

ORIGINAL ARTICLE

A common deletion in the haptoglobin gene associated with blood cholesterol levels among Chinese women

Neil S Zheng^{1,2}, Lisa A Bastarache¹, Julie A Bastarache³, Yingchang Lu⁴, Lorraine B Ware^{3,5}, Xiao-Ou Shu⁴, Joshua C Denny¹ and Jirong Long⁴

Haptoglobin (HP) protein plays a critical role in binding and removing free hemoglobin from blood. A deletion in the *HP* gene affects the protein structure and function. A recent study developed a novel method to impute this variant and discovered significant association of this variant with low-density lipoprotein (LDL) and total cholesterol levels among European descendants. In the present study, we investigated this variant among 3608 Chinese women. Consistent with findings from Europeans, we found significant associations between the deletion with lower cholesterol levels; women homozygous for the deletion allele (*HP1-HP1*), had a lower level of total cholesterol (-4.24 mg dl^{-1} , $P = 0.02$) and LDL cholesterol (-3.43 mg dl^{-1} , $P = 0.03$) than those not carrying the deletion allele (*HP2-HP2*). Especially, women carrying the *HP1S-HP1S*, had an even lower level of total cholesterol (-5.59 mg dl^{-1} , $P = 7.0 \times 10^{-3}$) and LDL cholesterol (-4.68 mg dl^{-1} , $P = 8.0 \times 10^{-3}$) compared to those carrying *HP2-HP2*. These associations remained significant after an adjustment for an established cholesterol level-related variant, rs2000999. Our study extends the previous findings regarding the association of *HP* structure variant with blood cholesterol levels to East Asians and affirms the validity of the new methodology for assessing *HP* structure variation. *Journal of Human Genetics* (2017) 62, 911–914; doi:10.1038/jhgc.2017.66; published online 29 June 2017

INTRODUCTION

High levels of circulating total cholesterol and low-density lipoprotein (LDL) cholesterol in the blood are major causes of cardiovascular diseases.¹ Blood cholesterol levels are determined by both genetic and lifestyle factors. Over the past decade, genome-wide association studies (GWAS) have identified more than 150 genetic loci associated with blood cholesterol levels, explaining ~13–15% of the variation of these traits.² However, a large majority of the previous studies on genetic contribution to cholesterol levels considered only single nucleotide polymorphisms (SNPs), which can miss the influence of structural variation in the human genome on cholesterol levels.²

Haptoglobin is a highly abundant protein in the blood that binds free hemoglobin to prevent iron loss and oxidative tissue damage during hemolysis.^{3,4} A common deletion of 1.7 kb (hg19, chr16:72,090,310–72,092,029) was identified in exons 3 and 4 of the *HP* gene, resulting in a variant allele, *HP1*, which encodes a protein product that forms a dimer in circulation. Individuals who are homozygous for the *HP2* allele (without the deletion) have haptoglobin multimers in circulation.^{4,5} Each of the *HP1* and *HP2* alleles can be further differentiated by nucleotide polymorphisms that cause haptoglobin to run faster or slower during gel electrophoresis,^{6,7} referred to as 'F' and 'S', respectively.⁵ Because of the complex genetic structure of the *HP* gene, it has been challenging to investigate the *HP* genetic polymorphisms, particularly the copy

number variant (CNV) defining the *HP1* and *HP2* alleles, in relation to human traits.

Recently, Boettger *et al.*⁵ reported a new method to analyze haptoglobin polymorphisms in a European-ancestry population by creating a reference panel to impute the *HP1/HP2* CNV. They also found the *HP1* allele was associated with lower LDL and total cholesterol levels. To evaluate the generalizability of this method and finding, we investigated the associations of *HP* gene polymorphisms with blood cholesterol levels in Chinese women.

