Novel common and rare genetic determinants of paraoxonase activity: *FTO, SERPINA12*, and *ITGAL*.

Daniel S. Kim\(^1,2\), Amber A. Burt\(^1\), David R. Crosslin\(^1,2\), Peggy D. Robertson\(^2\), Jane E. Ranchalis\(^1\), Edward J. Boyko\(^3\), Deborah A. Nickerson\(^2\), Clement E. Furlong\(^1,2\), Gail P. Jarvik\(^1,2\)

Address for Correspondence:

Gail P. Jarvik, M.D., Ph.D.
Medical Genetics, Box 357720
University of Washington
Seattle, WA 98195-7720

Email: pair@u.washington.edu

1. Department of Medicine, Division of Medical Genetics, University of Washington School of Medicine, Seattle, WA
2. Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA
3. Seattle Epidemiologic Research and Information Center, Veterans Affairs Puget Sound Health Care System, Seattle, WA

Word Count: Total: 11,130, Body: 5,104, References: 3,789, Tables: 778, Abstract: 198,
Supplemental Materials: 1,317

Total Number of Figures: 0; Tables: 3 (3 supplemental).
Abstract

**Background:** HDL-associated paraoxonase 1 (PON1) activity is associated with cardiovascular and other human diseases. As the role of genetic variants outside of the *PON* gene cluster on PON1 activity is unknown, we sought to identify common and rare variants in such loci.

**Methods:** We typed 33,057 variants on the CVD chip in 1,362 subjects to test for their effects on adjusted-PON1 activity. Three novel genes (*FTO, ITGAL, and SERPINA12*) and the *PON* gene cluster had SNPs associated with PON1 arylesterase (AREase) activity. These loci were carried forward for rare-variant analysis using Exome chip genotypes in an overlapping subset of 1,051 subjects using sequence kernel association testing.

**Results:** *PON1* (*p*=2.24x10\(^{-4}\)), *PON3* (*p*=0.022), *FTO* (*p*=0.019), and *SERPINA12* (*p*=0.039) had both common and rare-variants associated with PON1 AREase. *ITGAL* variants were associated with PON1 activity when using weighted-SKAT analysis (*p*=2.63x10\(^{-3}\)). When adjusting for the initial common variants, *SERPINA12* became marginally significant (*p*=0.09), while all other findings remained significant (*p*<0.05), suggesting independent rare-variant effects.

**Conclusions:** We present novel findings that common and rare variants in *FTO, SERPINA12,* and *ITGAL* predict PON1 activity. These results further link PON1 to diabetes and inflammation and may inform the role of HDL in human disease.
Introduction

Paraoxonase 1 (PON1) is a glycoprotein enzyme physically associated with high-density lipoprotein (HDL). PON1 is partially responsible for the cardioprotective effects of HDL through inhibition of low-density lipoprotein (LDL) peroxidation (1, 2) and inhibition of macrophage-endothelium interactions (2, 3), both of which are underlying changes in atherogenesis. Knockout studies in mice have demonstrated that a lack of PON1 enzyme activity leads to increased susceptibility to organophosphate toxicity and cardiovascular disease (CVD) (4), likely due to an inability to neutralize oxidized LDL (5). In light of the recent failures in clinical trials (6) and Mendelian randomization analyses (7, 8) to demonstrate a causal relationship between plasma HDL (HDL-C) and CVD, there is an increased emphasis on understanding the multiple biologic dimensions of HDL, such as PON1, that may be cardioprotective but are not reflected by measures of HDL-C.

PON1 has broad substrate specificity and metabolizes toxic organophosphorus insecticides (9). For biological purposes, PON1 activity is generally measured with regard to the rate of hydrolysis of paraoxon, diazoxon, and phenylacetate (arylesterase activity) (10, 11). These are termed POase, DZOase, and AREase activities, respectively. AREase enzymatic activity is less unaffected by the functional PON1_Q192R polymorphism, thus making it the best reflection of the levels of PON1 protein (12).

PON1 enzyme activity is influenced by both genetic and environmental factors. There are four well established functional PON1 mutations (13): two missense mutations (PON1_Q192R [rs662] and PON1_L55M [rs854560]) and two 5’ regulatory (PON1_108C/T [rs705379] and PON1_162A/G...
[rs705381]). *PON1-108C/T* has the largest effect on AREase activity, altering expression likely due to modification of an Sp1 binding site(14, 15). Recent findings within this Carotid Lesion Epidemiology and Risk (CLEAR) cohort attribute approximately 21% of PON1 AREase activity to these four functional *PON1* mutations and six additional common variants within the *PON* gene cluster (including those in *PON2* and *PON3*)(16). Rare deleterious variants have also been identified(17). In addition, we reported that dietary cholesterol, alcohol, and vitamin C were all positively predictive of PON1 activity, while dietary iron and folic acid predicted reduced PON1 activity in 1,402 subjects from the CLEAR cohort(18); these factors accounted for an additional 8.2% of PON1 activity when adjusting for age, sex, and current smoking status.

Despite these findings, a large portion of PON1 activity variance remains unexplained. This may be due to genetic variation either at other loci or due to rare single nucleotide variants (SNVs) in *PON1*. While rare SNVs are well known as causative mutations in Mendelian traits, they are only beginning to be studied in large datasets considering complex diseases or in quantitative phenotypes. Rare SNVs are expected to account for some of the “missing heritability” of traits such as PON1 activity(19-22). Moreover, rare coding SNVs often affect protein function, as compared to noncoding regulatory variants that affect the levels of protein. Accordingly, we used a two-step process to determine if common or rare SNVs at other loci contribute to variation in paraoxonase activity. We first considered the associations from the relatively common SNPs on the Illumina HumanCVD BeadChip, and then selected genes nominated by this analysis to consider the effects of rare SNVs typed utilizing the new Illumina HumanExome BeadChip. This was done to limit the number of hypotheses tested and maximize power. We also catalogued the 92 non-synonymous and 5’ and 3’-noncoding SNVs present in the *PON* gene
cluster in the Exome Sequencing Project (ESP) 6500 data, to report their frequencies and evolutionary conservation.
Methods

Ethics statement

Institutional Review Boards at the University of Washington, Virginia Mason Medical Center, and Veterans Affairs Puget Sound Health Care approved the study. Written, informed consent was obtained from all participants.

