nature genetics

Genome-wide association study identifies variants in *TMPRSS6* associated with hemoglobin levels

John C Chambers[1,13], Weihua Zhang[1,13], Yun Li[2], Joban Sehmi[3], Mark N Wass[4], Delilah Zabaneh[1], Clive Hoggart[1], Henry Bayele[5], Mark I McCarthy[6], Leena Peltonen[7], Nelson B Freimer[8], Surjit K Srai[5], Patrick H Maxwell[9], Michael J E Sternberg[4], Aimo Ruokonen[10], Gonçalo Abecasis[2], Marjo-Riitta Jarvelin[1,11,12], James Scott[3,13], Paul Elliott[1,13] & Jaspal S Kooner[3,13]

We carried out a genome-wide association study of hemoglobin levels in 16,001 individuals of European and Indian Asian ancestry. The most closely associated SNP (rs855791) results in nonsynonymous (V736A) change in the serine protease domain of TMPRSS6 and a blood hemoglobin concentration 0.13 (95% CI 0.09–0.17) g/dl lower per copy of allele A ($P = 1.6 \times 10^{-13}$). Our findings suggest that *TMPRSS6*, a regulator of hepcidin synthesis and iron handling, is crucial in hemoglobin level maintenance.

One-quarter of the world's population has anemia, with the highest burden in India and Southeast Asia[1]. Although iron deficiency is the principal cause of low hemoglobin levels worldwide, genetic factors also make an important contribution. Mutations in the globin genes, red cell fragility syndromes and defects in iron metabolism cause severe hereditary anemias[2,3]. Common variants at the *HBB* and *HBA1* loci have been associated with hemoglobin levels in a genetically isolated Sardinian population with high prevalence of β-thalassemia[4]. We carried out a genome-wide association study to identify common genetic variants influencing hemoglobin among individuals of European and Indian Asian ancestry.

Genome-wide association for hemoglobin levels was performed in 6,316 Europeans and 9,685 Indian Asians participating in the London Life Sciences Population (LOLIPOP) study and the North Finland Birth Cohort of 1966 (NFBC1966). All LOLIPOP participants were resident in London, UK. Within this study, participants of European ancestry self-reported as white and born in Europe on a questionnaire; those of Indian Asian ancestry reported having all four grandparents born on the Indian subcontinent. All participants in the NFBC were of European ancestry. Clinical characteristics of participants, and the genotyping platforms used, are summarized in the **Supplementary Methods** and **Supplementary Table 1**. Imputation was used to infer missing genotypes, using the HapMap CEU sample as reference for Europeans and pooled founder haplotypes from all three HapMap populations as reference for Indian Asians (HapMap build35, dbSNP build 125)[5]. Imputed SNPs with minor allele frequency (MAF) <0.01 or low quality score ($r^2 < 0.30$) were removed.

Single SNP marker tests were performed for association with hemoglobin level using linear regression under an additive genetic mode and adjusting for age and sex. Population substructure was characterized using principal components analyses[6] and was included as covariates in the regression models. Results for NFBC1966 Europeans, LOLIPOP Europeans and LOLIPOP Indian Asians were analyzed separately; results were then combined between studies using *z* scores weighted to the square root of sample size. Quantile-quantile plots showed good adherence to null expectations (**Supplementary Fig. 1**). The genome-wide association study had 80% power to identify SNPs associated with ~0.5% of population variation in hemoglobin in either ethnic group, or ~0.3% in combined analysis, at $P < 5 \times 10^{-8}$.

We found four SNPs among Europeans and three SNPs among Indian Asians that showed association with hemoglobin at a genome-wide significance threshold of $P < 5 \times 10^{-8}$ (ref. 7, **Table 1** and **Supplementary Fig. 2**). All seven SNPs identified are located in

Table 1 Genomic context and association test results for SNPs linked with hemoglobin levels

