where the CPK becomes elevated (>4xULN), with the most extreme reactions being rhabdomyolysis with renal impairment (3). A systematic review suggested that the incidence of statin-induced mild muscle pain is 190 cases/100,000 patient years with myopathy and rhabdomyolysis at 5 and 1.6 cases/100,000 patient years, respectively (4).

A number of genetic studies (5-9) have identified a non-synonymous polymorphism (p.V147L/c.521C>T) in the *SLCO1B1* gene (rs4149056), encoding an hepatic uptake transporter protein as a predisposing factor for statin myopathy. Our pilot proof-of-principle candidate gene study which analysed a subset of the cohort in this study (77 cases and 372 statin-tolerant controls) replicated the association between the *SLC01B1* gene polymorphism and statin myopathy (10) showing the validity of our recruitment strategy via the UK Clinical Practice Research Datalink (CPRD), an electronic health record database. The association with the *SLCO1B1* gene has biological plausibility in that it leads to impaired hepatic uptake of statins by the transporter (11, 12), causing increased circulating drug concentrations (13, 14). However, to date no other clinically relevant, reproducible genetic variants have been identified.

Other genetic markers have been associated with statin-myopathy including polymorphisms in the coenzyme Q2 4-hydroxybenzoate polyprenyltransferase (*COQ2*) (15) and human eyes shut ortholog (*EYS*) (16) genes, but have not been independently

replicated. The genetic association of statin myopathy with the *GATM* gene (17) has also not been replicated (18, 19).

Utilising statin-myopathy patients recruited via CPRD (20), the aims of our study were two-fold: to undertake a genome-wide association study to identify novel genetic risk factors predisposing individuals to statin-induced myopathy; and to validate any association in independent patient groups and perform meta-analysis of any association signals. Taking all cohorts together, this represents the largest genome-wide association study of the pharmacogenetics of statin myopathy undertaken to date.

RESULTS

Case-Control Discovery GWAS

A total of 128 out of 135 myopathy case samples and 654,642 SNPs passed the predefined genotyping QC criteria. Of the 7 individuals excluded, 3 failed sample call-rate criteria, 3 failed the gender identity check (due to sample mislabelling) and 1 individual was excluded as a population outlier after principal component analysis (figure S1). The individual statins responsible for the muscle toxicity are shown in table S1.

From the case-control discovery GWAS, a total of 21 SNPs were initially identified as notionally significant ($p<5x10^{-5}$) (12 from the all myopathy analysis and 9 from the severe myopathy analysis) (Figure 2A and B, respectively). However, only one signal reached genome-wide significance: rs4149056 in the *SLCO1B1* locus ($p=2.5x10^{-9}$) (Table 1). Sensitivity analysis of discovery cases (all myopathy) for simvastatin cases only (figure S3 and table S4) showed no genome-wide significant association signals ($p>5x10^{-8}$) though *SLCO1B1* was amongst the top associated loci. Similarly sensitivity

analysis of atorvastatin cases only (figure S4 and table S5) also showed no genome-wide significant associations. *SLCO1B1* was not identified in the top associated loci.

Univariate analysis of non-genetic variables for myopathy (n=128) and severe myopathy (n=32) cases versus statin tolerant controls (n=585) was undertaken (Table S1). Age, gender, BMI, antihypertensive co-medication, occurrence of cramps and previous history of hypertension showed an association with $p<0.10$ for all myopathy and mean daily dose, age and occurrence of cramps showed an association with $p<0.10$ for "severe myopathy". These variables were incorporated into the replication cohort logistic regression model for case-control analysis.

Replication cohort analysis

Twenty-one SNPs below a threshold of $p<5x10^{-5}$ in the discovery GWAS analysis were carried forward for genotyping in the "Replication Cohort" (consisting of the CPRD statin-exposed controls (n=585) and the EUDRAGENE statin-myopathy cases (n=19)). A total of 9 SNPs (2 from the severe myopathy analysis and 7 from the all myopathy analysis) however were subsequently excluded because of inability to design TaqMan or MassArray assays due to proximal sequence constraints (n=4), low genotyping call rate (n=4) and Hardy-Weinberg deviation (n=1). Thus, a total of 12 SNPs was genotyped (5 for all myopathy and 7 for severe myopathy) (Table 1).

Data from simvastatin and atorvastatin cases only sensitivity analyses for the 12 SNP initially identified in the overall discovery cohort is reported in Table S3. None of the 12 signals showed a statistically significant association with atorvastatin myopathy. Eight loci were notionally associated with simvastatin "all myopathy" albeit not to genome-wide significance.