Sample

The study cohort for these analyses consisted of 1,362 subjects from the previously described CLEAR study (18, 23-26) who had PON1 phenotype, covariate, and genetic data for the common variant analysis and 1,051 partially-overlapping subjects for the rare variant analysis. Ancestry was confirmed using STRUCTURE with three ancestral groups (27). Only Caucasian subjects were analyzed due to under-representation of minority samples in this cohort. Medications, including use of statins, were ascertained from pharmacy and medical records, as well as subject report. Current smoking status was by self-report and review of medical records. Descriptive statistics for the cohort are presented in Table 1.

Genotyping

The four known functional PON1 SNPs, PON1Q192R, PON1L55M, PON1-108C/T and PON1-162A/G were genotyped using previously described methods (14, 28). We genotyped 48,742 SNPs relevant to cardiovascular disease in 1,362 subjects using the Illumina HumanCVD BeadChip (CVD chip) (29). Duplicate genotyping for 34 individuals showed 99.7% consistency in calls for the CVD chip. Subjects were filtered out at an individual call rate of < 97%. SNPs were filtered with cutoffs of: minor allele frequency < 1% (for power considerations), Hardy-Weinberg
equilibrium $p < 10^{-4}$, or call-rate $< 97\%$. After SNP filtering, 33,057 polymorphic SNPs remained for analyses. A subset of the CLEAR cohort enriched for CAAD cases ($n = 1,051$ subjects, with 933 overlapping with the Illumina HumanCVD BeadChip data) underwent further genotyping of rare coding variants and common variants implicated in genome-wide association studies using the Illumina HumanExome BeadChip (Exome chip)(30). Of 251,336 Exome chip variants present, 92,253 were polymorphic in these data. All genotyping was performed blinded to phenotype.

**PON1 phenotyping**

The PON1 AREase activity was measured by a continuous spectrophotometric assay as previously described(28). AREase activity was measured in triplicate and averaged. AREase activity was utilized as the primary measured outcome of $PON$ gene cluster variation due to its closer correlation with protein levels than hydrolysis rates of the toxic substrates paraoxon or diazoxon.

**Common variant analysis**

Regression analyses of common variants genotyped on the CVD chip were performed using the PLINK analysis package(31). PON1 AREase activity was first adjusted for age, sex, current smoking status, and the four major functional $PON1$ polymorphisms ($PON1_{Q192R}$, $PON1_{L55M}$, $PON1_{-108C/T}$ and $PON1_{-162A/G}$) through linear regression. These adjusted PON1 activity residuals were further evaluated by linear regression under an additive genetic model.

**Rare variant analysis**
Methods that pool variants across loci have been utilized for rare SNV association testing, due to the limited power of traditional single-marker association analyses. Such a method, sequence kernel association testing (SKAT), was utilized for testing of SNVs genotyped by the Exome chip(32) utilizing an R-plugin (http://r-project.org). SKAT utilizes score-based variance-component tests to test for association between SNV sets within a region (with the ability to include both common and rare variants) and a phenotype, while adjusting for potentially confounding covariates in the model. The covariates adjusted for in the primary SKAT analyses of PON1 AREase activity were: age, sex, current smoking status, and the four functional PON1 mutations: PON1_Q192R, PON1_L55M, PON1_-108C/T and PON1_-162A/G.

Because SKAT is a regional DNA window-based test of association and we wanted to limit our tests, we chose to perform analyses on candidate genes (described in Table 2) identified by common variant analyses and with a false discovery rate (FDR) < 0.10 (PON1, FTO, ITGAL, and SERPINA12). For these gene regions found to not be significant for association with PON1 AREase activity, a second SKAT analysis was performed using adjusted weights that give higher weights to rare variants(32). In order to determine if the rare variant associations were independent of the common variant associations in the same genes, a separate sensitivity analysis was performed in 933 subjects with both Exome and CVD chip genotypes, with the SKAT analysis model adjusted for the common variants in addition to the other covariates.

In addition, we performed a secondary analysis considering the entire PON gene cluster (PON1, PON2, and PON3), as variants in PON2 and PON3 have been previously identified to be predictive of PON1 AREase activity in this cohort(26). Variants in the entire PON gene cluster
were considered in a SKAT analysis, with the model also adjusting for six common variants previously reported to predict PON1 activity (16) (rs854567, rs2299257, rs223783, rs2375005, rs3917486, and rs11768074) in addition to the other covariates.

**Exome Sequencing Project (ESP) PON region variant data**

The ESP6500 dataset is composed of individuals obtained from a number of large-scale NHLBI cohorts. It includes exome sequence data from individuals from the Women's Health Initiative, Framingham Heart Study, Jackson Heart Study, Multi-Ethnic Study of Atherosclerosis, Atherosclerosis Risk in Communities, Coronary Artery Risk Development in Young Adults, Cardiovascular Health Study, Genomic Research on Asthma in the African Diaspora, Lung Health Study, Pulmonary Arterial Hypertension population, Acute Lung Injury cohort, and the Cystic Fibrosis cohort. Exome sequencing was performed either at the University of Washington or the Broad Institute. Single base variant data presented in the Supplemental Materials are available on the ESP Exome Variant Server (http://evs.gs.washington.edu/EVS/).

