Published in final edited form as:

Genet Epidemiol. 2014 December ; 38(8): 709–713. doi:10.1002/gepi.21868.

Genetic Markers Associated with Plasma Protein C Level in African Americans: the Atherosclerosis Risk in Communities (ARIC) Study

M. Shahzeb Munir^{1,2}, Lu-Chen Weng², Weihong Tang², Saonli Basu³, James S. Pankow², Nena Matijevic⁴, Mary Cushman⁵, Eric Boerwinkle⁶, and Aaron R. Folsom²

¹Division of Preventive Medicine, Mayo Clinic, Rochester Minnesota, USA

²Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, USA

³Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, USA

⁴Department of Surgery, Medical School, University of Texas Health Science Center at Houston, Houston, Texas, USA

⁵Departments of Medicine and Pathology, University of Vermont, Burlington, Vermont, USA

⁶Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas, USA

Abstract

Protein C is an endogenous anticoagulant protein with anti-inflammatory properties. Single-nucleotide polymorphisms (SNPs) affect the levels of circulating protein C in European Americans. We performed a genome-wide association (GWA) scan of plasma protein C concentration with approximately 2.5 million SNPs in 2,701 African Americans in the Atherosclerosis Risk in Communities Study. Seventy-nine SNPs from the 20q11 and 2q14 regions reached the genome-wide significance threshold of 5×10^{-8} . A missense variant rs867186 in the PROC gene at 20q11 is known to affect protein C levels in individuals of European descent and showed the strongest signal ($P = 9.84 \times 10^{-65}$) in African Americans. The minor allele of this SNP was associated with higher protein C levels ($\beta = 0.49 \mu\text{g/ml}$; 10% variance explained). In the 2q14 region, the top SNPs were near or within the PROC gene: rs7580658 ($\beta = 0.15 \mu\text{g/ml}$; 2% variance explained, $P = 1.7 \times 10^{-12}$) and rs1799808 ($\beta = 0.15 \mu\text{g/ml}$; 2% variance explained, $P = 2.03 \times 10^{-12}$). These two SNPs were in strong linkage disequilibrium (LD) with another SNP rs1158867 that resides in a biochemically functional site and in weak to strong LD with the top PROC variants previously reported in individuals of European descent. In addition, two variants outside the PROC region were significantly and independently associated with protein C levels: rs4321325 in CYP27C1 and rs13419716 in MYO7B. In summary, this first GWA study for plasma protein C levels in African Americans confirms the associations of SNPs in the PROC and

Corresponding author: Weihong Tang, MD, PhD, Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, 1300 South Second Street, WBOB 300, Minneapolis, MN 55454, Fax: (612) 624-0315, Phone: (612) 626-9140, tang0097@umn.edu.

Conflict-of-interest disclosure: The authors declare no competing financial interests or other conflicts of interest.

PROCR regions with circulating levels of protein C across ethnic populations and identifies new candidates for protein C regulation.

INTRODUCTION

Protein C, a vitamin K-dependent plasma glycoprotein synthesized in the liver, is one of the most important endogenous anticoagulants. Upon activation by the thrombin-thrombomodulin complex, it inactivates factor Va and factor VIIIa, consequently reducing thrombin generation. Hereditary protein C deficiency, characterized by reduction of protein C levels or activity, is due to rare genetic mutations and contributes to familial venous thrombosis [Broekmans et al., 1983; Griffin et al., 1981]. In the general population, a low level of circulating protein C as well as common variants in the protein C gene are associated with increased risk of venous thromboembolism [Folsom et al., 2002; Koster et al., 1995; Smith et al., 2007]. Activated protein C has other physiologic effects including anti-inflammatory and antiapoptotic activities and endothelial barrier stabilization [Jackson and Xue 2008]. Plasma levels of protein C in individuals of European descent are influenced by single nucleotide polymorphisms (SNPs) in or near the *PROC* and *PROCR* genes [Aiach et al., 1999; Athanasiadis et al., 2011; Oudot-Mellakh et al., 2012; Pomp et al., 2009; Reiner et al., 2008; Spek et al., 1995; Tang et al., 2010]. However to date we are aware of no studies available assessing genome-wide markers with protein C levels among African Americans, which represent an admixed population of European and African ancestry and may provide additional information in assessing genetic associations due to different genetic background.