Replication cohort: Candidate gene analysis of the 19 EUDRAGENE myopathy cases (13 severe) with the 585 statin tolerant controls identified 2/12 associations of nominal significance ($p<0.05$) both of which were for the severe myopathy phenotype (Table 1). These were *SLCO1B1* rs4149056 (p=0.001, OR 3.98, 95%CI 1.75-9.03) and *SLCO1A2* rs4149000 (p=0.05, OR 2.53, 95%CI 1.00-6.39). No other statistically significant associations were observed in the other 9 SNPs ($p<0.05$). Minor allele frequencies for the 12 SNPs were comparable for both simvastatin and atorvastatin controls

Association Signal Validation

"Simvastatin" Cohort: Summary statistics from this cohort (Table 1) showed a genome wide significant association for 2 of 12 SNPs, both of which were associated with the severe myopathy phenotype in the initial analysis. Both rs4149056 in *SLCO1B1* and rs4149000 in *SLCO1A2* were significantly associated with both definite myopathy ($p=7.30\times10^{-14}$ and $p=7.62\times10^{-11}$ respectively) and the incipient or definite myopathy phenotype ($p=1.33\times10^{-11}$ and 6.49×10^{-12}). None of the other SNPs were significantly associated with either myopathy phenotype ($p>0.05$).

*"Cerivastatin" Cohort***:** Summary statistics from the cerivastatin rhabdomyolysis validation cohort (Table 1) showed that the same 2 SNPs (rs4149056 and rs4149000) showed a significant (albeit not to a genome-wide threshold) association (3.90×10^{-4} and 0.007 respectively). None of the other SNPs showed an association ($p>0.05$).

Further analysis of the SLCO1B1 and SLCO1A2 signals

The 2 SNPs which appear to be strongly associated with statin myopathy in the discovery and replication cohorts are within 2 gene loci (*SLCO1B1* and *SLCO1A2*) that are in a strong block of LD in the discovery cohort (data not shown). Conditional

analysis correcting for the *SLCO1B1* rs4149056 genotype was undertaken. The analysis of the 32 discovery severe myopathy cases vs 585 statin tolerant controls abolished the rs4149000 genotype association (p=0.934), as well as for the all myopathy phenotype (p=0.368) indicating that the 2 risk alleles are not acting in *cis* on the same haplotype and that the *SLCO1A2* association is not acting independently of *SLCO1B1*.

Meta-analysis of SLCO1B1 and SLCO1A2 association signals

Meta-analysis combining the discovery and all replication cohorts (limited to severe cases only) (271 cases vs. 7,493 controls) yielded a meta-analytic genome-wide significant P value for SLCO1B1, rs4149056 ($p=2.63x10^{-18}$; OR 2.99, 95% 2.34-3.82) (Table 1), highlighting the predominant role of *SLCO1B1* in predisposing to myopathy caused by a variety of statins. An increased statistical significance of the association signal within the *SLCO1A2* was also observed but no other meta-analysis demonstrated an increased statistical significance of the signal initially identified in the discovery case-control study (Table 1). Meta-analysis of the *SLCO1B1* signal for rs4149056 limited to the simvastatin-exposed cases and controls only (Figure 3) led to a p-value of $1.46x10^{-21}$ (OR 5.91, 95% CI 4.10-8.51; $I^2=1.00$) for the severe myopathy phenotype, and a p-value of $2.01x10^{-14}$ (OR 2.75, 95% CI 2.12-3.56, $I^2=0.78$) for the all myopathy phenotype.

DISCUSSION

Our discovery GWAS identified 12 SNPs which were nominally associated with either all myopathy (CK>4xULN ± muscle symptoms) or severe myopathy (CK>10xULN or rhabdomyolysis). Replication was undertaken in 3 separate patient cohorts, which

showed that only the previously identified (7, 10) and widely replicated c.521C>T variant (rs4149056) in *SLCO1B1,* and an intronic SNP in the *SLCO1A2* gene, were risk factors for statin myopathy though the latter. The latter however was not significant after adjustment for *SLCO1B1* genotype. Our data concur with a previous statin myopathy GWAS (10), as well as our own pilot data (10) and other candidate gene studies (5, 6, 9, 21), that the *SLCO1B1* c.521C>T polymorphism (rs4149056) is the predominant genetic risk factor for statin-induced myopathy. Our finding is also consistent with a recent meta-analysis of 14 studies comprising 3265 myopathy patients and 7743 controls (22). Additionally, previously reported associations in the *GATM* (17), *COQ2* (15), and *EYS* (16) gene loci were not replicated.

Our discovery cohort was heterogeneous in terms of the severity of myopathy and statin implicated. The association with the rs4149056 variant in *SLCO1B1* was stronger in patients with the severe form of myopathy (CPK>10x ULN or rhabdomyolysis) irrespective of the statin involved, reaching genome-wide significance (Figure 3). The lower effect size observed in patients with the less severe form of statin myopathy (defined in our discovery cohort as CK>4xULN) may reflect multiple causes in the mildly affected cases, and the difficulty in attributing causality to statins in all cases.

Of the different statins implicated, simvastatin was the most common accounting for 66% of our cases, and 69% of the severe cases. Meta-analysis of our discovery severe myopathy cohort with the simvastatin definite myopathy cohort did strengthen the association ($p=7.17x10^{-19}$) with little evidence of heterogeneity between the two ($I^2=1$). Further incorporation of the cerivastatin cohort marginally weakened the association in keeping with the different effect sizes of the *SLCO1B1* locus for different statins. These data are consistent with the fact the pharmacokinetic effect of the *SLCO1B1* variant is

greatest for simvastatin. The AUC for simvastatin acid is increased by 221% in CC homozygotes compared with those individuals who are TT homozygotes for the *SLCO1B1* c.521C>T polymorphism (23). Corresponding values for the other statins are as follows: atorvastatin (145%) (24), fluvastatin (19%) (25), lovastatin acid (186%) (26), pitavastatin (208%) (27), pravastatin (91%) (25), and rosuvastatin (65%) (24). No similar data is available for cerivastatin. Based on these pharmacokinetic data, and the results of our data, together with the recent meta-analysis (22), it might be suggested that *SLCO1B1* locus is important for all statins, but the effect size will likely vary being greatest for simvastatin and lowest for fluvastatin.

