McLachlan, S; Giambartolomei, C; White, J; Charoen, P; Wong, A; Finan, C; Engmann, J; Shah, T; Hersch, M; Podmore, C; Cavadino, A; Jefferis, BJ; Dale, CE; Hypponen, E; Morris, RW; Casas, JP; Kumari, M; Ben-Shlomo, Y; Gaunt, TR; Drenos, F; Langenberg, C; Kuh, D; Kivimaki, M; Rueedi, R; Waeber, G; Hingorani, AD; Price, JF; Walker, AP; UCLEB Consortium, (2016) Replication and Characterization of Association between ABO SNPs and Red Blood Cell Traits by Meta-Analysis in Europeans. PloS one, 11 (6). e0156914. ISSN 1932-6203 DOI: https://doi.org/10.1371/journal.pone.0156914

Downloaded from: http://researchonline.lshtm.ac.uk/2550797/

DOI: 10.1371/journal.pone.0156914

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
RESEARCH ARTICLE

Replication and Characterization of Association between ABO SNPs and Red Blood Cell Traits by Meta-Analysis in Europeans

Stela McLachlan1*, Claudia Giambartolomei2, Jon White3, Pimphen Charoen4,5, Andrew Wong6, Chris Finan7, Jorgen Engmann8,9, Tina Shah6, Micha Hersch10,11, Clara Podmore12, Alana Cavadinio1,3,14, Barbara J. Jefferis15, Caroline E. Dale9, Elina Hypponen14,16,17, Richard W. Morris15,18, Juan P. Casas19, Meena Kumari20,21, Yoav Ben-Shlomo18, Tom R. Gaunt22, Fotios Drenos22,23, Claudia Langenberg12,19, Diana Kuh6, Mika Kiwimaki21, Rico Ruedei10,11, Gerard Waever24, Aroon D. Hingorani7,9, Jacqueline F. Price1, Ann P. Walker23, UCLEB Consortium¶

1 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom, 2 Department of Psychiatry, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, The Leon and Norma Hess Center for Science and Medicine, New York, New York, United States of America, 3 University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, United Kingdom, 4 Department of Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom, 5 Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, 6 MRC Unit for Lifelong Health and Ageing at UCL, London, United Kingdom, 7 Institute of Cardiovascular Science, University College London, London, United Kingdom, 8 Genetic Epidemiology Group, Institute of Cardiovascular Science, University College London, London, United Kingdom, 9 Farr Institute of Health Informatics Research, Department of Epidemiology & Public Health, University College London, London, United Kingdom, 10 Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland, 11 Swiss Institute of Bioinformatics, Lausanne, Switzerland, 12 MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, United Kingdom, 13 Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, United Kingdom, 14 Population, Policy and Practice, UCL Institute of Child Health, University College London, London, United Kingdom, 15 Department of Primary Care & Population Health, UCL Institute of Epidemiology & Health Care, University College London, London, United Kingdom, 16 Centre for Population Health Research, School of Health Sciences and Sansom Institute of Health Research, University of South Australia, Adelaide, Australia, 17 South Australian Health and Medical Research Institute, Adelaide, Australia, 18 School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, 19 Institute of Health Informatics, University College London, London, United Kingdom, 20 Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom, 21 Department of Epidemiology & Public Health, UCL Institute of Epidemiology & Health Care, University College London, London, United Kingdom, 22 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, 23 Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, London, United Kingdom, 24 Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

¶ Membership of the UCLEB Consortium is provided in the Acknowledgments.

* stela.mclachlan@ed.ac.uk

Abstract

Red blood cell (RBC) traits are routinely measured in clinical practice as important markers of health. Deviations from the physiological ranges are usually a sign of disease, although...
collaborative projects and data sharing (http://www.ucl.ac.uk/pchh/research-groups-themes/bhrs-pub/).
For general data sharing enquiries, please contact Lucy Lennon (l.leennon@ucl.ac.uk). BWHHS: All
BWHHS data collected is held by the research team based at London School of Hygiene and
Tropical Medicine, for ongoing analysis. If you would like to collaborate with the BWHHS team,
contact the study coordinator, Antoinette Amuzu (antoinette.amuzu@lshtm.ac.uk) Data and biological
samples provided to the collaborators can only be used for the purposes originally stated and must not
be used in any other way without re-application to the BWHHS team. No data should be passed on to any
third party unless they were specified in the original application. CaPS: Data used for the Caerphilly
Prospective study (CaPS) was made available by the CaPS access committee. More information about its
managed access procedure is available on the study website (http://www.brhs.ac.uk/social-community-
medicine/projects/caerphilly/colaboration/). CoLaus: Data from the CoLaus/PsyCoLaus study can be
requested according to the procedure described on the CoLaus website (http://www.colaus.ch/en/cls_
home/cls_pro_home/cls-research-3.htm). ELSA: ELSA data are made available through the ESDS
website (http://www.elsa-project.ac.uk/availableData). EAS and ET2DS: Edinburgh Artery Study and
Edinburgh Type 2 Diabetes Study data are available to researchers upon request, subject to approval by the
data sharing committee. Data request forms can be obtained by emailing Stela McLachlan
(stela.mclachlan@ed.ac.uk). MRC NSHD: The NSHD data are made available to researchers who submit
data requests (mrcna.swiffinfo@ucl.ac.uk). More information is available in the full policy documents
(http://www.nshd.mrc.ac.uk/data.aspx). Managed access is in place for this study to ensure that use of the
data are within the bounds of consent given previously by participants, and to safeguard any
potential threat to anonymity since the participants are all born in the same week. Whitehall II: Data from the
Whitehall II study are made publicly available as described in the Whitehall II data sharing policy
(https://www.ucl.ac.uk/whitehallII/data-sharing).

