SUPPLEMENTARY INFORMATION

doi:10.1038/nature11677

Page

Supplementary Figures

Supplementary Note

Results of genome-wide association and replication testing

Genome-wide association was done amongst 62,553 people of European ancestry and 9,308 people of South Asian ancestry from 30 separate studies, using 2,644,161 autosomal and 67,645 X chromosome SNPs (**Supplementary Tables 1-2**). Data for Europeans and South Asians were analysed separately, followed by combined analysis of results for the two populations. Genomic control inflation factors are shown in **Supplementary Table 3.**

There were 2,484 SNPs associated with one or more red blood cell phenotype at $P<10^{-8}$ amongst Europeans (**Figure 1**), these were distributed between 63 genomic loci. We found a further 17 loci with SNPs showing suggestive evidence of association to red blood cell phenotypes ($P>10^{-8}$ and $P<10^{-7}$); at these loci we identified the SNP with the lowest P value against any trait and carried out additional replication testing using a combination of in-silico data and direct genotyping amongst 63,506 people of European ancestry (**Supplementary Table 1**). At 8 of the 17 loci the lead SNP showed replication ($P<0.05$ after Bonferroni correction for multiple testing, and $P<1x10^{-8}$ in combined analysis with their discovery GWA data). Taken together the genome-wide and replication data from Europeans identified 71 loci associated with red blood cell phenotypes at $P<10^{-8}$ (**Table 1**).

In the genome-wide association study of 9,308 people of South Asian ancestry, we found 43 SNPs associated with red blood cell phenotypes at $P<10^{-7}$, these were located at 6 genomic loci already identified in the Europeans. We found evidence for shared genetic effects between Europeans and South Asians at both known and novel loci. At the 59 loci associated with red blood cell phenotypes in Europeans that were successfully genotyped or imputed in South Asians, 49 showed directional consistency and 20 showed nominal replication in South Asians, with little evidence for heterogeneity of effect between the two population groups.

We therefore carried out a final meta-analysis of genome-wide association results from the two populations and identified five genomic loci associated with RBC traits at $P<10^{-7}$, that had not been not found in the separate European or South Asian specific analyses (**Table 1**). Four of these additional five loci all replicated in further testing ($P<0.05$ in replication samples, and $P<10^{-8}$ in combined analysis with GWA data).

Genome-wide significance and correction for multiple phenotypes

Our choice of statistical threshold was grounded on the guidelines derived from studies of the ENCODE regions which suggest that $P<5x10^{-8}$ is the appropriate threshold for genome-wide significance in Europeans, but was designed to provide us additional adjustment for the multiple phenotypes tested. The six red blood cell parameters studied are inter-related: correlation coefficients between the phenotype pairs range from r=0.07 to 0.96, and between their respective genome-wide association test results from r=0.14 to 0.79 (**Supplementary Table 23**). Based on this correlation matrix, a simple Bonferroni correction for six phenotypes is overly conservative.

To control for the multiple testing of the 6 phenotypes while accounting for the correlation between them, we first performed an eigenvalue decomposition of the correlation matrix (**Supplementary Table 23**) of the phenotypes and used the variance of the eigenvalues to estimate the effective number of independent phenotypes tested;[1] this indicated that the phenotypes correspond to approximately 4.7 independent phenotypes. We then used permutation testing to provide a further estimate of the appropriate correction for multiple phenotypes. The genome-wide association study for association of SNPs with red blood cell traits was run 100 times in the LOLIPOP study EW610 and IA160 samples, with randomisation of genotype and phenotype data to simulate expectations under the null hypothesis. The minimum P value (P_{min}) for association with any phenotype was determined for each SNP, and after 100 runs, the number of SNPs with P_{min} reaching suggestive statistical significance determined ($P<10^{-6}$, $P<10^{-7}$ or $P<5x10^{-8}$). In the first 100 runs, the phenotype data was left intact to assess the number of associations expected under the null hypothesis for 6 related phenotypes ("Related"). In the second 100 runs, the phenotype data for the six red blood cell

traits were also randomised to assess the number of associations expected under the null hypothesis for 6 unrelated phenotypes ("Unrelated"). Results of permutation testing show that when phenotype correlations between the red blood cell traits are preserved, the number of SNPs reaching high levels of statistical significance is ~5/6 the number found when using randomised, unrelated phenotype data (**Supplementary Table 24**). This was true amongst both Europeans and South Asians. Our findings indicate that studying 6 related red blood cell traits is equivalent to analysis of 5 independent phenotypes. Based on these observations we therefore adopt $P<1.0x10^{-8}$ to indicate genome-wide significance, to provide correction for the effective number of independent phenotypes studied.