We performed a genome-wide association (GWA) scan for plasma protein C concentration with approximately 2.5 million SNPs in a large African American sample from the Atherosclerosis Risk in Communities (ARIC) study.

MATERIALS AND METHODS

Study Population and Phenotype measurement

The ARIC study is a prospective population-based study of risk factors for atherosclerosis and cardiovascular diseases. Recruitment by probability sampling included 15,792 adults (4,266 of them were African Americans, 27%), aged 45 to 64 years at baseline in 1987 through 1989 from 4 U.S. communities (suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi), with the African Americans coming mostly from Jackson and Forsyth County. Four follow-up exams and surveillance for hospitalization and death were conducted to ascertain the development of cardiovascular diseases.

After informed consent, baseline measures of demographic and clinical characteristics, including anthropometry, lifestyle variables, medical history, and medication use, were collected by standardized protocols during a home interview and clinical examination in which fasting blood was drawn. Aliquots of citrated plasma were obtained by centrifugation at 4°C and stored at -70°C for protein C antigen measurement a few weeks after blood

collection using a commercial enzyme-linked immunosorbent assay (ELISA) (Asserachrom Protein C, Diagnostica Stago) at a central laboratory. The laboratory coefficient of variation was 12%; the overall reliability coefficient obtained from repeated blood drawing and independent laboratory testing of a sample of individuals over several weeks was 0.56 [Chambless et al., 1992]. DNA samples were extracted from stored buffy coats.

A total of 3,967 African American participants had protein C measurement available after exclusion of participants who used warfarin at the time of protein C measurement or were from Minneapolis and Washington County centers where only a small number of African American participants were recruited. After taking into account appropriate informed consent, availability of adequate amounts of high quality DNA and genome-wide genotype data, and genotyping quality control and assurance procedures, the final sample consisted of 2,701 African American participants for this genetic analysis.

The institutional review board for each field center approved the ARIC study.

Genotyping and imputation

The cohort was genotyped at the Broad Institute using Affymetrix Genome-Wide Human SNP array 6.0 according to the manufacturer's recommendations. The assay interrogates 906,600 SNPs. Several quality control procedures (QC) were performed on the genotype data [Lettre et al., 2011]. DNA samples with a genome-wide genotyping success rate <95%, duplicate discordance or sex mismatch, SNPs with genotyping success rate <95%, monomorphic SNPs, SNPs with minor allele frequency (MAF) < 1%, and SNPs that map to several genomic locations were removed from the analyses. The Hardy-Weinberg equilibrium goodness-of-fit test was performed for all SNPs, but SNPs were not excluded uniquely on this criterion given the admixed nature of the cohort genotyped. Imputation was performed using MACH 1.0.16 (<http://www.sph.umich.edu/csg/abecasis/MaCH/>). A combined CEU+YRI reference panel from HapMap phase 2 (release 22, build 36) [Huang et al., 2009] was used. Comparison of the GWAS imputation with SNPs on a candidate gene array indicates an allelic concordance rate of ~95.6%, comparable to rates calculated for individuals of African descent imputed with the HapMap 2 YRI individuals. [Huang et al., 2009] Imputation results were filtered using a minimum imputation quality score, indicated by an RSQ_HAT estimate of > 0.3 and a MAF threshold of >1%. This resulted in a total of 2,649,157 SNPs which were tested for association with protein C concentration.

Statistical analysis

Untransformed protein C values were analyzed as the distribution of protein C was approximately normal (skewness = 0.44, kurtosis = 0.65). The genetic association analysis was conducted in a linear regression model with ProbABEL v.0.1-0, which uses "allele dosage" for each SNP as a predictor assuming an additive genetic effect (<http://www.genabel.org/packages/ProbABEL>). The analysis was adjusted for age, gender and field center to reduce nongenetic variation in the distribution of protein C levels. To adjust for the influence of population stratification and global admixture, the first 10 principal components from EIGENSTRAT [Price et al., 2006] were also incorporated as covariates in the regression models. The *a priori* threshold of $P < 5.0 \times 10^{-8}$ was used to judge genome-wide

statistical significance for SNP associations. When more than one significant SNP clustered at a region, we conducted conditional analyses to additionally adjust for the top SNP from that region. In addition, the linkage disequilibrium (LD) value between SNPs from the HapMap YRI population sample, represented by r^2 , was used to evaluate the independence of associations at a region.