We explored and summarized the *PON1*, *PON2*, and *PON3* variants in the exomic data of 6503 ESP individuals, of which 4300 and 2203 were of European and African ancestry, respectively. Genomic positions (Hg19) and the reference and alternate allele on the forward strand were determined for each variant. Conservation for single base variants was assessed through PhastCons(33) and GERP(34) scores using SeattleSeq SNP annotation (http://snp.washington.edu/SeattleSeqAnnotation134/). Only variants that were non-synonymous, 5'-noncoding, or 3'-noncoding are presented, stratified by European ancestry (EA) and African ancestry (AA).
Results

Common Variant Association with PON1 AREase activity

Demographic measures of the CLEAR cohort are presented in Table 1. A total of 1,362 subjects had complete CVD chip genotype, covariate, and phenotype data for analyses of common variant association with PON1 AREase activity. The average age was 67.8 years, the subjects were 86.7% male, 18.2% current smokers, and 39.4% statin users. Of the 1,362 subjects, 460 (33.8%) were CAAD cases with >80% stenosis in one or both internal carotid arteries or were post-carotid endarterectomy, 687 (50.5%) were controls with <15% stenosis in both carotid arteries, 166 (12.2%) had moderate internal carotid obstruction with 50-79% stenosis, 27 (2.0%) had mild carotid obstruction with 15-49% stenosis, and 21 (1.5%) were ascertained for lower extremity peripheral vascular disease (ankle-brachial index of < 0.9).

Linear regression analyses across the 33,057 SNPs genotyped by the Illumina HumanCVD BeadChip in 1,362 subjects adjusting for age, sex, current smoking status, and the four functional PON1 mutations resulted in the detection of one additional PON1 SNP (rs2237583, p=3.88x10⁻⁸) reaching genome-wide significance (<5x10⁻⁸) for association with PON1 AREase activity (see Table 2). An additional 12 SNPs spanning seven different genes (PON1, FTO, SERPINA12, ITGAL, ALOX12, PDE4D, and FAM178A) were associated with PON1 AREase at the level of suggestive significance (p<10⁻⁴). As there were multiple SNPs located within PON1 and FTO, we evaluated the linkage disequilibrium (LD) between those SNPs in the CLEAR cohort data. For PON1, the three SNPs (rs2237583, rs3917478, and rs2299262) associating with PON1 AREase activity were in LD of r² < 0.60 with each other. In contrast, there were two significant
blocks of SNPs in LD with each other \((r^2 > 0.80)\) within \(FTO\). Rs7199363 and rs2004583 \((r^2 = 0.80)\); rs1136002 and rs2192872 \((r^2 = 0.96)\) and separately, rs2192872 and rs4784351 \((r^2 = 0.81)\) were all in strong pairwise LD.

**Exome Chip Analysis**

SKAT analyses of the Exome chip data considered an overlapping group of subjects \((n=933\) with both CVD and Exome chip data). The Exome chip genotyping was performed preferentially on CAAD cases \((70\%, \text{ see Table 1})\). The average age was 69.6 years, and these subjects were 89.8\% male, 14.9\% current smokers, and 52.5\% statin users. The mean PON1 AREase activity was 135.39 IU, nearly identical to that of the CVD chip subjects.

Using a FDR of \(< 0.1\) or \(p < 2.5 \times 10^{-5}\), four candidate genes were carried forward to rare variant testing SKAT analysis: \(PON1\), \(FTO\), \(ITGAL\), and \(SERPINA12\). These Exome chip data contained seven, four, four, and 10 polymorphic SNVs for \(PON1\), \(FTO\), \(ITGAL\), and \(SERPINA12\), respectively \((\text{see Table 3})\). We utilized SKAT to look for associations between each of these four groups of SNVs and PON1 AREase activity, adjusting for age, sex, current smoking status, and the genotypes of \(PON1_{Q192R}\), \(PON1_{L55M}\), \(PON1_{-108C/T}\) and \(PON1_{-162A/G}\).

The \(PON1\) region SNVs were highly predictive \((p = 2.24 \times 10^{-4})\) of PON1 activity, while \(FTO\) \((p = 0.019)\) and \(SERPINA12\) \((p = 0.0391)\) SNVs were more weakly associated with PON1 activity \((\text{Table 3})\). The \(ITGAL\) gene region SNVs were not associated with PON1 AREase activity using default weights in SKAT \((p = 0.169)\); however, when using a weighting algorithm that gave
more predictive power to the rarest SNVs (MAF < 1%), *ITGAL* SNVs were significantly associated with PON1 activity ($p = 2.63 \times 10^{-3}$).

To test whether the rare variant association results were independent of the common SNP associations detected in the CVD chip analysis, we ran a secondary analysis in the overlapping subjects with both CVD and Exome chip genotypes ($n=933$), adjusting the SKAT model for the genotypes of SNPs from the common variant analysis with FDR < 0.10 (see Table 2) in addition to age, sex, current smoking status, and the four functional *PON1* genotypes. From this analysis, adjusting for common variants, *PON1* ($p = 3.42 \times 10^{-3}$) and *FTO* ($p = 9.52 \times 10^{-3}$) SNVs remained significant for association with adjusted PON1 AREase activity. *SERPINA12* became marginally significant ($p = 0.09$) when adjusting for common *SERPINA12* variants previously identified in this cohort. When adjusting for its common variants, *ITGAL* SNVs remained significant for association with PON1 activity using adjusted weights ($p = 9.88 \times 10^{-4}$).