Replication of previously published findings.

Of the 38 loci previously reported to be associated with red blood cell traits,[2-6] we replicate 32 loci at $P<10^{-8}$, and a further 3 at $P<0.05$ (**Supplementary Table 6**). There are three loci reported to be associated with red blood cell phenotypes that did not replicate in our sample; all three were discovered in an East Asian GWAS[5]. The 3 SNPs have similar allele frequency amongst Europeans and East Asians suggesting that our findings do not result from loss of power. SNP rs7843479 is in moderate LD ($r^2=0.25$ [CEU], 0.72 [CHB/JPT]) with rs10503716 that is associated with MCV in the present study ($P=5.0x10^{-7}$) consistent with a causal variant that is in LD with both these SNPs. For two variants (rs6684514 and rs12127588), there are no SNPs nearby closely associated with respective phenotype in Europeans (**Supplementary Figure 4**), suggesting either population specific genetic variants, or gene-environment interactions.

Simulations of RNAi silencing in *D. melanogaster*

To inform selection of a threshold for reporting an RNAi silencer model as affected, and to assess the statistical significance of our findings, we carried out permutation testing in a genome-wide phenotypic screen of 5,658 *D.melanogaster* genes carried using RNAi silencing (UE, JMP unpublished data). The genome-wide screen was carried out using the methodology described in the present study and in the same laboratory, with the exception that each line was scored once rather than twice.

We selected random sets of 121 human genes, identified their *D. melanogaster* orthologs, and counted the number of orthologs with a blood cell phenotype in the RNAi screen. This was repeated 1,000,000 times for each of the three possible calling-thresholds (1 to 3), to build up an expectation under the null hypothesis (**Supplementary Figure 7**). Next we determined which of the 121 candidate genes identified in the red blood cell GWAs had a blood cell phenotype in the genome-wide RNAi screen at each of the three calling thresholds. We found the 121 candidates genes to be enriched 2-3 fold for haematological phenotypes in the RNAi screen compared to mean observed in simulations of the null hypothesis, and that this enrichment was robust to the precise choice of threshold ($P<0.05$, **Supplementary Figure 7**).

Based on this evidence for enrichment we extended our analysis by creating RNAi silencer models for all *D. melanogaster* genes orthologous to the 121 candidate genes identified in the GWAS; specifically this allowed us i. to evaluate 24 candidate genes not studied in the genome-wide screen, and ii. to carry out measurements in duplicate, thereby improving the accuracy of results. We selected a phenotype score of >=2 to define abnormal blood cell phenotype in the RNAi models which revealed blood cell phenotypes for 19 of the candidate genes. This threshold was chosen to provide a balance of sensitivity and specificity. We additionally provide ortholog specific results (**Supplementary Table 26**) to enable our findings to be reassessed using more or less stringent approaches to calling.

We studied RNAi silencer models for 74 of the human candidate genes; 19 of these show a blood cell phenotype in Drosophila (Drosophila positive), 55 do not (Drosophila negative). Drosophila positive genes are ~50% more likely than Drosophila negative genes to have a phenotype in mammalian systems (5/19 vs 10/55, $P=0.44$). Our results are consistent with the view that the Drosophila positive genes are enriched for genes involved in blood cell formation.