RESULTS

Table 1 presents selected baseline characteristics of the study sample by quartile of protein C. Age, gender, BMI, and smoking status were significantly correlated with protein C level. Of the four covariates, only gender was significantly associated with the top significant variants for protein C (rs7580658 and rs1799808, $p=0.03$ for both), and gender has been adjusted for as covariate in the GWAS. Manhattan and quantile-quantile (Q-Q) plots of the p-value distribution from the GWA scan are shown in Supplementary Figures 1 and 2 respectively. The genomic inflation lambda coefficient was 1.026, suggesting negligible test statistic inflation by potential population stratification or other technical factors. A total of 79 SNPs from multiple genes exceeded the genome-wide significance threshold of 5×10^{-8} and marked 2 regions: chromosome 2q14 (spanning 831,685 bp) and 20q11 (spanning 770,625 bp). Details of the top SNPs in the two regions are presented in table 2.

In the 20q11 region, 31 SNPs covering 5 genes exceeded the genome-wide significance threshold of $p < 5 \times 10^{-8}$. The strongest signal was observed for rs867186 ($p=9.84 \times 10^{-65}$), which is a coding-non synonymous missense variant in the endothelial protein C receptor (*PROCR*) gene causing a S219G substitution. This SNP was associated with a 0.48 $\mu\text{g/mL}$ higher plasma protein C level per minor C allele and explained 10.4% of its variance. With the exception of the top SNP, all other significant SNPs in this region lack protein coding function. Supplementary Figure 3 shows the regional association plot for the 20q11 region. In the conditional analysis adjusting for the top SNP, rs867186, no SNPs remained genome-wide significant.

Forty-eight SNPs in the 2q14 region, covering 6 genes, were associated with plasma protein C levels at $p < 5 \times 10^{-8}$. No SNP was exonic. The strongest signal was observed for rs7580658, an intergenic SNP that was 16,735 bp upstream of the 5' terminus of the protein C (*PROC*) gene. Each copy of the minor A allele of rs7580658 was associated with a 0.15 $\mu\text{g/mL}$ higher plasma protein C concentration ($p=1.7 \times 10^{-12}$, 2% variance explained). Considering the top 2 SNPs from the ARIC EA cohort previously published [Tang et al., 2010], associations in African Americans did not reach genome-wide significance but had suggestive associations (rs1158867, $p=2.75 \times 10^{-5}$ and rs1799810, $p=0.0063$; $r^2=0.48$ between rs1158867 & rs1799810).

The second top SNP associated with protein C was rs1799808, 5' to the *PROC* gene. Rs1799808 was in LD with the top variant rs7580658 ($r^2=0.78$). The minor T allele of rs1799808 was associated with 0.15 $\mu\text{g/mL}$ higher protein C level ($p=2.03 \times 10^{-12}$, 2% variance explained). Supplementary Figure 4 shows the regional association plot for the 2q14 region. Conditional analysis adjusting for rs7580658 or rs1799808 reduced the number of genome-wide significant SNPs to two: rs4321325 (intronic to *CYP27C1*, $\beta=0.20$,

$p=3.60\times 10^{-9}$ or 4.81×10^{-9} , imputation quality score=0.94) and rs13419716 (intronic to the *MYO7B*, $\beta=0.26$, $p=4.14\times 10^{-9}$ or 3.26×10^{-9} , imputation quality score=0.86). These two SNPs were also among top 10 in the initial analysis. There was little LD between the two SNPs that emerged in conditional analysis and the top two SNPs at 2q14 from the main analysis (rs7580658 and rs1799808): $r^2=0.006$ to 0.014 . The imputation quality score of the top two SNPs, rs7580658 and rs1799808 was 0.94 and 0.94, respectively.

We also looked up associations for the other two significant regions that were previously reported in the European American GWAS in ARIC [Tang et al., 2010]: *GCKR* at 2p23 and *BAZ1B* at 7q11-23. The top two SNPs in the *BAZ1B* locus in European Americans, rs17145713 and rs1178977, were not significant in African Americans ($p=0.33$, $\beta=0.02$ $\mu\text{g/ml}$ per minor allele for both SNPs). In *GCKR*, the top SNP in European Americans reached borderline significance among African Americans (rs1260326, $\beta=0.05$ $\mu\text{g/ml}$ per minor allele, $p=0.06$).