Prior research in a male-only subset of this cohort identified SNPs in *PON2* and *PON3* that predicted of PON1 activity(26). Thus, we performed a secondary SKAT analysis of the entire *PON* gene cluster (*PON1, PON2, PON3*) on chromosome 7. In total, there were 13 polymorphic SNVs within the region represented in our cohort, including the seven *PON1* SNVs previously analyzed by SKAT (see Table 3). When testing for association with PON1 AREase in a SKAT model adjusting for age, sex, current smoking status, and the four functional *PON1* mutations, the *PON* gene cluster was highly associated ($p = 7.10 \times 10^{-4}$), though modestly less significant when compared to the *PON1* gene region alone ($p = 2.24 \times 10^{-4}$), suggesting that little predictive power was gained by considering *PON2* and *PON3* SNVs. To evaluate the role of *PON2* and
PON3 SNVs, we analyzed PON2 (2 SNVs in the gene) and PON3 (4 SNVs in the gene) separately and found that while PON3 SNVs were significantly associated with PON1 AREase activity (p = 0.022), PON2 SNVs were not (p = 0.76). As the two SNVs present in PON2 (rs12026 and rs7493) in our cohort are in LD (r^2 = 1.0, MAF = 0.246 for both), we ran a separate association analysis considering only rs12026. PON2 SNV rs12026 remained not significantly associated with PON1 activity (p=0.65).

Prior work in the CLEAR cohort identified six additional SNPs in the PON gene cluster(26) (four in PON1 and one each in PON2 and PON3), excluding the four commonly recognized functional SNPs, which were predictive of PON1 activity. We therefore performed a separate SKAT analysis of PON1 and the PON gene cluster SNV associations with PON1 AREase activity to insure that our results were not due to LD with these previously identified common variants. In this SKAT model, the model included the covariates of the genotypes of these six previously identified SNPs (rs854567, rs2299257, rs223783, rs2375005, rs3917486, and rs11768074), as well as the four functional PON1 mutations, age, sex, and current smoking status. We found that the p-values for both PON1 (1.92 x 10^{-3}) and the PON gene cluster (9.30 x 10^{-3}) were less significant when adjusting for these variants, though still highly predictive of PON1 activity. Only PON1 SNV rs3917503 from the Exome chip data was in LD of r^2 > 0.60 with one of the six previously identified common variants, rs2299257 (r^2 = 0.74). These results suggest that, except for rs2299257, the SNVs reported to be associated with PON1 activity here are independent of previously reported common variants.

Description of the PON gene cluster SNVs in the ESP6500 data
Selection of only nonsynonymous and 5’- and 3’-noncoding SNVs from the ESP6500 data identified 35 SNVs within \textit{PON1}, which are presented in \textbf{Supplemental Table 1}. Of these 35 SNVs, two and three are 3’-noncoding and 5’-noncoding, respectively. Of the 30 remaining nonsynonymous SNVs (nsSNVs), there are two nonsense mutations (at protein positions 21 and 194, respectively), one missense-near-splice, and 27 missense variants. Included in the 27 nsSNVs are two of the four functional \textit{PON1} mutations, \textit{PON1}_{Q192R} (rs662) and \textit{PON1}_{L55M} (rs854560). These two nsSNPs are also the only nsSNPs with MAF > 1\% in the ESP6500 data for \textit{PON1}. Ten and 16 rare SNVs are present only in AA and EA populations, respectively.

Of the 24 SNVs in \textit{PON2}, there were 21 nonsynonymous, three 3’ noncoding, and zero 5’ noncoding SNVs (see \textbf{Supplemental Table 2}). Of the 21 nsSNVs, there are two missense-near-splice and 19 missense variants. Of these 24 variants, only two nsSNPs are common: rs7493 (Ser299Cys; MAF=0.247) and rs12026 (Ala136Gly; MAF=0.246). Seven and 14 rare SNVs are present only in AA and EA populations, respectively.

Of the 33 SNVs in the ESP6500 dataset, 31 were nonsynonymous, one was 5’ noncoding, and one was 3’ noncoding (see \textbf{Supplemental Table 3}). Of the 31 nsSNVs, there was one stop mutation (at position 32), three missense-near-splice, and 27 missense SNVs. Only one identified SNV had a MAF greater than 0.01: rs17883013 (Ala179Asp; MAF=0.018). Twelve and 14 rare SNVs are present only in AA and EA populations, respectively.

As expected, we did not observe all of the reported ESP SNVs in our data set. Considering the EA ESP variants, we observed 12.2\% (10/82) of the coding and none of the 3’- and 5’-noncoding
variants identified in the ESP6500 data. Thus, larger studies are needed to determine whether the further variation in the PON gene cluster identified in the EVS data is functional.
Discussion

PON1 is a glycoprotein enzyme physically associated with HDL. It demonstrates wide substrate specificity: from oxidized LDL and its resulting atheroprotective effects, to the metabolism of toxic organophosphates, statins(35), and the quorum-sensing factor of *Pseudomonas aeruginosa*(36, 37). Perhaps due to this extensive substrate specificity, PON1 also has been implicated in numerous human diseases in addition to vascular disease, such as Parkinson’s disease(38-42), systemic lupus erythematosus(43-45), breast cancer(46, 47), age-related macular degeneration(48-50), and diabetes(51), among other disorders(52). Given the recent failures of HDL-C to prevent atherosclerotic outcomes in recent trials and the varied effects of PON1 on numerous important human diseases, a fuller understanding of PON1 activity may shed light on the cardioprotective role of HDL. To this end, we have expanded the search for genetic variation that predicts PON1 activity to loci outside of the PON1 coding region and to rare variants.

In this study, we report novel findings that common variants in *FTO, ITGAL* and *SERPINA12* predict PON1 activity. In addition, we find that rare SNVs within the *PON* gene cluster are predictive of PON1 AREase activity. For *FTO, ITGAL* and *PON1* there were significant effects of rare SNVs on PON1 activity that were independent of the effects of the common variants, while for *SERPINA12* the independent effects of the rare variants were marginal (p=0.09). Variation in *FTO, ITGAL* and *SERPINA12* has not previously been explored for effects on PON1 activity.

