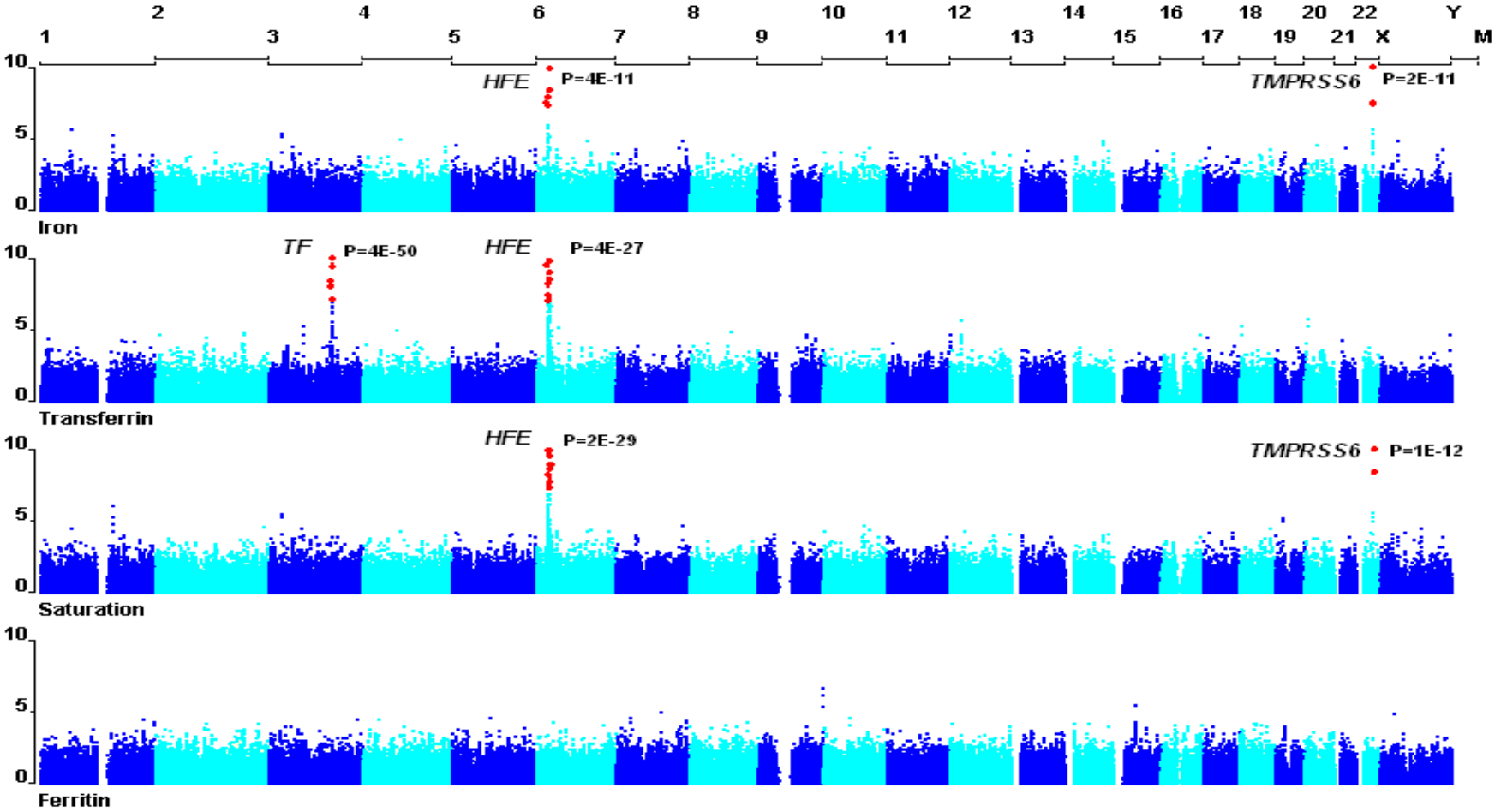
SUPPLEMENTARY INFORMATION

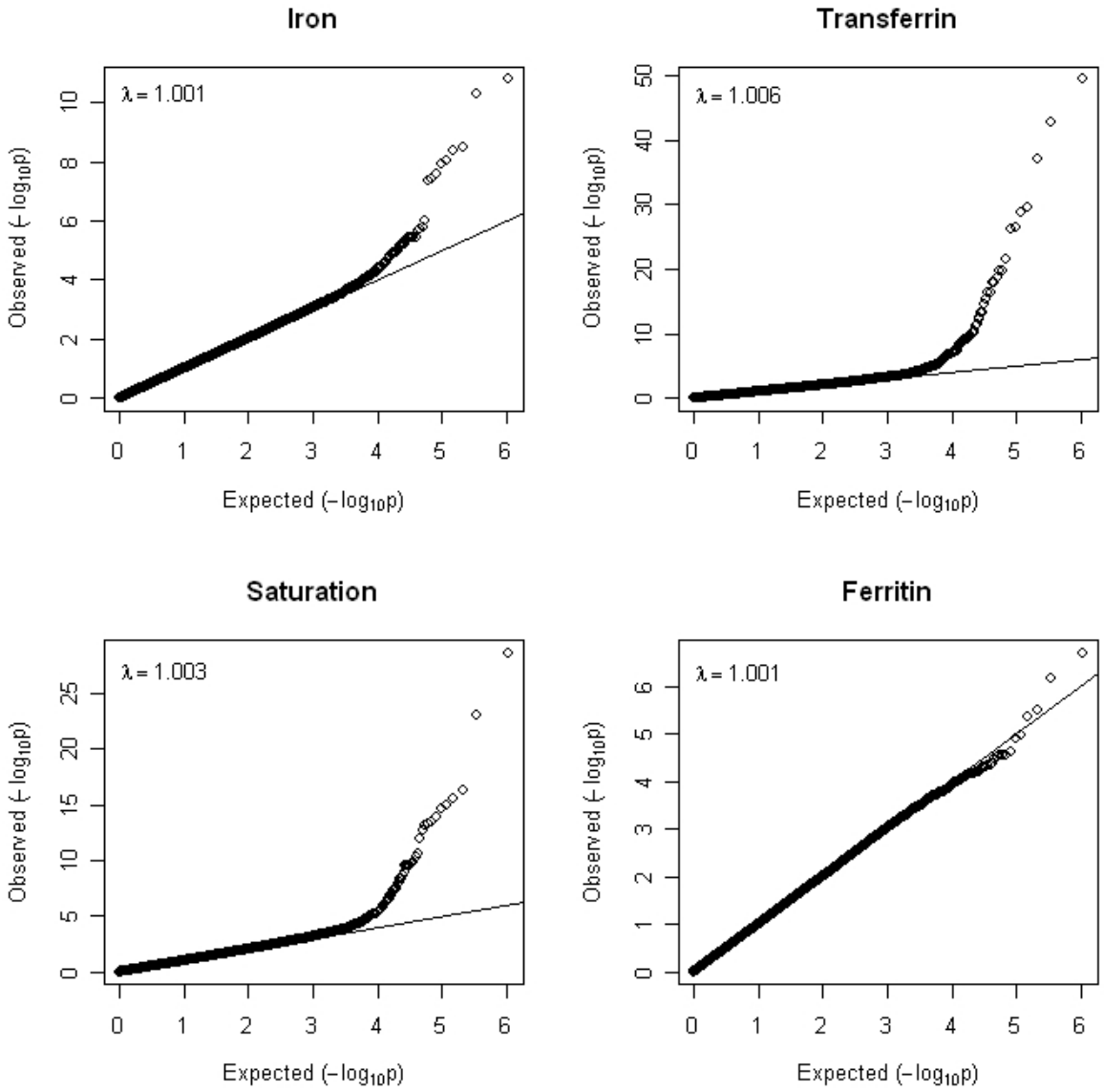
Common variants in *TMPRSS6* are associated with iron status and erythrocyte volume

Beben Benyamin, Manuel A. R. Ferreira, Gonneke Willemsen, Scott Gordon, Rita P. S. Middelberg, Brian P. McEvoy, Jouke-Jan Hottenga, Anjali K. Henders, Megan J. Campbell, Leanne Wallace, Ian H. Frazer, Andrew C. Heath, Eco J. C. de Geus, Dale R. Nyholt, Peter M. Visscher, Brenda W. Penninx, Dorret I. Boomsma, Nicholas G. Martin, Grant W. Montgomery, John B. Whitfield

Supplementary Figure 1. GWAS results for iron status in the Adolescent cohort. On the *x* and *y* axis are chromosome number and $-\log_{10}(P)$ value, respectively. Red dots indicate the SNPs that reached genome-wide significance ($P < 0.05$) after a Bonferroni correction for multiple testing.

