Title
Characterization of Statin Dose Response in Electronic Medical Records

Journal
Clinical Pharmacology & Therapeutics, 95(3)

ISSN
0009-9236

Authors
Wei, W-Q
Feng, Q
Jiang, L
et al.

Publication Date
2014-03-01

DOI
10.1038/clpt.2013.202

Peer reviewed

Published in final edited form as:
Clin Pharmacol Ther. 2014 March ; 95(3): 331–338. doi:10.1038/clpt.2013.202.

Characterization of Statin Dose-response within Electronic Medical Records

Wei-Qi Wei, MMed, PhD[1], **Qiping Feng, PhD**[2], **Lan Jiang, MS**[3], **Magarya S. Waitara, BA**[2], **Otito F. Iwuchukwu, PhD**[2], **Dan M. Roden, MD**[2,4,5,6], **Min Jiang, MS**[7], **Hua Xu, PhD**[7], **Ronald M. Krauss, MD**[8], **Jerome I. Rotter, MD**[9], **Deborah A. Nickerson, PhD**[10], **Robert L. Davis, MD, MPH**[11], **Richard L. Berg, MS**[12], **Peggy L. Peissig, MBA**[12], **Catherine A. McCarty, PhD**[13], **Russell A. Wilke, MD, PhD**[14,*], and **Joshua C. Denny, MD, MS**[1,*]

Wei-Qi Wei: wei-qi.wei@vanderbilt.edu; Qiping Feng: Qiping.feng@vanderbilt.edu; Lan Jiang: jiang@chgr.mc.vanderbilt.edu; Magarya S. Waitara: magarya.s.waitara@vanderbilt.edu; Otito F. Iwuchukwu: otito.f.iwuchukwu@vanderbilt.edu; Dan M. Roden: dan.roden@vanderbilt.edu; Min Jiang: min.jiang@uth.tmc.edu; Hua Xu: hua.xu@uth.tmc.edu; Ronald M. Krauss: RKrauss@chori.org; Jerome I. Rotter: jerome.rotter@cshs.org; Deborah A. Nickerson: debnick@u.washington.edu; Robert L. Davis: Robert.L.Davis@kp.org; Richard L. Berg: Berg.Richard@mcrf.mfldclin.edu; Peggy L. Peissig: Peissig.Peggy@securityhealth.org; Catherine A. McCarty: CMcCarty@eirh.org; Russell A. Wilke: russell.a.wilke@gmail.com; Joshua C. Denny: josh.denny@vanderbilt.edu

[1]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN

[2]Department of Medicine, Vanderbilt University, Nashville, TN

[3]Department of Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN

[4]Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN

[5]Oates Institute for Experimental Therapeutics, Vanderbilt University School of Medicine, Nashville, TN

[6]Office of Personalized Medicine, Vanderbilt University School of Medicine, Nashville, TN

[7]Department of Biomedical Informatics, University of Texas, Houston, TX

[8]Children's Hospital Oakland Research Institute, Oakland, CA

[9]Medical Genetics Institute, Cedars-Sinai Medical Center, West Los Angeles, CA

[10]Department of Genome Sciences, University of Washington, Seattle, WA

[11]Kaiser Permanente Georgia, Center for Health Research Southeast, Atlanta, GA

[12]Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI

[13]Essentia Institute of Rural Health, Duluth, MN

[14]Department of Internal Medicine, Sanford Healthcare, Fargo, ND

Abstract

Efforts to define the genetic architecture underlying variable statin response have met with limited success possibly because previous studies were limited to effect based on one-single-dose. We leveraged electronic medical records (EMRs) to extract potency (ED_{50}) and efficacy (E_{max}) of statin dose-response curves and tested them for association with 144 pre-selected variants. Two

[*]contributed equally

CONFLICT OF INTEREST
None

large biobanks were used to construct dose-response curves for 2,026 (simvastatin) and 2,252 subjects (atorvastatin). Atorvastatin was more efficacious, more potent, and demonstrated less inter-individual variability than simvastatin. A pharmacodynamic variant emerging from randomized trials (*PRDM16*) was associated with E_{max} for both. For atorvastatin, E_{max} was 51.7 mg/dl in homozygous for the minor allele versus 75.0 mg/dl for those homozygous for the major allele. We also identified several loci associated with ED_{50}. The extraction of rigorously defined traits from EMRs for pharmacogenetic studies represents a promising approach to further understand of genetic factors contributing to drug response.

INTRODUCTION

In 2010, U.S. federal legislators set an aggressive timeline for the widespread implementation of electronic medical records (EMRs).[1] According to the National Center for Health Statistics, physician adoption rates for basic EMR systems have risen to 72% in 2012 from 48% in 2009 and 42% in 2008.[2,3] The deployment of EMRs is not only improving patient care, it is generating huge clinical practice-based datasets ideal for the conduct of observational research.[4,5] Several EMR-derived observational datasets have been linked to secure biological repositories containing DNA. [6–9] These clinical practice-based biobanks offer a previously unavailable opportunity for evaluating genetic findings from randomized clinical trials (RCTs).[10,11]

Statins (HMG-CoA reductase inhibitors) reduce circulating levels of low density lipoprotein cholesterol (LDL-C), and the cardiovascular benefits of these drugs are well-established.[12,13] At present, simvastatin and atorvastatin are the two most commonly prescribed statins in the U.S.[14,15] As such, there is great interest in defining the genetic architecture underlying treatment outcome for both drugs.[16,17] Early studies conducted using archived DNA from RCTs have revealed a number of candidate gene variants with small but reproducible effects on treatment-induced change in low density lipoprotein cholesterol (ΔLDL-C).[18,19] Because these efforts were successful for pharmacodynamic genes (e.g., *HMGCR*) as well as pharmacokinetic genes (e.g., *SLCO1B1*), genotyping efforts have been expanded in an attempt to define additional loci contributing to lipid lowering response. Genome wide association studies (GWAS) conducted using the same RCTs indicated that several previously unrecognized loci (e.g., *CLMN1* and *PRDM16*) may contribute to the lipid lowering response observed during exposure to these drugs.[20] To date, however these findings have not been replicated in a practice-based setting.