Of these three novel loci, both *FTO* and *SERPINA12* are related to the cardiovascular risk factors diabetes and BMI. *FTO* encodes a nuclear protein of the AlkB-related non-heme iron
superfamily and has been associated with type-2 diabetes (53-55) as well as obesity and body mass index (BMI) in humans (56, 57). Its mechanism is not understood. Interestingly, FTO variants have been reported to have gene-by-gene interaction with PON1 SNP rs854560 in the prediction of BMI (58). However, none of the 5’ PON1 activity-predicting FTO SNPs in our study are in significant LD ($r^2 < 0.60$) with the FTO SNPs implicated in genome-wide association studies for diabetes or BMI, or for gene-by-gene interactions with PON1. Moreover, PON1 activity and BMI are not significantly associated ($p=0.56$) in our data. When we tested all FTO SNPs present on the CVD chip, we did find significant ($p<0.05$) associations with BMI in SNPs not predictive for PON1, but no associations in any FTO SNP with diabetes in our data (data not shown). FTO variants have not previously been tested for association with PON1 activity. SERPINA12, or visceral adipose tissue-derived serpin (VASPIN), encodes an insulin-sensitizing adipocytokine (59). Like FTO, SERPINA12 variants have been implicated in human type-2 diabetes (60, 61) and obesity (59, 62). However, when testing all SERPINA12 variants present in our data, we did not find a significant association with either BMI or diabetes (data not shown). Interestingly, PON1 activity has long been known to be lower in those with diabetes (51) and PON1 has also been reported as an anti-diabetic enzyme that stimulates insulin release from pancreatic beta cells (63). However, SNPs within PON1 have not been found to predict diabetes (64). The SERPINA12 SNP rs7152296 minor allele was associated with a decrease in PON1 AREase (beta coefficient = -4.8); in contrast, all FTO SNP minor alleles identified were positively associated with PON1 activity. Further functional studies will be required to determine both the relationships of FTO and SERPINA12 to paraoxonase biology; these results suggest that the relationship of PON1 activity to diabetes may be driven by loci
other than \textit{PON1} and be more complex than the prior hypothesis of reduced PON1 in diabetes due to oxidative stress\cite{55, 63}.

In contrast to \textit{FTO} and \textit{SERPINA12}, \textit{ITGAL} is not directly related to cardiovascular disease and instead encodes the integrin alpha L chain, which combines with the beta-2 chain to form the integrin lymphocyte function-associated antigen (LFA-1), which is expressed on all leukocytes\cite{65}. LFA-1 is a receptor for intercellular adhesion molecules (ICAMs) and plays a key role in leukocyte intercellular adhesion through interaction with ICAMs. Like HDL\cite{66, 67}, PON1 is an acute phase reactant and has been associated with the immune response\cite{68, 69} and inflammation\cite{70}. Specifically, PON1 has been demonstrated to be a host modulator of \textit{Pseudomonas aeruginosa} quorum sensing\cite{36}, with transgenic expression of \textit{PON1} being protective in drosophila from infection lethality\cite{37}. Therefore, the finding of \textit{ITGAL} SNP rs1557672 being associated with a decrease in PON1 activity (beta coefficient = -7.16) may indicate a further role of immune regulation of HDL, and specifically, PON1 activity.

Of the SNPs found to be associated with PON1 activity at a level of suggestive significance, SNPs in two genes, \textit{ALOX12} and \textit{PDE4D}, and near \textit{FAM178A} did not meet our FDR < 0.10 cutoff for further consideration. Notably 12-lipoxygenase (12-LO), like paraoxonase, can oxidize esterified fatty acids in lipoproteins and phospholipids and has been implicated in atherosclerosis in animal models\cite{71, 72}. Paraoxonase, on the other hand, appears to metabolize oxidized lipids\cite{2}. Further, \textit{ALOX12} variation has been associated with inflammatory markers CRP and ICAM\cite{73} and poor glycemic control\cite{74} in type 2 diabetes patients. Interestingly, 12-LO is a non-heme, yet iron-containing, enzyme and we have recently shown that dietary iron
is associated with reduced PON1 activity (18). One could speculate that the inhibitory effect of iron may be through increased 12-LO. We observe that the SNPs in ALOX12 and PDE4D, and nearby FAM178A, are among those with the largest effect size, but the smallest minor allele frequencies. Thus, it is possible that these represent true positives that we are underpowered to detect a significant association for due to low minor allele frequency. Further investigation of these loci, especially of ALOX12, due to its biology, is suggested by our findings.

Utilizing SKAT to analyze both the rare and common variants represented on the Illumina Exome chip, we were able to detect significant associations for PON1, PON3, or the PON gene cluster as a whole, even when known functional SNPs were considered as covariates. PON3 is also an HDL-associated enzyme, but is expressed in much lower levels than PON1 (75). In contrast to both PON1 and PON3, PON2 is ubiquitously expressed (76). Given that SNPs in PON2 have been previously reported in this cohort to predict PON1 enzyme activity, the failure of additional PON2 SNPs to predict PON1 activity in SKAT analyses may reflect a lack of rare SNVs in this sample, as only 2 PON2 SNPs from the Exome chip were polymorphic in the CLEAR cohort; these two SNVs, rs7493 and rs12026, were common and in tight LD (MAF of 24.6% for both and r² = 1.0).