DISCUSSION

To the best of our knowledge, this is the first report of a GWA scan for plasma protein C levels in African Americans. The top SNPs from the 79 significant markers were found in the *PROC* and the *PROCR* genes or nearby regions that are in LD with markers in these genes. Our GWAS in African Americans not only replicated the *PROC* and the *PROCR* loci reported from the previously published GWAS or candidate gene studies in individuals of European descent [Aiach et al., 1999; Athanasiadis et al., 2011; Oudot-Mellakh et al., 2012; Pomp et al., 2009; Reiner et al., 2008; Spek et al., 1995; Tang et al., 2010], but also identified new associations, including those in *CYP27C1* and *MYO7B*, yielding new candidates for the study of protein C regulation.

Three GWAS studies, all in populations of European descent, have previously evaluated protein C levels [Athanasiadis et al., 2011; Oudot-Mellakh et al., 2012; Tang et al., 2010]. All of the three confirmed the already known association with rs867186 in *PROCR* [Reiner et al., 2008]. In addition, the largest one, based on 8,048 individuals of the ARIC, reported additional signals at *PROC* (top SNP rs1158867), *EDEM2* (rs6120849), and *GCKR* (rs1260326) [Tang et al., 2010]. The locus *PROC* has also been previously reported in candidate gene studies [Aiach et al., 1999; Pomp et al., 2009; Reiner et al., 2008; Spek et al., 1995], while *GCKR*, *EDEM2*, and *BAZ1B* were newly identified [Tang et al., 2010]. Furthermore, the independent association with *EDEM2* was supported in the second GWAS after adjusting for *PROCR* rs867186 [Oudot-Mellakh et al., 2012].

Rs867186 is a missense mutation in the *PROCR* gene that encodes a receptor for activated protein C, endothelial protein C receptor (EPCR), which is an N-glycosylated type I membrane protein. The EPCR enhances activation of protein C [Ireland et al., 2005]. The missense mutation in rs867186 leads to a serine to glycine substitution (Ser219Gly) [Ireland et al., 2005; Reiner et al., 2008], which increases solubility relative to membrane bound EPCR levels and explains 75% of the variability of plasma soluble EPCR levels [Saposnik et al., 2004]. Previous studies showed a higher plasma protein C level in association with the minor allele of rs867186, which explained 10% to 20% of variance in protein C levels

[Athanasiadis et al., 2011; Oudot-Mellakh et al., 2012; Reiner et al., 2008; Tang et al., 2010]. Consistent with these reports, our findings showed, for the first time in African Americans, a higher plasma protein C level in association with the rs867186 minor allele and a comparable effect size ($\beta=0.49$ vs. $\beta=0.47$) and amount of variance explained. The mechanism underlying the association between the *EPCR* Ser219Gly variant and higher soluble EPCR and plasma protein C is unclear. It may relate to increased shedding of membrane bound EPCR into soluble EPCR, associated with increased solubility of EPCR [Qu et al., 2006]. This might in turn effectively increase the amount of circulating protein C by stabilizing the protein. Conversely, increased shedding of the EPCR from the endothelial surface may result in less cell-bound EPCR to bind protein C, leading to higher levels of protein C in circulation [Reiner et al., 2008].

EDEM2, the second independent locus in the 20q11 region reported in ARIC European Americans, did not reach genome-wide significance in our African American population after adjusting for rs867186. The top *EDEM2* SNP in European Americans, rs6120849 [Tang et al., 2010], showed a borderline association in our African American sample after adjusting for rs867186: $\beta=-0.04$, $p=0.056$. Furthermore, another *EDEM2* SNP, rs3746429, that results in amino acid change (T456A substitution) and was also one of the top independent *EDEM2* SNPs in European Americans [Tang et al., 2010], showed a stronger association in African Americans: $\beta=-0.13$, $p=0.01$. It is the strongest one among four *EDEM2* SNPs that showed association at $p<0.05$ in the conditional analysis. These data in African Americans support the existence of the second independent locus at the *PROCR/EDEM2* region and point to the *EDEM2* functional variant rs3746429 as a possible, underlying contributor to the second independent association. In addition, we also replicated the signal at *GCKR* at borderline, nominal significance ($p=0.06$ for rs1260326). The effect size (β) for rs1260326 was similar between European American and African American cohorts while its minor allele was less common in African Americans (0.1 vs 0.4). Therefore, the lack of significant replication for the association at *GCKR* in African Americans is likely due to reduced power because of lower allele frequency in African Americans.