Thalassaemia studies

We investigated whether genetic variation at the 75 loci identified might impact beta-thalassaemia phenotype, a genetic disorder characterised by defects in haemoglobin synthesis, anaemia and abnormal red blood cell indices, in both heterozygous carriers and affected individuals. Clinical severity of beta-thalassaemia is variable, ranging from severe transfusion dependent thalassaemia major to the mild thalassaemia intermedia, and is in part influenced by genetic modifiers.[7, 8]

First, we tested the association of each sentinel SNP with its respective discovery phenotype amongst 460 carriers for β-thalassaemia mutation, and 3,876 controls (without known β-thalassaemia mutations). We confirmed association of several of the sentinel SNPs with respective phenotype (**Supplementary Table 21**). There was little evidence for heterogeneity, although the association of rs17616316 (*EIF5*) with MCH was ~10-fold stronger amongst the beta-thalassaemia heterozygotes than amongst controls (heterogeneity $P=5.3x10^{-4}$, **Supplementary Table 21**). We then analysed 495 β-thalassaemia patients (375 β0/β0, 80 β0/β+ and 40 β+/β+) from the general Italian population, to assess the contribution of associated loci in anticipating the age of first transfusion. Evidence for association (P=0.01) was detected at SNP rs9386796, within the *CCDC162P* gene, where the allele associated with increased MCH levels anticipates the age at first transfusion, indicator of greater clinical severity. The functional role of this locus is at present unclear and requires additional replication in independent β-thalassemia cohorts. We also found that the weighted genetic risk score predicted time to first transfusion ($P=6.9x10^{-4}$). However this was determined entirely by genetic variation at the *MYB-HBS1L* locus,[9] and variation at the other 74 loci did not independently predict time to transfusion (P=0.17).

GWAS cohort methods

ALSPAC: Avon Longitudinal Study of Parents and Children. ALSPAC is a population-based birth cohort study consisting initially of over 13,000 women and their children recruited in the county of Avon, UK in the early 1990s[10]. Both mothers and children have been extensively followed from the 8th gestational week onwards using a combination of self-reported questionnaires, medical records and physical examinations. Biological samples including DNA have been collected for 10121 of the children from this cohort. Ethical approval was obtained from the ALSPAC Law and Ethics committee and relevant local ethics committees, and written informed consent provided by all parents. Haemoglobin levels were measured using the Haemocue system using blood collected from a 7.5ml EDTA tube.

Amish. The Old Order Amish individuals included in this study were participants of several ongoing studies of cardiovascular health carried out at the University of Maryland. Participants were relatively healthy volunteers from the Old Order Amish community of Lancaster County, Pennsylvania and their family members[11, 12]. Examinations were conducted at the Amish Research Clinic in Strasburg, PA. The Institutional Review Board at the University of Maryland approved all protocols and informed consent was obtained, including permission to use their DNA for genetic studies. Study participants were enrolled within the 2000-2008 time period. Of the total phenotyped participants, a total of 1578 had CBC measures (Quest Diagnostics, Horsham, PA) and genotype information (Affymetrix 500K or 6.0). The clinical protocol used for blood collection and processing has been described in detail previously[11]. Briefly, venous blood samples from all participants were collected for haematological assessment and DNA genotyping. CBC processing was completed within 24h after venesection. Association analysis was performed using Mixed models Analysis for Pedigrees and Populations (MMAP) software developed by J.R. O'Connell (http://edn.som.umaryland.edu/mmap/index.php)

CoLaus: Cohorte Lausannoise. The design of the CoLaus study has been described previously[13]. Briefly, it is a population-based study conducted between 2003 and 2006, which

recruited over 6,000 subjects in Lausanne, Switzerland. The following inclusion criteria were applied: a) voluntary participation in the examination, including blood sample, b) aged 35-75 years, and c) Caucasian origin defined as having both parents and grand-parents Caucasian (determined by birth place). The Institutional Review Board of the Centre Hospitalier Universitaire Vaudois (CHUV) in Lausanne and the Cantonal Ethics Committee approved the study protocol and signed informed consent was obtained from participants. Starting in 2009 all participants were invited for a follow-up visit 5 years after the initial study (expected completion of the study 2012). This follow-up study was approved by the local ethics committee. During the follow-up visit, similar variables as in the cross-sectional study are measured with in addition a hemogram. The latter was measured on a haematology Sysmex XE2100 analyser (TOA Medical Electronics, Kobe, Japan) according to the manufacturer's indications.

DESIR. DESIR is a French cohort from the general-population: 716 individuals were genotyped, 178 men and 538 women[14]. Written informed consent was obtained from every participant to the study. Blood was anticoagulated with EDTA. Blood count measurements were performed using either a Technicon H3RTX (Bayer Diagnostics), Puteaux, France or a JT2 analyser (Beckman/Coulter), Roissy, France or an Argos from ABX, Montpellier, France.