MATERIALS AND METHODS

Study populations and genetic data

Subjects for the current analysis were control participants in the Shanghai Breast Cancer Study (SBCS), a population-based case-control study, and participants of the Shanghai Women's Health Study (SWHS), a population-based cohort study, both conducted in urban Shanghai. The SBCS is described in detail elsewhere.^{8,9} Control subjects were recruited from permanent residents of Shanghai and a peripheral blood sample was obtained from 74% participants. The age range of participants was 20–70 years with an average age of 49.9 years. The SWHS is an ongoing study of ~75 000 women recruited between 1997 and 2000, between 40 and 70 years of age at recruitment and a permanent resident of one of seven communities in urban Shanghai. In-person interviews, anthropometrics and biological sample collection were carried out by trained interviewers. Approximately 75% of study participants provided a blood sample to the study. Details of the SWHS design and study

¹Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; ²Yale College, Yale University, New Haven, CT, USA; ³Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; ⁴Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN, USA and ⁵Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA
Correspondence: Dr J Long, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville 37203, TN, USA.
E-mail: jirong.long@vanderbilt.edu

Received 22 December 2016; revised 28 April 2017; accepted 20 May 2017; published online 29 June 2017

Table 1 Selected characteristics of study subjects, Shanghai Breast Cancer Study and the Shanghai Women's Health Study

Characteristics ^a	All (N = 3608)	SBCS (N = 1875)	SWHS (N = 1733)
<i>Demographic factors</i>			
Age	53.2 ± 9.6	49.9 ± 8.7	56.7 ± 9.1
BMI	23.9 ± 3.5	23.3 ± 3.4	24.6 ± 3.5
Education (%)			
<High school	59.4	55.2	64.0
High school	30.0	35.0	24.5
>High school	10.6	9.8	11.5
<i>Blood lipids (mg dl⁻¹)</i>			
Total cholesterol	183.0 ± 40.1	179.0 ± 39.8	187.2 ± 40.0
LDL cholesterol	102.0 ± 32.0	100.3 ± 31.6	103.8 ± 32.4
HDL cholesterol	50.6 ± 12.8	51.0 ± 12.6	50.2 ± 13.0
Triglycerides	163.6 ± 127.0	150.0 ± 128.5	178.2 ± 123.8

Abbreviations: BMI, Body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SBCS, Shanghai Breast Cancer Study; SWHS, Shanghai Women's Health Study.
^aUnless otherwise specified, mean ± s.d. are presented.

implementation have been described elsewhere.¹⁰ Participants of these studies provided written informed consent, and the Institutional Review Boards of all participating institutions approved the study protocols.

In both SBCS and SWHS, peripheral blood samples were collected from study participants during the in-person interview into EDTA-containing BD Vacutainer tubes, which were temperature controlled during transportation by blue ice packs and processed within 6 h of the blood draw. The samples were stored at -80 °C before biomarker assessment. The lipid profiles were measured using an ACE Clinical Chemistry System following the manufacture's protocol. For subjects whose triglyceride levels were under 400 mg dl⁻¹, LDL cholesterol levels were calculated using the Friedwald equation, LDL = TC - HDL - (TG/5); for all others, the LDL levels were directly measured. Genomic DNA extracted from buffy coats was genotyped in previous GWASs using primarily the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6.0) or Illumina Human660W BeadChip. Details of the GWAS methodology, including genotyping protocol, filtering and data cleaning procedures have been previously described.^{9,11,12} Included in this study were 3,608 participants from the SBCS (N = 1875) and SWHS (N = 1733) whose lipid and genetic data were generated in previous studies. Among the 3608 subjects, we have fasting information for 2433 participants, with 44.55% samples being fasting and 55.45% non-fasting. Selected characteristics of the study population are shown in Table 1.

Haptoglobin CNV imputation

The *HP* genetic structural variations were imputed within each GWAS dataset using Beagle (v.3.2.2), which uses localized haplotype cluster models for efficient and accurate imputation.¹³ A phased reference panel developed by Boettger *et al* was used to impute the *HP* alleles.^{14,15} In brief, the phased reference panel was constructed using the *HP* structural data and SNP data for the subjects included in the 1000 Genomes Project. The *HP* structural data were generated based on droplet digital PCR. The SNP data were generated using the Omni 2.5 SNP array. SNP located in ~2 Mb region (hg19, chr16:71,088,193-73,097,663), flanking 1 Mb of the *HP* CNV (hg19, chr16:72,090,310-72,092,029) were used to build the reference. SNPs within the *HP* CNV region were excluded. Two reference panels were built, the European and African panels. For the European reference panel, data for 1277 genetic markers in 548 individuals from the CEU, TSI and IBS populations were included. For the African reference panel, data for 1276 genetic markers in 198 individuals from the YRI population were included. The genetic structure of East Asians is more similar with Europeans than with Africans. Therefore, in the present study, we used the European reference panel to impute the data. From our GWAS data sets, we extracted genotype data for SNPs located within a 1 Mb flanking region of the CNV (hg19, chr16:71,070,878-73,097,663), excluding those within the *HP* CNV region. The *HP* CNV alleles (*HP1* or *HP2*) were then imputed to the reference panel using Beagle (v.3.2.2)¹³ within each data set. Alleles *HP1S* and *HP1F* within the *HP1*, and *HP2FS* and

