

https://dspace.org/

Candidate Gene Analysis Reveals Strong Association of CETP Variants With High Density Lipoprotein Cholesterol and PCSK9 Variants With Low Density Lipoprotein Cholesterol in Ghanaian Adults An AWIGen SubStudy

Godfred Agongo; L Amenga-Etego; E Nonterah; C Debpuur; Ananyo Choudhury ...

https://hdl.handle.net/10539/36507
WOS

Downloaded from a DSpace repository. https://dspace.org/

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fgene.2020.456661

Candidate Gene Analysis Reveals Strong Association of *CETP* Variants With High Density Lipoprotein Cholesterol and *PCSK9* Variants With Low Density Lipoprotein Cholesterol in Ghanaian Adults: An AWI-Gen Sub-Study

Godfred Agongo[1,2,3], Lucas Amenga-Etego[1,4], Engelbert A. Nonterah[1,5], Cornelius Debpuur[1], Ananyo Choudhury[2], Amy R. Bentley[6], Abraham R. Oduro[1], Charles N. Rotimi[6], Nigel J. Crowther[7], Michèle Ramsay[2,3], as members of AWI-Gen and H3Africa*

[1] Navrongo Health Research Centre, Navrongo, Ghana, [2] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [3] Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [4] West African Centre for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Legon, Ghana, [5] Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands, [6] Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States, [7] Department of Chemical Pathology, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

OPEN ACCESS

Edited by:
Solomon Fiifi Ofori-Acquah, University of Ghana, Ghana

Reviewed by:
Cesar Fortes-Lima, Uppsala University, Sweden
Francesco Montinaro, University of Tartu, Estonia

***Correspondence:**
Godfred Agongo
Godfred.Agongo@navrongo-hrc.org

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *26 February 2019*
Accepted: *30 September 2020*
Published: *30 October 2020*

Citation:
Agongo G, Amenga-Etego L, Nonterah EA, Debpuur C, Choudhury A, Bentley AR, Oduro AR, Rotimi CN, Crowther NJ and Ramsay M (2020) Candidate Gene Analysis Reveals Strong Association of CETP Variants With High Density Lipoprotein Cholesterol and PCSK9 Variants With Low Density Lipoprotein Cholesterol in Ghanaian Adults: An AWI-Gen Sub-Study. Front. Genet. 11:456661. doi: 10.3389/fgene.2020.456661

Variations in lipid levels are attributed partly to genetic factors. Genome-wide association studies (GWASs) mainly performed in European, African American and Asian cohorts have identified variants associated with LDL-C, HDL-C, total cholesterol (TC) and triglycerides (TG), but few studies have been performed in sub-Saharan Africans. This study evaluated the effect of single nucleotide variants (SNVs) in eight candidate loci (*ABCA1*, *LCAT*, *LPL*, *PON1*, *CETP*, *PCSK9*, *MVK*, and *MMAB*) on lipid levels among 1855 Ghanaian adults. All lipid levels were measured directly using an automated analyser. DNA was extracted and genotyped using the H3Africa SNV array. Linear regression models were used to test the association between SNVs and log-transformed lipid levels, adjusting for sex, age and waist circumference. In addition Bonferroni correction was performed to account for multiple testing. Several variants of *CETP*, *LCAT*, *PCSK9*, and *PON1* (MAF $>$ 0.05) were associated with HDL-C, LDL-C and TC levels at $p < 0.05$. The lead variants for association with HDL-C were rs17231520 in *CETP* ($\beta = 0.139$, $p < 0.0001$) and rs1109166 in *LCAT* ($\beta = -0.044$, $p = 0.028$). Lower LDL-C levels were associated with an intronic variant in *PCSK9* (rs11806638 [$\beta = -0.055$, $p = 0.027$]) and increased TC was associated with a variant in *PON1* (rs854558 [$\beta = 0.040$, $p = 0.020$]). *In silico* functional analyses indicated that these variants likely influence gene function through their effect on gene transcription.

We replicated a strong association between *CETP* variants and HDL-C and between *PCSK9* variant and LDL-C in West Africans, with two potentially functional variants and identified three novel variants in linkage disequilibrium in *PON1* which were associated with increasing TC levels in Ghanaians.

Keywords: candidate gene, lipid, single nucleotide variant, Ghanaians, AWI-Gen

INTRODUCTION

Cardiovascular disease (CVD) is a major health risk accounting for over 17 million deaths (about 30% of all deaths) globally each year. A major proportion (80%) of these deaths occurs in low and middle-income countries with the total number of annual deaths expected to reach 23.6 million by 2030 (Wang et al., 2016). One major risk factor for CVD is dyslipidemia which is a metabolic derangement that predisposes an individual to atherosclerosis (O'Donnell and Elosua, 2008; Reinikainen et al., 2015). There is variation in the prevalence of dyslipidemia across populations with the global adult prevalence of raised total cholesterol (TC) in 2008 estimated at 9.7% (WHO, 2011). The distribution of serum lipids is known to differ among individuals with African ancestry compared to better-studied world populations. These differences persist in populations of substantial African ancestry living in different environmental backgrounds, supporting the expectation that these differences in distribution are genetically determined (Willer et al., 2013; Bentley and Rotimi, 2017). It has been documented that approximately 25 to 80% of the inter-individual variation in lipid phenotypes is heritable (Perusse et al., 1997; Beekman et al., 2002). Single nucleotide variants (SNVs) in the following genes have been shown to be associated with differences in lipid levels: lipoprotein lipase (*LPL*) (Pirim et al., 2014), cholesteryl ester transport protein (*CETP*) (Ridker et al., 2009), paraoxonase 1 (*PON1*), (Durrington et al., 2001), ATP-binding cassette A1 (*ABCA1*) (Kelishadi et al., 2014), lecithin-cholesterol acyltransferase (*LCAT*) (Hovingh et al., 2005), protein convertase subtilisin/kexin type 9 (*PCSK9*) (Kotowski et al., 2006), methylmalonic aciduria cb1B (*MMAB*) (Sun et al., 2016) and mevalonate kinase (*MVK*) (Miao et al., 2017). Evidence from candidate gene and genome-wide association studies (GWASs) on the influence of genetic polymorphisms on lipid level variations comes primarily from studies in cohorts of European origin, and an increasing number in Asians and populations of African descent, but few in sub-Saharan Africans (Willer et al., 2013; Dron and Hegele, 2016). Identifying the common genetic variants associated with lipid levels in populations in Africa will assist in the detection of individuals at higher risk for dyslipidemia.

The ability to identify gene loci that associate with serum lipid (high density lipoprotein cholesterol [HDL-C], low density lipoprotein cholesterol [LDL-C], triglycerides [TG] and total cholesterol [TC]) levels has been advanced by the use of genotyping arrays that were developed from several projects including the HapMap project (The International HapMap Consortium et al., 2007; Wu et al., 2013), the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2010; Wood et al., 2013) and the Genome of the Netherlands Project (Boomsma et al., 2014; van Leeuwen et al., 2015). These arrays are generally Eurocentric and under-represent common variants found in African populations. Our study is novel in that we used data from the H3Africa SNV array (Mulder et al., 2018) enriched with common SNVs in African populations and improved imputation using an African reference panel, to investigate the influence of SNVs at selected loci on lipid levels in our study population from Ghana.

The aim of our study was to examine the genetic association of SNVs and identify novel variants in the transcribed regions of eight genes previously associated with LDL-C, HDL-C, TC, and TG and to perform a replication study for the associated SNVs in another African cohort. This is the first African study to use an Afrocentric gene array to perform and replicate a candidate gene analysis of genetic associations with serum lipid levels.

MATERIALS AND METHODS

Study Design and Population

This is a candidate gene study that was conducted in the Kassena-Nankana districts of northern Ghana as part of the Africa Wits-INDEPTH Partnership for Genomic research (AWI-Gen) project (Ramsay et al., 2016) under the broader Human Heredity and Health in Africa initiative (The H3Africa Consortium, 2014). The study population consisted of men and women aged 40–60 years who were selected using stratified random sampling from the two districts of northern Ghana. The east, west, north and south zones were selected and 25 clusters were randomly selected from each zone using data from the Navrongo Health and Demographic Surveillance System (NHDSS) (Oduro et al., 2012). A list of 2200 individuals (including 10% for non-response or refusal) was generated from the sampled clusters, with the sample size in each cluster being proportional to its population distribution in the age group 40 to 60 years. Individuals who agreed to participate in the study and who provided informed consent were assigned unique AWI-Gen identification numbers to ensure anonymity (Agongo et al., 2018).

Given this sample size, a power analysis was performed using Quanto software version 1.2.4 (Gauderman, 2002). Using information from a previous study that reported a mean HDL-C level in West Africans of 1.016 $\pm$ 0.321 mmol/l with an effect size of 0.388 mmol/l for the rs328 variant in the *LPL* gene at ∼5% minor allele frequency (MAF) (Bentley et al., 2014), we determined that our study had $>$ 90% power to detect an effect size of at least 0.129 mmol/l for lipid traits at MAF $>$ 0.05 for a sample size of 1800 individuals. This study was carried out in accordance with the recommendations of the National Institutes of Health

guidelines, and approvals from the Human Research Ethics Committee (HREC) of the University of the Witwatersrand (ID No. M12109 and M170880), the Ghana Health Service Ethics Review Committee (ID No. GHS-ERC:05/05/2015) and the Navrongo Health Research Centre Institutional Review Board (ID No. NHRCIRB178) with community engagement and informed consent. All participants gave written informed consent in accordance with the approved documents.