EGCUT: Estonian Genome Center of University of Tartu. The EGCUT cohort is from the population-based biobank of the Estonian Genome Project of University of Tartu[15]. The project was conducted according to the Estonian Gene Research Act and all participants signed the broad informed consent15. The current cohort size is over 50,000, from 18 years of age and up, which reflects closely the age distribution in the adult Estonian population. Participants were randomly selected from individuals visiting GP offices or hospitals and were recruited by general practitioners (GP) and physicians. Each participant filled out a Computer Assisted Personal interview, which included personal data (place of birth, place(s) of living, nationality etc.), genealogical data (family history, three generations), educational and occupational history and lifestyle data (physical activity, dietary habits, smoking, alcohol consumption, women´s health, quality of life). Venous blood was anticoagulated with EDTA and FBCs were performed using XE2100 automated haematology analyser (Sysmex, Kobe, Japan).

EPIC. The EPIC Obesity study used a case-cohort design which included 1284 participant whose body mass index was above 30 and a random sample of 2566 from the EPIC-Norfolk Study, a population-based cohort study of 25663 men and women of European descent aged 39-79 years recruited in Norfolk, between 1993 and 1997[16]. Blood sample was taken during the day in the GPs' surgeries or EPIC clinic, were held overnight. Early the following morning, samples were collected from GP surgeries by technicians. Some assays were performed on fresh blood samples and the remaining blood was stored in straws. A 1 x 2ml EDTA sample provided blood for full blood count. Two x 10 ml citrated samples provided twelve straws of plasma, four straws of red blood cells plus preservation buffer and four straws of buffy coat and saline. A Coulter MD18 haematology analyser was used for the measurement of full blood counts. Quality controls were carried out on the Coulter scheme daily. In addition, the Haematology Department of Addenbrooke's Hospital included the EPIC Laboratory in a monthly quality control scheme.

GeneBank. GeneBank is a single site (Cleveland Clinic), hospital-based registry and connecting sample repository comprised of approximately 10,000 sequential consenting subjects undergoing elective cardiac evaluation through either coronary angiography or cardiac computed tomography. The GeneBank cohort has been used previously for discovery and replication of novel genes and risk factors for atherosclerotic CVD[17]. Subject recruitment into GeneBank occurred between 2001 and 2006 and provides an on-going focus for analysing the association of biochemical and genetic factors with coronary atherosclerosis in a consecutive cohort of patients undergoing elective cardiac evaluation. Enrolment criteria included stable patients undergoing elective coronary evaluation and the ability to give informed consent.

Extensive clinical, demographic, laboratory and angiographic data were collected from electronic medical records. Ethnicity information was self-reported. All patients provided written informed consent prior to being enrolled in GeneBank and the Institutional Review Board of the Cleveland Clinic approved the study. Fasting blood samples were collected prior to heparin administration for subjects undergoing elective diagnostic coronary angiography. Blood cell phenotypes were determined within 6 hr of blood draw using an ADVIA 2120 haematology analyser, which is a flow cytometry-based system that provides a complete blood cell count and a white blood cell differential.

INGI CARL: The INGI Carlantino cohort study. The Carlantino cohort includes approximately 1000 samples from an isolated village of Southern Italy (Carlantino) settled 5 centuries ago by few founders in a remote area[18]. Genealogical data are available since XVII° century. Participants have been deeply phenotyped (hundreds of quantitative and qualitative traits). After DNA extraction, genotyping data have obtained with High Density SNPs arrays from Illumina. Data have been imputed to HapMap map. Ethics approval was obtained from the Ethics Committee of the Burlo Garofolo children hospital in Trieste. Written informed consent was obtained from every participant to the study.

INGI-Cilento. The INGI-Cilento is a population-based study of isolated populations located in the area of the National Park of Cilento e Vallo di Diano[19]. A total of 2,137 individuals were available with FBCs. The study design was approved by the ethics committee of Azienda Sanitaria Locale Napoli 1. The study was conducted according to the criteria set by the declaration of Helsinki and each subject signed an informed consent before participating to the study. Blood was anticoagulated with EDTA and FBCs were performed using the automated particle counters Max M analyser (Coulter Electronics, Miami, USA) (on average within 24 hours from venesection).

