Rare coding variation in paraoxonase-1 is associated with ischemic stroke in the NHLBI Exome Sequencing Project

Cover Title: Rare PON1 variation is associated with stroke

Daniel Seung Kim¹,², David R. Crosslin¹,², Paul L. Auer³,⁴, Stephanie M. Suzuki¹, Judit Marsillach¹,², Amber A. Burt¹, Adam S. Gordon², James F. Meschia⁵, Mike A. Nalls⁶, Bradford B. Worrall⁷,⁸,⁹, W. T. Longstreth Jr.¹⁰,¹¹, Rebecca F. Gottesman¹², Clement E. Furlong¹,², Ulrike Peters³,¹¹, Stephen S. Rich⁷, Deborah A. Nickerson², Gail P. Jarvik¹,² on behalf of the NHLBI Exome Sequencing Project.

Address for Correspondence:
Gail P. Jarvik, M.D., Ph.D.
Medical Genetics, Box 357720
University of Washington
Seattle, WA 98195-7720
Phone: 1-206-221-3974
Fax: 1-206-543-3050
Email: pair@u.washington.edu

1. Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, WA

2. Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA
3. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

4. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA.

5. Department of Neurology, Mayo Clinic, Jacksonville, FL

6. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD

7. Center for Public Health Genomics, University of Virginia, Charlottesville, VA

8. Department of Neurology, University of Virginia, Charlottesville, VA

9. Department of Public Health Sciences, University of Virginia, Charlottesville, VA

10. Department of Neurology, University of Washington, Seattle, WA

11. Department of Epidemiology, University of Washington, Seattle, WA

12. Cerebrovascular Division, Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD.

Tables: 3

Color Figures: 1

**Keywords:** Paraoxonase, rare variation, genetics, stroke, atherosclerosis

**Word Count:** 5,604 total (including references)

**References:** 20
Abstract

HDL-associated paraoxonase-1 (PON1) is an enzyme whose activity is associated with cerebrovascular disease. Common PON1 genetic variants have not been consistently associated with cerebrovascular disease. Rare coding variation that likely alters PON1 enzyme function may be more strongly associated with stroke. The NHLBI Exome Sequencing Project (ESP) sequenced the coding regions (exomes) of the genome for heart, lung, and blood-related phenotypes (including ischemic stroke). In this sample of 4,204 unrelated participants, 496 had verified, non-cardioembolic ischemic stroke. After filtering, 28 non-synonymous PON1 variants were identified. Analysis with the Sequence Kernel Association Test (SKAT), adjusted for covariates, identified significant associations between PON1 variants and ischemic stroke (p=3.01x10^{-3}). Stratified analyses demonstrated a stronger association of PON1 variants with ischemic stroke in African ancestry (AA) participants (p=5.03x10^{-3}). Ethnic differences in the association between PON1 variants with stroke could be due to the effects of PON1 Val109Ile (overall p=7.88x10^{-3}; AA p=6.52x10^{-4}), found at higher frequency in AA participants (1.16% vs. 0.02%) and whose protein is less stable than the common allele. In summary, rare genetic variation in PON1 was associated with ischemic stroke, with stronger associations identified in those of AA. Increased focus on PON1 enzyme function and its role in cerebrovascular disease is warranted.

Abstract Word Count: 200 / 200
Background

Recent results from a large-scale Mendelian randomization study(1) and randomized clinical trial(2) investigating high density lipoprotein (HDL) have raised doubt on the long-held belief that total HDL cholesterol (HDL-C) is cardioprotective. In light of these findings, research has shifted to the individual components of HDL, whose activities are not reflected by usual measures of HDL-C. Paraoxonase 1 (PON1), encoded by the PON1 gene, is a liver-produced glycoprotein enzyme whose enzyme activity is strongly cardioprotective, particularly for carotid artery disease(3), a risk factor for ischemic stroke.