The aim of our GWAS was to identify other loci associated with statin myopathy. Apart from the association with *SLCO1B1*, we also identified an association with a SNP located in the 5'UTR of the *SLCO1A2* locus (Figure S2). This signal however was not independent of the *SLCO1B1* signal. *SLCO1A2* encodes SLCO1A2, a hepatic-expressed efflux transporter (28), which is responsible for the sodium-independent transport of organic anions such as bromosulfophthalein, taurocholate and unconjugated cholate bile acids (29, 30). SLCO1A2 has also been shown to have substrate specificity for pitavastatin (31) and rosuvastatin (32), but to date, there is no evidence of its ability to transport simvastatin or atorvastatin. Determination of whether the *SLCO1A2* locus can act as an independent risk factor for statin myopathy will require a much larger sample size. None of the other loci identified in the discovery cohort were replicated in any of the cohorts, and meta-analysis did not provide any indication that these loci acted as predisposing factors for statin myopathy.

For a number of years, the c.521T>C *SLCO1B*1 variant has been recognised as a clinically important risk factor for statin-induced myopathy, particularly with regards to

simvastatin, and to a lesser extend atorvastatin. Indeed, summary of product characteristics labelling for both drugs (33, 34) highlights the increased risk of myopathy in individuals who are carriers of the low-activity C allele. It has been suggested that the maximum dose of simvastatin, pitavastatin and atorvastatin should be reduced by 4-fold in individuals who are CC homozygotes based on pharmacokinetic calculations(35). Interestingly, a recent small randomised trial (n=159) of patients not on statins because of prior myalgia attributed to a statin showed that providing information on the *SLCO1B1* genotype improved statin re-initiation and LDL-cholesterol lowering, but not adherence, when compared with the usual care arm (36).

It is clear that whilst S*LCO1B1* is a key risk factor for statin myopathy, it does not explain a significant proportion of the inter-individual variability in statin toxicity. Genetic studies to date have been limited by recruitment of significant numbers of cases of what is a rare ADR. As such many studies lack statistical power to detect small effect sizes and have in fact only identified the "low hanging fruit". It is possible that much of the heritability of statin myopathy risk may lie either with rare variants of large effect sizes or with other common genetic loci with small to modest effect sizes which will require much larger patient numbers. A major issue is that we do not fully understand the mechanism of statin myopathy and muscle damage, apart from the fact that high doses or high statin concentrations, increase risk. Further functional studies to uncover the mechanisms of muscle damage induced by statins will be important in elucidating further predisposing factors.

In conclusion, our data further confirm the predominant role of the *SLCO1B1* c.521 C>T polymorphism in predisposing to statin-induced myopathy in a "real-world" patient population, in particular with simvastatin. Moreover, the data failed to identify other

statin-myopathy associated genetic risk factors. However, this meta-analysis is relatively small and may lack statistical power. The additional signal identified in the *SLCO1A2* locus was not independent of *SLCO1B1*, but may require further investigation from a functional perspective to determine the role of this transporter in statin transport.

METHODS

The study design is summarised in Figure 1. Briefly, a discovery case-control genome-wide associations study was undertaken, followed by candidate variant replication in a second case-control cohort. The same association signals were validated in existing data from 2 independent validation case control studies.

CPRD Case-Control Recruitment

From a cohort of approximately 600,000 patients receiving statins identified in the CPRD (www.cprd.com), a case-control design was used to identify suitable patients for the study as previously described (20, 37). Participation was restricted to Caucasians ≥18 years of age and with the first ever statin prescription at least 1 year after the start of CPRD data collection.

All cases conformed to internationally agreed standards for statin induced myopathy and rhabdomyolysis (3). Cases were categorised into two: (a) myopathy: patients who discontinued their implicated statin with a rise in CPK >4x ULN; and (b) severe myopathy: individuals with a history of rhabdomyolysis or CPK>10x ULN after statin exposure. Controls were defined as individuals receiving statins for at least 3 months with no previous history of abnormal serum CPK measurements.

GPs were contacted with a list of potential cases and/or controls identified from their practice. They were first asked to review the medical records of listed individuals and remove any patients they considered to not fulfil the case or control criteria. They were then asked to contact suitable patients by letter requesting participation. Individuals who gave written informed consent were invited to provide either a saliva sample (by post) or a blood sample (by visiting the practice). All samples were then forwarded onto The University of Liverpool for processing. To preserve anonymity, patient and practice identifier codes were used throughout the recruitment process and all patient contact was via the GP only. A total of 149 myopathy cases and 585 controls were recruited between April 2010 and June 2013 though only 135 cases were available at the time of genotyping (20). Relevant clinical and demographic data (summarised in Table 1) was retrospectively obtained from the CPRD.