Funding: UCLEB: This work was supported by a British Heart Foundation Programme Grant (RG/10/
1228456). Individual studies were supported as follows. 1958BC: DNA collection of the 1958 birth
cohort was funded by the UK Medical Research Council (G0000934); and the Wellcome Trust (Grant
068545/Z/02). Genotyping was supported by a contract from the European Commission Framework
Programme 6 (018996); and grants from the French Ministry of Research. BRHS: The authors
acknowledge the British Regional Heart Study team for data collection. The British Regional Heart study is
variation between healthy individuals also occurs, at least partly due to genetic factors.
Recent large scale genetic studies identified loci associated with one or more of these traits;
further characterization of known loci and identification of new loci is necessary to better
understand their role in health and disease and to identify potential molecular mechanisms.
We performed meta-analysis of Metabochip association results for six RBC traits—hemoglo-
bin concentration (Hb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean
corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV) and red
cell count (RCC)—in 11 093 Europeans from seven studies of the UCL-LSHTM-Edin-
burgh-Bristol (UCLEB) Consortium. We identified 394 non-overlapping SNPs in five loci at
gene-wide significance: 6p22.1-6p21.33 (with HFE among others), 6q23.2 (with HBS1L
among others), 6q23.3 (contains no genes), 9q34.3 (only ABO gene) and 22q13.1 (with
TMPRSS6 among others), replicating previous findings of association with RBC traits at these loci and extending them by imputation to 1000 Genomes. We further characterized
associations between ABO SNPs and three traits: hemoglobin, hematocrit and red blood
cell count, replicating them in an independent cohort. Conditional analyses indicated the
independent association of each of these traits with ABO SNPs and a role for blood group O
in mediating the association. The 15 most significant RBC-associated ABO SNPs were also
associated with five cardiometabolic traits, with discordance in the direction of effect
between groups of traits, suggesting that ABO may act through more than one mechanism
to influence cardiometabolic risk.

Introduction
Red blood cell (RBC) traits are routinely measured in clinical practice. These traits, hemoglobin
concentration (Hb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean
corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV) and red
cell count (RCC) are tightly regulated so that deviations from the physiological ranges are usually a
sign of disease, such as hematological, infectious and immune disorders and cancer. Therefore,
they serve as markers not only of specific hematological conditions but also of the general
health of an individual. However, physiological values for the traits also vary between healthy
individuals and between different ethnic groups, at least in part due to genetic factors [1–13].

Heritability varies between the different RBC traits, with estimates ranging from 40–80% for
Hb and RCC, although values as low as 14% for MCH and as high as 96% for MCV have been
reported [14–17]. Only a few large scale genetic association studies of RBC traits have been per-
fomed to date, the latest one identifying 75 loci potentially influencing one or more of these
traits [13]. The largest studies were of Caucasian/European cohorts, followed by studies of Afri-
can Americans and Asians, with several of the RBC-associated loci consistently identified
across different ethnic groups. The recent use of large scale meta-analyses has accelerated the
detection of loci and has emphasized the need for additional studies to identify still undetected
contributors to the underlying genetic basis of these traits [1–13]; reviewed in [18, 19]. Further
characterization of known loci is equally important in order to better understand their roles in
erthropoiesis, susceptibility to and development of different disorders and their impact on the clinical
characteristics and prognosis in those disorders [18–20].

In this study, RBC traits were associated with five loci, including 11 new SNPs for ABO (the
gene encoding ABO blood group), further refining these associations. Conditional analyses
supported by a British Heart Foundation grants (RG/08/03/25942, RG/13/16/30528). The British Heart Foundation had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. BWHHS: The British Women’s Heart and Health Study is supported by funding from the British Heart Foundation; and the Department of Health Policy Research Programme (England), CaPS: The Caerphilly Prospective study was conducted by the former MRC Epidemiology Unit (South Wales) and funded by the Medical Research Council of the United Kingdom. The Department of Social and Community Medicine, University of Bristol now maintains the archive. EAS: The Edinburgh Artery Study is funded by the British Heart Foundation (Programme Grant RG/98002), with Metabochip genotyping funded by a project grant from the Chief Scientist Office of Scotland (Project Grant CB84/672). ELSA: Samples from the English Longitudinal Study of Ageing DNA Repository (EDNAR), received support under a grant (AS/176440S1) awarded by the National Institute on Ageing (NIA). ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute of Fiscal Studies. The data were collected by the National Centre for Social Research. ET2DS: The Edinburgh Type 2 Diabetes Study is funded by the Medical Research Council (Project Grant G0500877); the Chief Scientist Office of Scotland (Programme Support Grant CZQ1/138); Pfizer plc (Unrestricted Investigator Led Grant); and Diabetes UK (Clinical Research Fellowship 10/0003895). Research clinics were held at the Wellcome Trust Clinical Research Facility and Princess Alexandra Eye Pavilion in Edinburgh. MRC NHSD: The MRC National Survey of Health and Development 1946 is funded by the UK Medical Research Council (MC_UU_120191/1). WHII: The Whitehall II study is supported by grants from the Medical Research Council (K013351); British Heart Foundation (RG/07/008/23674); Stroke Association; National Heart Lung and Blood Institute (SRO1 HL036310); National Institute on Aging (SRO1AG13196); Agency for Health Care Policy Research (HS06516); and the John D. and Catherine T. MacArthur Foundation; and the National Institute on Aging (SRO1AG13196). The authors also express their gratitude to the investigators who have contributed to the point to the role of ABO SNPs predicting blood group O in the association with Hb, Hct and RCC and to the independent association of all three traits with the ABO locus. ABO is a highly pleiotropic locus; direct comparison of results for the 15 most significant RBC-associated ABO SNPs versus additional cardiometabolic traits showed that RCC (Hct, Hb, RCC) and liver (logALP) traits exhibited a negative direction of effect whereas some lipid (LDL, TC) and coagulation factor traits (logVWF, FVIII) had a positive effect, indicating the wide influence of the ABO locus upon cardiometabolic risk.

**Results**

**Meta-analysis in the UCLEB consortium**

The study characteristics are given in Table 1. The analysis was done on ~4.2 million genotyped and imputed SNPs. An identical analysis was performed for each study and trait, accounting for the covariates available for that specific study, before combining the results by meta-analysis. Genomic-control inflation factors (λ(G)) varied between 0.987 for Hb in MRC NHSD to 1.041 for RCC in BWHHS in individual studies; these were used to correct the test statistics in the meta-analysis (S2 Table). Manhattan plots for the meta-analysis of each trait are shown in S2 Fig. The corresponding Q-Q plots are shown in S3 Fig.

The 736 statistically significant single SNP-trait associations are listed in S3 Table. In total, there were 394 non-overlapping SNPs in five loci which reached the genome-wide significance threshold of $5 \times 10^{-8}$ used in this study (116 SNPs associated with Hb in three loci; 15 with Hct in one locus; 371 with MCH in three loci; six with MCHC in one locus; 188 with MCV in four loci; and 40 with RCC in two loci). 338 of these SNPs (86%) were significantly associated with more than one RBC trait. The five loci with SNPs reaching statistical significance were: 6p22.1-6p21.33 (locus includes the HFE gene) and 22q13.1 (includes the HBS1L-MYB region); 9q34.3 (includes the HBS1L gene) and 371 with MCH in three loci; six with MCHC in one locus; 188 with MCV in four loci; and 40 with RCC in two loci). 338 of these SNPs (86%) were significantly associated with more than one RBC trait. The five loci with SNPs reaching statistical significance were: 6p22.1-6p21.33 (locus includes HFE as a primary candidate gene), 6q23.3 (includes HBS1L gene among others), 6q23.3 (locus does not contain any genes), 9q34.3 (includes only the ABO gene) and 22q13.1 (includes the TMPPS6 gene among others). Overlap of the significant loci for all six RBC traits is shown in S4 Fig.