Demographic Data, Anthropometric Measurements and Lipid Analyses

Since many of the participants did not know their exact birthdates, their ages were mainly estimated using information from the Navrongo Health and Demographic Surveillance System (Oduro et al., 2012). Fasting whole blood samples were taken from participants for lipid measurements and DNA extraction for genetic analyses. Serum HDL-C, LDL-C, TC and TG were all measured directly using an automated chemistry analyser and waist circumference was measured as described in detail elsewhere (Ali et al., 2018).

Selection of Candidate Genes

Genes were selected based on evidence from genome-wide association and candidate gene studies on lipid levels in peoples of African descent (Adeyemo et al., 2012; Elbers et al., 2012; Bentley et al., 2014) and sub-Saharan Africans (Abessolo et al., 2014; Niemsiri et al., 2015). We analysed polymorphic markers in 8 candidate genes: *ABCA1*, *LCAT*, *LPL*, *PON1*, *CETP*, *PCSK9*, *MVK*, and *MMAB* by selecting SNVs in the transcribed region of each gene (genome coordinates shown in **Table 1**).

DNA Extraction and Genotyping

The DNA was extracted using a modified protocol of the salting out method (Diego Chacon-Cortes, 2014). Briefly, this involved thawing the frozen samples before lysing the red blood cells with cold sucrose-Triton X lysing buffer. Following this was the centrifugation and washing of the pellets with the buffer. The white blood cells were lysed and proteins degraded by adding 20 mM TRIS 5 mM ethylenediaminetetraacetic acid (EDTA), 500 μl of proteinase-K and 200 μl of 10% sodium dodecyl sulphate. The lysate was precipitated by adding NaCl and centrifuging the solution. Absolute ethanol was added and agitated and the DNA spooled into a cryo vial. The DNA was washed in 70% ethanol to remove excess salt after which the DNA was suspended in 10 mM Tris, 1 mM EDTA buffer. Prior to genotyping the DNA samples were normalised by assessing the DNA concentration using a NanoDrop DN-100 spectrophotometer (Thermo Fisher Scientific, MA, United States). The ratio of the absorbance at 260/280 nm indicated the estimated DNA purity with an acceptable range between 1.8 and 2.0. A ratio below 1.5 was an indication of protein contamination. Genotyping was performed using the H3Africa single nucleotide variant (SNV) array designed on the Illumina platform. The array is enriched for common variation in African populations and contains >2.3 million SNVs (Mulder et al., 2018).

Imputation and Quality Control Processes

The following pre-imputation quality control (QC) filters were applied to the genotyping data of the entire AWI-Gen dataset of 10,903 individuals from six different sites of which our study site is one. Individuals with a missing call rate >0.05 and SNVs with a missing call rate >0.05, MAF < 0.01 and Hardy-Weinberg equilibrium (HWE) p-value < 0.0001 were removed from the data set. SNVs that did not match the GRCh37 reference alleles or strands were also removed. The filtered dataset (with 1,729,661 SNVs and 10,903 individuals) was, pre-phased with EAGLE2 (Loh et al., 2016) and imputed using the Sanger Imputation Server with the African reference panel. The default Positional Burrows-Wheeler transform (PBWT) algorithm was used for imputation. At post imputation QC, poorly imputed SNVs (SNVs with IMPUTE2 information score < 0.6), SNVs with MAF < 0.01 and HWE p-value < 0.00001 were excluded resulting in the final quality controlled imputed dataset containing 10,903 individuals and 13,980,000 SNVs. The candidate gene data for this study was extracted from the larger dataset. The final data set included 1855 participants from Ghana. Only SNVs in the candidate genes with MAF > 0.05 in the regions described in **Table 1** were used for data analysis in this study. To assess population structure, principal component analysis was done using the genotype data from the array.

Data Analysis

Lipid levels were presented as medians with interquartile ranges and compared between men and women using Mann Whitney U

TABLE 1 | Number of variants tested within selected genes following quality control of data.

Gene symbol	Chromosome position	Start transcript	End transcript	Number of SNVs
LPL	8	19,759,228	19,824,769	286
LCAT	16	67,973,653	67,978,034	6
CEPT	16	111,682,249	111,727,724	117
PON1	7	94,926,988	94,954,019	94
ABCA1	9	107,543,283	107,690,518	632
MVK	12	110,011,060	110,035,922	55
MMAB	12	109,991,542	110,011,679	84
PCSK9	1	55,505,221	55,530,525	118

test. The lipid levels were log-transformed and presented across genotypes of lead SNVs using analysis of covariance (ANCOVA) with adjustment for age, sex and waist circumference. PLINK version 1.90 was used for association analyses[1] (Purcell et al., 2007; Chang et al., 2015). Data on SNVs in the transcribed regions of selected genes were extracted from the H3Africa AWI-Gen genotyped data for further analyses. Pearson χ^2 test was used to assess deviation from HWE equilibrium by comparing observed to expected frequencies. Linear regression analyses were used to test for association between log-transformed lipid traits (Agongo et al., 2018) and SNVs. Standardised β values and confidence intervals were calculated using the major allele as reference. All *p*-values were corrected for multiple testing using the Bonferroni method (Bland and Altman, 1995) and adjusted for covariates (age, sex and waist circumference). LocusZoom plots were drawn using an online analysis tool at http://csg.sph.umich.edu/locuszoom by the University of Michigan, Department of Biostatistics, Centre for Statistical Genetics (Pruim et al., 2010). All Bonferroni-adjusted *p*-values at 5% significance level after covariate adjustment were considered significant.

Functional Analysis of Significant SNVs

Functional variant analyses included localisation of the variant within the gene region and combined annotation dependent depletion (CADD) scores to predict the damaging effect of the variant on protein structure and function. Another annotation was loss of function tool (LoFtool) score which predicted genic intolerance and consequent susceptibility to disease. A CADD score above 10 implied a deleterious effect and a low LoFtool score indicated damaging effect of the mutations on the gene. The RegulomeDB (RDB) score (Boyle et al., 2012) was used to assess the regulatory potential of the variants. All of these were analysed using Variant Effect Predictor (McLaren et al., 2016). The 1000 Genomes database (accessed using[2]) was used to compare the MAFs of the SNVs analysed in the current study with those observed in other populations.

Linkage Disequilibrium Among Significantly Associated Variants

Linkage disequilibrium (LD) among significantly associated SNVs was assessed using LDlink, a web-based tool (Machiela and Chanock, 2015).

[1]http://pngu.mgh.harvard.edu/purcell/plink/

[2]https://www.ncbi.nlm.nih.gov/dbvar

Replication

All variants with $p < 0.05$ in either the covariate-adjusted or unadjusted models ($n = 21$) were selected for replication using the Africa America Diabetes Mellitus study (AADM), which has been previously described (Rotimi et al., 2001). Briefly, AADM is a genetic epidemiology study of type 2 diabetes (T2D), enrolling participants from university medical centres in Nigeria (Enugu, Lagos, and Ibadan), Ghana (Accra and Kumasi), and Kenya (Eldoret). Analyses were conducted using linear mixed models of the log transformations of serum lipids in up to 4317 participants with available data. The first 2 PCs of the genotypes were included in the model. All models also included a genetic relationship matrix to account for the random effect of relatedness, as related individuals were included in AADM, and were adjusted for T2D, as this is a case-control study. Models were run using EPACTS (Efficient and Parallelizable Association Container Toolbox)[3] with and without adjustment for additional covariates including age, sex, waist circumference and Bonferroni-corrected *p* values. Replication was defined as an association in a consistent direction with Bonferroni-corrected $p < 0.0024$ (i.e., 0.05/21).

RESULTS

Characteristics of the Study Population

The demographic characteristics and lipid levels of the study population, stratified by sex, are shown in **Table 2**. Participants who had no genetic data were excluded from the analyses resulting in a sample size of 1855. The average age of the study participants was 51 years with women being significantly older than men ($p = 0.0001$). Waist circumference among women was significantly higher than that among men ($p < 0.0001$). There was no significant difference in LDL-C ($p = 0.427$), TC ($p = 0.093$) and TG ($p = 0.854$) levels between men and women but HDL-C levels were significantly lower ($p = 0.0009$) among women. Though not presented in the results, there was no self-reported dyslipidemia or self-reported lipid lowering treatment among the study participants. The genetic principal component analysis showed that there was no significant genetic structure that would influence the results, as illustrated in **Supplementary Figure S1**, and therefore genetic structure was not corrected for in the analyses.

[3]http://genome.sph.umich.edu/wiki/EPACTS

TABLE 2 | Basic characteristics of the study population in northern Ghana.