	Genomic context			Alleles		Minor allele frequency		Europeans	Indian Asian	Combined
SNP	Chr	Position	Locus	Minor	Alt	EW	IA	*P*	*P*	*P*
rs855791	22	35792882	*TMPRSS6*	A	G	0.34	0.53	1.8×10^{-5}	1.8×10^{-9}	1.6×10^{-13}
rs4820268	22	35799537	*TMPRSS6*	G	A	0.43	0.55	4.7×10^{-5}	7.2×10^{-10}	2.0×10^{-13}
rs2413450	22	35800170	*TMPRSS6*	T	C	0.46	0.54	5.6×10^{-5}	5.8×10^{-10}	2.0×10^{-13}
rs228918	22	35836626	*TMPRSS6*	C	T	0.47	0.48	1.2×10^{-8}	7.6×10^{-4}	4.4×10^{-10}
rs228919	22	35836659	*TMPRSS6*	T	G	0.40	0.48	1.2×10^{-8}	7.4×10^{-4}	4.4×10^{-10}
rs228921	22	35836822	*TMPRSS6*	G	A	0.41	0.48	1.2×10^{-8}	7.3×10^{-4}	1.9×10^{-10}
rs5756520	22	35838453	*TMPRSS6*	A	G	0.41	0.48	1.6×10^{-8}	4.6×10^{-4}	2.7×10^{-10}

Results among people of European (EW) or Indian Asian (IA) ancestry in the genome-wide association study at $P < 5 \times 10^{-8}$. Effect sizes are shown as unit change (95% CI) in hemoglobin per copy of minor allele (selected based on allele frequencies in Europeans).

[1]Department of Epidemiology and Public Health, Imperial College London, London, UK. [2]Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. [3]National Heart and Lung Institute and [4]Structural Bioinformatics Group, Imperial College London, London, UK. [5]Department of Structural & Molecular Biology, University College London, London, UK. [6]Oxford Centre for Diabetes, Endocrinology and Metabolism and Oxford NIHR Biomedical Research Centre, Oxford, UK. [7]Wellcome Trust Sanger Institute, Cambridge, UK. [8]Center for Neurobehavioral Genetics, University of California, Los Angeles, California, USA. [9]Division of Medicine, University College of London, London, UK. [10]Department of Clinical Chemistry, University Hospital Oulu, Oulu, Finland. [11]Institute of Health Sciences and [12]Biocenter Oulu, University of Oulu, Oulu, Finland. [13]These authors contributed equally to this work. Correspondence should be addressed to J.C.C. (john.chambers@ic.ac.uk) or J.S.K. (j.kooner@ic.ac.uk).

Received 24 March; accepted 22 June; published online 11 October 2009; doi:10.1038/ng.462

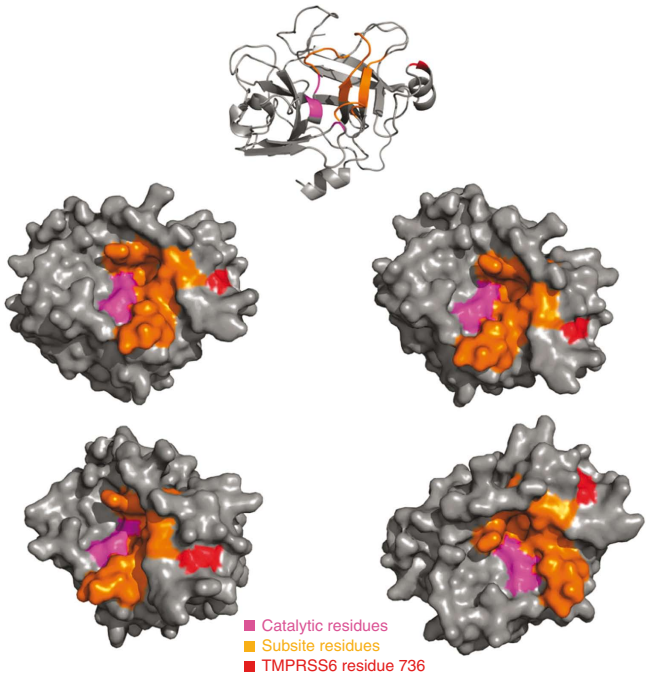

Figure 1 Molecular model of the serine protease domain of TMPRSS6 showing binding site residues (orange), catalytic residues (magenta) and the location of the V736A amino acid substitution caused by SNP rs855791 (red).