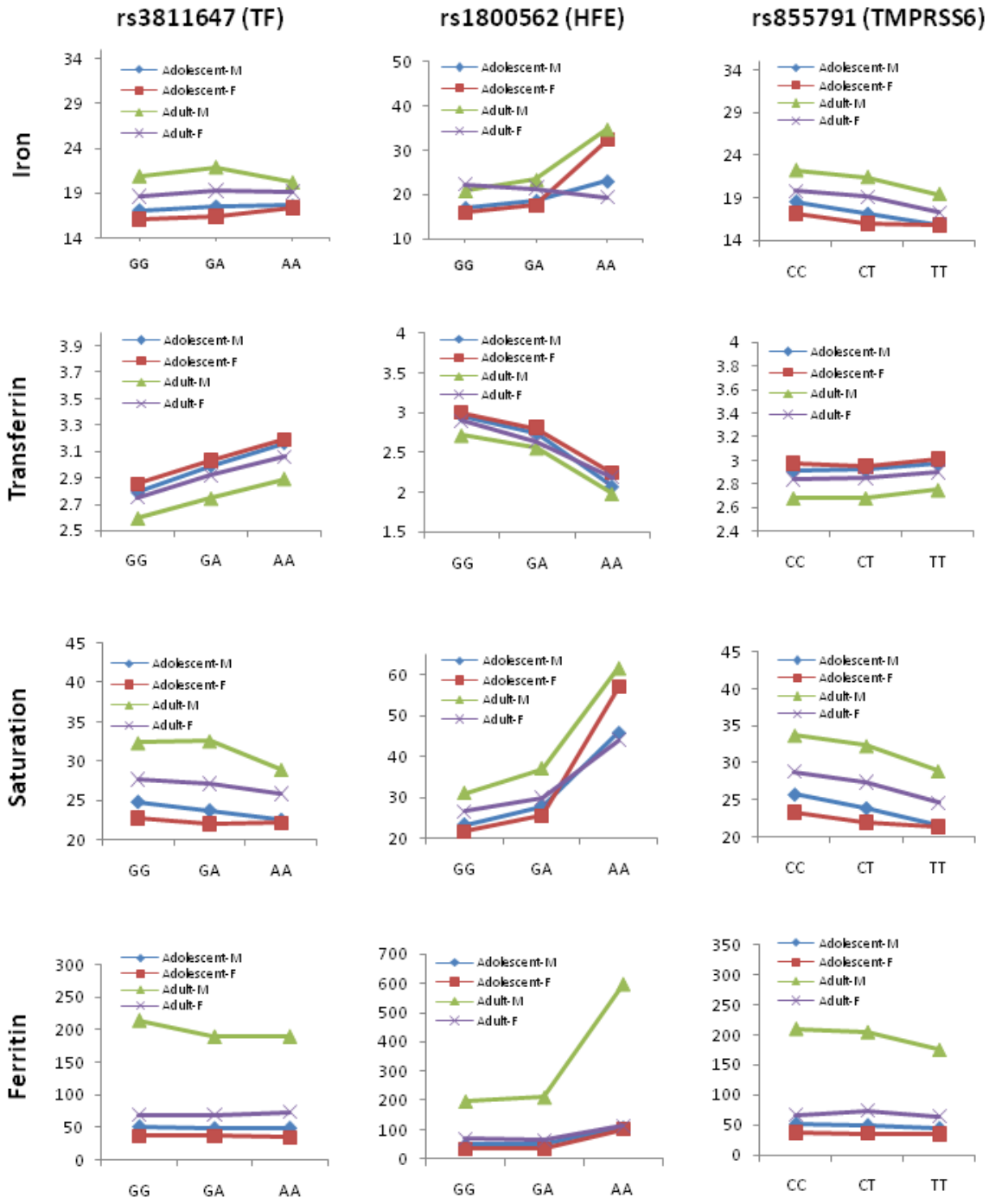


Supplementary Figure 2. Q-Q plots of observed and expected $-\log_{(10)}P$ of the associations between SNPs and serum iron, serum transferrin, transferrin saturation and serum ferritin in the Adolescent cohort.