Variable	Men (*n* = 866)	Women (*n* = 989)	Total (*n* = 1855)	*p* value
Age (years)	50 (46–56)	52 (47–56)	51 (46–56)	0.0001
Waist circumference (cm)	72 (69–77)	75 (71–82)	74 (69–79)	<0.0001
HDL-C (mmol/l)	1.14 (0.91–1.40)	1.08 (0.89–1.30)	1.11 (0.90–1.35)	0.0009
LDL-C (mmol/l)	1.60 (1.15–2.19)	1.61 (1.20–2.15)	1.61 (1.18–2.16)	0.6599
TC (mmol/l)	3.11 (2.59–3.70)	3.17 (2.68–3.78)	3.15 (2.62–3.75)	0.0929
TG (mmol/l)	0.56 (0.43–0.73)	0.55 (0.43–0.74)	0.56 (0.43–0.73)	0.8541

Data given as median (interquartile range).

Variants Associated With Lipid Levels

We tested the association between variants within the transcribed regions of selected genes (*ABCA1*, *CETP*, *LCAT*, *LPL*, *MMAB*, *MVK*, *PCSK9*, and *PON1*) and each lipid fraction as a continuous variable (HDL-C, LDL-C, TC, and TG). There were associations between variants of *CETP*, *LCAT*, *PCSK9*, and *PON1* and lipid levels at $p < 0.05$ after adjustment for age, sex and waist circumference (**Table 3**). The lead associated SNVs were the following: rs17231520 and rs34065661 in *CETP* and rs1109166 in *LCAT* for HDL-C, rs11806638 in *PCSK9* for LDL-C and rs854558 in *PON1* for TC. For *CETP*, there were 11 SNVs that were significantly associated with HDL-C levels, with the strongest associations being observed at rs17231520 and rs34065661 (**Table 3**). All but one of these SNVs (rs3816117) showed a positive association between the minor allele (relative to the major allele) and HDL-C levels. One SNV within the *LCAT* gene (rs1109166) showed a significant negative association with HDL-C levels. With regards to LDL-C, 4 SNVs within the *PCSK9* gene were identified with significant associations with serum LDL-C levels. However, only one of these (rs11806638) remained significant ($p < 0.05$) after adjustment for age, sex and waist circumference. All these variants demonstrated negative associations with LDL-C levels. Five SNVs in the *PON1* gene were significantly associated with serum TC levels, but only 3 (rs854558, rs854564, and rs854565) of these remained significant after adjustment for age, sex and waist circumference. All these SNVs were positively associated with TC serum levels. No significant gene associations were found for serum TG levels.

The LocusZoom plots for the significant associations are shown in **Figures 1A,B** and lead SNVs from *CETP*, *LCAT*, *PCSK9*, and *PON1* are shown. **Figure 2** shows the median serum lipid levels given each genotype for each lead SNV. The rs17231520 SNV within *CETP* shows an additive effect, with serum HDL-C levels increasing in the presence of the minor allele (A). The SNV within *LCAT*, rs1109166, showed a dominant and negative effect of the minor allele (T) on serum HDL-C levels. The minor allele (A) at rs11806638 in the *PCSK9* gene demonstrated a negative effect on LDL-C levels, in a recessive manner. The serum levels of TC were positively associated with the minor allele (T) at rs854558 in the *PON1* gene, with the T allele acting in a dominant fashion.

Functional Annotation of Significantly Associated Variants

Results of functional analyses of the significantly associated variants are summarised in **Table 4**. The functional annotation shows that the leading associated variant (rs17231520: $p = 4.44e\text{-}09$) in the *CETP* region was an upstream variant with an

TABLE 3 | Association of single nucleotide variants of selected candidate genes with lipid levels among the total study group ($n = 1855$).

					Without covariates			With covariates[2]	
SNV(Gene)	**A1/A2**	**Position[1]**	**MAF**	**pHWE**	**β (SE)**	**p_R value**	**p_B value**	**β (SE)**	**p_R value**	**p_B value**
HDL-C										
rs17231520(*CETP*)	A/G	56995827	0.089	1.000	0.1386 (0.0209)	4.19e-11	4.90e-09	0.1387 (0.0209)	3.80e-11	4.44e-09
rs34065661(*CETP*)	G/C	56995935	0.090	1.000	0.1372 (0.0207)	4.47e-11	5.23e-09	0.1371 (0.0207)	4.30e-11	5.03e-09
rs711752(*CETP*)	A/G	56996211	0.246	0.573	0.0833 (0.0137)	1.35e-09	1.58e-07	0.0845 (0.0137)	7.45e-10	8.72e-08
rs708272(*CETP*)	A/G	56996288	0.245	0.531	0.0832 (0.0137)	1.42e-09	1.66e-07	0.0844 (0.0137)	7.82e-10	9.14e-08
rs891142(*CETP*)	T/C	57003977	0.098	1.000	0.0984 (0.0200)	9.29e-07	1.09e-04	0.0983 (0.0200)	9.55e-07	1.11e-04
rs891143(*CETP*)	T/C	57003980	0.097	0.692	0.0976 (0.0200)	1.16e-06	1.36e-04	0.0973 (0.0200)	1.24e-06	1.45e-04
rs4784740(*CETP*)	G/C	57001085	0.091	0.780	0.1008 (0.0208)	1.35e-06	1.58e-04	0.1009 (0.0208)	1.31e-06	1.53e-04
rs3816117(*CETP*)	T/C	56996158	0.378	0.459	−0.0539 (0.0124)	1.48e-05	1.73e-03	−0.0551 (0.0124)	9.17e-06	1.07e-03
rs158478(*CETP*)	C/A	57007734	0.300	0.294	0.0487 (0.0129)	1.59e-04	0.01858	0.0488 (0.0129)	1.49e-04	0.01740
rs891141(*CETP*)	G/T	57003723	0.168	0.115	0.0576 (0.0157)	2.46e-04	0.02875	0.0589 (0.0157)	1.76e-04	0.02053
rs289719(*CETP*)	T/C	57009941	0.440	0.963	0.0430 (0.0120)	3.62e-04	0.04236	0.0443 (0.0120)	2.30e-04	0.02685
rs1109166(*LCAT*)	T/C	67977382	0.219	0.946	−0.0406 (0.0145)	0.00510	0.03060	−0.0408 (0.0145)	0.00476	0.02853
LDL-C										
rs11806638(*PCSK9*)	A/C	55518160	0.396	0.961	−0.0601 (0.0152)	7.87e-05	9.29e-03	−0.0549 (0.0149)	2.27e-04	0.02683
rs11804420(*PCSK9*)	G/A	55520445	0.298	0.346	−0.0612 (0.0165)	2.02e-04	0.02389	−0.0513 (0.0161)	0.00147	0.1736
rs11800231(*PCSK9*)	A/G	55517940	0.215	0.149	−0.0682 (0.0184)	2.21e-04	0.02601	−0.0635 (0.0180)	4.25e-04	0.05017
rs45545732(*PCSK9*)	T/A	55519231	0.273	0.860	−0.0605 (0.0168)	3.08e-04	0.03634	−0.0521 (0.0164)	0.00147	0.17400
TC										
rs854558(*PON1*)	T/C	94945374	0.385	0.659	0.0435 (0.0110)	7.40e-05	6.95e-03	0.0404 (0.0109)	2.13e-04	0.02005
rs854564(*PON1*)	G/T	94948182	0.386	0.807	0.0412 (0.0110)	1.76e-04	0.01654	0.0383 (0.0109)	4.68e-04	0.04398
rs854565(*PON1*)	A/G	94948344	0.386	0.807	0.0412 (0.0110)	1.76e-04	0.01654	0.0383 (0.0109)	4.68e-04	0.04398
rs854566(*PON1*)	A/G	94948749	0.377	0.401	0.0405 (0.0110)	2.22e-04	0.02090	0.0374 (0.0109)	6.23e-04	0.05858
rs854567(*PON1*)	A/G	94948784	0.377	0.401	0.0405 (0.0110)	2.22e-04	0.02090	0.0374 (0.0109)	6.23e-04	0.05858

A1/A2, Minor/major allele; MAF, Minor allele frequency in the study population; pHWE, Hardy-Weinberg p value; N, total number of participants; SE, standard error;
[1]Genomic position and chromosome in build 37; p_R, raw p value; p_B, Bonferroni-adjusted p value;
[2] Covariates were sex, age and waist circumference.