The 6p22.1-6p21.33 locus includes 181 genes over ~2.3 Mb and was significantly associated with Hb, MCH and MCV. The lead SNP (with the lowest P value) for all three traits was rs80215559 in SLC17A2 (P = 8.90×10^{-14}, P = 2.42×10^{-21} and P = 1.85×10^{-15}, respectively). However, the HFE C282Y functional variant has an extended founder haplotype, therefore, HFE is considered to be the primary candidate gene in this locus, previously associated with all three RBC traits [3–9, 12, 13].

MCH, MCV and RCC were significantly associated with SNPs in the 6q23.2 locus, which is approximately 272 Kb in size and contains two genes, one pseudogene and a microRNA (ALDH8A1, HBS1L, MEMO1P2 and MIR3662, respectively). Lead SNPs in this locus were rs34164109 for MCH (P = 4.52×10^{-17}), rs35959442 for MCV (P = 2.44×10^{-15}) and rs35786788 for RCC (P = 2.47×10^{-15}). One of the genes in this locus (HBS1L) has been repeatedly reported to be associated with all three traits as part of the HBS1L-MYB region [5–9, 11–13].

The signal at 6q23.3 had only two SNPs reaching the significance threshold, (rs592423 and rs590856) and the lead association for both was with MCV (P = 1.90×10^{-8} and P = 2.45×10^{-8}, respectively). Both SNPs were also associated with MCH and a further 23 SNPs were associated with MCV at the suggestive threshold of statistical significance. These findings confirm the previous association of this locus with MCH and MCV [6, 7, 9, 13]. There are no known genes in this locus, however the closely located CITED2 has been reported as a potential candidate gene.

The 9q34.3 locus contains only the ABO gene and is approximately 16 Kb in size. SNPs in this locus were associated with Hb, Hct and RCC. rs507666 was the lead SNP for Hb (P = 1.84×10^{-10}), rs8176643 for Hct (P = 9.64×10^{-12}) and rs8176685 for RCC (P = 1.98×10^{-8}).
SNPs in the 22q13.1 locus (~ 86 Kb in size) were associated with Hb, MCH, MCHC and MCV. This locus contains four genes, however, TMPRSS6 is a primary candidate gene and has been previously identified as associated with all four RBC traits [2, 3, 5–9, 11–13, 21]. The lead SNPs associated with these four traits in the present study were all in TMPRSS6: rs877908 for Hb (P = 1.27×10⁻⁸), rs855791 for MCH and MCV (P = 1.57×10⁻²⁴ and P = 5.40×10⁻¹⁶, respectively), and rs2413450 for MCHC (P = 2.15×10⁻⁸).

There was a strong positive correlation between hemoglobin, hematocrit and red blood cell count as well as between mean corpuscular hemoglobin and mean corpuscular volume (r > 0.75; S6 Table). All other RBC trait combinations show weak to moderate positive or negative correlation with each other (S6 Table).

Replication analysis of ABO locus in the CoLaus study

For replication, 47 SNPs were selected in/around the ABO gene. All of these were below the suggestive threshold for Hb, Hct, and RCC in the meta-analysis, out of which 15 were below the significance threshold. The CoLaus study had 2848 participants for whom both genotypes and phenotypes were available. Using a nominal significance cut-off of P < 0.05 for replication, we confirmed the association between four significant SNPs and Hb, and between 10 significant SNPs and Hct (Table 2). Two significant associations of SNPs with RCC were not replicated. Likewise, associations between five SNPs (rs600038, rs651007, rs579459, rs649129 and rs495828) and Hb and Hct were not replicated. However, CoLaus was previously part of a

Table 1. Study characteristics.

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>F, %</th>
<th>Age, year</th>
<th>T2D, %</th>
<th>Ever smoked, %</th>
<th>eGFR, mL/ min/1.73m²</th>
<th>Hb, g/dL</th>
<th>Hct, %</th>
<th>MCH, pg</th>
<th>MCHC, g/dL</th>
<th>MCV, fl</th>
<th>RCC, x10¹²/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRHS</td>
<td>2310</td>
<td>0</td>
<td>68.89 (5.62)</td>
<td>12</td>
<td>73</td>
<td>63.13 (11.15)</td>
<td>14.56 (1.21)</td>
<td>45.09 (3.49)</td>
<td>30.13 (1.89)</td>
<td>32.33 (1.16)</td>
<td>93.23 (5.36)</td>
<td>4.84 (0.41)</td>
</tr>
<tr>
<td>BWHHS</td>
<td>1884</td>
<td>100</td>
<td>70.73 (5.31)</td>
<td>15</td>
<td>47</td>
<td>66.81 (11.53)</td>
<td>13.45 (1.07)</td>
<td>41.70 (3.20)</td>
<td>29.51 (1.79)</td>
<td>32.30 (1.19)</td>
<td>91.40 (5.05)</td>
<td>4.57 (0.38)</td>
</tr>
<tr>
<td>CaPS</td>
<td>1268</td>
<td>0</td>
<td>56.78 (4.46)</td>
<td>12</td>
<td>80</td>
<td>69.32 (11.93)</td>
<td>14.94 (1.07)</td>
<td>44.65 (3.17)</td>
<td>30.88 (1.83)</td>
<td>33.50 (0.81)</td>
<td>92.16 (4.78)</td>
<td>4.85 (0.36)</td>
</tr>
<tr>
<td>ELSA</td>
<td>1881</td>
<td>47</td>
<td>73.69 (9.45)</td>
<td>14</td>
<td>67</td>
<td>63.12 (11.83)</td>
<td>14.25 (1.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ET2DS</td>
<td>996</td>
<td>49</td>
<td>67.89 (4.24)</td>
<td>100</td>
<td>61</td>
<td>70.42 (19.03)</td>
<td>14.06 (1.49)</td>
<td>40.62 (4.18)</td>
<td>31.31 (1.96)</td>
<td>34.61 (0.67)</td>
<td>90.43 (4.87)</td>
<td>4.50 (0.48)</td>
</tr>
<tr>
<td>MRC NSHD</td>
<td>1672</td>
<td>52</td>
<td>63.29 (1.11)</td>
<td>8a</td>
<td>70 a</td>
<td>92.96 (22.19)</td>
<td>14.21 (1.22)</td>
<td>41.87 (3.41)</td>
<td>30.63 (1.73)</td>
<td>34.11 (1.15)</td>
<td>90.26 (4.72)</td>
<td>4.65 (0.41)</td>
</tr>
<tr>
<td>WHII</td>
<td>1082</td>
<td>22</td>
<td>43.78 (5.87)</td>
<td></td>
<td></td>
<td></td>
<td>14.60 (1.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total N</td>
<td>11</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>11 093</td>
<td>8127</td>
<td>8125 b</td>
<td>8125 b</td>
<td>5817</td>
<td>8128</td>
</tr>
</tbody>
</table>