Numerous single nucleotide variants (SNVs), including rare protein truncating(4) and promoter SNVs(5) that alter gene expression have been described for PON1. Three specific PON1 variants (PON1-108C/T, PON1L55M, and PON1Q192R) have been extensively studied for their strong effects on PON1 expression, enzyme activity, or both. Despite the strong association between PON1 enzyme activity and cerebrovascular disease, common PON1 SNVs (minor allele frequency (MAF) greater than 5%) have not been consistently associated with atherosclerotic end-organ damage(6, 7). Moreover, meta-analyses of PON1Q192R have found only a weak association with coronary artery disease (CAD), while PON1-108C/T and PON1L55M have no demonstrated evidence for CAD association(8, 9).

Rare coding SNVs are often unique to an individual or family and likely alter protein function, possibly accounting for a greater portion of genetic risk and missing heritability than common SNVs(10). By focusing on the putative deleterious coding SNVs in PON1 that result in a change or loss in PON1 enzyme activity, a stronger association between PON1 variation and
cerebrovascular disease may be revealed. The goal of this study was to determine whether the burden of rare coding variation in the \textit{PON1} gene was associated with ischemic stroke in participants of the NHLBI Exome Sequencing Project (ESP) and to functionally characterize the most strongly associated rare variant with non-cardioembolic ischemic stroke.
Methods

Ethics Statement:
Institutional review boards at each individual site involved in the ESP approved the study, and each study participant at each study site provided written, informed consent.

Participants:
The National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project (ESP) is a multi-center study to deeply sequence the exomes of individuals with a variety of heart, lung, and blood disorders. The participants in the ESP were ascertained from different studies, defined by three consortia: WHISP (Women’s Health Initiative Sequencing Project), HeartGO (Framingham Heart Study, Jackson Heart Study, Multi-Ethnic Study of Atherosclerosis, Atherosclerosis Risk in Communities, Coronary Artery Risk Development in Young Adults, and Cardiovascular Health Study), and LungGO (Genomic Research on Asthma in the African Diaspora, Lung Health Study, Pulmonary Arterial Hypertension population, Acute Lung Injury cohort, and the Cystic Fibrosis cohort). Additional participants with ischemic stroke were identified from two independent studies, Siblings With Ischemic Stroke Study (SWISS) and the Ischemic Stroke Genetics Study (ISGS). Other ischemic stroke participants represented by these analyses were identified within HeartGO and the WHISP. Previous analysis of a large subset (n=2,440) of the ESP found no systematic biases in patterns and characteristics of SNVs attributable to cohort or technical sources of variation(11).

Unrelated ischemic stroke cases (non-cardioembolic small/lacunar and large/atherosclerotic vessel subtypes), diagnosed <65 years of age or with a positive family history of stroke, were
selected for inclusion from HeartGo (n=250) and WHISP (n=250). Additional affected sib-pairs (n=50) with ischemic stroke were selected from SWISS. Subjects with hemorrhagic stroke were excluded from all analyses.

**Exome Sequencing and Variant Calling:**

Exome sequencing was performed at the University of Washington and the Broad Institute of MIT/Harvard University. Library construction, exome capture, sequencing, and mapping were performed as previously described(11). Multi-sample variant calling was conducted at the University of Michigan; detailed information on the calling methods can be found in the Supplemental Materials.

**Single Nucleotide Variant Filtering:**

Genetic variants within the *PON1* gene cluster were extracted from variant call format (VCF) files. SNVs were filtered for a minimum read-depth of 8x, 97% overall site call-rate, and a Hardy-Weinberg equilibrium rejection cut-off $P = 10^{-6}$. Only non-synonymous coding SNVs that are predicted to alter protein residues (missense), splicing of mRNA transcripts (splice), or prematurely truncate proteins (nonsense) were included for analyses. After applying these criteria, a total of 28 SNVs remained for *PON1*. Description of these SNVs can be found in Supplemental Table S1 and can also be found online on the Exome Variant Server (http://evs.gs.washington.edu/EVS/). Conservation for single-base variants was assessed through genomic evolutionary rate profiling (GERP)(12) scores using SeattleSeq SNP annotation (http://snp.washington.edu/SeattleSeqAnnotation137).
Analyses:

Subject Filtering:

Of the 6,823 participants in the ESP dataset, 4,224 were used for analyses of PON gene cluster variation association with ischemic stroke. Exclusion criteria for this specific study included: relatedness up to the 3rd degree (first cousins – as described in the Supplemental Materials), sex mismatch, low concordance with prior genotype data, and individual genotype call rate < 90%.

As SWISS recruited sibships with ischemic stroke that were then sequenced as part of the NHBLI ESP, only one sibling from each pair was used for analyses (n=49 cases). For ischemic stroke controls, additional phenotype exclusion criteria excluded participants with other cardiovascular or potentially confounding phenotype (e.g., myocardial infarction, chronic obstructive pulmonary disease, and ventilator use) and cystic fibrosis. Participants who were collected for high levels of cardiac risk factors (high blood pressure, high low-density lipoprotein levels, high body mass index) but who had not had any noted cardiovascular outcomes (e.g., stroke or myocardial infarction) were included as “controls” for the purposes of this study.

Genetic Ancestry:

Genetic ancestry was determined through principal component analysis (PCA). PCA was performed using the SNPRelate R statistical computing package(13). Prior to inclusion into the correlation matrix, SNVs were selected after LD pruning at r=0.5, and a MAF > 0.03. For the sample of 4,204 ESP participants, genetically determined European ancestry was assigned to all participants with eigenvectors 1 and 2 values less than and greater than four (±2) SD from the medians of eigenvectors 1 and 2 of self-identified European ancestry participants (n=2,414). For genetically determined African ancestry, we identified all participants with values less than and
greater than two (±2) SD from the medians of eigenvector 1 and 2 of self-identified African
ancestry participants (n=1677). The process of calculating principal component eigenvectors was
then repeated within the European and African ancestry groups, to obtain ancestry-specific
eigenvectors.

**Statistical Analyses:**

The optimized Sequence Kernel Association Test (SKAT-O)(14) was used for testing association
of SNVs in each of the \textit{PON} genes with ischemic stroke, using an R plugin (http:// r-project.org).
SKAT pools variants across loci, thereby addressing the problem of limited statistical power with
rare variants. It then applies score-based variance-component tests to assess association between
SNV sets within the \textit{PON} gene and ischemic stroke, while adjusting for potentially confounding
covariates in the model. The covariates adjusted for in SKAT analyses of ischemic stroke were
age, sex, current smoking status, and the first three PCA eigenvectors to adjust for population
stratification. Default settings, including small sample size correction when n < 2000, were used
for SKAT analyses. Single variant score test association results were calculated using skatMeta
(http://cran.r-project.org/web/packages/skatMeta/index.html) to identify potential single variant
associations driving the observed \textit{PON1} association with ischemic stroke.

To determine whether one genetic ancestry group was responsible for the observed association,
stratified analyses were performed in AA (n=1,677) and EA (n=2,414) subsets. For these
analyses, genetic ancestry specific PCA eigenvectors were calculated considering only those of a
certain genetic ancestry group to adjust for potential population substructure. These ancestry-
specific PCA eigenvectors were used to adjust for population stratification, in addition to age,
sex, and current smoking status.