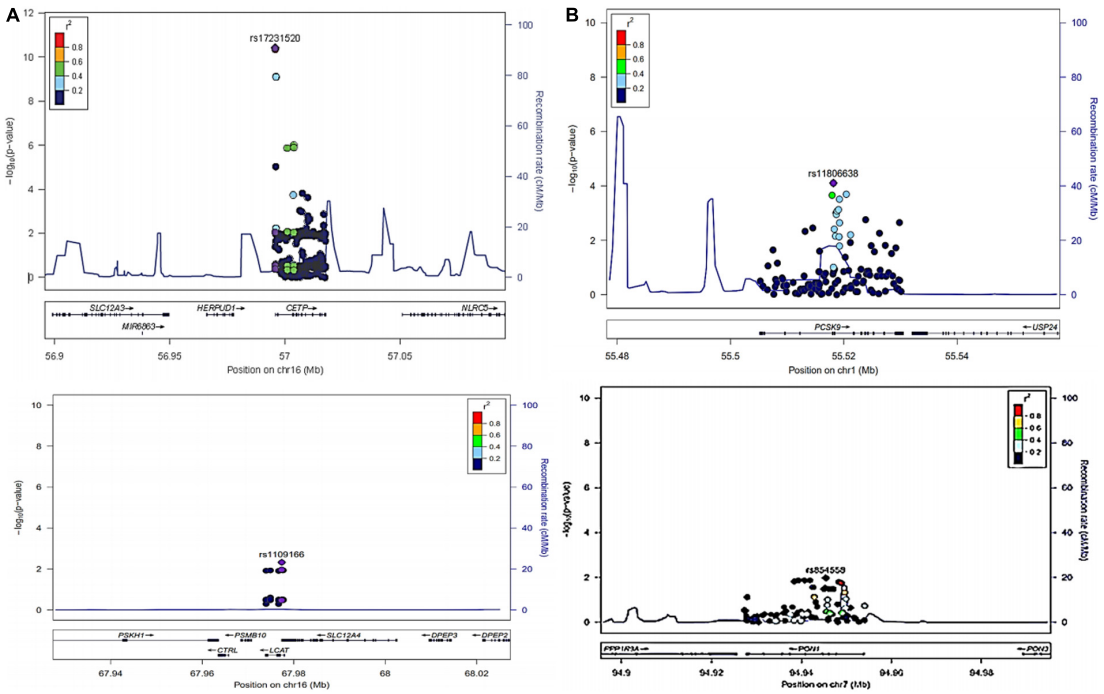

FIGURE 1 | (A) LocusZoom plots of rs17231520 and rs34065661 (*CETP*) and rs1109166(*LCAT*) associated with higher and lower levels of HDL-C respectively after adjustment for sex, age and waist circumference. **(B)** LocusZoom plots of rs11806638(*PCSK9*) and rs854558(*PON1*) associated with lower levels of LDL-C and higher levels of TC respectively after adjustment for sex, age and waist circumference.

RDB score of 5 suggesting possible regulation of the rate of transcription. The second lead SNV in *CETP* (rs34065661: p = 5.03e-09) was a missense variant which also had an RDB score of 5 but with a CADD score > 10 which suggested a possible deleterious effect on gene structure and protein function. All other significantly associated variants were located in the introns of four genes (*CETP, LCAT, PCSK9*, and *PON1)* and had RDB scores of 4 or 5 which indicated possible effects on binding sites and regulation of transcription. The LoFtool score of all mutations in *PON1* was 0.787 suggesting a mild effect of these mutations to high TC levels. The LoFtool scores of rs1109166 in *LCAT* and rs11806638 in *PCSK9* were 0.127 sand 0.467, respectively, suggesting possibly damaging effects of the mutations on gene function.

The frequencies of the minor alleles for our lead SNVs were similar to what have been observed in 1KG African populations (**Table 4** and **Figure 3**). Notably, some of our associated variants, including rs17231520 and rs34065661 in *CETP*, are observed only in those with African ancestry. For the SNVs that were common to Africans and European ancestry all the SNVs in the study population, except rs316117-T, rs711752-A, rs708272-A, and rs158478-C, had higher variant allele frequencies than those of the SNVs in the European ancestry populations.

Linkage Disequilibrium Among Significant Variants

The assessment of LD among significantly associated SNVs was performed using LDlink, a web-based tool (Machiela and Chanock, 2015). The value of r^2 in *CETP* ranged from 0.000 to 1.000 (**Supplementary Table S1**) while that of *PON1* ranged from 0.993 to 1.000 (**Supplementary Table S2**). In *CETP*, there was complete LD between two pairs of SNVs (rs711752 with rs708272 and rs17231520 with rs34065661) and strong LD between three pairs of SNVs (rs891142 with rs891143, rs4784740 with rs891142 and rs4784740 with rs891143) (**Supplementary Table S1**). All the significant variants in *PON1* were in complete LD with each other (**Supplementary Table S2**).

Replication Results

Replication results of significant variants in our population using the AADM study are shown in **Table 5**. All the *CETP* variants before and after adjustment for covariates, except rs158478, were replicated with the two lead signals in the replication analyses, after adjustment for covariates, being rs34065661 (β = 0.109, p = 5.20e-10) and rs17231520 (β = 0.107, p = 1.52e-10). Similarly, all signals in *PCSK9* before adjustment for covariates were replicated, and the rs11806638 variant after adjustment for covariates was replicated ($\beta = -0.031$, $p = 0.00040$). All replicated associations showed the same direction of allelic effect. None of the significant SNVs in *LCAT* and *PON1* was replicated.

DISCUSSION

This study on adults in northern Ghana evaluated the associations between lipid levels and common nucleotide variants

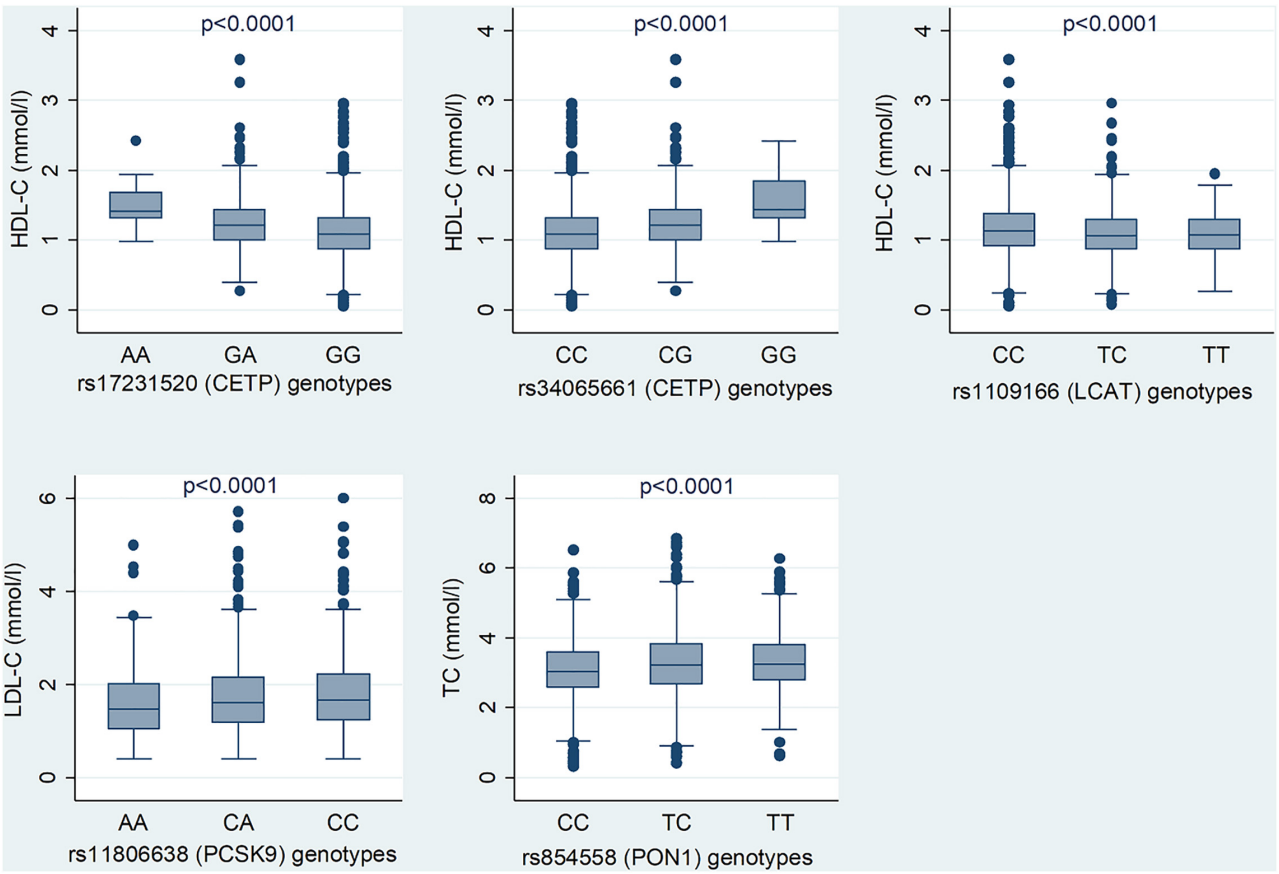

FIGURE 2 | Median lipid levels for the genotypes of the lead SNVs among Ghanaian adults. The *p* value is adjusted for age, sex and waist circumference.

in the transcribed regions of eight candidate genes involved in lipid pathways. After adjustment for sex, age and waist circumference, SNVs in four genes (*CETP*, *LCAT*, *PCSK9*, and *PON1*) were significantly associated with lipid levels. Variants in *CETP* and *LCAT* were significantly associated with HDL-C levels. One variant in *PCSK9* was significantly associated with LDL-C and another in *PON1* associated with TC levels.