Table 2 *HP* gene allele and genotype frequency distribution in European descendants and Chinese

<i>HP</i> genotypes and alleles	European descendants ^a (%)	Chinese descendants (%)
<i>HP</i> genotypes		
<i>HP1-HP1</i>		
<i>HP1F-HP1F</i>	0.84	0.19
<i>HP1F-HP1S</i>	5.88	2.30
<i>HP1S-HP1S</i>	5.04	9.95
Sub-total	11.76	12.44
<i>HP1-HP2</i>		
<i>HP1F-HP2FS</i>	20.17	5.43
<i>HP1S-HP2FS</i>	27.73	39.08
<i>HP1S-HP2SS</i>	1.68	0.00
Sub-total	49.58	44.51
<i>HP2-HP2</i>		
<i>HP2FS-HP2FS</i>	34.45	43.02
<i>HP2FS-HP2SS</i>	4.20	0.03
Sub-total	38.65	43.05
<i>HP</i> alleles		
<i>HP1</i>		
<i>HP1F</i>	13.87	4.06
<i>HP1S</i>	22.69	30.64
Sub-total	36.56	34.70
<i>HP2</i>		
<i>HP2FS</i>	60.50	65.29
<i>HP2SS</i>	2.94	0.02
Sub-total	63.44	65.31

^aDerived from Supplementary Table 10 in Boettger *et al*.⁵

HP2SS within the *HP2*, were also imputed by running Beagle.⁵ Haplotypes surrounding the *HP*, including *HP1A*hap (for *HP1S*), *hapHP1B* (for *HP1F*), *HP2A*hap (for *HP2FS*), and *HP2B*hap (*HP2SS*) were extracted from the phased data in Beagle as described in Boettger *et al*. (2016).⁵ On basis of these data, we derived the *HP* genotypes, that is, *HP1-HP1* (*HP1F-HP1F*, *HP1F-HP1S*, *HP1S-HP1S*), *HP1-HP2* (*HP1F-HP2FS*, *HP1S-HP2FS*, *HP1S-HP2SS*) and *HP2-HP2* (*HP2FS-HP2FS* and *HP2FS-HP2SS*). Participants of from the SBCS and SWHS have a similar genotype distribution.

Statistical analyses

Using geometric mean differences and *P* values derived from multiple linear regression models, we assessed associations between *HP* polymorphisms and blood cholesterol levels, including total cholesterol, LDL cholesterol, high-density lipoprotein (HDL) cholesterol and triglycerides. The linear regression models were run on log-transformed blood lipid data and adjusted for age and data source.

RESULTS

The imputation quality in our study was high for all of the *HP* alleles for both data sets (*r*² ranged from 0.84 to 0.92; Supplementary Table 1) except for the *HP2SS* allele, which has a near zero frequency in our study population (Table 2). Haplotypes were composed of 19 SNPs flanking 20 kb of the deletion and three major haplotypes, A, B and C, were observed (Supplementary Table 2). The *HP1F* allele was primarily observed on haplotype B background, while *HP1S* allele was observed almost exclusively on haplotype A background. The *HP2FS* allele was observed on the backgrounds of both haplotypes A and B. When comparing the genotype frequencies of populations of European descent to Chinese populations, the frequencies for *HP1-HP1*,