When adjusting for PON gene cluster variants that have been previously associated with PON1 activity (26), the p-values for association of the previously unidentified variants remained significant, but not as strongly associated with PON1 AREase for both PON1 and the PON gene cluster. One possible explanation for the change is non-independence of previously identified SNPs and the SNVs tested from the Exome chip due to LD. Such LD was found only for PON1
SNP rs3917503 from the Exome chip data and the previously identified SNP, rs2299257 ($r^2 = 0.74$). Rs2299257 was reported to predict 0.85% of PON1 enzyme activity in the CLEAR study(30). The MAFs for rs3917503 and rs2299257 are 0.46 and 0.39, respectively, with both minor alleles being associated with an increase in PON1 AREase activity (data not shown). Both SNPs are intronic (rs3917503 is in the third and rs2299257 is in the fourth of eight introns of PON1). It is also notable that 2 of the 7 SNVs present in PON1 (Leu90Pro, MAF=0.0010, and Trp194STOP, MAF=0.0005) were first described as functional in the CLEAR study in 2003 by an early application of sequencing based on extreme or atypical phenotype(17). Overall, our results support the expectation of rare functional alleles.

The current catalogue of PON gene cluster nonsynonymous and 5’/3’-noncoding SNVs in the ESP6500 data, both rare and common, are also summarized here. Of the 93 such SNVs presented in PON1, PON2, and PON3 in the ESP data, only 7, 2, and 4 variants, respectively, are present in these data. With regard to SNVs present in EA stratified populations, we had data on 4 missense, 2 intronic, and 1 nonsense SNVs within PON1. To the best of the authors’ knowledge, this marks the first publication analyzing the effects of the PON1 nonsense SNV, Trp194STOP. Of the 14 PON1 EA variants with a Genomic Evolutionary Rate Profiling (GERP) conservation score greater > 3 present, 3 were present here. No rare PON2 variants were present in our data, while 10 PON2 SNVs with MAF < 0.01 and GERP conservation score > 3 were present in EA subjects in the ESP6500 data. These data had 4 PON3 SNVs: 1 intronic, 1 missense, 1 missense-near-splice, and 1 nonsense. Of the 17 EA variants with a GERP conservation score > 3, only 2 are present in this study.
Some limitations of this study must be considered. First, this cohort is composed entirely of subjects of European ancestry, which limits the generalization to other ancestral groups. This is particularly of interest as there are many rare SNVs in the PON gene cluster that are only present in individuals of African ancestry. Similarly, although our study is large in size in the context of paraoxonase epidemiology, it remains small when considering rare genetic variants; thus, true positives may be underpowered or unsampled here. Additionally, our analysis strategy for rare variants utilized a candidate gene approach based on findings from common variant analyses. The common variant analyses were performed with the Illumina HumanCVD BeadChip, which was selective for genes that were potentially involved in cardiovascular phenotypes. The CVD chip, and thus this study, could have excluded consideration of loci relevant to PON1 biology. Finally, while replication in an independent sample is ideal, given the need for larger datasets to identify rare variant effects, we considered all available data, rather than stratifying to provide a replication set. Cohorts in which PON1 activity can be measured are limited due to the fact that PON1 activity is calcium dependent and, thus, cannot be measured from plasma stored in EDTA tubes due to calcium chelation and irreversible inactivation of PON1. Therefore, replication data were not available. However, our finding of independent effects of both common and pooled rare variants in the same loci for the same trait provides some validation of the associations of variation at these loci and PON1 activity.

In conclusion, we describe the novel findings that common and rare genetic variants in FTO, ITGAL and SERPINA12 are associated with PON1 enzyme activity. Additionally, an association with ALOX12 variants does not meet our experiment-wide cut-off, but suggests the need for further study. These associations strengthen the tie of PON1 activity to diabetes and
inflammation. We also identify rare variants in these loci and the PON1 gene cluster that predict PON1 activity. In addition, we describe 92 non-synonymous, 5’- and 3’-noncoding variants from the ESP6500 data for the PON gene cluster. Given the importance of PON1 to HDL biology and human disease and the potential for rare variants to explain a large portion of the “missing heritability” of complex genetic traits, future studies will be needed to further elucidate the genetic determinants of PON1 enzymatic activity and their relationship to cardiovascular disease risk, including BMI and inflammation.

**Disclosures:** None

**Conflicts of Interest:** None

**Acknowledgements:** We would like to thank the study participants. We would also like to thank the following people for their technical assistance: Tamara J Bacus, Thomas Hatsukami, Ella R. Jarvik, Julieann Marshall, Laura McKinstry, and Karen Nakayama. This work was funded by NIH RO1 HL67406 and a State of Washington Life Sciences Discovery Award to the Northwest Institute of Genetic Medicine. DSK was supported in part by a Sarnoff Cardiovascular Research Fellowship for Medical Students Award. This work utilized resources of SeattleSNPs; NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, WA (http://pga.gs.washington.edu). Past work in this cohort and specifically Dr. Boyko’s involvement, was supported in part by resources from the VA Puget Sound Health Care System, Seattle, Washington, including the Veteran Affairs Epidemiologic Research and Information Center (award CSP 701S).
References


55. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.


Table 1. Description of the CLEAR cohort.

<table>
<thead>
<tr>
<th>Baseline Characteristics</th>
<th>Common Variant Analysis (n=1,362)</th>
<th>SKAT Analysis (n=1,051)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1181 (86.7%)</td>
<td>944 (89.8%)</td>
</tr>
<tr>
<td>Female</td>
<td>181 (13.3%)</td>
<td>107 (10.2%)</td>
</tr>
<tr>
<td>Age, mean ± SE, years</td>
<td>67.80 ± 9.51</td>
<td>69.63 ± 9.08</td>
</tr>
<tr>
<td>CAAD Case-Control Status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe CAAD (≥80% stenosis)</td>
<td>460 (33.8%)</td>
<td>535 (50.9%)</td>
</tr>
<tr>
<td>Moderate CAAD (50-79% stenosis)</td>
<td>166 (12.2%)</td>
<td>204 (19.4%)</td>
</tr>
<tr>
<td>Mild CAAD (15-49% stenosis)</td>
<td>27 (2.0%)</td>
<td>0</td>
</tr>
<tr>
<td>Control (&lt;15% stenosis bilaterally)</td>
<td>687 (50.5%)</td>
<td>312 (29.7%)</td>
</tr>
<tr>
<td>Lower Extremity Vascular Disease</td>
<td>21 (1.5%)</td>
<td>0</td>
</tr>
<tr>
<td>Current smoking status, n (%)</td>
<td>248 (18.2%)</td>
<td>157 (14.9%)</td>
</tr>
<tr>
<td>Statin use, n (%)</td>
<td>536 (39.4%)</td>
<td>552 (52.5%)</td>
</tr>
<tr>
<td>PON1 AREase activity, mean ± SE, IU</td>
<td>135.3 ± 52.5</td>
<td>135.4 ± 49.7</td>
</tr>
</tbody>
</table>