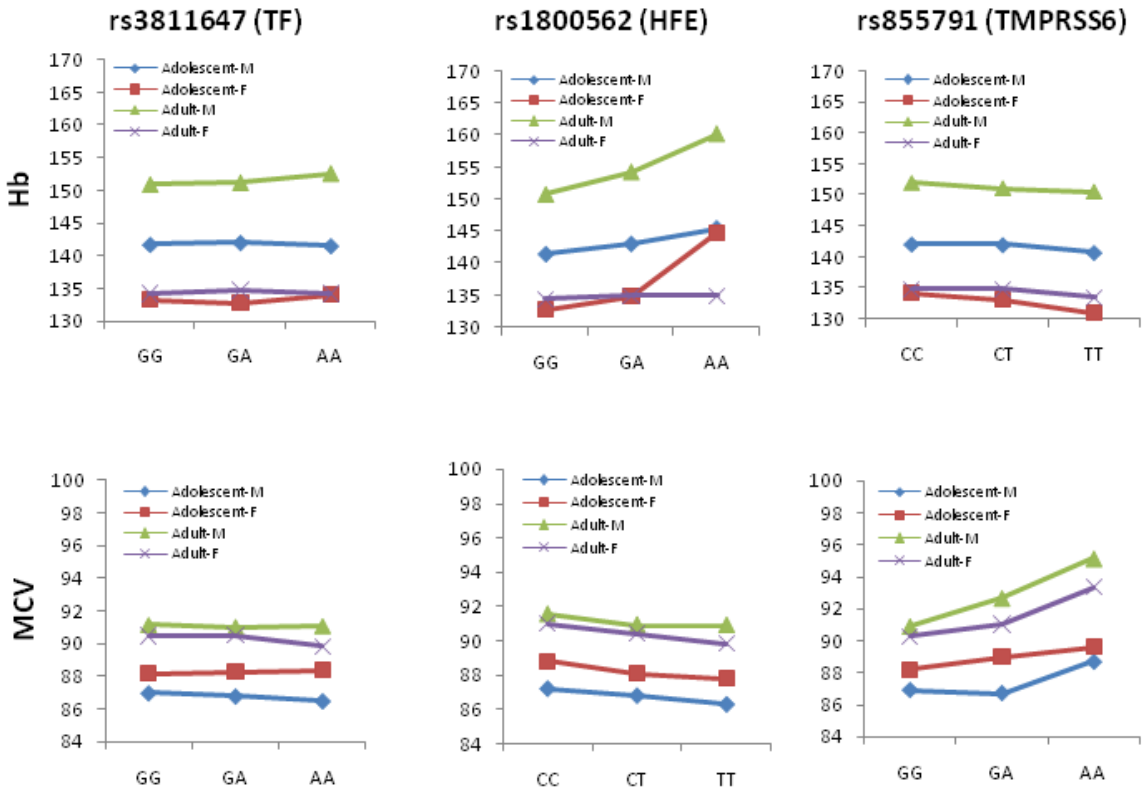


Supplementary Figure 3. The genotypic means of 3 associated SNPs for (**A**) iron markers (serum iron, transferrin, transferrin saturation and serum ferritin); (**B**) Hb and MCV.