In addition to the above, 5 supplemental cases of statin-induced myopathy conforming to our phenotype criteria were identified in the tertiary adult muscle clinic run through Salford Royal NHS Foundation Trust, UK, and recruited into the UKMYONET genetic study (1).

Additional Cohorts and Studies

Three replication cohorts were utilised to validate associations identified in the case-control discovery GWAS (summarised in Figure 1)

EUDRAGENE Cohort: A total of 19 Adult (>18 years of age) statin-exposed myopathy cases (5 simvastatin, 5 atorvastatin, 2 rosuvastatin, 1 fluvastatin) matching the case phenotype (defined above) were recruited using spontaneous adverse drug reaction reports and laboratory CPK results from the European pharmacovigilance centre and UK primary care practices from 2006 to 2012 via the EUDRAGENE collaborative

network (38). All cases were adjudicated using internationally accepted criteria for myopathy and severe myopathy (3) by an independent panel , consisting of clinicians and pharmacovigilance experts. Of the 19 cases, 13 were categorised as having severe myopathy. EUDRAGENE cases were combined with the 585 CPRD statin-tolerant controls to form the "Replication cohort".

The "Simvastatin" validation case-control study: A total of 141 simvastatin myopathy patients were recruited via the SEARCH collaborative group (7). These consisted of 54 "definite" statin myopathies (defined as otherwise unexplained muscle symptoms with CK>10x ULN) and 87 "incipient" statin myopathies (defined as ALT>1.7x ULN and CK both >5x baseline and >3x ULN). For the purpose of meta-analysis, the "definite" myopathy phenotype was aligned to this studies severe myopathy phenotype with both "incipient and definite" aligning to the all myopathy phenotype. A total of 4,046 statin-tolerant controls were included from the SEARCH and Heart Protection Study (HPS) groups (7, 39).

The "Cerivastatin" validation case-control study The study sample consisted of data from 172 cases of cerivastatin rhabdomyolysis and 361 statin-using controls from the Heart and Vascular Health Study (HVH) as previously reported (8). All cases and controls were of European ancestry. Controls with creatine kinase levels >10X ULN were excluded.

Study approvals

Ethical approval for recruitment via CPRD was obtained from the National Research Ethics Committee North West 2 – Liverpool Central. Furthermore, approval to use the CPRD data was obtained from the Independent Scientific Advisory Committee (ISAC) at the Medicines and Healthcare products Regulatory Agency. In addition, site-specific

approval (SSI) to contact the GP practices was obtained for each of 132 primary care trusts across the UK, as described previously (40). Written informed consent was obtained from all study subjects or their guardians. The UKMYONET study was approved by the North West Research Multi-Centre Research Ethics Committee (MREC 98/8/86), and all participants gave written informed consent.

Multi-Centre ethics approval was obtained from the South East Research Ethics Committee for the SEARCH study, and from the local ethics committees covering each of the 69 UK hospitals involved in the Heart Protection Study.

The recruitment of cerivastatin case subjects was approved by the University of Washington Institutional Review Board and the use of the HVH study subjects was approved by the Group Health Subjects Review Committee.

DNA Extraction and Genotyping

Case-Control "Discovery" Cohort: For the CPRD recruits (cases and controls), genomic DNA was extracted from 5ml whole blood or 2ml Saliva (collected using the Oragene DNA Sampling kit , DNAGenotek, Ontario, Canada) using the Chemagic Magnetic Module (MSM) 1 system as per the manufacturer's protocol (Chemagen Biopolymer-Technologie AG, Baesweiler, Germany).

At the time of analysis, DNA samples from a total of 135 myopathy cases from the discovery cohort were available. At least 1.5μg DNA from myopathy cases was genotyped for a total of 982,958 SNPs by ARK-Genomics, University of Edinburgh (Edinburgh, UK) using the Illumina OmniExpress Exome v1.0 Bead Chip array according to the manufacturers protocol (Illumina Inc., San Diego, CA).

Discovery case-control study- population controls: Population control genotype data for the initial discovery case-control GWAS was obtained from the Wellcome-Trust Case-Control consortium 2 (WTCCC2) cohort of 2,501 individuals from the UK Blood Service.

Replication Study: All 585 CPRD statin-tolerant controls and 19 myopathy patients from the EUDRAGENE cohort were genotyped for statistically significant association signals identified in the case-control discovery GWAS ($p<5x10^{-5}$) using either the Agena MassArray iPLEX platform (Agena Biosciences Inc., San Diego CA) or TaqMan real-time PCR SNP genotyping assay (Life Technologies, Paisley, UK) according to the manufacturer's protocols.

Genotyping QC and Imputation

Case-Control "Discovery" Cohort: Cases identified via CPRD recruitment were excluded if they failed to meet the following criteria: a) gender as determined by the "Sex Check" function within PLINK (41) differed from that reported in the clinical data; b) genotype call-rate <90%; and c) principle component analysis (PCA) (using SNPRelate (42)in R v3.01) demonstrated that the individual did not cluster with the HapMap CEU (Utah residents with European ancestry) population (Figure S1).