the *TMPRSS6* locus on chromosome 22. The SNPs identified among Europeans replicated among Indian Asians, and those identified in Indian Asians replicated among Europeans (all $P < 0.001$). In combined analysis of European and Indian Asian data, rs855791 (G→A) in *TMPRSS6* showed the strongest association with hemoglobin level. The association of rs855791 with hemoglobin level was replicated in a further sample of 5,187 Europeans ($P = 4.3 \times 10^{-7}$) and 6,721 Indian Asians ($P = 1.4 \times 10^{-11}$) from the LOLIPOP study (**Supplementary Tables 2** and **3**); effect sizes were similar among Europeans and Indian Asians, with no evidence for heterogeneity ($P > 0.1$). The proportion of population variance in hemoglobin explained by rs855791 was 0.25% in Europeans and 0.31% among Indian Asians. In addition, rs855791 was strongly associated with erythrocyte mean cell volume (MCV), mean cell hemoglobin (MCH) and mean cell hemoglobin concentration (MCHC), key indices of hemoglobin synthesis (**Supplementary Table 3**). The A allele of rs855791 associated with lower hemoglobin levels was more frequent in Indian Asians than in Europeans in the replication sample (0.52% versus 0.43%, $P = 3.5 \times 10^{-33}$). The 19% of individuals of European ancestry and the 27% of Indian Asian ancestry with AA genotype at rs855791 had hemoglobin concentrations on average 0.2 g/dl lower than did persons with GG genotype.

The linkage disequilibrium (LD) structure of the *TMPRSS6* locus was similar among Europeans and Indian Asians (**Supplementary Fig. 3**). The NFBC1966 and LOLIPOP data indicate that SNPs rs855791 and rs4820268 are in high LD (r^2: Europeans 0.83; Indian Asians 0.65) but that rs228918 is in low LD with these two SNPs ($r^2 < 0.1$ in Europeans and Indian Asians). In a stepwise analysis conditioned on rs855791, SNP rs228918 was independently associated with hemoglobin levels in Europeans and Indian Asians ($P < 0.05$).

In a combined analysis of genome-wide data from Europeans and Indian Asians, ten SNPs in the *TMPRSS6* locus, and a further six SNPs in the *HFE* locus on chromosome 6, were associated with hemoglobin at genome-wide significance (**Supplementary Table 4** and **Supplementary Fig. 2**). At the *HFE* locus, rs198846 (G→A) was most strongly associated with hemoglobin levels. SNP rs198846 is located near the mutation in *HFE* that results in a C282Y substitution in the HFE protein (rs1800562), a variant that causes hereditary hemochromatosis and influences hemoglobin levels[8]. Based on the HapMap CEU population, rs198846 is in weak LD ($r^2 = 0.006$) with rs1800562; in regression analysis, the relationship of rs198846 with hemoglobin was independent of rs1800562 ($P < 0.001$).

The association of rs198846, near *HFE*, with hemoglobin level was confirmed in the replication sample (**Supplementary Table 3**); the proportion of population variance in hemoglobin level explained by rs198846 was 0.32% among Europeans and 0.03% among Indian Asians. SNP rs198846 was also associated with MCV, MCH and MCHC (**Supplementary Table 3**). The major allele (G) of SNP rs198846 was associated with lower hemoglobin and was more frequent in Indian Asians than in Europeans (92% versus 84%, $P = 6.9 \times 10^{-81}$). The effects of rs855791 and rs198846 on hemoglobin levels were independent and additive. Hemoglobin levels were ~0.4g/dl lower among the 23% of Indian Asians and the 13% of Europeans homozygous for allele A of rs855791 and allele G of rs198846, compared with persons with ≤1 copy of these alleles.

SNPs in the *HBB* and *HBA1* loci, reported to be associated with hemoglobin in a founder population[4], were not related to hemoglobin, MCV, MCH or red blood cell count in Europeans or Indian Asians (all $P > 0.05$ after Bonferroni correction).

We report that the association of rs855791 in *TMPRSS6* with hemoglobin levels in Europeans and Indian Asians. SNP rs855791 is nonsynonymous and causes a valine-to-alanine amino acid change at position 736 of TMPRSS6, a type II plasma membrane serine protease expressed mainly in liver (**Fig. 1**)[9]. TMPRSS6 has a key role in iron homeostasis[10,11], inhibiting hepatic hepcidin production. Hepcidin is a direct inhibitor of ferroportin, a membrane iron transport protein present on enterocytes and macrophages, and thereby inhibits intestinal iron absorption and the release of iron from cellular stores[10,12]. Rare *TMPRSS6* mutations result in unregulated hepcidin synthesis, reduced iron absorption, and iron-deficiency anemia refractory to oral iron therapy[3]. Recent studies show that rs855791 and rs4820268 in *TMPRSS6* are associated with reduced iron and transferrin saturation[13], consistent with the hypothesis that rs855791 influences hepcidin-regulated iron homeostasis.