EMRs not only provide a powerful approach to the replication and refinement of these observations[21,22], but they also hold a number of distinct advantages over RCTs, including access to rich environmental data and medication histories stored in structured and unstructured format.[10] We have previously shown that data within EMRs can be used to define inter-individual variability in statin response, by extracting accurate measures of potency (ED_{50}) and lipid-lowering efficacy (E_{max}) from dose-response curves for patients exposed to atorvastatin during the course of routine clinical care[23,24]. We now expand this approach to simvastatin, in a second biobank, and we utilize the resulting dose-response traits to quantify the reproducibility of candidate gene associations previously identified in population-based and treatment-based cohorts. We demonstrate a novel benefit of accessing banked clinical data from EMRs[25] to confirm associations of candidate gene loci with ED_{50} and E_{max} for statin lipid response in a clinical practice-based setting.

RESULTS

This study includes data from patients exposed to multiple doses of simvastatin or multiple doses of atorvastatin during the course of routine clinical care. It is a common feature of

EMR-based studies, that regional differences in prescribing patterns can make it difficult to study identical traits for two different drugs in a single cohort. The data in the current study were therefore extracted from two separate biobanks. A total of 2,026 subjects exposed to two or more doses of simvastatin were identified from BioVU, a biobank located on the campus of Vanderbilt University Medical Center (VUMC) in central Tennessee. Another 2,252 subjects exposed to two or more doses of atorvastatin were then identified from a second biobank, the Personalized Medicine Research Project (PMRP) located at Marshfield Clinic, in central Wisconsin.

The distribution of dose-response traits for simvastatin is shown in Figure 1, and the distribution of dose-response traits for atorvastatin is shown in Figure 2. Consistent with results from randomized trials,[26] atorvastatin was shown to be more potent and more efficacious than simvastatin in our two study cohorts drawn from separate biobanks. The baseline characteristics of our study cohorts are summarized in Table 1. The age, gender, and race distributions of each cohort reflect those of the surrounding community.[7,27]

Prior to initiation of statin therapy, baseline LDL-C (E_0) was 138.0 ± 34.2 mg/dl in the simvastatin study subjects extracted from BioVU (n = 2,026) and 158.7 ± 28.7 mg/dl in the atorvastatin study subjects extracted from PMRP (n = 2,252). Our ability to extract pretreatment LDL-C levels (prior to the initiation of any statin) reflects subtle differences in the degree of chart fragmentation between our two EMR-linked biobanks.[28] Because VUMC represents one of the largest tertiary referral centers in the Southeastern United States, a large fraction of their patient base receives primary care outside the Vanderbilt system of care. Conversely, Marshfield Clinic provides primary care for nearly all of the patients served by their EMR. This difference introduces variability in access to pretreatment lipid levels, and corresponding variability in the power to identify gene variants associated with E_0 for each cohort. Variants at 15 loci were associated with E_0 extracted from the atorvastatin dose-response curves in PMRP (rs12916, rs541041, rs174546, rs602633, rs635634, rs646776, rs1367117, rs2053302, rs2290159, rs2954029, rs4299376, rs4686228, rs6511720, rs6588480, rs9987289), whereas only 4 variants (rs514230, rs6511720, rs17091962, rs12527253) were associated with E_0 from the simvastatin dose-response curves in BioVU. Our observation that a previously described variant in the LDL receptor gene (rs6511720) was associated with E_0 in both of our study cohorts ($p < 0.005$) serves as a means of internal validation, since variability in this gene is known to influence LDL-C levels across multiple geographic settings.[29–31]

Several variants previously shown to influence statin response in RCTs were associated with lipid lowering efficacy in our clinical practice-based cohorts. Table 3 lists seven variants nominally associated with E_{max} for simvastatin ($p < 0.05$), and Table 4 lists seven variants nominally associated with E_{max} for atorvastatin ($p < 0.05$). It is noteworthy that rs11807862 in *PRDM16* was associated with E_{max} for both simvastatin and atorvastatin. *PRDM16* had previously been associated with statin-induced change in clinical lipid profiles in our combined analysis of three randomized treatment trials (p-value of 2.1×10^{-6}).[20] Therefore, our observation validates the previous finding and makes it highly unlikely that the association of this variant with E_{max} in our current cohorts has occurred by chance. In addition, our data further reveal that this variant has a substantial effect size. For simvastatin, E_{max} was 53.2 mg/dl in subjects homozygous for the minor allele versus 60.9 mg/dl in subjects homozygous for the major allele (Table 3); for atorvastatin, E_{max} was 51.7 mg/dl in subjects homozygous for the minor allele versus 75.0 mg/dl in subjects homozygous for the major allele (Table 4). While our model inherently adjusts for baseline LDL-C, the difference in magnitude between E_{max} for simvastatin and E_{max} for atorvastatin may be due to biobank-specific differences in our ability to estimate pretreatment LDL-C.

Table 3 and Table 4 also list all variants nominally associated with statin potency defined as ED_{50}. This trait has not previously been studied as an endpoint in any genetic assessment of statin response. In BioVU, eight variants were associated with ED_{50} for simvastatin ($p < 0.05$), and in PMRP, eight variants were associated with ED_{50} for atorvastatin ($p < 0.05$). The strongest determinants of atorvastatin ED_{50} were two variants in partial linkage disequilibrium near the *SORT1* gene locus, rs602633 and rs646776 ($p < 0.005$). Although rs646776 was also associated with baseline LDL-C level (E_0) in this same study cohort, E_0 and ED_{50} were not correlated in this dataset ($r^2 = 0.016$), supporting the inference that the observed association between *SORT1* and ED_{50} is specifically related to statin response. In a meta-analysis of > 100,000 individuals of European ancestry, the *SORT1* gene is significantly associated with plasma LDL-C with p-value 1×10^{-107}.[29] Multiple studies on animal models have also disclosed that the *SORT1* gene can influence both hepatic apoB secretion[32] and cellular LDL uptake[46] and it therefore represents a plausible candidate for mediating statin treatment effects on LDL-C.