Data are mean (SD). The seven studies included in the analysis are: British Regional Heart Study (BRHS), British Women’s Heart and Health Study (BWHHS), Caerphilly Prospective Study (CaPS), English Longitudinal Study of Ageing (ELSA), Edinburgh Type 2 Diabetes Study (ET2DS), MRC National Survey of Health and Development (MRC NSHD) and the Whitehall II Study (WHII). Study characteristics are given for: number of participants with at least one red blood cell trait available (N), percentage of females in the study (F), participants’ mean age (Age), percentage of participants with confirmed type 2 diabetes (T2D), percentage of people who currently smoke or have previously smoked (Ever smoked), estimated glomerular filtration rate (eGFR), hemoglobin (Hb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV) and red blood cell count (RCC).

a Data from a different data collection wave conducted between six to nine years before.

b Calculated traits are shown in red; the MRC NSHD results shown as “b” include 419 participants for whom an observed MCHC value was not available, instead it was calculated.

doi:10.1371/journal.pone.0156914.l001
larger effort to identify loci associated with RBC traits in 71,861 individuals of European or South Asian ancestry [13]. Four out of five SNPs were present and highly significant in their analysis for Hb, Hct and RCC, respectively: rs651007 $P = 3.82 \times 10^{-14}$, $1.59 \times 10^{-11}$ and $1.61 \times 10^{-11}$; rs579459 $P = 1.35 \times 10^{-15}$, $7.59 \times 10^{-14}$ and $9.25 \times 10^{-18}$; rs649129 $P = 1.46 \times 10^{-15}$, $1.54 \times 10^{-13}$ and $9.77 \times 10^{-16}$; rs495828 $P = 1.59 \times 10^{-15}$, $1.48 \times 10^{-13}$ and $1.39 \times 10^{-15}$, giving confidence that the association between these SNPs and traits is genuine. For Hct and RCC, 29 and

<table>
<thead>
<tr>
<th>rs#</th>
<th>allele</th>
<th>chr</th>
<th>bp</th>
<th>trait</th>
<th>$\beta$</th>
<th>SE</th>
<th>$P$</th>
<th>$\beta$</th>
<th>SE</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8176685</td>
<td>d</td>
<td>9</td>
<td>136138765</td>
<td>Hb</td>
<td>-0.1328</td>
<td>0.0232</td>
<td>$1.08 \times 10^{-8}$</td>
<td>-0.1000</td>
<td>0.0543</td>
<td>$0.06547$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs149092047</td>
<td>d</td>
<td>9</td>
<td>136139907</td>
<td>Hb</td>
<td>-0.1008</td>
<td>0.0173</td>
<td>$5.29 \times 10^{-9}$</td>
<td>-0.0782</td>
<td>0.0365</td>
<td>$0.03198$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2519093</td>
<td>t</td>
<td>9</td>
<td>136141870</td>
<td>Hb</td>
<td>-0.1217</td>
<td>0.0193</td>
<td>$2.67 \times 10^{-10}$</td>
<td>-0.0754</td>
<td>0.0406</td>
<td>$0.06335$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs28850884</td>
<td>d</td>
<td>9</td>
<td>136145242</td>
<td>Hb</td>
<td>-0.1035</td>
<td>0.0178</td>
<td>$6.26 \times 10^{-9}$</td>
<td>-0.0963</td>
<td>0.0418</td>
<td>$0.02129$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9411378</td>
<td>a</td>
<td>9</td>
<td>13614525</td>
<td>Hb</td>
<td>-0.1033</td>
<td>0.0178</td>
<td>$6.47 \times 10^{-9}$</td>
<td>-0.0962</td>
<td>0.0417</td>
<td>$0.02120$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs550057</td>
<td>t</td>
<td>9</td>
<td>136146597</td>
<td>Hb</td>
<td>-0.0982</td>
<td>0.0172</td>
<td>$4.52 \times 10^{-9}$</td>
<td>-0.0793</td>
<td>0.0361</td>
<td>$0.02795$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs507666</td>
<td>a</td>
<td>9</td>
<td>136149399</td>
<td>Hb</td>
<td>-0.1228</td>
<td>0.0193</td>
<td>$1.84 \times 10^{-10}$</td>
<td>-0.0716</td>
<td>0.0399</td>
<td>$0.07251$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8176643</td>
<td>d</td>
<td>9</td>
<td>136149709</td>
<td>Hb</td>
<td>-0.1253</td>
<td>0.0199</td>
<td>$3.37 \times 10^{-10}$</td>
<td>-0.0802</td>
<td>0.0442</td>
<td>$0.06930$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs532436</td>
<td>a</td>
<td>9</td>
<td>136149830</td>
<td>Hb</td>
<td>-0.1219</td>
<td>0.0193</td>
<td>$2.44 \times 10^{-10}$</td>
<td>-0.0717</td>
<td>0.0399</td>
<td>$0.07242$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs600038</td>
<td>c</td>
<td>9</td>
<td>136151806</td>
<td>Hb</td>
<td>-0.1021</td>
<td>0.0185</td>
<td>$3.47 \times 10^{-8}$</td>
<td>-0.0320</td>
<td>0.0381</td>
<td>$0.40180$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs651007a</td>
<td>t</td>
<td>9</td>
<td>136153875</td>
<td>Hb</td>
<td>-0.1046</td>
<td>0.0185</td>
<td>$1.62 \times 10^{-8}$</td>
<td>-0.0329</td>
<td>0.0380</td>
<td>$0.38610$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs579459a</td>
<td>c</td>
<td>9</td>
<td>136154168</td>
<td>Hb</td>
<td>-0.1045</td>
<td>0.0185</td>
<td>$1.63 \times 10^{-8}$</td>
<td>-0.0326</td>
<td>0.0382</td>
<td>$0.39290$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs649129a</td>
<td>t</td>
<td>9</td>
<td>136154304</td>
<td>Hb</td>
<td>-0.1053</td>
<td>0.0185</td>
<td>$1.29 \times 10^{-8}$</td>
<td>-0.0324</td>
<td>0.0382</td>
<td>$0.39550$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs495828a</td>
<td>t</td>
<td>9</td>
<td>136154867</td>
<td>Hb</td>
<td>-0.1051</td>
<td>0.0185</td>
<td>$1.37 \times 10^{-8}$</td>
<td>-0.0324</td>
<td>0.0382</td>
<td>$0.39660$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs635634</td>
<td>t</td>
<td>9</td>
<td>136155000</td>
<td>Hb</td>
<td>-0.1205</td>
<td>0.0193</td>
<td>$3.86 \times 10^{-10}$</td>
<td>-0.0722</td>
<td>0.0399</td>
<td>$0.07080$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All SNPs were significantly associated ($P < 5 \times 10^{-8}$) with one or more red blood cell traits in seven studies from the UCLEB consortium. For every SNP (rs#) minor (risk) allele, chromosome (chr) and position (bp) on the human genome build 19 are given. Effect size ($\beta$), corresponding standard error (SE) and $P$ value are shown for each trait associated with the SNP. For replication analysis, SNPs meeting nominal significance cut-off of $P < 0.05$ are marked in bold.