The top 2 SNPs at 2q14, rs7580658 and rs1799808, are in high LD ($r^2=0.78$) and showed similar strength of association with protein C. While they were not the top *PROC* variants previously reported in the GWAS of ARIC European Americans (rs1158867 and rs1799810), they were in LD with these variants ($r^2=0.65 \sim 0.22$) and were significant in ARIC European Americans (rs7580658: $p=2.10 \times 10^{-14}$; rs1799808: $p=6.03 \times 10^{-17}$). Moreover, rs1799808 was also previously reported to be associated with plasma protein C levels in a candidate gene study of an European population [Pomp et al., 2009]. These data suggest that the associations observed at the *PROC* region across the different ethnic populations were attributable to the same underlying locus.

However, Protein C level was found to be significantly lower in African Americans than in European Americans [Folsom et al., 1992]. In addition to environmental factors, genetics may contribute to the ethnicity-related difference. We have compared the allele frequency of the top significant signals between African Americans and European Americans [Tang et al., 2010] in ARIC. While the effect size and allele frequency for rs867186 in *PROCR* were

similar between these two ethnic groups, we noticed that at the *PROC* region, the frequency of protein C-increasing allele (A1 in European Americans and A2 in African Americans) was two times higher in European Americans than in African Americans. This signal, along with other undiscovered genetic as well as environmental factors, might contribute to the difference in protein C level between the two ethnic groups.

In the *PROC* region, novel SNPs (rs4321325 in *CYP27C1* and rs13419716 in *MYO7B*) remained significantly associated with protein C after adjustment for the top SNPs (rs7580658 and rs1799808). It is unlikely that the novel associations were due to residual signals of the top SNPs because they were imputed with almost perfect quality. Interestingly, these two novel SNPs did not emerge as independent and significant signals in the GWAS of ARIC European Americans [Tang et al., 2010]. It is noted that rs13419716 was monomorphic in HapMap CEU but polymorphic in HapMap YRI (MAF=7%). Also, LD between the secondary variants and the top *PROC* SNPs was lower in HapMap YRI ($r^2=0.006\sim0.014$) than in CEU populations ($r^2=0.08\sim0.09$). It is possible that the lower LD at this region in subjects of African descent helped segregate the secondary signals from the top ones. It is unclear what role the novel SNPs play in modulating protein C level, as no report has previously been published about the variants or genes they reside in. Confirmation and interpretation of the novel associations awaits further replication from other populations and research that focus on the biology and physiological regulation of protein C.

Close to 90% of SNPs reported to have disease associations are intronic (45%) or intergenic (43%) [Hindorff et al., 2009]. However, the ENCODE project has assigned function to 80% of the genome that may affect protein production through gene regulating mechanisms. These biochemically functional genomic sites include transcription factor binding sites and chromatin access sites, i.e. DNAase I hypersensitive sites [Dunham et al., 2012]. We reviewed all SNP/plasma protein C associations reported to date to find whether these SNPs reside within biochemically functional genomic sites. We found 2 SNPs within these regions that showed suggestive signals in the African American GWAS (rs1158867 and rs6120849; $p=2.75\times10^{-5}$ and $p=8.64\times10^{-5}$, respectively) [Dunham et al., 2012]. Rs1158867 is located within the *PROC* gene and in strong LD with the top *PROC* SNPs rs7580658 and rs1799808 ($r^2=0.62\sim0.65$), whereas rs6120849 resides within the *EDEM2* gene and was significantly and independently associated with protein C level in ARIC European Americans [Tang et al., 2010]. These data suggest that the associations observed at the 2q14 region might be attributable to mechanisms related to the regulation of *PROC* expression.

In summary, we conclude that SNPs in the 2q14 and 20q11 regions, previously reported to be associated with plasma protein C levels among individuals of European descent, were significant in African Americans. We also identified novel signals at *CYP27C1* and *MYO7B* in this African American population. Data from this study improves understanding of the role that genetic loci in 2q14 and 20q11 regions play in regulating circulating levels of protein C across ethnic populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the University of Minnesota Supercomputing Institute for use of the blade supercomputers. The authors thank the staff and participants of the ARIC study for their important contributions.