In BioVU, the strongest determinant of ED_{50} for simvastatin was rs1555926 in *ZNF217* ($p < 0.0004$), although the effect size for this association was modest (reflecting a shift in the required dose < 1 mg per day). Two other notable variants associated with simvastatin ED_{50} were rs4149056 in *SLCO1B1*, rs35599367 in *CYP3A4*, and rs8014194 in *CLMN*. The first two loci (*SLCO1B1 and CYP3A4*) are well-known predictors of statin response.[33,34] Rs4149056 in *SLCO1B1* is significantly associated with statin-induced myopathy in a previous GWAS of 175 subjects taking 80 mg simvastatin daily (85 cases and 90 controls), and the observation has been further validated in a 20,000 subject cohort (the odd ratio for myopathy was 4.5 (95% CI, 2.6–7.7) per copy of C-allele). CYP3A4 is notable in atorvastatin metabolization. Previous study has reported that an inhibition of CYP3A4 can result in severe drug-induced myopathy.[35] The third locus (*CLMN*) has also been reported as a determinant of statin-induced change in total cholesterol (p-value 1.9×10^{-8}) in our prior combined GWAS using data from RCTs.[10]

Because BioVU contains study subjects of diverse ancestry, we further stratified our findings for dose-response using race as a categorical trait. We previously reported that geographic race is highly accurate in this EMR-linked biobank, when electronically extracted and compared to a panel of ancestry informative markers.[36] When our findings were stratified by race, the association between simvastatin ED_{50} and rs8014194 in *CLMN1* remained significant only in African Americans ($p = 0.015$, $n = 296$). Conversely, the association between simvastatin ED_{50} and rs4149056 in *SLCO1B1* remained significant only in European Americans ($p = 0.035$, $n = 1,338$). Stratification by race also yielded new associations not previously recognized in this cohort; for example, in European Americans, simvastatin ED_{50} was further associated with rs6708136 in *UGT1A1* ($p = 0.032$). Because of the known pharmacokinetic importance of genes like *UGT1A1* and *SLCO1B1*, the race specificity of these associations warrants further study.

In addition, we combined the data from both cohorts and performed a meta analysis. Most SNPs found in separated analyses (rs646776, rs7584099, rs4149056, rs35599367, and rs1564348 for ED50; rs11807862, rs6588480, rs2053302, rs7564379, rs9605146, rs17091962, and rs1800961 for Emax) remained significant at the $p<0.05$ level, though none exceeded a Bonferroni correction. Cochran's Q test showed no statistically different between outcomes from two cohorts for these SNPs. We also observed even lower P-values for rs11807862 in *PRDM16* (9×10^{-4}) and rs4149056 (0.01) in *SLCO1B1* than previous tests.

DISCUSSION

Leveraging routine care data for pharmacogenetic research offers a previously unavailable possibility to evaluate treatment effectiveness in contrast to treatment efficacy which is all that is available from randomized clinical trial. In this study, we demonstrate that EMRs can be used to efficiently extract dose response traits representing potency (ED_{50}) and efficacy (E_{max}) for two commonly used drugs. Our data confirm that atorvastatin is both more potent and more efficacious than simvastatin using real-world clinical data. We also observed that the distribution for atorvastatin potency (Fig 2) is much *narrower* than the distribution for simvastatin potency (Fig 1). For simvastatin, the wide variability in potency observed in our clinical practice-based data is consistent with prior observations that some patients do not get to target LDL-C while using this drug, even if followed up regularly in an effort to titrate to their LDL-C downward.[26] Since high dose simvastatin is no longer recommended as initial therapy[37], this phenomenon cannot simply be overcome by dose escalation. Patients genetically predisposed to lower potency with simvastatin (Panel C, Fig 1) may in fact need a more potent statin earlier in the course of their care.

This study also replicates several well-known associations between candidate gene variants and statin response within the context of routine clinical practice, and it extends our understanding of these relationships by exploring the use of potency as a novel phenotypic trait. For example, simvastatin ED_{50} (the daily dose of simvastatin needed to bring a subject's LDL-C level to half maximal effect) is associated with a regulatory variant in *CYP3A4* (rs35599367) and a non-synonymous coding variant in *SLCO1B1* (rs4149056) (Table 2). The pharmacokinetic impact of these variants has been thoroughly evaluated *in vitro* and *in vivo*. For atorvastatin (Table 3), ED_{50} is associated with a functional variant in *SORT1* (rs646776). While this finding requires replication, studies conducted in humans and animal models have shown that reduced expression of the *SORT1* gene product, sortilin, preferentially increases levels of very small dense LDL particles[32], a particle subclass known to exhibit decreased binding to LDL receptors.[38]

Lastly, this study advances our understanding of the biology underlying statin response. Several genetic loci previously shown to alter statin-mediated lipid changes in randomized treatment trials now show an association with statin efficacy in EMRs. For example, a common variant at the *PRDM16* gene locus (rs11807862) has previously been associated with statin-mediated lipid changes in our combined GWAS analysis of 3,932 subjects exposed to simvastatin, pravastatin, and atorvastatin[20]. In the current study, this same variant was associated with E_{max} for both simvastatin and atorvastatin. Thus, our findings confirm the relationship between rs11807862 and the lipid-lowering efficacy of statins, and they underscore the importance of this association in the context of routine clinical care. *PRDM16* may influence adipocyte maturation[39], and further studies are needed to characterize the link between this gene locus and lipid homeostasis *in vitro*.

In summary, our findings demonstrate that highly informative drug response traits can be extracted from EMR-linked biobanks, and they indicate these traits can be used to further our understanding of the genetic determinants of drug response in the context of routine clinical practice. Unique features of our approach include access to multiple doses, a reduction in phenotypic misclassification through the extraction of full dose-response curves, and scale.

METHODS

Study Settings

This study includes data from patients exposed to multiple doses of simvastatin at VUMC or multiple doses of atorvastatin at PMRP during the course of routine clinical care.