AREase = PON1 arylesterase enzyme activity. CAAD = carotid artery disease; CLEAR = Carotid Lesion Epidemiology and Risk study. PON1 = paraoxonase.

1. SKAT = sequence kernel association testing.
Table 2. Top CVD chip SNP results for association with PON1 AREase activity in the CLEAR cohort (n=1,362).  

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position</th>
<th>Closest Reference Gene</th>
<th>Minor/Major Allele</th>
<th>MAF</th>
<th>Beta Coefficient</th>
<th>FDR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2237583d</td>
<td>7</td>
<td>94,788,113</td>
<td>PON1</td>
<td>A/G</td>
<td>0.3991</td>
<td>13.76</td>
<td>0.000284</td>
<td>3.88x10^-8*</td>
</tr>
<tr>
<td>rs3917478d</td>
<td>7</td>
<td>94,789,505</td>
<td>PON1</td>
<td>A/G</td>
<td>0.2835</td>
<td>17.81</td>
<td>0.007168</td>
<td>1.14x10^-6*</td>
</tr>
<tr>
<td>rs7199363e</td>
<td>16</td>
<td>52,596,515</td>
<td>FTO</td>
<td>A/G</td>
<td>0.357</td>
<td>11.36</td>
<td>0.0406</td>
<td>7.37x10^-6*</td>
</tr>
<tr>
<td>rs2003583e</td>
<td>16</td>
<td>52,657,507</td>
<td>FTO</td>
<td>G/A</td>
<td>0.248</td>
<td>11.11</td>
<td>0.07104</td>
<td>1.48x10^-5*</td>
</tr>
<tr>
<td>rs1136002e</td>
<td>16</td>
<td>52,618,642</td>
<td>FTO</td>
<td>G/A</td>
<td>0.3363</td>
<td>9.899</td>
<td>0.07404</td>
<td>1.87x10^-5*</td>
</tr>
<tr>
<td>rs7152296e</td>
<td>14</td>
<td>94,030,137</td>
<td>SERPINA12</td>
<td>A/G</td>
<td>0.1574</td>
<td>-4.763</td>
<td>0.07701</td>
<td>1.99x10^-5*</td>
</tr>
<tr>
<td>rs1557672</td>
<td>16</td>
<td>30,418,072</td>
<td>ITGAL</td>
<td>G/A</td>
<td>0.3382</td>
<td>-7.162</td>
<td>0.07701</td>
<td>1.99x10^-5*</td>
</tr>
<tr>
<td>rs2299262d</td>
<td>7</td>
<td>94,787,864</td>
<td>PON1</td>
<td>G/A</td>
<td>0.1185</td>
<td>9.625</td>
<td>0.1454</td>
<td>3.96x10^-5</td>
</tr>
<tr>
<td>rs4784351e</td>
<td>16</td>
<td>52,633,199</td>
<td>FTO</td>
<td>G/A</td>
<td>0.297</td>
<td>9.609</td>
<td>0.1866</td>
<td>5.63x10^-5</td>
</tr>
<tr>
<td>rs11571365</td>
<td>17</td>
<td>6,854,579</td>
<td>ALOXI2</td>
<td>A/G</td>
<td>0.0154</td>
<td>33.96</td>
<td>0.1866</td>
<td>5.93x10^-5</td>
</tr>
<tr>
<td>rs2192872e</td>
<td>16</td>
<td>52,632,128</td>
<td>FTO</td>
<td>A/G</td>
<td>0.2423</td>
<td>9.216</td>
<td>0.1893</td>
<td>6.44x10^-5</td>
</tr>
<tr>
<td>rs27168</td>
<td>5</td>
<td>58,648,889</td>
<td>PDE4D</td>
<td>A/G</td>
<td>0.0858</td>
<td>15.37</td>
<td>0.2237</td>
<td>8.13x10^-5</td>
</tr>
<tr>
<td>rs11595116</td>
<td>10</td>
<td>102,632,697</td>
<td>(FAM178A)</td>
<td>A/G</td>
<td>0.0934</td>
<td>13.81</td>
<td>0.2256</td>
<td>8.71x10^-5</td>
</tr>
</tbody>
</table>