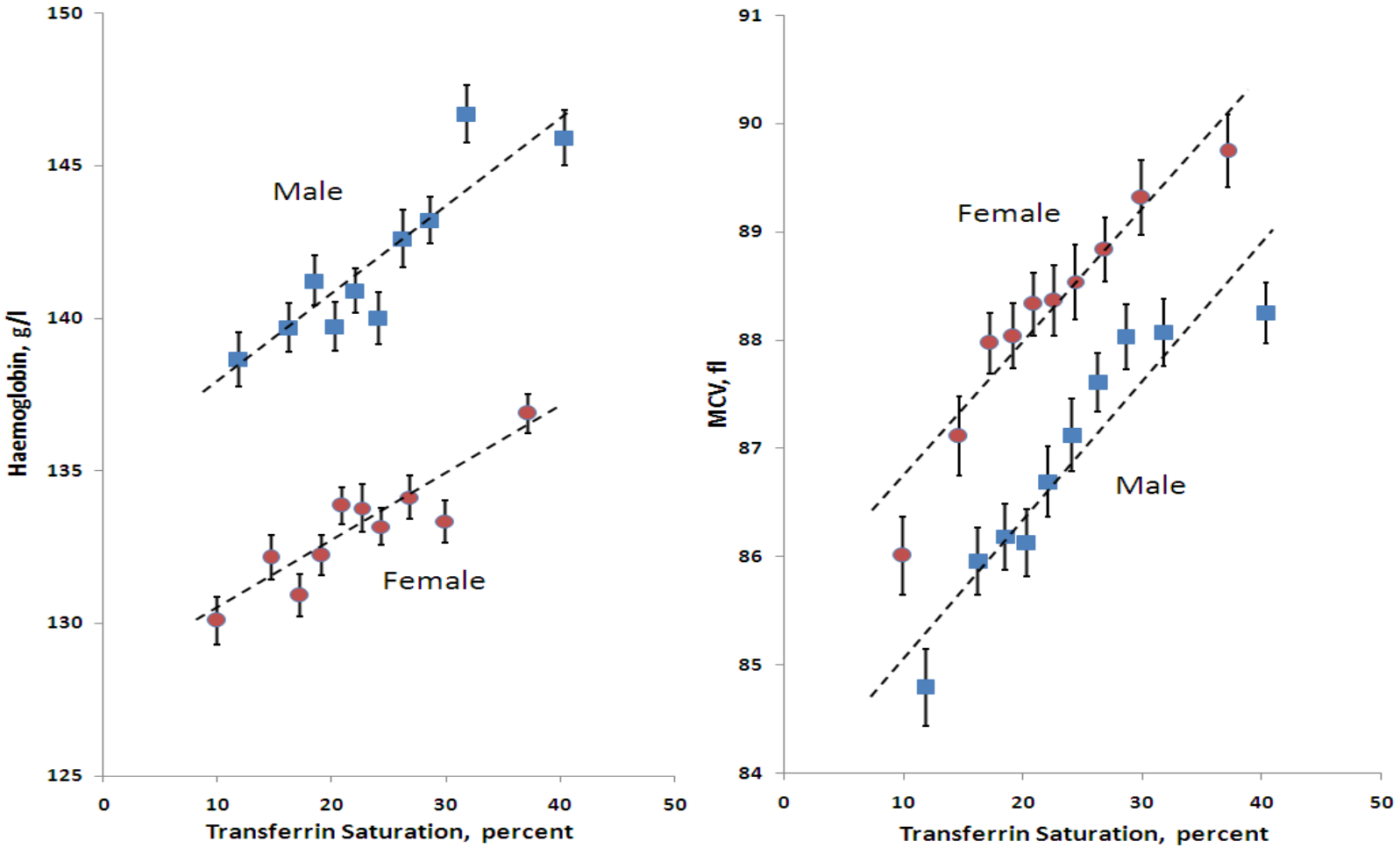
(**A**)



(B)


rs3811647 (TF)
rs1800562 (HFE)
rs855791 (TMPRSS6)
Hb
MCV
Adolescent-M
Adolescent-F
Adult-M
Adult-F
GG
GA
AA
CC
CT
TT

Supplementary Figure 4. Relationships between transferrin saturation and Hb (left panel) or MCV (right panel) in male and female Australian adolescents. Mean transferrin saturation results were divided into deciles for males and females, and the mean transferrin saturation for each decile group (x-axis) were plotted against mean Hb or MCV.



Supplementary Table 1. Descriptive Statistics of the adolescent and adult cohorts.