Table 3 Association of *HP* genotypes with blood lipid levels in Chinese

Blood lipids	No. of subjects	Geometric mean ^a (mg dl ⁻¹)	Difference (mg dl ⁻¹)	P ^b
<i>Total Cholesterol</i>				
<i>HP2-HP2</i>	1553	179.81 (177.93–181.71)	Reference	
<i>HP1-HP2</i>	1606	179.15 (177.37–180.94)	-1.27	0.31
<i>HP1-HP1</i>	449	175.36 (171.85–178.95)	-4.24	0.02
<i>HP1S-HP1S</i>	359	173.97 (170.25–177.78)	-5.59	0.007
<i>LDL Cholesterol</i>				
<i>HP2-HP2</i>	1552	97.38 (95.82–98.97)	Reference	
<i>HP1-HP2</i>	1605	97.83 (96.37–99.31)	0.11	0.92
<i>HP1-HP1</i>	449	93.83 (90.93–96.82)	-3.43	0.03
<i>HP1S-HP1S</i>	359	92.55 (89.37–95.85)	-4.68	0.008
<i>HDL Cholesterol</i>				
<i>HP2-HP2</i>	1552	48.95 (48.35–49.56)	Reference	
<i>HP1-HP2</i>	1606	49.31 (48.71–49.90)	0.38	0.38
<i>HP1-HP1</i>	449	48.53 (47.35–49.75)	-0.49	0.45
<i>HP1S-HP1S</i>	359	48.13 (46.85–49.45)	-0.87	0.23
<i>Triglycerides</i>				
<i>HP2-HP2</i>	1552	137.24 (133.08–141.54)	Reference	
<i>HP1-HP2</i>	1606	132.09 (128.45–135.84)	-6.51	0.01
<i>HP1-HP1</i>	449	134.19 (126.69–142.14)	-2.18	0.59
<i>HP1S-HP1S</i>	359	134.94 (126.51–143.92)	-1.62	0.72

^aGeometric mean and 95% CI.

^bDerived from linear regression models using log-transformed data with adjustments for age and data sources, compared to the *HP2-HP2* genotype.

HP1-HP2 and *HP2-HP2* genotypes were similar (Table 2). However, the *HP1S* allele is much more common in the Chinese population (30.64%) than in the European population (22.69%) and the *HP1F* allele is less common in Chinese (4.06%) than in Europeans (13.87%) (Table 2). The *HP2SS* allele frequency is 2.94% of the European population and 0.02% in the Chinese population (Table 2). In the present study, both the *HP1F* and *HP1S* alleles are 100% linked to the G allele of rs2000999, a SNP identified in a previous GWAS to be associated with cholesterol levels, while the *HP2FS* allele was observed on the background of both alleles of the SNP rs2000999, with the frequencies being 43.02% for the *HP2FS-G* haplotype and 22.27% for the *HP2FS-A* haplotype (Supplementary Table 2).

Given the small sample size for some of the allele sub-groups, the association analysis was performed primarily for *HP1-HP1* and *HP1-HP2* genotypes compared with *HP2-HP2*, the most common genotype (freq. = 43.05%; Table 2).

The *HP1-HP1* genotype was significantly associated with lower total cholesterol and LDL cholesterol levels. Compared to women with the *HP2-HP2* genotype, individuals homozygous for the *HP1* allele, on average, had a 4.24 mg dl⁻¹ lower level of total cholesterol ($P=0.02$) and a 3.43 mg dl⁻¹ lower level of LDL cholesterol ($P=0.03$) (Table 3). Furthermore, the *HP1S-HP1S* genotype, the most common *HP1-HP1* genotype (Table 2), was more strongly associated with a reduced level of total cholesterol and LDL cholesterol, with a mean difference of -5.59 mg dl⁻¹ for total cholesterol ($P=7.0 \times 10^{-3}$) and -4.68 mg dl⁻¹ for LDL cholesterol ($P=8.0 \times 10^{-3}$) when compared to the *HP2-HP2* genotype (Table 3).

We also found that the G allele of SNP rs2000999 was associated with decreased cholesterol, with a mean difference of -2.25 mg dl⁻¹ for total cholesterol ($P=0.02$) and -1.60 mg dl⁻¹ for LDL cholesterol ($P=0.06$) per G allele compared to the A/A genotype (Supplementary Table 3). These associations were no longer statistically significant after

adjusting for the *HP1/2* variant, while the associations for the *HP1S-HP1S* genotype remained statistically significant for total cholesterol ($P=0.04$) and marginally significant for LDL cholesterol ($P=0.06$) after adjusting for rs2000999.

A significant association of the *HP1-HP2* genotype with lower triglyceride levels was also observed; women carrying the *HP1-HP2* genotype had a -6.51 mg dl⁻¹ lower triglyceride level than those homozygous for the *HP2* allele ($P=0.01$; Table 3). We did not find any significant association of *HP* alleles with HDL cholesterol levels.