SNPs, in both the discovery and replication cohorts, were excluded if: a) minor allele frequency (MAF) <0.01, b) Hardy-Weinberg Equilibrium (HWE) p<0.0001 and c) genotype success rate<95%. All QC analysis was undertaken using PLINK v1.07 (41) unless otherwise stated.

For the purpose of the discovery case-control study, the CPRD case genotype dataset was merged with the WTCCC dataset prior to SNP phasing using SHAPEIT (43) and

imputation using IMPUTE2 (44, 45) was undertaken using 1000 genome phase 3 reference panel.

The "simvastatin" validation case-control study: Genotype imputation was undertaken using minimac (46) with the 1000 genomes European reference panel. Data for the association signal SNPs were provided for the validation and meta-analysis.

The "cerivastatin" validation case-control study: Samples were excluded from analysis for sex mismatch or call rate <95%. The following variant exclusions were applied to obtain a cleaned set of variants for imputation: call rate <97%, HWE $P<10^{-5}$, >2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency = 0. MaCH (46) was used to pre-phase the genotypes. The phased genotypes were imputed into a reference panel of 1,092 individual of multiple ethnicities from the Phase 1 (version 3) haplotypes of 1000 Genomes project using minimac (46). Genotyping and SNP calling was performed using the Illumina 370CNV Bead Chip as previously described (8). Data for the association signal SNPs were provided for the validation and meta-analysis.

Statistical Analysis

The study design and statistical analysis is summarised in Figure 1. In the discovery phase, cases passing genotype QC (n=128) recruited via CPRD were compared with WTCCC2 controls (n=2,501) using a logistic regression analysis undertaken in SNPTest (47) and adjusting for the first two principle components as covariates. All SNPs giving a p-value for association of $<5x10^{-5}$ were genotyped in the statin-tolerant cohort (n=585) and EUDRAGENE cohort (n=19) which together formed the replication cohort

A univariate analysis of non-genetic covariates (Chi-square for categorical outcomes and Student's T-Test for continuous variables) was undertaken (Table 1) using SPSS version 17.0. Variables demonstrating a p-value <0.10 between the discovery cohort cases and tolerant controls were carried forward and adjusted for in the SNP association analyses (Table S1). Logistic regression analysis of the candidate SNPs in the cases (discovery and replication) and statin-tolerant controls was undertaken using SNPTest. Meta-analysis of the combined discovery and replication cohorts along with the 2 validation studies was undertaken using a fixed-effects model with inverse-variant effect size weighting in GWAMA (48). Forest plots were prepared using the 'forestplot' function in R.

STUDY HIGHLIGHTS

What is the current knowledge on the topic?

Risk of statin-induced myopathy is associated with variation of the *SLCO1B1* gene which encodes the OATP1B1 hepatic uptake transporter, of which statins are substrates. To date no other validated genetic risk factors have been identified.

What question did this study address?

Undertaking a genome-wide association study in a "real-world" patient cohort recruited via the Clinical Practice Datalink (CPRD), this study aimed to determine whether any other novel genetic risk loci for statin myopathy could be identified.

What knowledge does this study add to our knowledge?

The study suggests that aside from SLCO1B1, no other risk loci for statin myopathy are apparent. The unexplained statin myopathy risk is likely due to non-genetic risk factors or the influence of rare genetic variants analysed in this study

How might this change clinical pharmacology or translational science?

Common genetic variants do not appear to explain statin myopathy risk. The data presented seem to suggest that future translational work in this field should focus on rare variant analysis and on identifying non-genetic risk factors.

ACKNOWLEDGEMENTS

The authors thank the following people:

- Sir Rory Collins, Jane Armitage, Jemma Hopewell and Sarah Parish from UK Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford for provision of summary statistical data from their case-control analysis of 141 myopathy cases and 4,054 controls (The "simvastatin" validation case-control study)
- Jennifer Campbell, Mark Hobbs and Gerard McCann and the Medicines and Healthcare products Regulatory Agency for their help in facilitating the recruitment of patients to the STAGE study via the Clinical Practice Research Datalink.
- All contributors to the CPRD, the recruiting general practice physicians and patients who participated in the study

- All EUDRAGENE and SEARCH collaborators, patients and staff who supported this study.

AUTHOR CONTRIBUTIONS

DFC, BF, ALJ, BMP, MP wrote the manuscript. MP, TvS, BMP designed research. DFC, EZ, HC, SRH, JCB, JAB, JF, BMP, MM, ML-M, AC, AA, TvS, MP performed research. DFC, BF, ALJ, AA, TvS, MP analyzed data.

Table 1. SNPs suggested to be associated with A) all statin-induced myopathy and B) severe myopathy from the discovery case-control analysis; replication analysis; independent simvastatin and cerivastatin study analyses and the combined meta-analysis. Data indicates p-values and odds ratios (95% CI) (per-allele) derived from logistic regression for discovery cohort vs. WTCCC cohort (n=2,501). Only associations <$5x10^{-5}$ in the initial discovery cohort are shown with those reaching genome-wide significance ($p<5x10^{-8}$) highlighted in **bold**. NA denotes where data is not available.