Animal and *in vitro* studies have shown that deletion of the serine protease domain of TMPRSS6 eliminates inhibition of hepcidin levels[3]. Comparison with other serine proteases, and molecular modeling using PHYRE (**Supplementary Methods**)[14], showed that the amino acid altered by rs855791 is located close to both the catalytic and the specificity site of the serine protease (**Fig. 1**), suggesting that rs855791 may be a causal variant, acting through altered protease activity or substrate binding.

The genome-wide association and replication data from Europeans and Indian Asians additionally demonstrate association of genetic variants in the *HFE* locus with hemoglobin. This association is also likely to be mediated through iron metabolism. HFE is a key component of the signaling pathway through which iron-loaded transferrin stimulates hepcidin synthesis[15], and genetic variants in *HFE* are well known to be associated with abnormal iron status and with hemoglobin levels[8,13].

We report here the association of common genetic variants in *TMPRSS6* with hemoglobin levels among individuals of both European and Indian Asian ancestry. This association may be mediated through alteration of protease function and hepcidin-mediated control of iron

homeostasis. Our findings could provide new insight into the genetic factors influencing anemia and related blood disorders.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the participants and research teams involved in LOLIPOP and NFBC1966, including P. Rantakallio, O. Tornwall and M. Jussila. The LOLIPOP study was supported by the British Heart Foundation (SP/04/002) and the Wellcome Trust. NFBC1966 was supported by the Academy of Finland (project grants 104781, 120315 and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, National Heart, Lung, and Blood Institute (US) grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, the Medical Research Council (UK) (studentship grant G0500539, centre grant G0600705), the Wellcome Trust (UK) (project grant GR069224) and the Research Council UK fellowship. We thank the National Public Health Institute, Biomedicum Helsinki, Finland, supported financially by the Academy of Finland and Biocentrum Helsinki, for DNA extraction, sample quality controls, biobank upkeep and aliquotting.

AUTHOR CONTRIBUTIONS

J.C.C., P.E., G.A., J. Scott and J.K. designed the study. J.C.C., J. Sehmi, M.-R.J., A.R. and J.K. supervised patient recruitment. J.C.C., P.E., J. Scott, G.A., M.I.M., H.B., L.P., N.B.F., M.J.E.S., P.H.M., S.K.S., M.-R.J., A.R. and J.K. supervised the experiments. J.C.C., W.Z., D.Z., Y.L., C.H. and M.N.W. performed data analysis. J.C.C., P.E., J. Scott and J.S.K. wrote the manuscript. All authors commented on and approved the manuscript.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. de Benoist, B., McLean, E., Egli, I. & Cogswell, M. (eds.) Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia (World Health Organization, Geneva, 2008).
2. Weatherall, D.J. & Provan, A.B. *Lancet* **355**, 1169–1175 (2000).
3. Finberg, K.E. *et al. Nat. Genet.* **40**, 569–571 (2008).
4. Uda, M. *et al. Proc. Natl. Acad. Sci. USA* **105**, 1620–1625 (2008).
5. Huang, L. *et al. Am. J. Hum. Genet.* **84**, 235–250 (2009).
6. Price, A.L. *et al. Nat. Genet.* **38**, 904–909 (2006).
7. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. *Genet. Epidemiol.* **32**, 381–385 (2008).
8. Beutler, E., Felitti, V., Gelbart, T. & Waalen, J. *Br. J. Haematol.* **120**, 887–893 (2003).
9. Ramsay, A.J., Reid, J.C., Velasco, G., Quigley, J.P. & Hooper, J.D. *Front. Biosci.* **13**, 569–579 (2008).
10. Du, X. *et al. Science* **320**, 1088–1092 (2008).
11. Folgueras, A.R. *et al. Blood* **112**, 2539–2545 (2008).
12. Nemeth, E. *et al. Science* **306**, 2090–2093 (2004).
13. Benyamin, B. *et al. Am. J. Hum. Genet.* **84**, 60–65 (2009).
14. Kelley, L.A. & Sternberg, M.J. *Nat. Protoc.* **4**, 363–371 (2009).
15. Gao, J. *et al. Cell Metab.* **9**, 217–227 (2009).