* Four SNPs previously reported as associated with RBC traits.

doi:10.1371/journal.pone.0156914.t002

Table 2. Results of meta-analysis and replication analysis for 15 most significant SNPs in the ABO locus.
two suggestive associations, respectively, were also replicated. We used the same nominal significance cut-off of $P < 0.05$ to call these associations replicated as they were in a locus already known to be associated with the traits (S4 Table).

**Conditional analyses in HFE locus**

Conditional analyses were performed using the two major HFE mutations (C282Y and H63D) as covariates, to test if the signal in this locus was due to HFE alone. C282Y and H63D lie on different ancestral haplotypes, which in case of C282Y may extend over 6 Mb and to around 700 kb or less for H63D [22–24]. The strong peak in this region, containing many SNPs, could be due to the association of HFE mutations, particularly C282Y, with long founder haplotypes. For all three traits, the signal in this locus disappeared in conditional analysis, indicating that these two mutations of HFE are responsible for all of the observed significant association (S5 Fig). For Hb however, a suggestive signal (spanning ~480 Kb and containing 28 genes) was seen in the main analysis within the major histocompatibility complex locus at 6p21.32-6p21.31, centromeric to HFE. This locus remained at the suggestive significance threshold in the conditional analysis (lead SNP rs2734573, $P = 1.33 \times 10^{-6}$; S5 Fig).

**Conditional analyses in ABO locus**

**Conditioning on blood groups.** At the ABO locus, using Hb as an example, analysis conditional on O blood group haplotype caused the signal at ~135.154 Mb (with rs507666 as lead SNP) to drop under the suggestive threshold while some SNPs at ~136.131 Mb and proximal became significant at the suggestive level (Fig 1B) compared to the unconditional analysis. Similarly, conditioning on AO blood group haplotype caused the same signal at ~135.154 Mb to drop around and below the suggestive threshold (Fig 1C). Analyses conditional on AA and B blood group haplotypes did not differ from the unconditional analysis (Fig 1D and 1E), while in analysis conditional on AB blood group haplotype a group of SNPs directly proximal to rs507666 increased in significance (Fig 1F). However, just as in the unconditional analysis, only three SNPs from this group were above the significance threshold: rs28850884, rs9411378 and rs550057. Analyses run separately on BO and BB blood group haplotypes did not differ from those in unconditional analysis and so were combined. The distribution of the three traits

![Fig 1. Meta-analysis association results for hemoglobin in unconditional and conditional analyses in the ABO locus. Regional plots show (A) unconditional analysis and analysis conditional on (B) O, (C) AO, (D) AA, (E) B and (F) AB blood group haplotype. The most significant SNP in the unconditional analysis, rs507666, is highlighted throughout to facilitate comparison of results. The color coding of the LD between the SNPs ranges from dark blue for $r^2 = 0$–0.2 to red for $r^2 = 0.8$–1, and is grey where LD information was not available. Blue line represents suggestive and red line significant threshold.](http://www.example.com/fig1.png)
associated with the ABO locus according to blood groups is shown in Table 3, with each trait showing highly significant ($p \leq 0.0001$) variation across the blood groups.

**Conditioning on associated RBC traits.** In analyses conditional on RBC traits we expect to see reduction in the association signal if the trait we are conditioning on plays a role in the association. Conditional analyses of all three traits, taking each one in turn as the dependent variable with the other two as covariates, showed total loss of signal at the ABO locus (Fig 2).

When conditioned on Hct, the signal for Hb at the HFE locus disappeared (Fig 2A). However, when the analysis was conditional on RCC, the signals for Hb and Hct at both HFE and TMPRSS6 loci rose (Fig 2B and 2D). The same happened when the analysis for RCC was conditional on Hb and Hct, with both HFE and TMPRSS6 loci becoming significant (Fig 2E and 2F). The signal at the HBS1L locus remained unaffected by conditional analyses (Fig 2E and 2F).

**Association of ABO SNPs with other cardiometabolic traits**

Due to the richly phenotyped studies within the UCLEB consortium we were able to access the results of prior large scale meta-analyses of a large number of ‘intermediate’ continuous traits,
predominantly cardiometabolic traits, including lipids/lipoproteins, coagulation factors and demographic factors [25]. We identified five cardiometabolic traits that were significantly associated with the ABO locus (S7 Fig): factor VIII (FVIII), von Willebrand factor (log transformed, logVWF), low-density lipoprotein cholesterol (LDL), total cholesterol (TC), and alkaline phosphatase (log transformed, logALP), all of which have been previously reported to be associated with ABO SNPs [26–32]. The study characteristics for these traits are given in S5 Table. We directly compared the effect sizes of 15 RBC-associated ABO SNPs across the three RBC and five additional cardiometabolic traits in the UCLEB consortium dataset using standardized regression coefficients (β; Fig 3). Direct comparison showed variation in effect size from β = -0.30 for logALP on rs507666 to β = 0.52 on the same SNP for logVWF, and from β = -0.25 to β = 0.47 on the widely cited rs651007. The direction of the minor allele effects on RBC traits and logALP were negative for all SNPs while the same alleles had positive effects on LDL- and total-cholesterol and coagulation factors, most notably logVWF. Only FVIII showed weak correlation with some of the RBC traits, all other traits were not correlated with RBC traits.