**Permutation Testing/Statistical Significance:**

As the NHBLI ESP represents the largest available collection of phenotyped exome sequences, replication of our rare variant results was not possible. Moreover, dividing the existing sample set into discovery and replication groups has been shown to be less powerful than combined analysis; thus, we analyzed all 4,204 subjects together(15). Phenotype permutation testing iterated 100,000 times was used to determine significance. In brief, ischemic stroke and control phenotypes and covariate data were randomly assigned to each of the 4,204 subjects (or 2,414 and 1,677 for EA and AA specific analyses, respectively) and analyses were repeated to obtain a p-value, using the “bootstrap” command in SKAT-O. This permutation process was repeated 100,000 times to obtain a histogram of p-values from phenotype permutation. Using the resulting permutation p-value histogram, a two-sided p-value is reported. All significant gene associations with stroke in each genetic ancestry subgroup (EA and AA) with a p ≤ 0.05 were carried forward to permutation testing. Gene associations with a permutation p-value ≤ 0.05 in conjunction with a prior adjusted p ≤ 0.05 were declared significant. As this was an evaluation of a specific candidate gene (PON1) based upon strong *a priori* data, no attempts at identifying associations across the genome or genome-wide corrections to p-values were performed.

**Functional Characterization of Individual PON1 Variants**

Additional information on the expression, purification, measurement of PON1 arylesterase hydrolysis (AREase) rate, and mass spectrometry confirmation of expressed peptide are detailed in the **Supplemental Materials**.
**PON1 Protein Heat Resistance Assay:**

Assays measuring purified PON1 resistance to heat were performed to assess the function of PON1 protein including the amino acid change most associated with ischemic stroke. These were carried out at 55°C in a GeneMate Mini Dry Bath incubator block. Protein samples were diluted with a Tris-calcium buffer (20 mM Tris-HCl pH 8.0, 1 mM CaCl₂, 150 mM NaCl) to obtain nearly equivalent rates of AREase activity, or about 0.711 Units/ml AREase activity. The PON1R192 protein was diluted 1:20 and the PON1R192-V109I sample was diluted 1:140. Total sample (750 µl) was mixed and 125 µl removed prior to heating. Samples were taken at the timepoints indicated and stored on ice. Samples (20 µl) were assayed in triplicate for each timepoint, with 6 replicates assayed for the starting zero min timepoint. Samples were assayed using a SpectraMax Plus³⁸⁴ Microplate Reader (Molecular Devices) with transparent UV 96-well plates (Greiner Bio-One) for 4 min at 270 nm, with sample absorbance readings taken every 15 s. AREase activity was assayed with 3.26 mM phenyl acetate in a no salt buffer (9 mM Tris-HCl pH 8.0, 0.9 mM CaCl₂). The average activity at the time zero min reading (no heat) was assigned the 100% value and residual AREase activity was expressed as a percentage of the 100% value.
Results

Demographic information of the ESP participants in this analysis is presented in Table 1. A total of 4,204 participants had phenotype, genotype, covariate information, and passed quality control measures. The average age was 57.5 years, 32.1% of the studied population was male, and 21.1% reported being current smokers. Ischemic stroke cases were older and were comprised of proportionally more females, as WHI was a major contributor of stroke cases. Cases had an average age of 61.9 years and 19.2% were male, compared to 56.8 years and 33.8% male for controls. Rates of smoking were similar between the ischemic stroke case and non-stroke control group (21.2 and 21.1, respectively). Genetic ancestry of the cohort was 57.4% EA, 39.9% AA, and 2.7% other ancestry (including Hispanic, Asian, and Native American ancestry). Participants of EA comprised a larger proportion of stroke cases (82.7%) compared to controls (54.0%).

Using SKAT regression methods adjusting for age, sex, current smoking status, and the first three PCA eigenvectors, PON1 (p=1.29x10^{-3}) was associated with ischemic stroke at nominal levels of statistical significance in pooled analyses (Table 2). Permutation testing established the significant association of PON1 with ischemic stroke of p=3.01x10^{-3}.

To explore whether an individual ancestral group was responsible for the observed PON gene cluster associations, we stratified analyses within AA and EA subgroups (Table 2). Using ethnic-specific PCA eigenvectors in addition to age, sex, and current smoking status, PON1 was found to be nominally significant for association with ischemic stroke in the ESP AA-subset (p=5.73x10^{-4}), while the EA-subset was only marginally significant (p=0.07). With permutation
testing of 100,000 iterations, only the association of \( PON1 \) and stroke in the ESP AA-subset remained significant (permutation \( p = 5.03 \times 10^{-3} \)).