Funding

The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C) and grants R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Part of the work was supported by grant R01-HL095603.

References

- Aiach M, Nicaud V, Alhenc-Gelas M, Gandrille S, Arnaud E, Amiral J, Guize L, Fiessinger JN, Emmerich J. Complex association of protein C gene promoter polymorphism with circulating protein C levels and thrombotic risk. *Arterioscler Thromb Vasc Biol.* 1999; 19:1573–1576. [PubMed: 10364092]
- Athanasiadis G, Buil A, Souto JC, Borrell M, Lopez S, Martinez-Perez A, Lathrop M, Fontcuberta J, Almasy L, Soria JM. A genome-wide association study of the Protein C anticoagulant pathway. *PLoS One.* 2011; 6:e29168. [PubMed: 22216198]
- Broekmans AW, Veltkamp JJ, Bertina RM. Congenital protein C deficiency and venous thromboembolism. A study of three Dutch families. *N Engl J Med.* 1983; 309:340–344. [PubMed: 6688122]
- Chambless LE, McMahon R, Wu K, Folsom A, Finch A, Shen YL. Short-term intraindividual variability in hemostasis factors. The ARIC Study. *Atherosclerosis Risk in Communities Intraindividual Variability Study.* *Ann Epidemiol.* 1992; 2:723–733. [PubMed: 1342324]
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Fietze S, Harrow J, Kaul R, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489:57–74. [PubMed: 22955616]
- Folsom AR, Aleksic N, Wang L, Cushman M, Wu KK, White RH. Protein C, antithrombin, and venous thromboembolism incidence: a prospective population-based study. *Arterioscler Thromb Vasc Biol.* 2002; 22:1018–1022. [PubMed: 12067914]
- Folsom AR, Wu KK, Conlan MG, Finch A, Davis CE, Marcucci G, Sorlie PD, Szklo M. Distributions of hemostatic variables in blacks and whites: population reference values from the Atherosclerosis Risk in Communities (ARIC) Study. *Ethn Dis.* 1992; 2:35–46. [PubMed: 1458214]
- Griffin JH, Evatt B, Zimmerman TS, Kleiss AJ, Wideman C. Deficiency of protein C in congenital thrombotic disease. *J Clin Invest.* 1981; 68:1370–1373. [PubMed: 6895379]
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362–9367. [PubMed: 19474294]
- Huang L, Li Y, Singleton AB, Hardy JA, Abecasis G, Rosenberg NA, Scheet P. Genotype-imputation accuracy across worldwide human populations. *Am J Hum Genet.* 2009; 84:235–250. [PubMed: 19215730]
- Ireland H, Konstantoulas CJ, Cooper JA, Hawe E, Humphries SE, Mather H, Goodall AH, Hogwood J, Juhan-Vague I, Yudkin JS, et al. EPCR Ser219Gly: elevated sEPCR, prothrombin F1+2, risk for coronary heart disease, and increased sEPCR shedding in vitro. *Atherosclerosis.* 2005; 183:283–292. [PubMed: 15921688]