Discussion

Seven studies in the UCLEB Consortium with a total of 11 093 individuals of European ancestry contributed data for this meta-analysis of six red blood cell traits. There were 394 non-overlapping SNPs in five loci associated with the RBC traits at genome-wide significance level, out of which 86% were significantly associated with more than one trait. The five loci (with candidate gene) identified were: 6p22.1-6p21.33 (HFE), 6q23.2 (HBS1L), 6q23.3 (contains no genes), 9q34.3 (ABO) and 22q13.1 (TMPRSS6). Associations at the HFE, HBS1L and TMPRSS6 genes with RBC traits are well established across different ethnic groups, adults and/or children and associated with one or more red blood cell traits [2–4, 6–10, 12, 13, 21, 33]. Both HFE and TMPRSS6 are involved in iron homeostasis by regulating the production of hepcidin, which is considered to be the "master" iron regulator; mutations in HFE cause the most common form of hereditary hemochromatosis while mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). SNPs in the HBS1L-MYB locus are associated with fetal hemoglobin levels and all hematological traits of the three major blood-cell lineages: white blood cells, RBC
and platelets. SNPs in the 6q23.3 locus show association with MCH and MCV in several studies. This locus does not contain any genes itself, however, the nearby CITED2 gene has been reported as a candidate gene [6, 7, 9, 13]. CITED2 inhibits transactivation of hypoxia-inducible factor 1-alpha (HIF1A)-induced genes affecting essential physiological processes, including iron homeostasis. Underlying correlation between the traits is reflected in the results. High correlation between Hb, Hct and RCC shows in the association of these traits with ABO locus, while high correlation between MCH and MCV shows at almost all other associated loci. Similar patterns were observed in the CHARGE Consortium [6], where it was suggested that the observed patterns characterize various clinical hematological diseases and provide context for interpretation of the results.

We have also replicated and confirmed the association of ABO SNPs with Hb, Hct and RCC in Europeans, previously reported in East Asians, Caucasians and African-Americans [5, 7, 9, 10, 13, 34]. By imputation to 1000 Genomes data, we identified 11 new SNPs significantly associated with these traits and replicated all but one in an independent population, further refining the associations. We then focused on the characterization of the ABO locus, looking to dissect its association with RBC traits by means of conditional analysis. First we investigated whether these three RBC traits were associated with the identified loci, especially ABO, independently. The results indicated that the signals at HFE and TMPRSS6 were mediated through the MCV component of Hct and through Hb. This is in agreement with findings that patients with hereditary hemochromatosis have higher mean Hb and MCV, and lower mean RCC than controls [35, 36], while in iron-deficiency states such as IRIDA, low values for Hb and MCV occur, with RCC being higher [2, 37]. In both conditions, changes in Hb and MCV are in the opposite direction to RCC, consistent with our observations in the conditional analysis. The HBS1L signal for RCC appeared not to be influenced by Hb or Hct, while the signal at the ABO locus was influenced by all three traits. Consistent with this association of ABO with RBC traits, a recent study of biomarkers of iron status found that the ABO locus was associated with serum ferritin concentration, an indicator of body iron stores, in Europeans [38]. A study conducted in South Eastern Nigeria found a significant difference in serum iron concentration and total iron binding capacity (TIBC) between the ABO blood groups, with group A having highest and group O the lowest values for both iron parameters [39]. Given that the other four loci we identified are well accepted as having a direct or indirect role in iron homeostasis, it is an intriguing possibility that the ABO locus may also influence RBC traits through an effect on iron.

Further, we looked at the distribution of the traits associated with the ABO locus in our study according to haplotypes predicting blood group, which showed a significant difference in trait means between the blood group types. Results from analyses conditional on the genetic determinants of blood groups suggested that the locus at ~135.154 Mb is greatly influenced by the blood group O deletion. They also indicated the possibility of two loci here: one at the 5′ end at ~135.15 Mb with rs507666 as the lead SNP, and another one proximal, between ~135.15 and ~135.14 Mb, with rs28850884 as lead SNP. However, these two loci are not entirely independent—SNPs in both loci are in moderate LD with each other and the signal on both of them disappears completely in analysis conditional on rs507666 (results not shown). Since the ABO locus is also associated with several other traits and diseases, including lipids, coagulation factors, markers of endothelial function, cardiovascular disease, infectious disease and cancer [26–32, 40–42], we investigated whether other traits available in the UCLEB Consortium are associated with it. This directly demonstrated the pleiotropic associations of the same ABO SNPs with other cardiometabolic traits: coagulation factors (FVIII, logVWF), lipids (LDL, TC) and a liver function trait (logALP). This is not surprising given that, in addition to their expression on red blood cells, ABO antigens are widely expressed on different cell types, including vascular endothelium and epithelium [42, 43]. Hence they have potential for wide influence upon
human traits. ABO determinants are also present in body fluids and are carried on the coagulation protein von Willebrand factor, consistent with the ABO locus associations we demonstrated with these traits [44]. The direction of effect on the ABO alleles associated with cardiometabolic traits in our study were consistent with previous reports. For different traits, these effect directions are widely reported to be associated with generally either beneficial or detrimental effects upon cardiometabolic risk. Our results show that for the 15 ABO SNPs most strongly associated with RBC traits, an increase in copies of the minor allele was associated with increased coagulation factor and lipid levels and decreased RBC traits and alkaline phosphatase. This effect was reversed for blood group O, resulting in decreased coagulation factor and lipid levels and increased Hb, Hct, RCC and ALP, both acting protective. These findings add to the mounting evidence for the role of the ABO blood group as a risk factor for development of cardiovascular disease, thrombosis, diabetes mellitus and cancer, with the O blood type being protective [45–49].

Studies such as ARIC [50] and HEIRS [51] showed that HFE C282Y homozygosity, which is associated with increased iron absorption, is also associated with lower total and LDL cholesterol. Equally, studies show that triglyceride, total cholesterol and LDL cholesterol are decreased in women with iron deficiency anemia [52, 53]. Both conditions have in common the diminished production of hepcidin, which in turn leads to iron-depleted macrophages which may negatively regulate systemic cholesterol levels and so offer protection from atherosclerosis [54]. Evidence from mouse studies suggests that macrophage iron depletion may have pro-inflammatory effects [55]; reviewed in [56], which could be relevant to O blood group individuals as discussed above. rs651007, close to the 5’ end of the ABO gene and significantly associated with RBC traits in our study, was the only SNP associated with decreased ferritin concentration and decreased atherosclerosis risk in a recent study of the role of iron and hepcidin in atherosclerosis [57]. Additionally, increased levels of IL-10 have been found in acute coronary syndrome cases compared to healthy controls, reflecting an underlying inflammatory state, with levels of IL-10 being higher in blood type O compared to individuals with other blood types and associated with poor outcome in these patients [58]. They suggested that inflammation is a more important risk factor in these individuals, while increased coagulation is more important in non-O blood type [58].