Previous studies have suggested that African populations and populations of African descent tend to have higher HDL-C levels than Europeans suggesting a more favourable profile for CVD risk (Seedat et al., 1992; Chaturvedi et al., 1994; Johnson et al., 2004; Miljkovic-Gacic et al., 2006; D'Adamo et al., 2010). In this study higher HDL-C levels were strongly associated with an upstream variant, a missense variant and intronic variants in the *CETP* gene. The lead SNV (rs17231520) was in the upstream regulatory region and functional prediction suggests an impact on the rate of transcription. The high CADD score of the associated missense variant (rs34065661) is interesting as it suggests a deleterious effect on protein function. The two lead SNVs in *CETP* may be the causal variants for high HDL-C in West Africans since they are functionally relevant, their effect sizes are larger than those of the others and they are common in African populations (>0.08), and were not observed in 1000 Genomes Project European and Asian populations. Additionally we identified several intronic SNVs in the *CETP* gene where the variant alleles were positively associated, except rs3816117, which was negatively associated, with HDL-C levels. Functional analysis suggests these intronic variants regulate the rate of gene transcription but their low CADD scores indicate that the effect of their regulatory activity on gene function is mild. The significantly associated variants in *CETP* probably impaired the ability of the gene to accelerate the transfer of cholesteryl esters (CE) from cholesteryl ester-rich HDL-C formed by *LCAT* to other lipoprotein particles resulting in increased HDL-C levels (Brown et al., 1989).

Most of the SNVs in the *CETP* gene replicate previous GWAS findings in peoples of African descent (Buyske et al., 2012; Elbers et al., 2012; Pirim et al., 2016). Our replication analysis, using the African American diabetes mellitus (AADM) study involving over 4000 East and West Africans (Rotimi et al., 2001), revealed that the direction and strength of effects of the SNVs in *CETP* were consistent, except for rs158478. The minor allele frequencies of the replicated variants in our population were similar to the allele frequencies in other Africans both in the 1000 Genomes data and replication population but the effect sizes of the variants were larger in our study (**Tables 3**, **5**). The rs17231520 (A) and

TABLE 4 | Functional annotations of lipid trait associated SNVs in Ghanaian adults and their minor allele frequencies in comparison to other population groups.

Variant(Gene)	Localisation	CADD score	RDB score	LoFtool score		Minor allele frequencies per population				
					Minor allele	Study population	Africa	Europe	East Asia	South Asia
rs17231520(*CETP*)	Upstream variant	4.38	5	0.970	A	0.089	0.082	0.000	0.000	0.000
rs34065661(*CETP*)	Missense variant	12.64	5	0.970	G	0.090	0.083	0.000	0.000	0.000
rs711752(*CETP*)	Intron variant	8.46	5	0.970	A	0.246	0.246	0.426	0.375	0.451
rs708272(*CETP*)	Intron variant	0.46	5	0.970	A	0.245	0.247	0.425	0.375	0.448
rs891142(*CETP*)	Intron variant	2.69	–	0.970	T	0.098	0.108	0.000	0.021	0.010
rs891143(*CETP*)	Intron variant	1.39	–	0.970	T	0.097	0.103	0.000	0.000	0.000
rs4784740(*CETP*)	Intron variant	0.86	6	0.970	G	0.091	0.096	0.000	0.000	0.000
rs3816117(*CETP*)	Intron variant	0.77	5	0.970	T	0.378	0.404	0.516	0.518	0.397
rs158478(*CETP*)	Intron variant	1.57	–	0.970	C	0.300	0.284	0.524	0.555	0.428
rs891141(*CETP*)	Intron variant	0.17	5	0.970	G	0.168	0.153	0.001	0.021	0.013
rs289719(*CETP*)	Intron variant	2.45	5	0.970	T	0.440	0.441	0.330	0.310	0.452
rs1109166(*LCAT*)	Intron variant	3.06	4	0.217	T	0.219	0.260	0.815	0.904	0.761
rs11806638(*PCSK9*)	Intron variant	3.86	5	0.467	A	0.396	0.356	0.064	0.027	0.064
rs854558(*PON1*)	Intron variant	1.92	–	0.787	T	0.385	0.340	0.714	0.702	0.614
rs854564(*PON1*)	Intron variant	3.01	5	0.787	G	0.386	0.341	0.286	0.299	0.387
rs854565(*PON1*)	Intron variant	2.23	–	0.787	A	0.386	0.341	0.286	0.298	0.387

CADD, combined annotation dependent depletion score predicts damaging effect of the variant on protein structure and function; LoFtool, loss of function tool score predicts gene intolerance and consequent susceptibility to disease; RDB, RegulomeDB score assesses the regulatory potential of the variant; the frequencies are for the 1000 Genomes data.

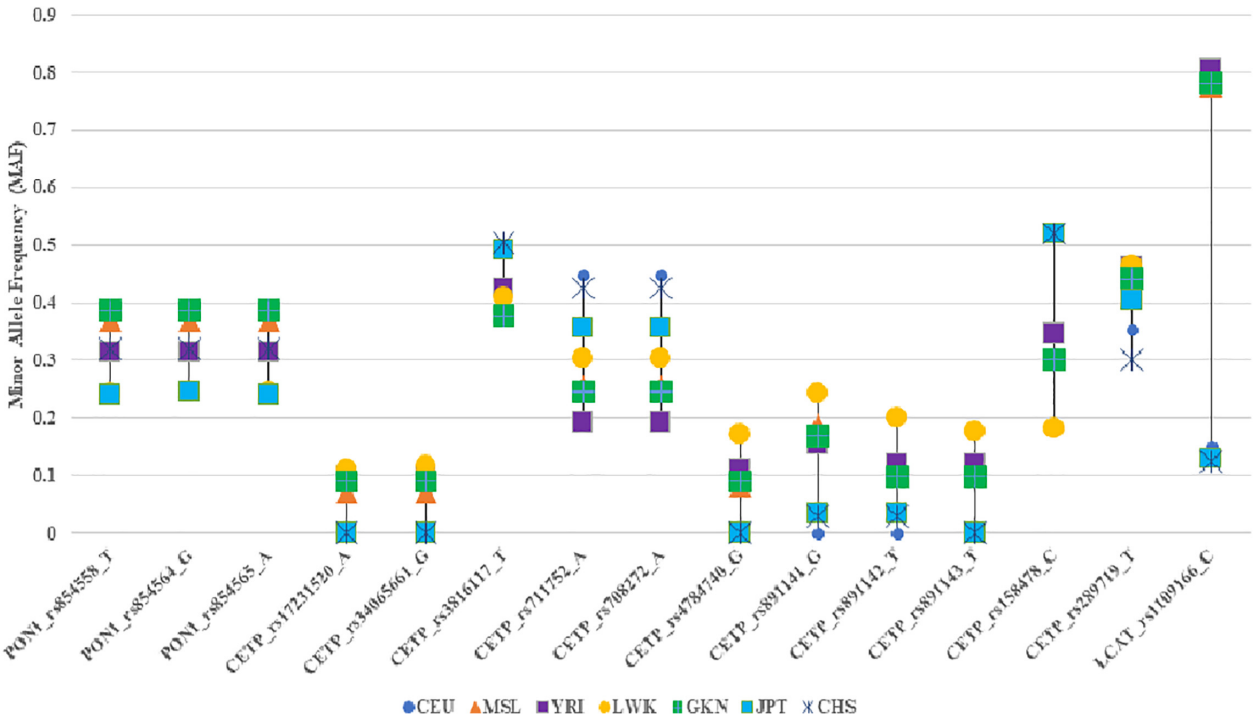

FIGURE 3 | Comparison of the minor allele frequencies of the significantly associated SNPs in the study population with those of the 1000 Genomes. CEU: Northern Europeans from Utah, MSL: Mende in Sierra Leone, YRI: Yoruba in Ibadan, Nigeria, LWK: Luhya in Webuye, Kenya, GKN: the study population, JPT: Japanese in Tokyo, Japan, CHS: Southern Han Chinese.

rs34065661 (G) alleles which are absent in European and Asian populations, were previously found to be positively associated with HDL-C levels among African Americans (Buyske et al., 2012; Elbers et al., 2012) and sub-Saharan Africans (Pirim et al., 2016). As in our study, the two variants were in strong LD in all these studies and the frequencies of the rs17231520 (A) and rs34065661 (G) alleles were similar to those in our populations but the effect sizes of these alleles were smaller in our study. Similarly our study replicated rs711752, rs708272, rs891141, rs891143 and rs289719 variants in *CETP* that were previously found to be associated with HDL-C in sub-Saharan Africans by Pirim et al. (2016). Though the MAFs of these variants in this African population were similar to those in our study, the effect sizes of the variants were generally larger in our population. Also, the MAFs of these variants in our population were similar to those in other African populations (MSL, YRI, and LWK) but were larger than the MAFs in Europeans and Asian populations (CEU, JPT, CHS).