- Jackson CJ, Xue M. Activated protein C--an anticoagulant that does more than stop clots. *Int J Biochem Cell Biol.* 2008; 40:2692–2697. [PubMed: 18249579]
- Koster T, Rosendaal FR, Briet E, van der Meer FJ, Colly LP, Trienekens PH, Poort SR, Reitsma PH, Vandenbroucke JP. Protein C deficiency in a controlled series of unselected outpatients: an infrequent but clear risk factor for venous thrombosis (Leiden Thrombophilia Study). *Blood.* 1995; 85:2756–2761. [PubMed: 7742536]
- Lette G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, Bennett F, Bowden DW, Chakravarti A, Dreisbach A, et al. Genome-wide association study of coronary heart disease and its risk factors in 8, 090 African Americans: the NHLBI CARE Project. *PLoS Genet.* 2011; 7:e1001300. [PubMed: 21347282]
- Oudot-Mellakh T, Cohen W, Germain M, Saut N, Kallel C, Zelenika D, Lathrop M, Tregouet DA, Morange PE. Genome wide association study for plasma levels of natural anticoagulant inhibitors and protein C anticoagulant pathway: the MARTHA project. *Br J Haematol.* 2012; 157:230–239. [PubMed: 22443383]
- Pomp ER, Doggen CJ, Vos HL, Reitsma PH, Rosendaal FR. Polymorphisms in the protein C gene as risk factor for venous thrombosis. *Thromb Haemost.* 2009; 101:62–67. [PubMed: 19132190]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909. [PubMed: 16862161]
- Qu D, Wang Y, Song Y, Esmon NL, Esmon CT. The Ser219-->Gly dimorphism of the endothelial protein C receptor contributes to the higher soluble protein levels observed in individuals with the A3 haplotype. *J Thromb Haemost.* 2006; 4:229–235. [PubMed: 16409473]
- Reiner AP, Carty CL, Jenny NS, Nievergelt C, Cushman M, Stearns-Kurosawa DJ, Kurosawa S, Kuller LH, Lange LA. PROC, PROCR and PROS1 polymorphisms, plasma anticoagulant phenotypes, and risk of cardiovascular disease and mortality in older adults: the Cardiovascular Health Study. *J Thromb Haemost.* 2008; 6:1625–1632. [PubMed: 18680534]
- Saposnik B, Reny JL, Gaussem P, Emmerich J, Aiach M, Gandrille S. A haplotype of the EPCR gene is associated with increased plasma levels of sEPCR and is a candidate risk factor for thrombosis. *Blood.* 2004; 103:1311–1318. [PubMed: 14576048]
- Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marcianti KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR, et al. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *Jama.* 2007; 297:489–498. [PubMed: 17284699]
- Spek CA, Koster T, Rosendaal FR, Bertina RM, Reitsma PH. Genotypic variation in the promoter region of the protein C gene is associated with plasma protein C levels and thrombotic risk. *Arterioscler Thromb Vasc Biol.* 1995; 15:214–218. [PubMed: 7749828]
- Tang W, Basu S, Kong X, Pankow JS, Aleksic N, Tan A, Cushman M, Boerwinkle E, Folsom AR. Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood.* 2010; 116:5032–5036. [PubMed: 20802025]

Characteristics of participants included in the GWAS for protein C in ARIC African Americans by quartile of protein C: mean $\pm$ standard deviation or percentage unless otherwise stated

	Q1 (N=748)	Q2 (N=541)	Q3 (N=769)	Q4 (N=643)	p-value
Protein C range, ug/ml	1.20 – 2.70	2.80 – 3.00	3.10 – 3.50	3.60 – 5.80	
Age, years	52.9 $\pm$ 5.9	53.1 $\pm$ 6.0	53.4 $\pm$ 5.8	53.8 $\pm$ 5.6	0.017
Body mass index, kg/m ²	28.6 $\pm$ 5.8	29.9 $\pm$ 6.1	30.1 $\pm$ 6.4	29.9 $\pm$ 5.4	<0.0001
Male (%)	47.6	36.2	34.7	29.2	<0.0001
Smoking					<0.0001
Current (%)	37.6	30.1	27.3	24.1	
Former (%)	22.1	25.8	23.6	25.1	
Never (%)	40.3	44.1	49.0	50.8	
Protein C, ug/ml	2.41 $\pm$ 0.28	2.91 $\pm$ 0.08	3.29 $\pm$ 0.14	4.01 $\pm$ 0.41	
Median protein C (IQR), ug/ml	2.50 (0.30)	2.90 (0.20)	3.30 (0.20)	3.90 (0.50)	

IQR=interquartile range.

Table 2

Top SNP associations for protein C at genome-wide significance ($P < 5 \times 10^{-8}$) in ARIC African Americans.

SNP	Position	Region	Gene	Function	A1/A2	AFA2	β /SE	p-value	Imput	Var%
rs7580658	128159261	2q14.3	Intergenic	-	G/A	0.23	0.15/0.02	1.70×10^{-12}	0.94	2.0
rs1799808	128175862	2q14.3	PROC	5KB upstream of gene	C/T	0.21	0.15/0.02	2.03×10^{-12}	0.95	2.0
rs867186	33764554	20q11.22	PROC	missense	T/C	0.09	0.49/0.03	9.84×10^{-65}	0.99	10.4

A1 indicates allele 1 (major allele); A2, allele 2 (minor allele); AFA2, allele frequency for A2; β , change in protein C level per 1 minor allele A2; SE, standard error; Imput, ratio of observed to expected variance as a measure of imputation quality; Var%, percentage variance explained by the SNP.