(A) Adolescent cohort

Traits	Sex	N	Min	Max	Mean	SD
Iron (µmol/L)	M	1470	3.80	42.50	17.34	5.46
	F	1501	3.45	46.40	16.35	5.35
	Total	2971	3.45	46.40	16.84	5.42
Transferrin (g/L)	M	1476	1.60	5.18	2.92	0.36
	F	1508	1.60	5.36	2.97	0.39
	Total	2984	1.60	5.36	2.95	0.37
Transferrin saturation (%)	M	1469	4.81	89.01	24.05	8.22
	F	1501	3.42	70.57	22.32	7.72
	Total	2970	3.42	89.01	23.17	8.02
Ferritin (log_{10}) (µg/L)	M	1476	0.76	2.90	1.70	0.23
	F	1506	0.00	2.37	1.56	0.26
	Total	2982	0.00	2.90	1.63	0.26
Hb (g/L)	M	1226	110.33	174.00	141.80	9.33
	F	1295	110.00	166.00	133.21	7.76
	Total	2521	110.00	174.00	137.39	9.58
MCV (fL)	M	1225	75.00	98.33	86.72	3.49
	F	1295	77.00	101.00	88.03	3.66
	Total	2520	75.00	101.00	87.39	3.64

Note: The values are the average of measurements at up to 4 visits

(B) Adult cohort

Traits	Sex	N	Min	Max	Mean	SD
Iron (µmol/L)	M	914	6.50	48.30	21.33	6.43
	F	1414	3.10	50.20	18.96	6.44
	Total	2328	3.10	50.20	19.89	6.54
Transferrin (g/L)	M	916	1.55	4.34	2.70	0.35
	F	1413	1.53	4.84	2.86	0.46
	Total	2329	1.53	4.84	2.79	0.43
Transferrin saturation (%)	M	913	7.22	81.89	32.13	10.58
	F	1411	3.73	65.82	27.21	10.04
	Total	2324	3.73	81.89	29.14	10.53
Ferritin (log_{10}) (µg/L)	M	917	0.55	3.26	2.30	0.36
	F	1414	0.30	2.90	1.84	0.40
	Total	2331	0.30	3.26	2.02	0.44
Hb (g/L)*	M	1195	104.74	185.31	151.18	9.69
	F	2281	77.34	167.58	134.50	9.96
	Total	3476	77.34	185.31	140.23	12.66
MCV (fL)*	M	1194	68.40	112.00	91.10	4.40
	F	2273	67.90	109.70	90.42	4.63
	Total	3467	67.90	112.00	90.66	4.57

* These measurements are from Dutch samples

Supplementary Table 2. Age-corrected additive effects (in SD) of the three SNPs in *TMPRSS6*, *HFE* and *TF* on serum iron, transferrin, transferrin saturation, ferritin, and blood haemoglobin and MCV stratified by sex. The significance of the difference between the estimates from males and females was tested using a two-tail t-test.

		Adolescent				Adult			
Trait	**Sex**	**N**	**Beta**	**SE**	***P***	**N**	**Beta**	**SE**	***P***
					rs855791 (TMPRSS6)				
Iron	M	1216	-0.222	0.039	0.077	901	-0.193	0.046	0.934
	F	1289	-0.127	0.037		1399	-0.188	0.039	
Transferrin	M	1219	0.085	0.041	0.684	901	0.101	0.047	0.461
	F	1293	0.062	0.039		1398	0.056	0.039	
Saturation	M	1214	-0.236	0.038	0.052	899	-0.206	0.046	0.815
	F	1288	-0.133	0.037		1396	-0.192	0.038	
Log_{10}(ferritin)	M	1220	-0.111	0.039	0.245	902	-0.107	0.047	0.080
	F	1289	-0.050	0.035		1399	0.000	0.039	
Hb	M	1210	-0.100	0.039	0.180	1106	-0.064	0.033	0.659
	F	1258	-0.174	0.039		2082	-0.046	0.024	
MCV	M	1209	-0.11	0.042	0.615	1105	-0.083	0.041	0.285
	F	1258	-0.141	0.045		2076	-0.138	0.031	
					r1800562 (HFE)				
Iron	M	1214	0.304	0.072	0.748	902	0.447	0.086	0.048
	F	1288	0.336	0.069		1399	0.231	0.067	
Transferrin	M	1217	-0.653	0.077	0.194	902	-0.572	0.087	0.644
	F	1292	-0.515	0.073		1398	-0.623	0.068	
Saturation	M	1212	0.560	0.073	0.843	900	0.678	0.087	0.067
	F	1287	0.540	0.07		1396	0.477	0.067	
Log_{10}(ferritin)	M	1218	0.172	0.073	0.522	903	0.211	0.087	0.117
	F	1288	0.109	0.066		1399	0.038	0.068	
Hb	M	1210	0.200	0.074	0.788	1188	0.295	0.066	0.002
	F	1258	0.228	0.073		2263	0.044	0.049	
MCV	M	1209	0.030	0.08	0.086	1187	0.406	0.083	0.049
	F	1258	0.228	0.083		2255	0.200	0.064	
					rs3811647 (TF)				
Iron	M	1216	0.062	0.04	0.744	902	0.045	0.05	0.853
	F	1289	0.080	0.038		1399	0.057	0.041	
Transferrin	M	1219	0.500	0.043	0.139	902	0.425	0.051	0.258