A) All Myopathy

			Discovery Case-Control Study CPRD cases (n=128) vs. WTCCC (n=2,501)		**Replication Study** EUDRAGENE cases (n=19) vs. CPRD statin-Tolerant (n=585)		**Simvastatin validation case-control study** Definite/Incipient myopathy cases (n=141) vs. controls (n=4,046)		**Cerivastatin validation case-control study** Cases (n=172) vs. controls (n=361)		**Combined Meta-analysis** Cases (n=460) vs. controls (n=7,493)		
rs#	**Chr**	**Gene**	**per allele OR (95%CI)**	**p**	**per allele OR (95% CI)**	**p**	**per allele OR (95% CI)**	**p**	**per allele OR (95% CI)**	**p**	**per allele OR (95% CI)**	**p**	**I^2**
rs36121096	5	*PDE4D*	3.82 (2.20-6.62)	$2.0x10^{-5}$	0 (0-∞)	1.00	1.37 (0.56-6.07)	0.61	1.51 (0.60-3.82)	0.38	2.01 (1.22-3.31)	0.006	0.00
rs55902659	5	*SLC12A2*	0.44 (0.31-0.66)	$4.9x10^{-6}$	0.42 (0.15-1.18)	0.10	0.81 (0.76-1.90)	0.24	1.00 (0.72-1.37)	0.99	0.74 (0.59-0.92)	0.008	0.84
rs17359612	9	*TLE1*	2.49 (1.71-3.64)	$1.1x10^{-5}$	1.59 (0.58-4.35)	0.36	1.25 (0.51-4.17)	0.54	1.21 (0.63-2.33)	0.56	1.67 (1.19-2.34)	0.003	0.00
rs79860430	14	*ATG14*	2.59 (1.76-3.82)	$8.4x10^{-6}$	0 (0-∞)	1.00	1.13 (0.34-8.16)	0.81	1.27 (0.62-2.58)	0.51	2.17 (1.48-3.17)	$7.61x10^{-5}$	0.53
rs77855582	16	*GALNS*	3.88 (2.25-6.69)	$1.9x10^{-5}$	3.60 (0.99-13.0)	0.05	1.61 (0.78-4.55)	0.45	NA	NA	NA	NA	NA

B) Severe Myopathy

			Discovery Case-Control Study CPRD cases (n=32) vs. WTCCC (n=2,501)		**Replication Study** EUDRAGENE cases (n=13) vs. CPRD Statin-Tolerant (n=585)		**Simvastatin validation case-control study** Definite myopathy cases (n=54) vs. controls (n=4,046)		**Cerivastatin validation case-control study** Cases (n=172) vs. controls (n=361)		**Combined Meta-analysis** Cases (n=271) vs. controls (n=7,493)		
rs#	**Chr**	**Gene**	**per allele OR (95%CI)**	**p**	**per allele OR (95% CI)**	**p**	**per allele OR (95% CI)**	**P**	**per allele OR (95% CI)**	**p**	**per allele OR (95% CI)**	**p**	**I2**
rs73089338	3	*CDCP1*	4.63 (2.70-7.96)	$1.9x10^{-7}$	1.07 (0.25-4.54)	0.92	0.70 (0.26-1.90)	0.49	1.02 (0.49-2.16)	0.95	1.94 (1.27-2.98)	0.002	0.86
rs504365	5	*RASGRF2*	0.18 (0.07-0.44)	$2.9x10^{-6}$	1.63 (0.73-3.65)	0.23	1.60 (0.99-2.57)	0.06	1.28 (0.94-1.75)	0.11	1.17 (0.92-1.50)	0.20	0.82
rs2247256	8	*ERICH1*	0.16 (0.06-0.43)	$1.9x10^{-6}$	1.81 (0.80-4.07)	0.15	0.58 (0.36-0.91)	0.02	1.03 (0.76-1.40)	0.85	0.78 (0.61-0.99)	0.04	0.82
rs117576073	11	*CYP2R1*	8.36 (3.66-19.06)	$2.8x10^{-5}$	0 (0-∞)	1.00	0.48 (0.03-8.68)	0.62	0.31 (0.04-2.31)	0.26	3.11 (1.25-7.78	0.01	0.79
rs4149056	12	*SLCO1B1*	5.15 (3.13-8.45)	**$2.5x10^{-9}$**	3.98 (1.75-9.03)	0.001	4.91 (3.09-7.77)	**$1.3x10^{-11}$**	1.86 (1.32-2.62)	$3.9x10^{-4}$	2.99 (2.34-3.82)	**$2.63x10^{-18}$**	0.87
rs4149000	12	*SLCO1A2*	3.94 (2.36-6.57)	$2.9x10^{-6}$	2.53 (1.00-6.39)	0.050	7.29 (4.13-12.8)	**$6.5x10^{-12}$**	1.74 (1.16-2.60)	0.007	2.81 (2.10-3.75)	**$3.31x10^{-12}$**	0.91
rs28447350	13	Intergenic	3.66 (2.23-6.00)	$5.3x10^{-7}$	0.55 (0.21-1.43)	0.219	0.88 (0.49-1.59)	0.676	1.06 (0.73-1.52)	0.76	1.32 (1.01-1.74)	0.04	0.84

REFERENCES

(1) Rothwell, S. *et al.* Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. *Annals of the rheumatic diseases* **75**, 1558-66 (2016).