\( PON1_{Q192R} \) and \( PON1_{L55M} \) are known determinants of PON1 enzyme activity and have previously been associated with cardiovascular disease(3, 8). To investigate whether the associations observed between \( PON1 \) and ischemic stroke were determined by these two functional \( PON1 \) variants, the SKAT analyses was repeated with the two variants removed. The significance of the association with ischemic stroke for all tested groups (EA, AA, pooled) remained largely unchanged and significant (pooled \( p = 0.00127 \), AA \( p = 5.70 \times 10^{-4} \), EA \( p = 0.07 \)), suggesting the two variants were not entirely responsible for our observed significant associations between \( PON1 \) and stroke.

Individual \( PON1 \) SNV associations with ischemic stroke were determined using a SKAT-based regression approach (Table 3). \( PON1 \) SNV 7:94944679 (Val109Ile missense variant, rs61736513) was positively associated with ischemic stroke in the pooled (\( p = 7.88 \times 10^{-3} \)) and AA-subset (\( p = 6.52 \times 10^{-4} \)) analyses. \( PON1_{V109I} \) MAF was higher in AA (1.19%) compared to EA (0.02%) participants. Two other \( PON1 \) SNVs, 7:94937419 (rs80019660, MAF = 0.23%) and 7:94953721 (rs146211440, MAF = 0.01%), were associated with ischemic stroke in both the pooled and EA-subset analyses (see Table 3). The \( PON1 \) SNV rs80019660 is an Ala201Val missense (found more frequently in EA subjects than AA) that is protective against ischemic stroke, while rs146211440 is a Ser23Ala variant found only in EA subjects that confers increased risk.
To determine the effects of \textit{PON1} SNV rs61736513 (\textit{PON1v109i}) on PON1 enzyme function, several functional tests were performed. First the \textit{PON1v109i} missense variant was inserted into a \textit{PON1} plasmid containing the \textit{PON1}_{192R} variant, forming the \textit{PON1}_{192R-109I} construct. Both \textit{PON1}_{192R-109I} and the \textit{PON1}_{192R} variants were then expressed and purified via liquid chromatography separately. Post-hoc mass spectrometry analysis of the PON1 proteins used in each experiment structurally confirmed both \textit{PON1}_{192R} and \textit{PON1}_{192R-109I} as the PON1 protein variants tested. While, assays of PON1 enzyme function at baseline showed that the two had approximately normal hydrolysis rates of arylester (AREase activity, data not shown), heat resistance assays for \textit{PON1}_{192R-109I} demonstrated significantly lower (p < 0.05) residual AREase activity at each successive time point when compared to the nonmutated \textit{PON1}_{192R} protein (see Figure 1). These data suggest that, under some circumstances, the function of PON1 is affected by this substitution.
Discussion

In light of the recent evidence that challenges the assertion that HDL-C levels mark the cardioprotective properties of HDL(1, 2), a more thorough understanding of PON1 and, specifically, how deleterious genetic SNVs might alter PON1 enzyme function, may provide new insights as to how HDL and its associated components act in concert to prevent atherosclerotic disease.

Within this context, we have completed the first large-scale study of the effects of rare coding variation in the PON gene cluster on the cardiovascular outcome of non-cardioembolic ischemic stroke. Rare coding variation in PON1, likely to alter function and be deleterious, is associated with ischemic stroke risk (permutation p-value=3.01x10^{-3}). Moreover, the association between this coding variation in PON1 and stroke is independent of the common functional PON1 variants, PON1_q192r and PON1_L55M. These effects of PON1 are more pronounced in participants of AA (permutation p=5.03x10^{-3}) compared to participants of EA, which may be attributed to the PON1_v109I mutation that is found more frequently in AA subjects. Finally, we have demonstrated that the PON1_v109I mutation results in a protein that is functionally compromised.