Finally, the current discovery analysis is somewhat biased towards loci already known to be associated with cardiometabolic disease due to the design of Metabochip [59]. Imputation to the 1000 Genomes expanded the regions covered by the chip but could not completely fill the gaps. Analysis based on GWAS coverage imputed to the latest 1000 Genomes phase could discover new loci associated with these traits. Most of the studies, including ours, focus on classic ABO blood group phenotypes. The complexity of the ABO blood group system extends to a number of different subtypes and the way that they may influence the traits and disease development requires more detailed investigation. However, we were able to further dissect the association between ABO locus and RBC traits as well as show how the same SNPs influence range of traits and put it in context of cardiometabolic risk. Our results support the suggestion by Johansson et al. [58] that the ABO blood group of an individual may determine which risk factors are more important for them.

Materials and Methods

Studies and phenotypes

The UCL-LSHTM-Edinburgh-Bristol (UCLEB) Consortium has been described in detail in Shah et al. [25]. Six RBC traits were analyzed: Hb, Hct, MCH, MCHC, MCV and RCC. We calculated missing traits as detailed in S1 Note. The total sample size in seven eligible studies was
The Cohorte Lausannoise [60] (CoLaus, N = 2848) study was approached for replication of the results due to the availability of both genotypes imputed to 1000 Genomes and the traits of interest. Discovery and replication study design are further explained in S1 Note. All participants were of European ancestry.

Ethics Statement

Each study of the UCLEB consortium was approved by the appropriate local research ethics committee and all participants gave written informed consent, with details as follow: 1958BC — Written consent was obtained from participants for the use of information in medical studies. The 45-year biomedical survey and genetic studies were approved by the South-East Multi-Centre Research Ethics Committee (ref: 01/1/44) and the joint UCL/UCLH Committees on the Ethics of Human Research (Ref: 08/H0714/40); BRHS — The National Research Ethics Service Committee for London provided ethical approval. Participants provided a written informed consent for the investigation, which was performed in accordance with the Declaration of Helsinki; BWHHS — Ethics approval was granted for the British Women’s Heart & Health Studies from the London Multi Centre Research Ethics Committee and 23 Local Research Ethics Committees (Harrogate: Harrogate Health Care, Shrewsbury: Shropshire Health Authority, Lowestoft: Great Yarmouth James Pagent Healthcare NHS Trust, Mansfield: North Nottinghamshire Health, Southport: North Sefton Research Ethics Committee, Merthyr Tydfil: Awdurdod Lechyd Bro Taf Health Authority, Guildford: South West Surrey Local Research Ethics Committee, Burnley: East Lancashire Health Authority, Newcastle-under-Lyme: North Staffordshire Health Authority, Exeter: Exeter Research Ethics Committee, Falkirk: Fife Health Board, Ipswich: East Suffolk Local Research Ethics Committee, Gloucester: Southmead Health Services, Carlisle: East Cumbria Local Research Ethics Committee, Dunfermline: Fife Health Board, Darlington: County Durham Local Research Ethics Committee, Ayr: Fife Health Board, Grimsby: South Humber Health Authority, Bedford: North Bedfordshire Health Authority, Wigan: Wigan & Leigh LREC, Scunthorpe: South Humber Health Authority, Hartlepool: Hartlepool Local Research Ethics Committee, Bristol: Southmead Health Services). Participants were asked for written informed consent to review their medical records and for permission to perform anonymized genetic tests; CaPS — Ethics approval was obtained from the South East Wales Local Research Ethics Committee, and each subject signed their agreement to be involved; EAS — Approval for the study was given by the Lothian Health Board ethics committee and written informed consent was obtained from each subject; ELSA — Ethical approval for ELSA was given by the National Research Ethics Service and all participants gave written consent; ET2DS — Ethical permission was obtained from the Lothian Medical Research Ethics Committee. All subjects attended a dedicated research clinic where they gave written informed consent for participation in the study; MRC NSHD — Ethical approval was obtained from the Central Manchester Research Ethics Committee (07/H1008/168 and 07/H1008/245) and North Thames Multi-Centre Research Ethics Committee (MREC 98/1/121). Informed written consent was obtained from the study members; WHII — In Whitehall II, participants provided their written informed consent and the National Health Services (NHS), Health Research Authority, National Research Ethics Service (NRES) Committee London—Harrow approved the consent procedure. Blood samples were collected and analyzed using standard methods and assays.

For the replication study, the Institutional Review Board of the Centre Hospitalier Universitaire Vaudois (CHUV) in Lausanne and the Cantonal Ethics Committee approved the CoLaus study protocol and signed informed consent was obtained from participants. Starting in 2009
all participants were invited for a follow-up visit five years after the initial study, completed in 2012. This follow-up study was approved by the local ethics committee.

Genotypes and imputation

All studies with RBC traits were genotyped using Metabochip and imputed to the 1000 Genomes (Phase1, CEU haplotype set). Genotyping, quality control (QC) procedures and imputation were performed in each study separately using the same protocol [25]. The exclusion criteria for samples and SNPs consisted of an Illumina GenCall score < 0.15, sample and SNP rates < 0.95, HWE P value ≤ 0.001, replicate discordance, discrepancy in sex, ethnicity and relatedness checks or against previous genotype data, where available. No MAF cut-off was applied at this stage. Imputation was done on cleaned genotypes in three phases—chunking, phasing and imputing—using the MACH1 and Minimac software [61, 62] and applying the strategy described at Minimac:1000 Genomes Imputation Cookbook. The resulting file contained ~4.5 million autosomal SNPs (R² > 0.3, MAF > 0.001). The data were collated in posterior probabilities from genotype dosage, incorporating imputation uncertainty, to be used by the R Package snpStats.

Association and meta-analysis

Based on descriptive statistics and visual inspection (Table 1; S1 Fig), association testing was done on untransformed data, separately for each study and on each trait. The associations between genotyped and imputed SNPs and traits were tested by linear regression using an additive genetic model. Adjustment was made for age, sex and diabetes status, where appropriate. Analyses were carried out using snpStats. Prior to meta-analysis, SNPs with MAF frequency < 0.005 were filtered out. Genomic-control inflation factors (λGC) were calculated and used to correct test statistics from each study for each trait in the meta-analysis. Study-specific estimates of effect size were combined by fixed effects meta-analysis weighted by inverse variance using the METAL program. Genomic-control inflation factors were calculated again after the meta-analysis, but the results were not further corrected. Significance threshold was set at $P < 5 \times 10^{-8}$, suggestive at $P < 5 \times 10^{-6}$. Locus (significant or suggestive) was defined as all SNPs below the threshold with SNPs located less than 500 kb apart. LocusZoom [63] was used to display results, using 1000 Genomes Phase 1 CEU haplotype set for linkage disequilibrium (LD) calculation. The positions of SNPs are given on human genome build 19.