DISCUSSION

In this study, we found that the *HP1-HP1* genotype, especially *HP1S-HP1S*, was significantly associated with lower total and LDL cholesterol levels in Chinese women. This result is consistent with the finding of Boettger *et al.*⁵, showing an inverse association of the *HP1* allele with total and LDL cholesterol in individuals of European ancestry. Our study provides additional evidence supporting an important role of haptoglobin in regulating blood cholesterol levels. Furthermore, it supports the validity of using SNP haplotype imputation to analyze *HP* structure variations.

The *HP* deletion is in low linkage disequilibrium with rs2000999, a cholesterol level associated SNP identified in a previous GWAS with $r^2=0.15$. However, the cholesterol lowering allele *HP1* is 100% linked to the G allele of rs2000999. Because a high proportion of the cholesterol increasing allele *HP2* is also linked to the cholesterol decreasing G allele of SNP rs2000999, SNP rs2000999 cannot capture entirely the association between *HP* alleles and blood cholesterol, which may explain our observation of a significant association of the *HP1S-HP1S* genotype with total and LDL cholesterol even after adjusting for SNP rs2000999.

While the most well-known function of haptoglobin is to bind free hemoglobin, it also interacts with apolipoprotein E (ApoE),^{5,16,17}

a plasma protein that participates in the removal of chylomicron and very LDL remnants in blood as a ligand for LDL receptors, LDL receptor-related protein 1 and cell surface heparan sulfate proteoglycans.^{18,19} Oxidation decreases the ability of ApoE to bind lipoproteins, impairing its ability to clear plasma lipids from the blood.²⁰ Moreover, oxidation also induces a conformational change in ApoE, leading ApoE to become unstable and prone to deposit on blood vessel walls.²⁰ When haptoglobin binds to ApoE, it protects ApoE from oxidative damage.¹⁶ Proteins encoded by the *HP1* allele have greater antioxidant capacity than *HP2*-encoded proteins,²¹ providing greater antioxidant protection when bound to ApoE proteins, potentially allowing ApoE proteins to more efficiently remove lipoproteins in blood.

We also observed a significant association of *HP1-HP2* genotypes with lower triglyceride levels in this study. This was consistent with a previous report of a reduced triglyceride level in *HP1* allele carriers compared to *HP2* carriers.⁵ However, the long tail in triglyceride density distribution shown in Supplementary Figure 1 indicates there were outliers with much higher triglyceride levels that may have affected the analysis of triglyceride levels with the *HP* CNV. The absence of a similar association with *HP1-HP1* genotypes also suggests that the *HP1-HP2* genotype association observed in our study might represent a false association due to chances. Thus this finding requires further confirmation.

There are some limitations in the present study. First, due to the low frequency of *HP1F* and *HP2SS*, we were unable to derive stable estimations for genotypes *HP1F-HP1F* and *HP2SS-HP2SS*. Therefore, we could not perform additional association analyses for these genotypes. It would be ideal to technically validate the imputed genotypes by the ddPCR technique. However, we were constrained by a lack of proper equipment and a limited budget. Nevertheless, any imputation error, if exists, is most likely to be differential and would lead an underestimation of the true effect. Second, only about half of the plasma samples included in our study are fasting samples, and other are non-fasting samples. The latter could have compromised our ability to evaluate the association of *HP* polymorphisms with circulating triglyceride levels. In addition, the effect sizes observed in the present study are small and a large sample size is needed to robustly validate these results. Finally, information on the use of lipid lowering medications, especially statins, was unavailable in our study. However, our participants were recruited between 1997 and 2000 when the use of statins and other lipid lowering medications was uncommon in China. Inability to adjust for use of lipid lowering medication in our analysis could reduce the statistical power of the study.

In summary, through SNP haplotype imputation, we extended the findings of associations of *HP1/2* CNV with total and LDL cholesterol levels reported previously in European descendants to East Asian populations, providing further support of a significant role of haptoglobin in blood cholesterol levels. Our study also extends the validity of the novel *HP* imputation method developed by Boettger *et al.* to individuals of Chinese ancestry, which could be used in future studies to use densely genotyped data to efficiently study potential associations of haptoglobin polymorphisms with complex human phenotypes in populations of Asian ancestry.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The Shanghai Breast Cancer Study is supported by a grant from the National Institutes of Health, R01CA064277; the Shanghai Women's Health Study is supported by the National Institutes of Health grants R37CA070867 and UM1CA182910. Additional support includes HL126671 to JAB. We would like to thank Dr. Boettger for providing the reference panel for SNP haplotype imputation of *HP* polymorphisms.