The finding that PON1 is more significantly associated with ischemic stroke in participants of AA than EA is interesting, although the finding requires replication. Previous investigations into PON1 SNVs and cardiovascular and cerebrovascular disease have largely focused on European and Asian cohorts(3, 7-9). However, relative to EA patients, those of AA have a higher rate of ischemic stroke in the United States(16), receive fewer evidence-based treatments when in hospital, and thus have a longer length-of-stay relative to white patients(17). Given these
considerations, an association of PON1 SNVs with ischemic stroke in patients of AA may have consequences for genetic risk prediction in this high-risk population and could potentially help reduce the high morbidity and mortality of stroke. Moreover, the finding that the PON1 V109I protein is less stable under heat stress testing warrants further functional testing within human cells. Although the PON1 V109I protein has normal baseline PON1 enzyme activity, it is possible that it more rapidly degraded in vivo, thus leading to lower levels of the cardioprotective PON1 protein and an increased risk of ischemic stroke.

Although rare variation could account for a large portion of complex trait inheritance, such as for ischemic stroke, alternative and potentially complementary hypotheses have been proposed. One of these hypotheses is that gene-by-environment interactions among common SNVs comprise a large portion of heritability(18, 19). Given the wide variety of pharmacologic and dietary determinants on PON1 expression and enzyme activity(20), the potential interaction of these environmental factors with PON1 variants could represent another important source of trait heritability.

Some limitations of this study should be considered. First, although the ESP data contained two coding PON1 functional SNVs (PON1 Q192 and PON1 L55M), PON1 -108C/T was not captured by the exome sequencing methods. PON1 -108C/T is a major determinant of PON1 activity, accounting for approximately 14% of PON1 activity variance(7, 20). However, as PON1 -108C/T has not been associated with heart disease in meta-analyses(8, 9) or carotid artery disease in smaller cohorts(3, 6, 7), and neither of the other PON1 functional SNVs affected results, it may not have accounted for increased risk of ischemic stroke in this study. Second, participants of African Ancestry
represented only a small portion of total ischemic stroke cases in this data (77 of 496 total cases). As replication data were not available, we permuted the phenotype 100,000 times and obtained a permutation p-value that remained significant (permutation p=5.03x10^{-3}) and suggestive of a true positive result. Separate replication using exome or whole genome sequence data that captures rare coding variation is needed to verify our result. Third, the cohort was comprised primarily of females for both the ischemic stroke cases and controls; this limits generalizability of our findings. Fourth, our definition of controls for ischemic stroke in this study included subjects with high cardiac risk factors, but no cardiovascular events. However, when we performed a smaller and more restrictive analysis using only subjects collected as “controls” or “deeply phenotyped resources” we found that the association between PON1 and ischemic stroke remained significant. We therefore believe that our definition of controls for ischemic stroke were valid, and may have more accurately represented the broader population. Finally, studies of paraoxonase and cardiovascular disease would optimally include measures of paraoxonase activity. Unfortunately, PON1 activity assays could not be completed for the purposes of this study. Most sites used specimens derived from their stored plasma in tubes containing ethylenediaminetetraacetic acid (EDTA); however, EDTA irreversibly inactivates PON1 by chelation of calcium. This also limited a potential source of functional validation of our findings through testing of participant plasma for PON1 enzyme activity.

In conclusion, we present the first known application of exome sequence data to the PON gene cluster and describe the strong association between rare coding variation in PON1 and non-cardioembolic ischemic stroke in 4,404 participants. We also present evidence that participants of AA have a stronger association between PON1 variation and stroke risk than those of EA, and
that the activity of the PON1\textsubscript{V109I} protein variant found almost exclusively in participants of AA is less stable compared to the common allele. These results strengthen the link between PON1 and cardiovascular disease by demonstrating that rare coding variation, which is likely to change PON1 protein function, is associated with non-cardioembolic ischemic stroke where common variant studies in the past have failed to find an association.