Simvastatin Cohort—VUMC admits more than 65,000 unique inpatients yearly, and provides comprehensive longitudinal care for the majority of these patients. In the outpatient arena, VUMC clinics host ~2 million patient encounters yearly. VUMC has previously constructed a de-identified version of its integrated (combined inpatient-outpatient) EMR for epidemiological research in a practice-based setting, and in 2007 this resource began linking DNA samples to clinical data at a rate of ~500 samples per week[7]. With DNA linked to the de-identified EMRs of more than 167,000 unique individuals, BioVU currently represents on of the nation's largest clinical practice-based biobanks.[7] BioVU reflects the racial makeup of the surrounding community, and the majority of the records in this database (80%) are from subjects of European ancestry.[40]

Atorvastatin Cohort—Marshfield Clinic, in Central Wisconsin, provides healthcare services for nearly 350,000 unique individuals (also ~2 million clinical visits per year). In 2002, the Center for Human Genetics (CHG) at the Marshfield Clinic began approaching the surrounding community (initially 19 zip codes around the city of Marshfield, Wisconsin) to offer participation in the first population-based biobank in the U.S., linking coded clinical data to DNA samples for large scale studies of genetic epidemiology and treatment outcome[41]. At present, this secure encrypted biobank (the PMRP database) provides access to DNA and comprehensive longitudinal clinical data for over 20,000 adult study subjects[42]. The vast majority of the subjects in this database (98%) are of Northern European ancestry[27].

Design

This study was conducted in accordance with the basic principles of the Declaration of Helsinki, and approved by the Institutional Review Boards of VUMC and Marshfield Clinic. BioVU (the source of our simvastatin cohort) and PMRP (the source of our atorvastatin cohort) follow different enrollment procedures, and both approaches to biobanking have been published[7]. BioVU follows an "opt-out" approach, using EMR-derived data that are completely de-identified. Work with the BioVU database has therefore been determined to represent *non-human-subject research* by the Federal Office of Human Subject Research Protection[7]. By comparison, the PMRP follows an "opt-in" approach, and all data are coded[27]. Within the PMRP database, all study subjects have provided written informed consent for large scale pharmacogenetic association studies.

Phenotyping—EMRs contain medication information in both structured and unstructured formats. Structured data (e.g., name-value pairs, such as "drug = simvastatin") can be easily retrieved and converted into a ready-to-analyze format by computational approaches. Unstructured data (e.g., free text within clinical narratives, such as "the patient takes atorvastatin 20 mg tablets, ½ tablet daily") is inherently rich in content but more difficult to extract than structured data[43]. We therefore leveraged our previously validated MedEx natural language processing (NLP) system[44] to extract and reconstruct retrospective drug exposure histories from unstructured data. This NLP pipeline for medication data has produced highly accurate output compared to manual chart review in BioVU (F-measures 93%–96%)[44] and PMRP (sensitivity 80–97%, specificity 95–99%)[45].

Clinical lipid data were then extracted directly from structured laboratory records. We extracted all clinical lipid panels, and LDL-C levels were plotted longitudinally alongside statin exposure so that each lipid panel could be linked to drug and dose[23]. Because LDL-C levels typically reach steady state within 4–6 weeks after initiating statin treatment or changing statin dose), we filtered all lipid data and only accepted LDL-C levels obtained in window beginning six weeks after the initiation of each dose and ending with the cessation of the drug or a change to a new dose. We commonly observed that more than one LDL-C result could be linked to a given statin dose, and a median LDL-C value was therefore calculated for each drug dose. We then linked statin exposure to lipid data and applied a maximum-effect model to construct individual dose-response curves as published[24]. Under this model, change in LDL-C is a function of statin dose, and each parameter is assumed to vary for individuals around a population average.

In order to characterize the dose-response relationships in detail, we limited our phenotyping efforts to individuals exposed to two or more doses of the same drug during the course of routine care. We also required that each individual had baseline LDL-C levels available within their electronic record (i.e., at 0 mg daily, prior to initiation of any statin). At the time of this analysis, BioVU contained 202,813 LDL-C results for 48,583 unique patients ever exposed to simvastatin, 10,280 of whom have had exposure to two or more doses. VUMC is a tertiary referral center and only 2,026 (approximately 20%) of these 10,280 patient records contain pretreatment LDL-C values. We then extracted data for the construction of atorvastatin dose-response curves from a 2nd biobank, the PMRP in central Wisconsin. In PMRP, 33,625 LDL-C results have previously been extracted for 3,644 unique patients exposed to atorvastatin, and 2,252 of these patients have had exposure to two or more doses (requiring 0 mg daily, prior to initiation of any statin).[23,24]

In both biobanks, we then derived phenotypic traits, for ED_{50} (potency) and E_{max} (maximal lipid-lowering efficacy), based on our published dose-response equation[46,47]:

$$\mathrm{LDL}_{\mathrm{Dose}} = \mathrm{E}_0 - \frac{\mathrm{E}_{\max} \times \mathrm{Dose}}{\mathrm{ED}_{50} + \mathrm{Dose}} \quad \text{(EQ 1)}$$

LDL_{Dose} represents the LDL-C value at each specific statin dose, E_0 represents baseline LDL-C level (prior to the administration of any statin), E_{max} represents the maximum modeled reduction in LDL-C level on simvastatin or atorvastatin, and ED_{50} represents the dose that causes half maximal reduction.

By applying a non-linear random coefficients model, where parameters from EQ 1 represent random coefficients, we were able to estimate dose-response parameters (E_0, ED_{50}, and E_{max}) for simvastatin for 1,953 unique individuals in BioVU using the same approach. Raw data for these 1,953 patients are plotted in Figure 1, along with the distribution for each trait. After removing those subjects who opted out of BioVU prior to initiation of this specific sub-study, 1,852 samples were submitted for genotyping. In PMRP, we were able to estimate all atorvastatin dose-response parameters for 2,213 unique individuals. Raw data are plotted for these 2,213 subjects in Figure 2, along with the distribution for each derived trait. All 2,213 samples were submitted for genotyping.