- Prospective Studies Collaboration, Lewington, S., Whitlock, G., Clarke, R., Sherliker, P., Emberson, J. *et al.* Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet Lond. Engl.* **370**, 1829–1839 (2007).
- Global Lipids Genetics Consortium, Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
- Allison, A. C. & Rees, W. A. The binding of haemoglobin by plasma proteins (haptoglobins); its bearing on the renal threshold for haemoglobin and the aetiology of haemoglobinuria. *Br. Med. J.* **2**, 1137–1143 (1957).
- Langlois, M. R. & Delange, J. R. Biological and clinical significance of haptoglobin polymorphism in humans. *Clin. Chem.* **42**, 1589–1600 (1996).
- Boettger, L. M., Salem, R. M., Handsaker, R. E., Peloso, G. M., Kathiresan, S., Hirschhorn, J. N. *et al.* Recurring exon deletions in the *HP* (haptoglobin) gene contribute to lower blood cholesterol levels. *Nat. Genet.* **48**, 359–366 (2016).
- Smithies, O. Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. *Biochem. J.* **61**, 629–641 (1955).
- Smithies, O., Connell, G. E. & Dixon, G. H. Inheritance of haptoglobin subtypes. *Am. J. Hum. Genet.* **14**, 14–21 (1962).
- Gao, Y. T., Shu, X. O., Dai, Q., Potter, J. D., Brinton, L. A., Wen, W. *et al.* Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. *Int. J. Cancer* **87**, 295–300 (2000).
- Zheng, W., Long, J., Gao, Y.-T., Li, C., Zheng, Y., Xiang, Y.-B. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324–328 (2009).
- Zheng, W., Chow, W.-H., Yang, G., Jin, F., Rothman, N., Blair, A. *et al.* The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am. J. Epidemiol.* **162**, 1123–1131 (2005).
- Cai, Q., Zhang, B., Sung, H., Low, S.-K., Kweon, S.-S., Lu, W. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat. Genet.* **46**, 886–890 (2014).
- Zeng, C., Matsuda, K., Jia, W.-H., Chang, J., Kweon, S.-S., Xiang, Y.-B. *et al.* Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology* **150**, 1633–1645 (2016).
- Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
- The 1000 Genome Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
- Salvatore, A., Cigliano, L., Carlucci, A., Buccì, E. M. & Abrescia, P. Haptoglobin binds apolipoprotein E and influences cholesterol esterification in the cerebrospinal fluid. *J. Neurochem.* **110**, 255–263 (2009).
- Cigliano, L., Pugliese, C. R., Spagnuolo, M. S., Palumbo, R. & Abrescia, P. Haptoglobin binds the antiatherogenic protein apolipoprotein E—impairment of apolipoprotein E stimulation of both lecithin:cholesterol acyltransferase activity and cholesterol uptake by hepatocytes. *FEBS J.* **276**, 6158–6171 (2009).
- Mahley, R. W. & Apolipoprotein, E. Cholesterol transport protein with expanding role in cell biology. *Science* **240**, 622–630 (1988).
- Ji, Z. S., Brecht, W. J., Miranda, R. D., Hussain, M. M., Innerarity, T. L. & Mahley, R. W. Role of heparan sulfate proteoglycans in the binding and uptake of apolipoprotein E-enriched remnant lipoproteins by cultured cells. *J. Biol. Chem.* **268**, 10160–10167 (1993).
- Yang, Y., Cao, Z., Tian, L., Garvey, W. T. & Cheng, G. VPO1 mediates ApoE oxidation and impairs the clearance of plasma lipids. *PLoS ONE* **8**, e57571 (2013).
- Melamed-Frank, M., Lache, O., Enav, B. I., Szafrane, T., Levy, N. S., Ricklis, R. M. *et al.* Structure-function analysis of the antioxidant properties of haptoglobin. *Blood* **98**, 3693–3698 (2001).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)