In our study the minor allele (T) of rs1109166 in *LCAT* was negatively associated with HDL-C levels. Since the T allele is the common non-African allele (0.76 to 0.90 in Europeans and Asians), its association with low HDL-C in Africans may not be functionally relevant, despite the fact that the low LoFtool score predicts a deleterious effect for this intronic variant (Le Hir et al., 2003; Chorev and Carmel, 2012). It may be tagging other functionally relevant variants through LD, which could interfere with the ability of *LCAT* to form CE and to promote cholesterol efflux from peripheral cells. Furthermore, this variant was not associated with HDL-C levels in the replication study. Our study is therefore the first to report this association in an African population and requires further investigation.

PCSK9 binds to the LDL receptor (LDLR) to form a complex which moves from the endosomal recycling pathway to the lysosome for degradation (Akram et al., 2010). Since LDLR removes cholesterol-rich LDL particles from plasma (Poirier et al., 2008), a loss of function or reduction in PCSK9 activity inhibits LDLR degradation and leads to the reduction of LDL-C in the blood (Park et al., 2004). In this study the rs11806638 (A) allele in *PCSK9* was negatively associated with LDL-C level. In addition to this variant being previously associated with lower LDL-C levels in African Americans (Musunuru et al., 2012) we replicated its association in other sub-Saharan Africans from the AADM study (Rotimi et al., 2001). The replication of this variant has clinical translational impact since it has the potential to act as a *PCSK9* inhibitor in Africans with high LDL-C levels leading to reduced CVD risk.

PON1 encodes an enzyme that is present in the HDL particle and protects HDL and LDL from peroxidation by degrading or hydrolysing specific oxidised cholesteryl esters and phospholipids contained in oxidised lipoproteins (Barter et al., 2004). This process contributes mainly to the antioxidant properties of the HDL particle which results in improved macrophage-mediated cholesterol efflux (Parthasarathy et al., 1990). Genetic variation in the gene is therefore associated with

increased TC levels. Our results show positive associations of rs854558-T, rs854564-G, and rs854565-A alleles in *PON1* with TC in the study population. To the best of our knowledge this study is the first to report the association of these variants with TC levels, although other studies have shown associations of other SNVs in the *PON1* gene with serum lipid levels (Chang et al., 2010; Bentley et al., 2014; Luo et al., 2018). The similar association observed in the lead variant and other signals (rs854564 and rs854565) in the gene with TC levels could be due to the complete LD between these signals and the lead variant. Though the frequency of the variant allele of rs854558 is substantial (0.385) in the study population, its effect size is small and it is unlikely to play a major role in hypercholesterolemia and CVD risk.

Strengths and Limitations of the Study

The study had several strengths. Firstly, individuals were genotyped using an array enriched for gene variants that are common in African populations and genotypes were imputed using and African reference panel. Secondly, it is one of very few studies to evaluate the association of SNVs with lipid levels in Africa thereby contributing knowledge on the influence of genetic factors on lipid levels in under-studied African populations. The study is the first of its kind to identify the association of three polymorphisms in LD (rs854558, rs854564, and rs854565) in *PON1* which were associated with higher TC levels. This population is relatively drug-naïve as they do not have easy access to lipid-lowering medication. The West and East African replication cohort (Rotimi et al., 2001) supported the association of *CETP* variants with high HDL-C levels and *PCSK9* variants with LDL-C levels.

In terms of limitations, the sample size was relatively small and therefore underpowered to detect variants with small effect sizes. However, power calculations suggested that there was sufficient power to replicate previously reported gene associations with lipids. Candidate genes were selected from studies involving people of African descent, but not resident in Africa, and thus genes chosen may not be appropriate for indigenous African populations. Furthermore, only eight genes were included in the study and therefore additional genes that contribute more strongly to modulating lipid levels may have been overlooked. However, a GWAS for lipid levels is underway for the entire AWI-Gen dataset. The Bonferroni method to reduce false discovery rate assumes that pairwise tests are independent and this is considered by many to be overly conservative, as many SNVs from the same gene locus would be in linkage disequilibrium. Therefore results from sets of tests may not be independent (Lewis, 2002). Lipid lowering medications may

TABLE 5 | Replication results of lipid trait associated variants in East and West Africans (AADM study).

					Without covariates		With covariates[2]	
SNV(Gene)	**A1/A2**	**Position[1]**	**N**	**MAF**	**β (SE)**	***p* value**	**β (SE)**	***p* value**
HDL-C								
rs17231520 (*CETP*)	**A/G**	**56995827**	**4317**	**0.076**	**0.1063 (0.01726)**	**8.07e-10**	**0.1069 (0.01716)**	**5.20e-10**
rs34065661 (*CETP*)	**G/C**	**56995935**	**4317**	**0.077**	**0.109 (0.0171)**	**2.00e-10**	**0.1091 (0.017)**	**1.52e-10**
rs711752 (*CETP*)	**A/G**	**56996211**	**4317**	**0.224**	**0.06399 (0.01083)**	**3.76e-09**	**0.0638 (0.01077)**	**3.41e-09**
rs708272 (*CETP*)	**A/G**	**56996288**	**4317**	**0.223**	**0.0643 (0.01087)**	**3.54e-09**	**0.0641 (0.0108)**	**3.21e-09**
rs891142 (*CETP*)	**T/C**	**57003977**	**4317**	**0.096**	**0.0679 (0.0157)**	**1.6e-05**	**0.06801 (0.01562)**	**1.37e-05**
rs891143 (*CETP*)	**T/C**	**57003980**	**4317**	**0.088**	**0.07841 (0.01635)**	**1.68e-06**	**0.07826 (0.01627)**	**1.57e-06**
rs4784740 (*CETP*)	**G/C**	**57001085**	**4317**	**0.081**	**0.08446 (0.01681)**	**5.30e-07**	**0.08399 (0.01673)**	**5.40e-07**
rs3816117 (*CETP*)	**T/C**	**56996158**	**4317**	**0.426**	**−0.04769 (0.009266)**	**2.8e-07**	**−0.04784 (0.009212)**	**2.159e-07**
rs158478 (*CETP*)	C/A	57007734	4317	0.301	0.01886 (0.01012)	0.06234	0.01819 (0.01006)	0.07064
rs891141(*CETP*)	**G/T**	**57003723**	**4317**	**0.142**	**0.06963 (0.01308)**	**1.08e-07**	**0.06754 (0.01302)**	**2.24e-07**
rs289719 (*CETP*)	**T/C**	**57009941**	**4317**	**0.434**	**0.03065 (0.009093)**	**0.00076**	**0.02986 (0.009044)**	**0.00097**
rs1109166 (*LCAT*)	T/C	67977382	4317	0.236	−0.01591 (0.01109)	0.1516	−0.01629 (0.01103)	0.1399
LDL-C								
rs11806638 (*PCSK9*)	**A/C**	**55518160**	**4287**	**0.378**	**−0.03089 (0.008969)**	**0.00058**	**−0.03106 (0.008801)**	**0.00042**
rs11804420 (*PCSK9*)	**G/A**	**55520445**	**4287**	**0.253**	**−0.04537 (0.00981)**	**3.9e-06**	−0.04365 (0.009631)	6.0e-06
rs11800231(*PCSK9*)	**A/G**	**55517940**	**4287**	**0.197**	**−0.04148 (0.01084)**	**0.00013**	−0.04165 (0.01065)	9.3e-05
rs45545732(*PCSK9*)	**T/A**	**55519231**	**4287**	**0.228**	**−0.03805 (0.01023)**	**0.0002**	−0.03612 (0.01004)	0.00033
TC								
rs854558(*PON1*)	T/C	94945374	4317	0.355	0.009671 (0.006512)	0.1376	0.01005 (0.006332)	0.1125
rs854564(*PON1*)	G/T	94948182	4317	0.357	0.009167 (0.006488)	0.1578	0.009591 (0.006309)	0.1285
rs854565(*PON1*)	A/G	94948344	4317	0.357	0.008752 (0.006486)	0.1773	0.009207 (0.006308)	0.1445
rs854566(*PON1*)	A/G	94948749	4317	0.342	0.008547 (0.006536)	0.1911	0.009499 (0.006355)	0.1351
rs854567(*PON1*)	A/G	94948784	4317	0.342	0.009227 (0.006547)	0.1588	0.01017 (0.006366)	0.1103

A1/A2, Minor/major allele; MAF, Minor allele frequency in AADM study; pHWE, Hardy-Weinberg p value; N, total number of participants; SE, Standard error;
[1]Genomic position and chromosome in build 37;
[2] Covariates were age, sex, waist circumference and T2D; results of replicated variants indicated in bold.

affect the direction of associations and mask the genetic effect on lipid levels, but since no person reported taking such medication this was not a concern for the study. Another limitation is that the study used SNV arrays, whereas sequencing data could have led to the discovery of new variants associated with lipid levels among Africans. We are mindful of the potential masking effect of comorbidity in our study sample on the signals of association with lipid levels. Further research is needed on gene-gene interaction and gene-environment interaction to fully elucidate the factors influencing lipid levels in this African population.

CONCLUSION

Our findings showed that several variants in *CETP*, *LCAT*, *PCSK9*, and *PON1* were significantly associated with HDL-C, LDL-C and TC with the strongest signal being that of rs117231520 (*CETP*) with serum HDL-C levels in this Ghanaian population. *CETP* and *PCSK9* variants replicate previously reported lipid trait GWAS loci in African Americans and strengthen the evidence of the influence of these signals on lipid levels in peoples of African descent. The novelty of this study lies in the identification of three genetic polymorphisms in LD (rs854558, rs854564, and rs854565) in *PON1* which were associated with increasing TC levels in Ghanaians.