Chr = Chromosome; FDR = false discovery rate; LD = linkage disequilibrium; MAF = minor allele frequency; SNP = single nucleotide polymorphism.  
* = FDR < 0.10. Locus carried forward for rare variant testing.  
= Position information based upon reference 36.3 build.  
= Parentheses indicate that rs11595116 is located 39,665 bases upstream of FAM178A, rather than at the locus.  
= Beta coefficients, FDR, and P-values based upon linear regression in the CLEAR cohort adjusting for age, sex, current smoking status, and the four highest variance functional PON1 mutations: PON1_A-162G, PON1_C-108T, PON1_L55M, and PON1_Q192R, see methods.  
= LD among rs2237583, rs3917478, and rs2299262 is r^2 < 0.60.  
= Pairwise LD in FTO SNPs are as follows: rs7199363 and rs2004583, r^2 = 0.80; rs1136002 and rs2192872, r^2 = 0.96; and rs4784351 and rs2192872, r^2 = 0.813.
Table 3. Pooled rare exome chip variant genotypes tested for effects on PON1 activity prediction in the CLEAR cohort.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Number of SNPs&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Variant(rsID)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>PhastCons</th>
<th>GERP</th>
<th>Protein Change</th>
<th>Minor Allele, CLEAR MAF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PON1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7:94930391(rs854555)</td>
<td>0.006</td>
<td>-1.12</td>
<td>Intrinsic</td>
<td>A, 0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7:94937419(rs80019660)</td>
<td>0.409</td>
<td>4.46</td>
<td>Ala201Val</td>
<td>A, 0.0033</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7:94937439(rs3917594)</td>
<td>1</td>
<td>4.46</td>
<td>Trp194STOP</td>
<td>T, 0.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7:94944735(rs72552788)</td>
<td>1</td>
<td>4.48</td>
<td>Leu90Pro</td>
<td>G, 0.0010</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7:94945453(rs3917503)</td>
<td>0</td>
<td>-1.13</td>
<td>Intrinsic</td>
<td>T, 0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7:94953733</td>
<td>0.002</td>
<td>1.93</td>
<td>Asn19Asp</td>
<td>C, 0.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7:94953771</td>
<td>0</td>
<td>-2.34</td>
<td>Ala6Val</td>
<td>A, 0.0014</td>
<td></td>
</tr>
<tr>
<td><strong>FTO</strong></td>
<td>4</td>
<td>16:53818708(rs3751813)</td>
<td>0.001</td>
<td>-0.268</td>
<td>Intrinsic</td>
<td>G, 0.47</td>
<td><strong>0.0190</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16:53860139</td>
<td>0.059</td>
<td>0.959</td>
<td>Ala163Thr</td>
<td>A, 0.0029</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16:53860253</td>
<td>0.911</td>
<td>2.99</td>
<td>Val201Ile</td>
<td>A, 0.0019</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16:54120952(rs11646512)</td>
<td>0.953</td>
<td>0.553</td>
<td>Intrinsic</td>
<td>A, 0.45</td>
<td></td>
</tr>
<tr>
<td><strong>ITGAL</strong></td>
<td>4</td>
<td>16:30490776</td>
<td>0.999</td>
<td>0.319</td>
<td>Synonymous</td>
<td>A, 0.0005</td>
<td><strong>0.169</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16:30492823(rs1064524)</td>
<td>0</td>
<td>-9.69</td>
<td>Arg214Trp&lt;sup&gt;e&lt;/sup&gt;</td>
<td>T, 0.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16:30518041(rs2230433)</td>
<td>0.001</td>
<td>-9.09</td>
<td>Arg791Thr&lt;sup&gt;e&lt;/sup&gt;</td>
<td>C, 0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16:30531182</td>
<td>0.38</td>
<td>1.42</td>
<td>Val1078Ala&lt;sup&gt;e&lt;/sup&gt;</td>
<td>C, 0.0005</td>
<td></td>
</tr>
<tr>
<td><strong>SERPINA12</strong></td>
<td>10</td>
<td>14:94953767</td>
<td>0</td>
<td>-0.112</td>
<td>Ala373Val</td>
<td>A, 0.0019</td>
<td><strong>0.0391</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14:94953823</td>
<td>1</td>
<td>2.39</td>
<td>His354Gln</td>
<td>C, 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14:94953832</td>
<td>0.991</td>
<td>4.84</td>
<td>3’ Splice</td>
<td>A, 0.0014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14:94954368(rs11626701)</td>
<td>0.003</td>
<td>0.969</td>
<td>Intrinsic</td>
<td>A, 0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14:94956030</td>
<td>0.695</td>
<td>4.78</td>
<td>Gly327Asp</td>
<td>T, 0.0014</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14:94962714</td>
<td>0</td>
<td>-0.659</td>
<td>Arg301Cys</td>
<td>A, 0.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14:94962887</td>
<td>0</td>
<td>-10.3</td>
<td>Arg243His</td>
<td>T, 0.0005</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>p-value</td>
<td>Log2 odds</td>
<td>Mutation</td>
<td>Allele</td>
<td>p-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>-----------</td>
<td>--------------</td>
<td>------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:94964104</td>
<td>0.0071</td>
<td>4.92</td>
<td>Arg211STOP</td>
<td>A</td>
<td>0.0071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:94964127</td>
<td>0.0048</td>
<td>3.99</td>
<td>Leu203Pro</td>
<td>G</td>
<td>0.0048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:94964220</td>
<td>0.0005</td>
<td>-6.14</td>
<td>Met172Thr</td>
<td>G</td>
<td>0.0005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AA = African Ancestry; EA = European Ancestry; ESP = Exome Sequencing Project. GERP = Genomic Evolutionary Rate Profiling; GERP conservation score > 3 is highly conserved. PhastCons = PHylogenetic Analysis with Space/Time models; PhastCons identifies highly evolutionary conserved elements (~1) in DNA.

a = Number of SNPs within gene reflects only the sites that vary (MAF > 0) in these data.
b = Positional information based upon NCBI 37 build.
c = When adjusting for six previously identified SNPs in the PON gene cluster (rs854567, rs2299257, rs223783, rs2375005, rs3914486, and rs11768074) and the 4 functional PON1 SNPs, age, sex, and current smoking status, the p-values for PON1 and the PON gene cluster are 1.92x10^{-3} and 9.30x10^{-3}, respectively.
d = SKAT using adjusted weights (see the Statistical Methods section), ITGAL p = 2.63x10^{-3}.
e = Protein change position information based on the longer transcript, of length = 1171 amino acid residues.