**Disclosures:** None.

**Acknowledgements:** The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute (NHLBI) and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. Additional NHLBI GO ESP authors are listed in the **Supplemental Materials.**

**Sources of Funding:** This work was funded in part by WA State Life Sciences Discovery Fund (265508, to the Northwest Institute of Genetic Medicine). The NHLBI funded the multicenter effort with the following grants: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010). SSM was supported in part by NIH Grants ES09883 and T32ES007032. MAN was supported in part by the Intramural Research Program of the National Institute of Aging, NIH project Z01 AG-000954-06. DSK was supported in part by the Benjamin and Margaret Hall Endowed Fellowship in Genome Sciences, a Markey Foundation award, and National Institutes of Health 5T31HG000035-18 and 1F31MH101905-01. Funding sources for specific cohorts are presented in the **Supplemental Materials.**
References


Figure Legend

Figure 1. PON1<sub>192R-109I</sub> protein variant is retains less PON1 arylesterase activity with heat stress compared to PON1<sub>192R</sub> protein. The x-axis represents time in minutes after the initiation of heat stress. The y-axis represents the % of original PON1 arylesterase (measured at t=0 min) activity retained. % PON1 arylesterase activity retained was significantly different (p < 0.05) for all time points comparing PON1<sub>192R-109I</sub> and PON1<sub>192R</sub> proteins.
Table 1. Description of studied subset of the NHLBI ESP (n=4204 participants).

<table>
<thead>
<tr>
<th></th>
<th>All Participants (n=4204)</th>
<th>Controls (n=3708)</th>
<th>Stroke cases (n=496)</th>
<th>EA Participants (n=2414)</th>
<th>AA Participants (n=1677)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, years</strong></td>
<td>57.48 ± 13.15</td>
<td>56.83 ± 13.16</td>
<td>61.85 ± 11.79</td>
<td>57.65 ± 14.11</td>
<td>56.71 ± 11.49</td>
</tr>
<tr>
<td><strong>Males, (%)</strong></td>
<td>1348 (32.1)</td>
<td>1253 (33.8)</td>
<td>95 (19.2)</td>
<td>922 (38.2)</td>
<td>399 (23.8)</td>
</tr>
<tr>
<td><strong>Current Smokers, (%)</strong></td>
<td>887 (21.1)</td>
<td>782 (21.1)</td>
<td>105 (21.2)</td>
<td>528 (21.9)</td>
<td>342 (20.4)</td>
</tr>
<tr>
<td><strong>EA, (%)</strong></td>
<td>2414 (57.4)</td>
<td>2004 (54.0)</td>
<td>410 (82.7)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>AA, (%)</strong></td>
<td>1677 (39.9)</td>
<td>1600 (43.2)</td>
<td>77 (15.5)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Other Ancestry, (%)</strong></td>
<td>113 (2.7)</td>
<td>104 (2.8)</td>
<td>9 (1.8)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Abbreviations: AA – African Ancestry; EA – European Ancestry; ESP = Exome Sequencing Project; NHLBI = National Heart, Lung, and Blood Institute
Table 2. Results from gene-based exomic variant burden testing for association with stroke.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Total Number of SNVs</th>
<th>EA Polymorphic SNVs</th>
<th>EA SKAT P-Value</th>
<th>AA Polymorphic SNVs</th>
<th>AA SKAT P-Value</th>
<th>Pooled Polymorphic SNVs</th>
<th>Pooled P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PON1</td>
<td>28</td>
<td>19</td>
<td>0.0716</td>
<td>13</td>
<td>5.73x10^-4</td>
<td>27</td>
<td>1.29x10^-3</td>
</tr>
</tbody>
</table>