(2) Collins, R. *et al.* Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* **388**, 2532-61 (2016).

(3) Alfirevic, A. *et al.* Phenotype standardization for statin-induced myotoxicity. *Clin Pharmacol Ther* **96**, 470-6 (2014).

(4) Law, M. & Rudnicka, A.R. Statin safety: a systematic review. *Am J Cardiol* **97**, 52C-60C (2006).

(5) Brunham, L.R. *et al.* Differential effect of the rs4149056 variant in SLCO1B1 on myopathy associated with simvastatin and atorvastatin. *Pharmacogenomics J* **12**, 233-7 (2012).

(6) Donnelly, L.A. *et al.* Common nonsynonymous substitutions in SLCO1B1 predispose to statin intolerance in routinely treated individuals with type 2 diabetes: a go-DARTS study. *Clin Pharmacol Ther* **89**, 210-6 (2011).

(7) Group, S.C. *et al.* SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med* **359**, 789-99 (2008).

(8) Marciante, K.D. *et al.* Cerivastatin, genetic variants, and the risk of rhabdomyolysis. *Pharmacogenet Genomics* **21**, 280-8 (2011).

(9) Voora, D. *et al.* The SLCO1B1*5 genetic variant is associated with statin-induced side effects. *J Am Coll Cardiol* **54**, 1609-16 (2009).

(10) Carr, D.F. *et al.* SLCO1B1 Genetic Variant Associated With Statin-Induced Myopathy: A Proof of Concept Study Using the Clinical Practice Research Datalink (CPRD). *Clin Pharmacol Ther* **96**, 695-791 (2013).

(11) Lau, Y.Y., Huang, Y., Frassetto, L. & Benet, L.Z. effect of OATP1B transporter inhibition on the pharmacokinetics of atorvastatin in healthy volunteers. *Clinical pharmacology and therapeutics* **81**, 194-204 (2007).

(12) Noe, J., Portmann, R., Brun, M.E. & Funk, C. Substrate-dependent drug-drug interactions between gemfibrozil, fluvastatin and other organic anion-transporting peptide (OATP) substrates on OATP1B1, OATP2B1, and OATP1B3. *Drug Metab Dispos* **35**, 1308-14 (2007).

(13) Niemi, M. *et al.* High plasma pravastatin concentrations are associated with single nucleotide polymorphisms and haplotypes of organic anion transporting polypeptide-C (OATP-C, SLCO1B1). *Pharmacogenetics* **14**, 429-40 (2004).

(14) Pasanen, M.K., Neuvonen, M., Neuvonen, P.J. & Niemi, M. SLCO1B1 polymorphism markedly affects the pharmacokinetics of simvastatin acid. *Pharmacogenetics and genomics* **16**, 873-9 (2006).

(15) Puccetti, L., Scarpini, F., Cappellone, R. & Auteri, A. Genetic influence in statin intolerance. *Clin Pharmacol Ther* **90**, 365 (2011).

(16) Isackson, P.J. *et al.* Association of common variants in the human eyes shut ortholog (EYS) with statin-induced myopathy: evidence for additional functions of EYS. *Muscle Nerve* **44**, 531-8 (2011).

(17) Mangravite, L.M. *et al.* A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. *Nature*, (2013).

(18) Carr, D.F., Alfirevic, A., Johnson, R., Chinoy, H., van Staa, T. & Pirmohamed, M. GATM gene variants and statin myopathy risk. *Nature* **513**, E1 (2014).