Conditional analyses

Conditional analysis was performed using the SNPs encoding the two major HFE mutations, C282Y (rs1800562) and H63D (rs1799945), as covariates to further investigate if the signal at this locus for Hb, MCH and MCV was due to HFE alone. Also, to further dissect the signal at the ABO locus, we used analyses conditional on the genetic determinants of blood groups (AA, AO, B, AB and O) with a blood group entered as a binary variable one at the time (e.g. O vs all other groups). The three SNP haplotypes used to derive blood groups are presented in Table 4. To investigate whether the signal in the ABO locus for Hb, Hct and RCC affects a trait independently of the other two traits, the subset of five studies which had data on all three traits was analyzed. For a given trait, the other two traits were included in the model as covariates one at a time [64].

Association with other traits within UCLEB

To assess pleiotropy by direct comparison within UCLEB, we ran an association study on standardized phenotypes (mean = 0, SD = 1). Effect sizes (standardized regression coefficients, $\beta$) were visualized using phenotypicForest 0.2.
Pearson’s $r$ was calculated to assess correlation between all the traits within the study.

URLs
- Minimac: 1000 Genomes Imputation Cookbook, http://genome.sph.umich.edu/wiki/Minimac:_1000_Genomes_Imputation_Cookbook
- METAL, http://www.sph.umich.edu/csg/abecasis/Metal/
- phenotypicForest 0.2, http://chrisladroue.com/phorest/
- European Genome-phenome Archive, http://www.ebi.ac.uk/ega

Supporting Information

S1 Fig. Histograms for six red blood cell traits. Please note that the hemoglobin values for WHII were rounded.
(TIF)

S2 Fig. Manhattan plots of meta-analysis association results for red blood cell traits in the UCLEB consortium. Line at $-\log_{10}(P\text{ value}) = 5.3$ represents suggestive threshold and line at $-\log_{10}(P\text{ value}) = 7.3$ significant threshold.
(TIF)

S3 Fig. Q-Q plots of meta-analysis association results for red blood cell traits in the UCLEB consortium.
(TIF)

S4 Fig. Venn diagram of overlap between loci above the significance threshold ($P < 5 \times 10^{-8}$). Traits include hemoglobin (Hb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV) and red blood cell count (RCC).
(TIF)

S5 Fig. Manhattan plots of meta-analysis association results for Hb, MCH and MCV in analyses conditional on two major HFE gene mutations, C282Y and H63D. Line at $-\log_{10}(P\text{ value}) = 5.3$ represents suggestive threshold and line at $-\log_{10}(P\text{ value}) = 7.3$ significant threshold.
(TIF)

S6 Fig. Comparison between van der Harst et al. and our study on overlapping 79 SNPs within +/- 10Kb from ABO gene. Regional plots for (A-B) Hb, (D-E) Hct and (E-F) RCC show two loci: (A,C,E) first locus at ~136.154 close to the 5' region of the gene and (B,D,F) second locus at ~136.131 Mb, using 1000 genomes (Phase 1, EUR haplotype set) for LD calculation. Blue line represents suggestive and red line significant threshold.
(TIF)
S7 Fig. Meta-analysis association results for three red blood cell and five cardiometabolic traits in the ABO locus. Regional plots show association results for (A) von Willebrand factor (log transformed, logVWF), (B) factor VIII (FVIII), (C) total cholesterol (TC), (D) low-density lipoprotein (LDL), (E) hemoglobin (Hb), (F) red blood cell count (RCC), (G) hematocrit (Hct) and (H) alkaline phosphatase (log transformed, logALP). The most significant associations for logVWF are capped at $1 \times 10^{-325}$. The most significant SNP for Hb, rs507666, is highlighted throughout to facilitate comparison of results. Blue line represents suggestive and red line significant threshold.

(TIF)

S1 Note. Supplementary note.

(DOCX)

S2 Note. Data access arrangements.

(DOCX)

S1 Table. Formulae for calculating missing erythrocyte traits.

(DOCX)

S2 Table. Genomic-control inflation factors ($\lambda_{GC}$) for each study and trait. $\lambda_{GC}$ were recalculated for each trait after meta-analysis.

(DOCX)

S3 Table. Meta-analysis results. All SNPs significantly associated with one or more red blood cell traits ($P < 5 \times 10^{-8}$) in seven studies from the UCLEB consortium (separate file).

(XLSX)

S4 Table. Results of replication in the CoLaus study. Results for 47 SNPs in/around ABO gene, which were either significantly or suggestively associated with Hb, Hct and RCC in seven studies from the UCLEB consortium. The rs numbers of the SNPs which were significantly associated with Hb, Hct and/or RCC ($P < 5 \times 10^{-8}$, N = 15) in the discovery study are highlighted in bold and the corresponding fields are marked in blue; these are additionally shown in Table 3. The significant replication results ($P < 0.05$) are marked in red (separate file).

(XLSX)

S5 Table. Study characteristics on other cardiometabolic traits associated with ABO SNPs and available in UCLEB.

(DOCX)

S6 Table. Pearson’s $r$ correlation coefficient between six red blood cell traits and five cardiometabolic traits.

(XLSX)

Acknowledgments

Jackie Cooper¹, Ian N. Day², Maneka De Silva², Frank Dudbridge³, Ghazaleh Fatemifar³, Victoria Garfield⁴, Steve E. Humphries¹, Debbie A. Lawlor², Teri-Louise Davies², Vincent Plagnol⁵, Christine Power⁶, Sonia Shah⁷, Reecha Sofat⁷, Daniel I. Swerdlow⁸, Philippa J. Talmud¹, Peter Whincup⁹, John C. Whittaker¹⁰, Delilah Zabaneh⁵

¹Centre for Cardiovascular Genetics, Department of Medicine, University College London, London, United Kingdom

²MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom
**Author Contributions**

Conceived and designed the experiments: SM. Performed the experiments: SM. Analyzed the data: SM CG JW PC AW CF JE TS MH. Contributed reagents/materials/analysis tools: SM CG JW PC AW CF JE TS RR. Wrote the paper: SM JFP APW CG MH CP BJJ RWM YB-S TRG FD CL M. Kivimaki. Collected and provided data for meta-analysis study: SM AW CP AC BJJ CED EH RWM JPC M. Kumari YB-S TRG FD CL DK M. Kivimaki ADH JFP. Collected and provided data for replication study: CP MH RR GW.

**References**