Abbreviations: AA – African Ancestry; EA – European Ancestry; ESP6500 – release of NHLBI ESP data containing genetic data from 6503 participants; Polymorphic – with a rare variant occurring at least once in the studied subset; PON1 – paraoxonase, SKAT – sequence kernel association testing; SNV – single nucleotide variation.

a Total number of coding SNVs within PON1 in the ESP6500 dataset.
b Only polymorphic SNVs are used in the regression-based SKAT analysis.
c Analyses adjusting for age, sex, current smoking status, and first 3 principal component eigenvectors (ancestry specific eigenvectors for AA/EA analyses to adjust for potential population substructure).
d P-value after 100,000 iteration phenotype permutation for AA participants’ PON1 rare coding variation and stroke is p=0.00503.
e P-value after 100,000 iteration phenotype permutation for all participants’ PON1 rare coding variation and stroke is p=0.00301.
### Table 3. PON1 single nucleotide variants associated with ischemic stroke.

<table>
<thead>
<tr>
<th>PON1 Variant</th>
<th>rsID</th>
<th>Missense Type</th>
<th>Pooled MAF</th>
<th>EA MAF</th>
<th>AA MAF</th>
<th>Beta ± SE</th>
<th>SKAT Meta P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pooled Analyses</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7:94944679</td>
<td>rs61736513</td>
<td>Val109Ile</td>
<td>0.49%</td>
<td>0.02%</td>
<td>1.16%</td>
<td>0.131 ± 0.049</td>
<td>7.88x10&lt;sup&gt;-3&lt;/sup&gt;</td>
</tr>
<tr>
<td>7:94953721</td>
<td>rs146211440</td>
<td>Ser23Ala</td>
<td>0.01%&lt;sup&gt;b&lt;/sup&gt;</td>
<td>b</td>
<td>b</td>
<td>0.762 ± 0.0312</td>
<td>0.0147</td>
</tr>
<tr>
<td>7:94937419</td>
<td>rs80019660</td>
<td>Ala201Val</td>
<td>0.23%</td>
<td>0.35%</td>
<td>0.06%</td>
<td>-0.172 ± 0.0719</td>
<td>0.0165</td>
</tr>
<tr>
<td>7:94947661</td>
<td>rs141665531</td>
<td>Pro40Leu</td>
<td>0.01%&lt;sup&gt;b&lt;/sup&gt;</td>
<td>b</td>
<td>b</td>
<td>0.730 ± 0.313</td>
<td>0.0195</td>
</tr>
<tr>
<td><strong>AA-only Analyses</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7:94944679</td>
<td>rs61736513</td>
<td>Val109Ile</td>
<td>1.16%</td>
<td>0.02%</td>
<td>1.16%</td>
<td>0.115 ± 0.0337</td>
<td>6.52x10&lt;sup&gt;-4&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>EA-only Analyses</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7:94937419</td>
<td>rs80019660</td>
<td>Ala201Val</td>
<td>0.35%</td>
<td>0.35%</td>
<td>0.06%</td>
<td>-0.193 ± 0.0890</td>
<td>0.0299</td>
</tr>
<tr>
<td>7:94953721</td>
<td>rs146211440</td>
<td>Ser23Ala</td>
<td>0.02%&lt;sup&gt;b&lt;/sup&gt;</td>
<td>b</td>
<td>b</td>
<td>0.718 ± 0.366</td>
<td>0.0494</td>
</tr>
</tbody>
</table>

Abbreviations: AA – African Ancestry; EA – European Ancestry; ESP – Exome Sequencing Project; MAF – minor allele frequency; PON – paraoxonase, SKAT – sequence kernel association testing.

<sup>a</sup> Analyses adjusting for age, sex, current smoking status, and first 3 principal component eigenvectors (ancestry specific for AA/EA analyses).

<sup>b</sup> PON1 SNV observed only for EA subjects in the ESP6500 data.
Figure 1.