(19) Floyd, J.S., Bis, J.C., Brody, J.A., Heckbert, S.R., Rice, K. & Psaty, B.M. GATM locus does not replicate in rhabdomyolysis study. *Nature* **513**, E1-3 (2014).
(20) O'Meara, H. *et al.* Electronic Health Records For Biological Sample Collection: Feasibility Study Of Statin-Induced Myopathy Using The Clinical Practice Research Datalink. *Br J Clin Pharmacol*, **77**, 831-8 (2013).
(21) Bakar, N.S., Neely, D., Avery, P., Brown, C., Daly, A.K. & Kamali, F. Genetic and Clinical Factors Are Associated With Statin-Related Myotoxicity of Moderate Severity: A Case-Control Study. *Clin Pharmacol Ther* **104**, 178-87 (2018).
(22) Xiang, Q. *et al.* Association between SLCO1B1 T521C polymorphism and risk of statin-induced myopathy: a meta-analysis. *Pharmacogenomics J* **18**, 721-9 (2018).
(23) Pasanen, M.K., Neuvonen, M., Neuvonen, P.J. & Niemi, M. SLCO1B1 polymorphism markedly affects the pharmacokinetics of simvastatin acid. *Pharmacogenet Genomics* **16**, 873-9 (2006).
(24) Pasanen, M.K., Fredrikson, H., Neuvonen, P.J. & Niemi, M. Different effects of SLCO1B1 polymorphism on the pharmacokinetics of atorvastatin and rosuvastatin. *Clin Pharmacol Ther* **82**, 726-33 (2007).
(25) Niemi, M., Pasanen, M.K. & Neuvonen, P.J. SLCO1B1 polymorphism and sex affect the pharmacokinetics of pravastatin but not fluvastatin. *Clin Pharmacol Ther* **80**, 356-66 (2006).
(26) Tornio, A., Vakkilainen, J., Neuvonen, M., Backman, J.T., Neuvonen, P.J. & Niemi, M. SLCO1B1 polymorphism markedly affects the pharmacokinetics of lovastatin acid. *Pharmacogenet Genomics* **25**, 382-7 (2015).
(27) Deng, J.W. *et al.* The effect of SLCO1B1*15 on the disposition of pravastatin and pitavastatin is substrate dependent: the contribution of transporting activity changes by SLCO1B1*15. *Pharmacogenet Genomics* **18**, 424-33 (2008).
(28) Kalliokoski, A. & Niemi, M. Impact of OATP transporters on pharmacokinetics. *British journal of pharmacology* **158**, 693-705 (2009).
(29) Franke, R.M., Scherkenbach, L.A. & Sparreboom, A. Pharmacogenetics of the organic anion transporting polypeptide 1A2. *Pharmacogenomics* **10**, 339-44 (2009).
(30) Kullak-Ublick, G.A. *et al.* Molecular and functional characterization of an organic anion transporting polypeptide cloned from human liver. *Gastroenterology* **109**, 1274-82 (1995).
(31) Shirasaka, Y., Suzuki, K., Shichiri, M., Nakanishi, T. & Tamai, I. Intestinal absorption of HMG-CoA reductase inhibitor pitavastatin mediated by organic anion transporting polypeptide and P-glycoprotein/multidrug resistance 1. *Drug metabolism and pharmacokinetics* **26**, 171-9 (2011).
(32) Liu, H. *et al.* Solute Carrier Family of the Organic Anion-Transporting Polypeptides 1A2- Madin-Darby Canine Kidney II: A Promising In Vitro System to Understand the Role of Organic Anion-Transporting Polypeptide 1A2 in Blood-Brain Barrier Drug Penetration. *Drug metabolism and disposition: the biological fate of chemicals* **43**, 1008-18 (2015).
(33) Medicines.org.uk. *Atorvastatin 10 mg film-coated tablets - Summary of Product Characteristics (SPC)*. <www.medicines.org.uk/emc/product/2959> (2016). Accessed 21. Aug 2018.
(34) Medicines.org.uk. *Simvastatin 40mg - Summary of Product Characteristics (SPC).* <www.medicines.org.uk/emc/product/5688/> (2018). Accessed 21 Aug. 2018.

(35) Niemi, M. Transporter pharmacogenetics and statin toxicity. *Clin Pharmacol Ther* **87**, 130-3 (2010).
(36) Peyser, B. *et al.* Effects of Delivering SLCO1B1 Pharmacogenetic Information in Randomized Trial and Observational Settings. *Circulation Genomic and precision medicine* **11**, e002228 (2018).
(37) van Staa, T.P., Carr, D.F., O'Meara, H., McCann, G. & Pirmohamed, M. Predictors and outcomes of increases in creatine phosphokinase levels or rhabdomyolysis risk during statin treatment. *Br J Clin Pharmacol*, **78**, 649-59 (2014).
(38) Molokhia, M. & McKeigue, P. EUDRAGENE: European collaboration to establish a case-control DNA collection for studying the genetic basis of adverse drug reactions. *Pharmacogenomics* **7**, 633-8 (2006).
(39) Hopewell, J.C. *et al.* Impact of common genetic variation on response to simvastatin therapy among 18 705 participants in the Heart Protection Study. *European heart journal* **34**, 982-92 (2013).
(40) O'Meara, H. *et al.* Electronic health records for biological sample collection: feasibility study of statin-induced myopathy using the Clinical Practice Research Datalink. *Br J Clin Pharmacol* **77**, 831-8 (2014).
(41) Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
(42) Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. & Weir, B.S. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-8 (2012).
(43) Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
(44) Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457-70 (2011).
(45) Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
(46) Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955-9 (2012).
(47) Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* **39**, 906-13 (2007).
(48) Magi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC bioinformatics* **11**, 288 (2010).

FIGURE LEGENDS

Figure 1. Schematic representation of the discovery, replication and validation cohort case-control analyses and subsequent meta-analysis performed. Patient numbers represent those included in analyses post-sample QC.

Figure 2. Manhattan plot of genome-wide association analysis of statin-induced myopathy. The data represents logistic regression derived log p-values (y-axis) of SNPs for the discovery case-control analysis of A) the "all myopathy" phenotype (n=128) and B) the "severe myopathy" sub-phenotype (n=32) with the WTCCC2 population controls (n=2,501). X-axis is the position of the SNP with the chromosome indicated.

Figure 3. Forest plot depicting meta-analysis for the *SLCO1B1* c.521T>C (rs4149056) polymorphism for both "all myopathy" (CK>4xULN) and "severe myopathy" (CK>10xULN/ rhabdomyolysis) phenotypes caused by all statins (upper panels) and simvastatin-only (lower panel).