	F	1293	0.413	0.04		1398	0.351	0.041	
Saturation	M	1214	-0.115	0.04	0.415	900	-0.101	0.05	0.851
	F	1288	-0.070	0.038		1396	-0.089	0.04	
Log_{10}(ferritin)	M	1220	-0.057	0.041	0.769	903	-0.097	0.051	0.130
	F	1289	-0.041	0.036		1399	0.002	0.041	
Hb	M	1210	0.010	0.041	0.569	1191	0.050	0.034	0.394
	F	1258	-0.023	0.041		2279	0.014	0.025	
MCV	M	1209	-0.060	0.044	0.683	1190	-0.015	0.043	0.444
	F	1258	-0.034	0.046		2271	-0.056	0.032	

Supplementary Table 3. Correlations between the mean values (from up to four visits) for each of the phenotypes in adolescents (in parentheses are numbers of subjects).

	Iron	**Transferrin**	**Transferrin Saturation**	**Ferritin (log_{10})**	**Hb**	**MCV**
Iron	(2971)					
Transferrin	0.01^{NS} (2970)	(2984)				
Transferrin Saturation	0.92*** (2969)	-0.35*** (2969)	(2969)			
Ferritin (log_{10})	0.11*** (2968)	-0.32*** (2982)	0.20*** (2967)	(2982)		
Hb	0.16*** (2264)	-0.02^{NS} (2263)	0.15*** (2263)	0.17*** (2262)	(2521)	
MCV	0.21*** (2262)	-0.15*** (2261)	0.24*** (2261)	0.16*** (2260)	0.13*** (2519)	(2520)

*** $P < 0.001$

Supplementary Methods

Subjects

Australian adolescent participants comprised twins and their non-twin siblings living in south-east Queensland. Most (98% by self-report) are of mixed European ancestry (mainly British Isles). The twins were invited to participate in a melanoma risk factors study, as close as possible to their 12th and 14th birthdays; and a cognition study, as close as possible to their 16th birthday. A small proportion returned for later studies at age 18 or more. Non-twin siblings (aged 10−18) were asked to participate at the same time. 2,984 subjects had relevant phenotype data and most participated in more than one study (Descriptive statistics are presented in **Supplementary Table 1A)**. Blood samples were collected at the end of testing sessions from participants and if possible from their parents; biochemical analyses were performed on blood serum from the adolescent twins and their siblings, but not from the parents. Pedigree relationship and zygosity were confirmed by genotype data.

Australian adult participants were from twin-family studies on alcohol use and its biological consequences, based on two cohorts of twins enrolled in the Australian Twin Registry. The first (Cohort 1) were born before 1964. Cohort 2 were born between 1964 and 1972. As before, twins were unselected with regard to personal or family history of alcoholism or other psychiatric or medical disorders. These twin cohorts had participated in previous postal questionnaire and telephone interview studies, and recruitment was extended through invitations to their parents, siblings, adult children and spouses. There were 917 men and 1414 women with relevant phenotype and genotyping data, aged 48.2 ± 11.4 years (range: 18-84 years) (Descriptive statistics are presented in **Supplementary Table 1B**).

The Dutch participants were drawn from the GAIN-MDD study, which is a case-control study of major depressive disorder in unrelated individuals aged 18-60 years[1].