Genotyping—Single nucleotide polymorphisms (SNPs) were preselected based on three criteria: (1) variants associated with baseline lipid levels (LDL-C or total cholesterol level) by the Global Lipids Consortium[29], (2) pharmacodynamic variants associated with change in total cholesterol, change in LDL-C, or change in HDL-C in our prior combined GWAS of 3,932 subjects exposed to either simvastatin, pravastatin or atorvastatin in RCTs [19,20], and

(3) variants of proven functional relevance in pharmacokinetic candidate genes[48]. These candidate gene loci (Table 2) were genotyped for 31 SNPs associated with LDL-C or total cholesterol ($p < 10^{-8}$) from Global Lipids Consortium, 93 pharmacodynamic SNPs (40 for Δtotal cholesterol, 36 for ΔLDL-C, 17 for ΔHDL-C), and 20 pharmacokinetic SNPs. Each variant was genotyped in both cohorts, on an Illumina BeadXpress array (Illumina, San Diego, CA). Genotyping was successful (call rate >99%) for 137 SNPs in BioVU (simvastatin dose-response) and 140 SNPs in PMRP (atorvastatin dose-response).

Statistical Analyses

Statistical analyses were conducted using the PLINK genetic analysis toolset version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink). Minor allele frequency and Hardy-Weinberg equilibrium (HWE) statistics were calculated for each SNP after stratifying by race[40]. Two SNPs in European ancestry subjects (rs7075971 and rs12916) and two SNPs in African ancestry subjects (rs12916 and rs17645290) deviated from HWE (p-value less than 0.01). These SNPs were removed from our analyses. Genotype-phenotype association tests were then conducted using an additive model in PLINK, for E_0, ED_{50} and E_{max}. Because our goal was to establish proof of principle, replicating prior findings from randomized trials, our results are presented with unadjusted p values. Simon et al. previously demonstrated that race, but not gender, affected change in LDL-C levels in response to simvastatin therapy.[49] Thus, we also stratified our results by race. We also combined the data from both cohorts and performed a meta analysis. The meta analysis was performed using METAL.[50]

Acknowledgments

FUNDING SOURCE

This work was funded by NIH grants UL1RR024975, U19HL069757, U19HL065962, and RC2GM092318, and AHA 13POST16470018.

References

1. Shea S, Hripcsak G. Accelerating the use of electronic health records in physician practices. The New England journal of medicine. 2010; 362:192–195.10.1056/NEJMp0910140 [PubMed: 20089969]
2. Marcotte L, et al. Achieving meaningful use of health information technology: a guide for physicians to the EHR incentive programs. Archives of internal medicine. 2012; 172:731–736.10.1001/archinternmed.2012.872 [PubMed: 22782203]
3. Statistics, N. C. f. H. National Ambulatory Medical Care Survey, 2008–2010. 2011.
4. McCarty CA, Wilke RA. Biobanking and pharmacogenomics. Pharmacogenomics. 2010; 11:637–641.10.2217/pgs.10.13 [PubMed: 20415552]
5. Ritchie MD, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. American journal of human genetics. 2010; 86:560–572.10.1016/j.ajhg.2010.03.003 [PubMed: 20362271]
6. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large, population-based biobank. Future Medicine. 2005; 2:49–79.
7. Roden DM, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clinical pharmacology and therapeutics. 2008; 84:362–369.10.1038/clpt.2008.89 [PubMed: 18500243]
8. Hoffmann TJ, et al. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. Genomics. 2011; 98:79–89.10.1016/j.ygeno.2011.04.005 [PubMed: 21565264]

9. Gottesman O, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genetics in medicine: official journal of the American College of Medical Genetics. 201310.1038/gim.2013.72
10. Hudson KL. Genomics, health care, and society. The New England journal of medicine. 2011; 365:1033–1041.10.1056/NEJMra1010517 [PubMed: 21916641]
11. Pulley JM, et al. Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. Clinical pharmacology and therapeutics. 2012; 92:87–95.10.1038/clpt.2011.371 [PubMed: 22588608]
12. Baigent C, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet. 2010; 376:1670–1681.10.1016/S0140-6736(10)61350-5 [PubMed: 21067804]
13. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. Lancet. 2012; 380:565–571.10.1016/S0140-6736(12)61190-8 [PubMed: 22883507]
14. Nichols GA, Wang F, Pedula KL. Comparison of evidence-based versus non-evidence-based pharmacotherapy on the risk of cardiovascular hospitalization and all-cause mortality among patients with established cardiovascular disease. The American journal of cardiology. 2010; 105:786–791. [PubMed: 20211320]
15. Health, United States. 2011: With Special Feature on Socioeconomic Status and Health. National Center for Health Statistics; Hyattsville, Maryland: 2011.
16. Roden DM, et al. Cardiovascular pharmacogenomics. Circulation Research. 2011; 109:807–820. [PubMed: 21921273]
17. Wilke RA, Dolan ME. Genetics and variable drug response. JAMA: The Journal of the American Medical Association. 2011; 306:306–307. [PubMed: 21771992]
18. Kajinami K, Brousseau ME, Ordovas JM, Schaefer EJ. CYP3A4 genotypes and plasma lipoprotein levels before and after treatment with atorvastatin in primary hypercholesterolemia. The American journal of cardiology. 2004; 93:104–107. [PubMed: 14697480]
19. Thompson JF, et al. Comprehensive whole-genome and candidate gene analysis for response to statin therapy in the Treating to New Targets (TNT) cohort. Circulation Cardiovascular Genetics. 2009; 2:173–181. [PubMed: 20031582]
20. Barber MJ, et al. Genome-wide association of lipid-lowering response to statins in combined study populations. PloS one. 2010; 5:e9763.10.1371/journal.pone.0009763 [PubMed: 20339536]
21. Wei WQ, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. Journal of the American Medical Informatics Association: JAMIA. 2012; 19:219–224. [PubMed: 22249968]
22. Wilke RA, et al. The emerging role of electronic medical records in pharmacogenomics. Clinical pharmacology and therapeutics. 2011; 89:379–386.10.1038/clpt.2010.260 [PubMed: 21248726]
23. Peissig P, et al. Construction of atorvastatin dose-response relationships using data from a large population-based DNA biobank. Basic & Clinical Pharmacology & Toxicology. 2007; 100:286–288. [PubMed: 17371534]
24. Wilke RA, et al. Characterization of low-density lipoprotein cholesterol-lowering efficacy for atorvastatin in a population-based DNA biorepository. Basic & Clinical Pharmacology & Toxicology. 2008; 103:354–359. [PubMed: 18834356]
25. Baron RJ. Meaningful use of health information technology is managing information. JAMA: the journal of the American Medical Association. 2010; 304:89–90.10.1001/jama.2010.910 [PubMed: 20606155]
26. Rogers SL, et al. A dose-specific meta-analysis of lipid changes in randomized controlled trials of atorvastatin and simvastatin. Clinical therapeutics. 2007; 29:242–252.10.1016/j.clinthera.2007.02.001 [PubMed: 17472817]
27. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. Future Medicine. 2005; 2:49–79.