DATA AVAILABILITY STATEMENT

The data is available under the AWIGen Study in EGA (EGAS00001002482).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the National Institutes of Health guidelines, Human Research Ethics Committees (HREC) of the University of the Witwatersrand (ID No. M12109), the Ghana Health Service Ethics Review Committee (ID No. GHS-ERC:05/05/2015) and the Navrongo Health Research Centre Institutional Review Board (ID No. NHRCIRB178) with community engagement and written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Human Research Ethics Committee (HREC) of the University of the Witwatersrand (ID No. M12109), the Ghana Health Service Ethics Review Committee (ID No. GHS-ERC:05/05/2015) and the Navrongo Health Research Centre Institutional Review Board (ID No. NHRCIRB178).

AUTHOR CONTRIBUTIONS

GA, LA-E, EN, CD, and AO collected the data. AC performed genetic data curation. GA conducted the data analysis and drafted the manuscript. GA, LA-E, NC, and MR interpreted the results. AB and CR performed the replication assessment. LA-E, EN, CD, AO, NC, AC, AB, CR, and MR edited the draft. MR was the principal investigator and team leader in the AWI-Gen project and provided scientific leadership in the development of the research protocol. All authors read and approved the final draft.

FUNDING

This study was funded by the National Institutes of Health (NIH) through the H3Africa AWI-Gen project (NIH grant number U54HG006938) and the Wits Non-Communicable Disease Research Leadership Programme (NIH Fogarty International Centre grant number D43TW008330). AWI-Gen is supported by the National Human Genome Research Institute (NHGRI), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Office of the Director (OD) at the National Institutes of Health. The replication study was supported by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology (CIT), and the Office of the Director at the National Institutes of Health (1ZIAHG200362).

ACKNOWLEDGMENTS

We are grateful to the chiefs and people of the Kassena-Nankana traditional area (Ghana) for their cooperation and provision of samples that made the study possible. We appreciate the support and contribution of the management and staff of the Navrongo Health Research Centre (NHRC) and the Sydney Brenner Institute for Molecular Bioscience (SBIMB). We are particularly grateful to Prof. Scott Hazelhurst of the SBIMB for playing a lead role in the AWI-Gen genetic data curation. We acknowledge the roles of Ms Immaculate Anati of the NHRC for processing the buffy coats and Ms Cassandra Soo and Dr. Stuart Ali both of the SBIMB for the DNA extraction and the management of the AWI-Gen phenotype data, respectively. We also appreciate the effort of Prof. Osman Sankoh, formerly Director of INDEPTH and co-principal investigator, on the AWI-Gen project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.456661/full#supplementary-material

Supplementary Figure 1 | Principal component analysis using data from the H3Africa SNV array (∼2.3 million SNVs) on the full AWI-Gen dataset, including the participants from each of the six study sites. This study was based only on data from the Navrongo study site (nvr) and the participants clustered closely together illustrating that there was little population structure in this group. The other five study sites were Nanoro in Burkina Faso (nan), Nairobi (nai) and three study sites in South Africa: Soweto (SWT), Agincourt (agt) and Dikgale (dkg).

REFERENCES

Abessolo, F. A., Bruno, M. J., Ma, N., and Yangou, M. (2014). Enzymatic and genetic polymorphisms of paraoxonase-1 in the Gabonese population: the relation to lipid parameters in patients with diabetes Enzymatic and genetic polymorphisms of paraox- onase-1 in the Gabonese population: the relation to lipid paramete. *Metab. Diabet. S. Afr.* 17, 92–99. doi: 10.1080/22201009.2012.10872283

Adeyemo, A., Bentley, A. R., Meilleur, K. G., Doumatey, A. P., Chen, G., Zhou, J., et al. (2012). Transferability and fine mapping of genome-wide associated loci for lipids in African Americans. *BMC Med. J.* 13:88. doi: 10.1186/1471-2350-13-88

Agongo, G., Nonterah, E. A., Debpuur, C., Amenga-Etego, L., Ali, S., Oduro, A., et al. (2018). The burden of dyslipidaemia and factors associated with lipid levels among adults in rural northern Ghana: an AWI-Gen sub-study. *PLoS One* 13:e0206326. doi: 10.1371/journal.pone.0206326

Akram, O. N., Bernier, A., Petrides, F., Wong, G., and Lambert, G. (2010). Beyond LDL cholesterol, a new role for PCSK9. *Arterioscler. Thromb. Vasc. Biol.* 30, 1279–1281. doi: 10.1161/atvbaha.110.209007

Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-etego, L., Boua, R. P., et al. (2018). Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Glob. Health Action* 11:1507133. doi: 10.1080/16549716.2018.1507133

Barter, P. J., Nicholls, S., Rye, K.-A., Anantharamaiah, G. M., Navab, M., and Fogelman, A. M. (2004). Antiinflammatory properties of HDL. *Circ. Res.* 95, 764–772. doi: 10.1161/01.res.0000146094.59640.13

Beekman, M., Heijmans, B. T., Martin, N. G., Pedersen, N. L., Whitfield, J. B., DeFaire, U., et al. (2002). Heritabilities of apolipoprotein and lipid levels in three countries. *Twin Res. Offic. J. Intern. Soc. Twin Stud.* 5, 87–97. doi: 10.1375/twin.5.2.87

Bentley, A. R., Chen, G., Shriner, D., Doumatey, A. P., Zhou, J., Huang, H., et al. (2014). Gene-based sequencing identifies lipid-influencing variants with ethnicity-specific effects in african Americans. *PLoS Genet.* 10:1004190. doi: 10.1371/journal.pgen.1004190

Bentley, A. R., and Rotimi, C. N. (2017). Interethnic differences in serum lipids and implications for Cardiometabolic disease risk in African ancestry populations. *Glob. Heart* 12, 141–150. doi: 10.1016/j.gheart.2017.01.011

Bland, J. M., and Altman, D. G. (1995). Multiple significance tests: the bonferroni method. *Br. Med. J.* 310:170. doi: 10.1136/bmj.310.6973.170

Boomsma, D. I., Wijmenga, C., Slagboom, E. P., Swertz, M. A., Karssen, L. C., Abdellaoui, A., et al. (2014). The genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.* 22, 221–227.

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797. doi: 10.1101/gr.137323.112

Brown, M. L., Inazu, A., Hesler, C. B., Agellon, L. B., Mann, C., Whitlock, M. E., et al. (1989). Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. *Nature* 342, 448–451. doi: 10.1038/342448a0

Buyske, S., Wu, Y., Carty, C. L., Cheng, I., Assimes, T. L., Dumitrescu, L., et al. (2012). Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS One* 7:e35651. doi: 10.1371/journal.pone.0035651

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8

Chang, M., Yesupriya, A., Ned, R. M., Mueller, P. W., and Dowling, N. F. (2010). Genetic variants associated with fasting blood lipids in the U.S. population: third national health and nutrition examination survey. *BMC Med. Genet.* 11:62. doi: 10.1186/1471-2350-11-62

Chaturvedi, N., McKeigue, P. M., and Marmot, M. G. (1994). Relationship of glucose intolerance to coronary risk in Afro-caribbeans compared with Europeans. *Diabetologia* 37, 765–772. doi: 10.1007/s001250050177

Chorev, M., and Carmel, L. (2012). The function of introns. *Front. Genet.* 3:55. doi: 10.3389/fgene.2012.00055

D'Adamo, E., Northrup, V., Weiss, R., Santoro, N., Pierpont, B., Savoye, M., et al. (2010). Ethnic differences in lipoprotein subclasses in obese adolescents: importance of liver and intraabdominal fat accretion. *Am. J. Clin. Nutr.* 92, 500–508. doi: 10.3945/ajcn.2010.29270

Diego Chacon-Cortes, L. R. G. (2014). Methods for extracting genomic DNA from whole blood samples: current perspectives. *J. Biorespos. Sci. Appl. Med.* 2, 1–9. doi: 10.2147/bsam.s46573

Dron, J. S., and Hegele, R. A. (2016). Genetics of lipid and lipoprotein disorders and traits. *Curr. Genet. Med. Rep.* 4, 130–141. doi: 10.1007/s40142-016-0097-y

Durrington, P. N., Mackness, B., and Mackness, M. I. (2001). Brief reviews: paraoxonase and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* 21, 473–480.