Subjects eligible for inclusion in the GAIN-MDD study came from two Dutch longitudinal projects: the Netherlands Study of Depression and Anxiety (NESDA, www.nesda.nl), and from the Netherlands Twin Registry (NTR, www.tweelingenregister.org) [2]. Before the start of the NESDA and NTR biological sample collection, processing and storage protocols were harmonized. Blood sampling for the NESDA participants took place during the baseline visit and was done between 0830–0930 h at one of the seven field sites. For NTR, biological samples were taken at the respondents' home between 0700 and 1000 h. Average age of participants (65% female) was 43.8 years (SD = 13.6) (Descriptive statistics are presented in **Supplementary Table 1B**).

Participants (and where appropriate their parents or guardians) gave informed consent to participation, and all studies were approved by appropriate ethics committees.

Laboratory methods

Blood samples from all participants were collected and serum was separated from the blood and stored at –70 ºC until analyzed. Serum iron, transferrin and ferritin were measured using Roche methods on a Hitachi 917 Analyzer. Transferrin saturation (%) with iron was calculated from iron and transferrin concentrations. Hb and MCV were measured on whole blood using a Coulter Model STKS blood counter. DNA was extracted from blood samples and genotyped with Illumina 610K chips (for the adolescents) or 370K chips (for the Australian adults). SNPs were included in the analyses if they met the following conditions: Hardy-Weinberg Equilibrium test $P \geq 10^{-6}$, minor allele frequency ≥ 0.01, call-rate ≥ 0.95 and the mean value of GenCall score ≥ 0.7.

Genotyping for 600,000 single nucleotide polymorphisms in the GAIN-MDD study was conducted by Perlegen Sciences (Mountain View, CA, USA) with the use of a set of four

proprietary, high-density oligonucleotide arrays. The SNP quality-control process is described in detail in Sullivan et al [3]. A total of 427,037 SNPs on Chromosome 1 to 22 met QC criteria.

Data analysis

Subjects found to be of non-European ancestry by principal components analysis (EIGENSTRAT)[4] of the genotyping data were also excluded. The principal component analysis used a set of 276,891 autosomal SNPs that were common to Australian samples, HapMap 3 (11 global populations) and 5 Northern European (Denmark, Finland, the Netherlands, the United Kingdom and Sweden) populations from the GenomEUtwin Consortium (http://www.genomeutwin.org). We excluded 104 individuals who were more than 2 standard deviations from the mean of principal component (PC) 1 and 2 derived from the European populations.

For each trait, the distribution of the phenotype was checked for the deviation from normality. Ferritin was not normally distributed, so we performed a $\log_{10}$ transformation. Next, we adjusted the phenotypes for the effects of sex, age, sex*age (and age^2, sex*age^2 in the adolescent data, except for Hb and MCV). The standardized residuals were retained and used for the association analyses. For any given phenotype, we removed outlying individuals who were more than 5 standard deviations from the mean. For the adolescent data, where up to 4 measurements were available, the means were used. Association analyses were conducted in MERLIN[5] using the score test (--fastAssoc option), which takes account of family relationships. For the top associated SNPs, the association results were confirmed using a more computationally intensive variance-component likelihood ratio test, which accounts for IBD sharing (--assoc option in MERLIN). The analyses of the Dutch data, which is composed of unrelated individuals, was performed using a linear regression in PLINK[6].

Combined effect sizes and *P* values from the adolescent and adult data were estimated by meta-analysis of the effect size weighted by inverse variance for each study in METAL (www.sph.umich.edu/csg/abecasis/metal/). The evidence for heterogeneity in effect sizes in the adolescent and adult data were also tested in METAL. Visualization and annotation of the GWAS results were performed in WGA Viewer[7]. Other analyses (correlations, estimation of genotypic means) were done using SPSS.

References

1. Boomsma, D.I. et al. *Eur J Hum Genet* **16**, 335-342 (2008).
2. Penninx, B.W. et al. *Int J Methods Psychiatr Res* **17**, 121-140 (2008).
3. Sullivan, P.F. et al. *Mol Psychiatry* **14**, 359-375 (2009).
4. Price, A.L. et al. *Nat Genet* **38**, 904-909 (2006).
5. Chen, W.M. & Abecasis, G.R. *Am J Hum Genet* **81**, 913-926 (2007).
6. Purcell, S. et al. *Am J Hum Genet* **81**, 559-575 (2007).
7. Ge, D. et al. *Genome Res* **18**, 640-643 (2008).