28. Wei WQ, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. Journal of the American Medical Informatics Association: JAMIA. 2012; 19:219–224.10.1136/amiajnl-2011-000597 [PubMed: 22249968]
29. Teslovich TM, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–713.10.1038/nature09270 [PubMed: 20686565]
30. Kathiresan S, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nature genetics. 2008; 40:189–197.10.1038/ng.75 [PubMed: 18193044]
31. Willer CJ, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature genetics. 2008; 40:161–169.10.1038/ng.76 [PubMed: 18193043]
32. Musunuru K, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010; 466:714–719. [PubMed: 20686566]
33. Link E, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. The New England journal of medicine. 2008; 359:789–799.10.1056/NEJMoa0801936 [PubMed: 18650507]
34. Voora D, et al. The SLCO1B1*5 genetic variant is associated with statin-induced side effects. Journal of the American College of Cardiology. 2009; 54:1609–1616. [PubMed: 19833260]
35. Wilke RA, Moore JH, Burmester JK. Relative impact of CYP3A genotype and concomitant medication on the severity of atorvastatin-induced muscle damage. Pharmacogenetics and genomics. 2005; 15:415–421. [PubMed: 15900215]
36. 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. 2010; 12:648–650.10.1097/GIM.0b013e3181efe2df [PubMed: 20733501]
37. FDA. 2011
38. Campos H, Arnold KS, Balestra ME, Innerarity TL, Krauss RM. Differences in receptor binding of LDL subfractions. Arteriosclerosis, thrombosis, and vascular biology. 1996; 16:794–801.
39. Seale P, et al. PRDM16 controls a brown fat/skeletal muscle switch. Nature. 2008; 454:961–967. [PubMed: 18719582]
40. 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. 2010; 12:648–650. [PubMed: 20733501]
41. Kaiser J. Biobanks. Population databases boom, from Iceland to the U.S. Science (New York, NY). 2002; 298:1158–1161.
42. McCarty CA, Chapman-Stone D, Derfus T, Giampietro PF, Fost N. Community consultation and communication for a population-based DNA biobank: the Marshfield clinic personalized medicine research project. American Journal of Medical Genetics. Part A. 2008; 146A:3026–3033. [PubMed: 19006210]
43. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. Journal of the American Medical Informatics Association: JAMIA. 2010; 17:514–518. [PubMed: 20819854]
44. Xu H, et al. MedEx: a medication information extraction system for clinical narratives. Journal of the American Medical Informatics Association: JAMIA. 2010; 17:19–24. [PubMed: 20064797]
45. Sirohi, E.; Peissig, P. Study of effect of drug lexicons on medication extraction from electronic medical records. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing; 2005. p. 308-318.
46. Wilke RA, et al. Characterization of low-density lipoprotein cholesterol-lowering efficacy for atorvastatin in a population-based DNA biorepository. Basic & clinical pharmacology & toxicology. 2008; 103:354–359.10.1111/j.1742-7843.2008.00291.x [PubMed: 18834356]
47. Peissig P, et al. Construction of atorvastatin dose-response relationships using data from a large population-based DNA biobank. Basic & clinical pharmacology & toxicology. 2007; 100:286–288.10.1111/j.1742-7843.2006.00035.x [PubMed: 17371534]
48. Feng Q, Wilke RA, Baye TM. Individualized risk for statin-induced myopathy: current knowledge, emerging challenges and potential solutions. Pharmacogenomics. 2012; 13:579–594.10.2217/pgs.12.11 [PubMed: 22462750]

49. Simon JA, et al. Phenotypic predictors of response to simvastatin therapy among African-Americans and Caucasians: the Cholesterol and Pharmacogenetics (CAP) Study. The American journal of cardiology. 2006; 97:843–850.10.1016/j.amjcard.2005.09.134 [PubMed: 16516587]

50. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191.10.1093/bioinformatics/btq340 [PubMed: 20616382]

Study Highlights

1. What is the current knowledge on the topic?

Efforts to define the genetic architecture underlying variable statin response have met with limited success possibly because previous studies were limited to effect based on a single dose.

2. What question this study addressed?

This study extracted rigorously defined phenotypes of statin dose-response curves (ED_{50} and E_{max}) from electronic medical records (EMRs) and tested them for association with 144 pre-selected variants.

3. What this study adds to our knowledge?

A variant in *PRDM16* was associated with E_{max} for both statins. For atorvastatin, E_{max} was 51.7 mg/dl in homozygous for the minor allele versus 75.0 mg/dl for those homozygous for the major allele. We also identified several loci associated with ED_{50}.

4. How this might change clinical pharmacology and therapeutics?

The extraction of rigorously defined traits from EMRs for pharmacogenetic studies represents a promising approach to further understand of genetic factors contributing to drug response.