Elbers, C. C., Guo, Y., Tragante, V., van Iperen, E. P. A., Lanktree, M. B., Castillo, B. A., et al. (2012). Gene-centric meta-analysis of lipid traits in African, East Asian and Hispanic populations. *PLoS One* 7:e50198. doi: 10.1371/journal.pone.050198

Gauderman, W. J. (2002). Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* 155, 478–484. doi: 10.1093/aje/155.5.478

Hovingh, G. K., Hutten, B. A., Holleboom, A. G., Petersen, W., Rol, P., Stalenhoef, A., et al. (2005). Compromised LCAT function is associated with increased atherosclerosis G. *Circulation* 112, 879–884. doi: 10.1161/circulationaha.105.540427

Johnson, J. L., Slentz, C. A., Duscha, B. D., Samsa, G. P., McCartney, J. S., Houmard, J. A., et al. (2004). Gender and racial differences in lipoprotein subclass distributions: the STRRIDE study. *Atherosclerosis* 176, 371–377. doi: 10.1016/j.atherosclerosis.2004.05.018

Kelishadi, R., Haghjooy, S., Hasan, M., and Ban, M. (2014). Genetic association with low concentrations of high density lipoprotein-cholesterol in a pediatric population of the Middle East and North Africa: the CASPIAN-III study. *Atherosclerosis* 237, 273–278. doi: 10.1016/j.atherosclerosis.2014.08.043

Kotowski, I. K., Pertsemlidis, A., Luke, A., Cooper, R. S., Vega, G. L., Cohen, J. C., et al. (2006). A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* 78, 410–422. doi: 10.1086/500615

Le Hir, H., Nott, A., and Moore, M. J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends B Biochem. Sci.* 28, 215–220. doi: 10.1016/s0968-0004(03)00052-5

Lewis, C. M. (2002). Genetic association studies: design, analysis and interpretation. *Brief. Bioinform.* 3, 146–153. doi: 10.1093/bib/3.2.146

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, A. Y., Finucane, H. K., et al. (2016). Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* 48, 1443–1448. doi: 10.1038/ng.3679

Luo, Z., Pu, L., Muhammad, I., Chen, Y., and Sun, X. (2018). Associations of the PON1 rs662 polymorphism with circulating oxidized low-density lipoprotein and lipid levels: a systematic review and meta-analysis. *Lipids Health Dis.* 17:281. doi: 10.1186/s12944-018-0937-8

Machiela, M. J., and Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557. doi: 10.1093/bioinformatics/btv402

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4

Miao, L., Yin, R., Huang, F., Chen, W., and Cao, X. (2017). The effect of MVK-MMAB variants, their haplotypes and G × E interactions on serum lipid levels and the risk of coronary heart disease and ischemic stroke. *Oncotarget* 8, 72801–72817. doi: 10.18632/oncotarget.20349

Miljkovic-Gacic, I., Bunker, C. H., Ferrell, R. E., Kammerer, C. M., Evans, R. W., Patrick, A. L., et al. (2006). Lipoprotein subclass and particle size differences in Afro-Caribbeans, African Americans, and white Americans: associations with hepatic lipase gene variation. *Metab. Clin. Exper.* 55, 96–102. doi: 10.1016/j.metabol.2005.07.011

Mulder, N., Abimiku, A., Adebamowo, S. N., de Vries, J., Matimba, A., Olowoyo, P., et al. (2018). H3Africa: current perspectives. *Pharmacogen. Pers. Med.* 11, 59–66. doi: 10.2147/pgpm.s141546

Musunuru, K., Romaine, S. P. R., Lettre, G., Wilson, J. G., Volcik, K. A., Tsai, M. Y., et al. (2012). Multi-ethnic analysis of lipid-associated loci: the NHLBI CARe project. *PLoS One* 7:e36473. doi: 10.1371/journal.pone.036473

Niemsiri, V., Wang, X., Pirim, D., Radwan, Z. H., Bunker, C. H., Barmada, M. M., et al. (2015). Genetic contribution of SCARB1 variants to lipid traits in African

blacks: a candidate gene association study. *BMC Med. Genet.* 16:106. doi: 10.1186/s12881-015-0250-6

O'Donnell, C. J., and Elosua, R. (2008). Cardiovascular risk factors. *Insights Framingham Heart* 61, 299–310.

Oduro, A. R., Wak, G., Azongo, D., Debpuur, C., Wontuo, P., Kondayire, F., et al. (2012). Profile of the Navrongo health and demographic surveillance system. *Int. J. Epidemiol.* 41, 968–976. doi: 10.1093/ije/dys111

Park, S. W., Moon, Y.-A., and Horton, J. D. (2004). Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver. *J. Biol. Chem.* 279, 50630–50638. doi: 10.1074/jbc.m410077200

Parthasarathy, S., Barnett, J., and Fong, L. G. (1990). High-density lipoprotein inhibits the oxidative modification of low-density lipoprotein. *Biochim. Biophys. Acta* 1044, 275–283.

Perusse, L., Rice, T., Despres, J. P., Bergeron, J., Province, M. A., Gagnon, J., et al. (1997). Familial resemblance of plasma lipids, lipoproteins and postheparin lipoprotein and hepatic lipases in the HERITAGE Family Study. *Arterioscler. Thromb. Vasc. Biol.* 17, 3263–3269. doi: 10.1161/01.atv.17.11.3263

Pirim, D., Wang, X., Niemsiri, V., Radwan, Z. H., Clareann, H., Hokanson, J. E., et al. (2016). Resequencing of the CETP gene in American whites and African blacks: association of rare and common variants with HDL-cholesterol levels. *Metabolism* 65, 36–47. doi: 10.1016/j.metabol.2015.09.020

Pirim, D., Wang, X., Radwan, Z. H., Niemsiri, V., Hokanson, J. E., Hamman, R. F., et al. (2014). Lipoprotein lipase gene sequencing and plasma lipid profile. *J. Lipid Res.* 55, 85–93. doi: 10.1194/jlr.m043265

Poirier, S., Mayer, G., Benjannet, S., Bergeron, E., Marcinkiewicz, J., Nassoury, N., et al. (2008). The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2. *J. Biol. Chem.* 283, 2363–2372. doi: 10.1074/jbc.m708098200

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337. doi: 10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., et al. (2016). The AWI-Gen collaborative centre: understanding the interplay between genomic and environmental risk factors for Cardiometabolic diseases in sub-Saharan Africa. *Glob. Health Epidemiol. Genom.* 1:e20. doi: 10.1017/gheg.2016.17

Reinikainen, J., Laatikainen, T., Karvanen, J., and Tolonen, H. (2015). Cardiovascular disease and cardiovascular risk factors lifetime cumulative risk factors predict cardiovascular disease mortality in a 50-year follow-up study in Finland. *Intern. J. Epidemiol.* 44, 108–116. doi: 10.1093/ije/dyu235

Ridker, P. M., Paré, G., Parker, A. N., Zee, R. Y. L., Miletich, J. P., and Chasman, D. I. (2009). Polymorphism in the CETP gene region, hdl cholesterol, and risk of future myocardial infarction: genomewide analysis among 18,245 initially healthy women from the women's genome health study. *Circ. Cardiovasc. Genet.* 2, 26–33. doi: 10.1161/circgenetics.108.817304

Rotimi, C. N., Dunston, G. M., Berg, K., Akinsete, O., Amoah, A., Owusu, S., et al. (2001). In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann. Epidemiol.* 11, 51–58.

Seedat, Y. K., Mayet, F. G., Latiff, G. H., and Joubert, G. (1992). Risk factors and coronary heart disease in Durban blacks–the missing links. *S. Afr. Med. J.* 82, 251–256.

Sun, J., Qian, Y., Jiang, Y., Chen, J., Dai, J., Jin, G. et al. (2016). Association of KCTD10, MVK, and MMAB polymorphisms with dyslipidemia and coronary heart disease in Han Chinese population. *Lipids Health Dis.* 15:171. doi: 10.1186/s12944-016-0348-7

The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073 doi: 10.1038/nature09534

The H3Africa Consortium (2014). Enabling the genomic revolution in Africa: H3Africa is developing capacity for health-related genomics research in Africa. *Science* 344, 1346–1348

The International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861. doi: 10.1038/nature06258

van Leeuwen, E. M., Karssen, L. C., Deelen, J., Isaacs, A., Medina-Gomez, C., Mbarek, H., et al. (2015). Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat. Commun.* 6:6065. doi: 10.1038/ncomms7065

Wang, H., Naghavi, M., Allen, C., Barber, R. M., Carter, A., Casey, D. C., et al. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1459–1544.

WHO (2011). *Global Atlas on Cardiovascular Disease Prevention and Control*. Geneva: World Health Organization.

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. *Nat. Rev. Genet.* 45, 1274–1283. doi: 10.1038/ng.2797

Wood, A. R., Perry, J. R. B., Tanaka, T., Hernandez, D. G., Zheng, H.-F., Melzer, D., et al. (2013). Imputation of variants from the 1000 genomes project modestly improves known associations and can identify low-frequency variant - phenotype associations undetected by HapMap based imputation. *PLoS One* 8:e64343. doi: 10.1371/journal.pone.0064343

Wu, Y., Waite, L. L., Jackson, A. U., Sheu, W. H.-H., Buyske, S., Absher, D., et al. (2013). Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet.* 9:e1003379. doi: 10.1371/journal.pgen.1003379

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor SO-A declared a shared affiliation, though no other collaboration, with one of the authors LA-E.

Copyright © 2020 Agongo, Amenga-Etego, Nonterah, Debpuur, Choudhury, Bentley, Oduro, Rotimi, Crowther and Ramsay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.