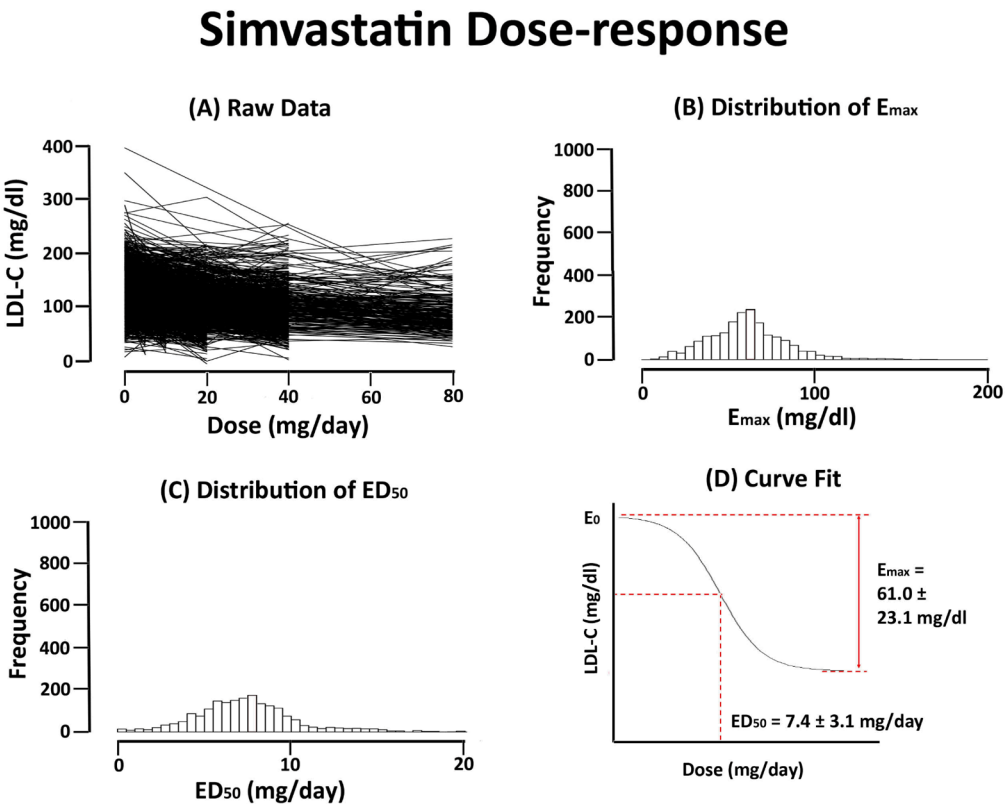

Figure 1.

Dose-response curves for simvastatin. **Panel A**: Raw dose-response data are plotted for 2,026 subjects with data sufficient to fit Equation 1 in BioVU. **Panel B**: Distribution of E_{max} for simvastatin. **Panel C**: Distribution of ED_{50} for simvastatin. **Panel D**: log dose-response curve showing mean ± standard deviation for each parameter.

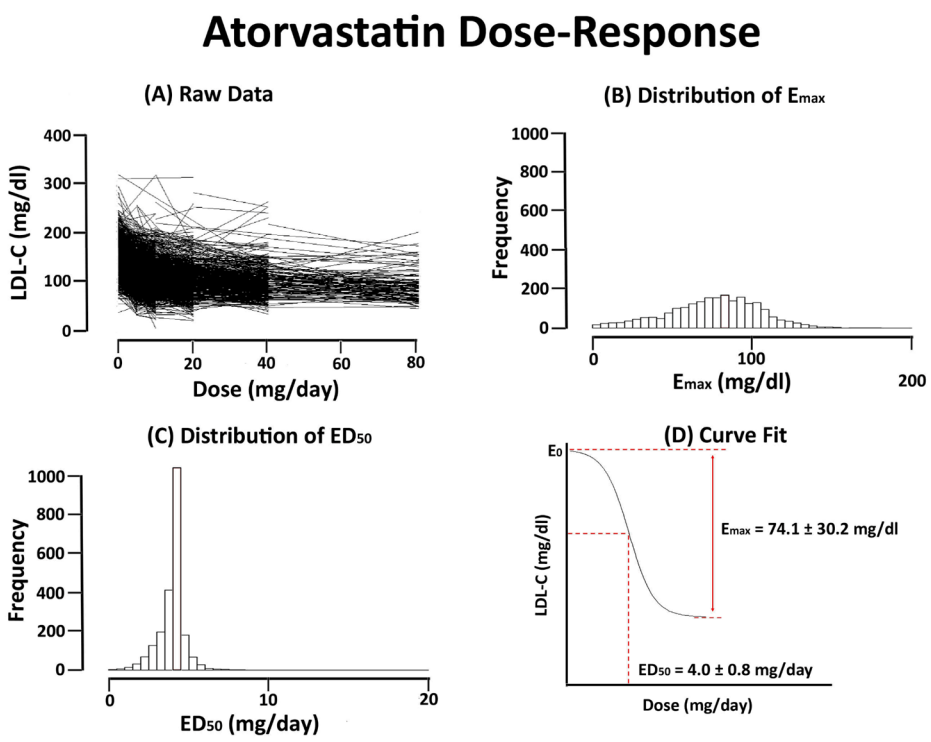

Figure 2.

Dose-response curves for atorvastatin. **Panel A**: Raw dose-response data are plotted for 2,213 subjects with data sufficient to fit Equation 1 in PMRP. **Panel B**: Distribution of E_{max} for atorvastatin. **Panel C**: Distribution of ED_{50} for atorvastatin. **Panel D**: log dose-response curve showing mean ± standard deviation for each parameter.

Table 1

The baseline characteristics of the two identified study cohorts exposed to statins.

	Simvastatin (BioVU)	Atorvastatin (PMRP)
N	2026	2252
Age	58.8 ± 12.4	62.21±12.84
Ancestry (European Americans)	77.05%	98.20%
Gender (Female)	52.12%	51.82%

Table 2

Loci containing markers evaluated for association with statin dose-response.

ABCA1	EYA2	LDLR	RHBDL3
ABCG8	F11	LIG4	RHCE
ABO	FADS1	LIPC	RNF175
ACVR2A	FAM5C	LOC285501	RNGTT
AGAP1	FEN1	LOC389249	ROR1
ANGPTL3	FGF14	LOC399988	RP11-49L2.2
ANKRD12	GALNT10	LOC642692	RYR3
ANXA2P3	GCKR	LOC645218	SFRP2
APOB	GKRP	LOC645453	SLC12A3
ASB18	GLIS1	LOC651746	SLC22A2
BCAS3	GPD2	LOC728241	SLCO1A2
BNC2	GPR149	LOC728727	SLCO1B1
BUD13	GRM3	LOC729397	SORT1
C21orf37	HACE1	LOC729750	SOSTDC1
C3orf53	hCG_1745121	LOC729913	ST3GAL4
C6orf106	HDLCQ10	LST3	STSL
CCDC113	HECW2	MAP3K1	TAGLN3
CCDC50	HFE	MGC99796	THSD4
CCDC85A	HINT3	MOSC1	THSD7A
CD82	HJURP	MYLIP	TIMD4
CDH9	HLA-DRA	NDUFV2	TMEM57
CELSR2	HLA-DRA1	NEGR1	TMPRSS7
CES7	HMG1L1	NLRC5	TMTC1
CETP	HMGCR	OSTN	TOPI
CLMN	HNF4A	p37	TRIB1
COL4A3BP	HPR	PAX1	TRPC4
COX17	HSPH1	PCSK9	TRPM7
CRBN	HUNK	PDCD12	TSC2

Clin Pharmacol Ther. Author manuscript; available in PMC 2014 March 06.

CSNK2A2	IDH3B	PGM1	TSEN2		
CYP3A4	IFT172	PHLDB2	TTC7A		
CYP3A5	IGF2R	PLD1	UGT1A1		
CYP7A1	IGSF11	POPDC2	UGT1A8		
dGK	INSIG2	PPM1G	UGT1A9		
DGUOK	IQGAP2	PPP1R3B	USP1		
DHX38	IRF4	PRDM16	USP50		
DISC1	ITPR2	PRO0628	USP8		
DLC1	KCNA4	PSRC1	UTS2D		
DNAH8	KIAA1324	RAB3C	XKR3		
DOCK7	KIAA1804	RAB3C	ZHX3		
EEPD1	KIAA1912	RAB3GAP1	ZNF217		
EPHB1	KLKBL4	RAF1	ZNF259		
ETS1	LDLCQ3	RBMS3	ZNF679		

Table 3

Variants associated with response to SIMVASTATIN

All association p<0.05 are reported. PRDM16 is associated with EMAX for both atorvastatin and simvastatin.

	SNP	Effect Size			*P* Value	Gene
		M/M	M/m	m/m		
E_{MAX} (mg/dL)	rs6588480	61.86±24.38	57.40±21.14	59.46±17.88	0.0022	GLIS1
	rs776746	61.22±23.72	60.06±22.21	55.71±21.70	0.0065	CYP3A5
	rs7564379	61.00±23.44	59.92±22.07	56.53±27.90	0.0090	DGUOK
	rs17091962	59.97±23.01	64.01±25.09	66.51±22.13	0.0127	NEGR1
	rs1800961	60.16±23.26	66.71±23.62	58.76±26.03	0.0141	HNF4A
	rs2740574	61.07±23.65	58.87±20.73	57.12±23.93	0.0316	CYP3A4
	rs11807862	**60.92±23.34**	**59.50±23.02**	**53.23±23.51**	**0.0443**	**PRDM16**
ED_{50} (mg/day)	rs1555926	7.63±3.27	7.00±2.82	7.08±2.86	0.0004	ZNF217
	rs17645290	7.33±2.99	7.57±3.47	8.53±3.63	0.0029	USP8
	rs4149056	7.55±3.20	7.12±3.02	7.06±2.30	0.0153	SLCO1B1
	rs35599367	7.49±3.18	6.78±2.69	10.15±0.00	0.0178	CYP3A4
	rs6495228	7.54±3.15	7.16±3.16	6.96±2.66	0.0202	RYR3
	rs8014194	7.34±3.08	7.40±2.95	8.04±4.11	0.0327	CLMN
	rs4438302	7.31±2.92	7.41±3.18	7.88±3.61	0.0343	LOC729913
	rs6029526	7.54±3.40	7.53±2.97	7.10±3.10	0.0432	PRO0628

M/M = homozygous major allele, M/m = heterozygous, m/m = homozygous minor allele

Clin Pharmacol Ther. Author manuscript; available in PMC 2014 March 06.

Table 4

Variants associated with response to ATORVASTATIN

All association p<0.05 are reported. PRDM16 is associated with EMAX for both atorvastatin and simvastatin.

	SNP	Effect Size			*P* Value	Gene
		M/M	M/m	m/m		
E_{MAX} (mg/dL)	rs11759246	74.87±29.90	69.35±32.56	52.59±47.87	0.0061	HINT3
	rs11807862	**74.96±29.71**	**71.66±32.43**	**51.65±33.87**	**0.0084**	**PRDM16**
	rs2053302	75.26±30.42	72.68±29.86	65.20±25.42	0.0088	FAM5C
	rs17632387	75.14±30.14	71.23±30.07	70.85±33.72	0.0177	THSD7A
	rs12527253	75.93±29.56	72.55±30.83	73.34±30.72	0.0277	RNGTT
	rs9939224	75.39±30.60	72.97±29.13	71.43±31.49	0.0464	CETP
	rs9605146	75.75±30.37	73.84±30.20	72.16±29.95	0.0499	XKR3
ED_{50} (mg/day)	rs646776	4.04±0.83	3.90±0.79	3.95±0.86	0.0014	SORT1
	rs602633	4.04±0.83	3.90±0.80	4.04±0.77	0.0043	SORT1
	rs7584099	4.05±0.82	3.99±0.81	3.92±0.86	0.0129	ACVR2A
	rs1568002	4.04±0.80	3.97±0.85	3.94±0.76	0.0271	LOC651746
	rs9513978	3.95±0.86	3.99±0.80	4.06±0.79	0.0331	FGF14
	rs2290159	4.03±0.82	3.93±0.82	3.97±0.71	0.0359	RAF1
	rs1564348	4.02±0.82	3.95±0.81	3.86±0.77	0.0398	SLC22A1
	rs6588480	4.02±0.81	3.95±0.85	3.89±0.70	0.0424	GLIS1

M/M = homozygous major allele, M/m = heterozygous, m/m = homozygous minor allele