INGI FVG: The INGI Friuli Venezia Giulia cohort study. The Friuli Venezia Giulia cohort is characterized by approximately 1700 samples from six isolated villages of Northern Italy (San Martino del Carso, Erto, Clauzetto, Sauris, Illegio, Resia)[20]. Isolation was in most cases due to a combination of a geographical barrier (mountains) plus a linguistic one. Participants have been deeply phenotyped (hundreds of quantitative and qualitative traits). After DNA extraction genotyping data have obtained with High Density SNPs arrays from Illumina (700K), and imputed using both HapMap and 1000 genome data. Ethics approval was obtained from the Ethics Committee of the Burlo Garofolo children hospital in Trieste. Written informed consent was obtained from every participant to the study.

INGI Val Borbera. The INGI-Val Borbera project was initiated in 2005 with the collection of phenotypic data from a geographically isolated population of North West Italy living in the Val Borbera Valley in Piedmont[21]. Inhabitants of the valley were invited to participate in the study by public advertisements through local authorities, televisions and newspapers as well as local physicians and mailings. Meetings were organized in all villages to present the project and its aims. The importance of the participation of entire families was underscored in all instances, nevertheless all people that volunteered to participate were included in the study, providing they had at least one grandparent from the valley. The study, including the overall plan and the informed consent form was reviewed and approved by the institutional review boards of San Raffaele Hospital in Milan and by the Regione Piemonte ethical committee. Information and biological samples were obtained from 1803 inhabitants between 18 and 102 years of age. 1664 DNAs were genotyped with the 370k Illumina chip. Only individuals aged 18 years or older were eligible. Venous blood measurements were done using either an SF3000 haematology analyser or a XE2100 haematology analyser (DASIT). The two instruments displayed no significant statistical differences in measurements range and association analyses were not adjusted for instruments type.

KORA F3. The study population for the KORA F3 GWAS was recruited from the KORA S3 survey (4,856 subjects, response 75%)[22]. It is an independent population-based sample from the general population living in the region of Augsburg, Southern Germany, examined in the years 1994/95 (KORA S3). The standardized examinations have been described in detail elsewhere 15. A total of 3,006 subjects participated in a follow-up examination of S3 in 2004/05 (KORA F3). For KORA F3 we selected 1,643 subjects of these participants then aged 35 to 79 years. Informed consent has been given. The local ethical committee has approved the study. DNA was extracted from fresh blood, and was stored at -80°C. FBCs were performed on fresh venous EDTA-anticoagulated blood using an automatic blood counter (Beckman Coulter STKS).

KORA F4. The KORA S4 survey, an independent population-based sample from the general population living in the region of Augsburg, Southern Germany, was conducted in 1999/2001[23]. The standardized examinations applied in the survey (4,261 participants, response 67%) have been described in detail elsewhere. A total of 3,080 subjects participated in a follow-up examination of S4 in 2006/08 (KORA F4). For KORA 1000K we selected 1,814 subjects of these participants. Informed consent has been given. The local ethical committee has approved the study. The KORA S3 and S4 samples do not overlap. DNA was extracted from fresh blood, and was stored at -80°C. FBCs were performed on fresh venous EDTA-anticoagulated blood using an automatic blood counter (Beckman Coulter LH 750).

LBC1921: Lothian Birth Cohort 1921. The LBC1921 cohort consists of 550 relatively healthy individuals, 316 females and 234 males, assessed on cognitive and medical traits at 79 years of age[24, 25]. They were born in 1921, most took part in the Scottish Mental Survey of 1932, and almost all lived independently in the Lothian region (Edinburgh City and surrounding area) in Scotland. When tested, the sample had a mean age of 79.1 years (SD = 0.6). A full description of participant recruitment and testing can be found elsewhere [25]. Ethics permission for the study was obtained from the Lothian Research Ethics Committee (LREC/1998/4/183). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent. Venous blood was collected in 2.7ml Sarstedt tubes and anticoagulated with EDTA. Full blood counts were performed on the same day using a Coulter LH 750 Haematology Analyser (Beckman Coulter Inc, Milton, UK).

LBC1936: Lothian Birth Cohort 1936. The LBC1936 consists of 1,091 relatively healthy individuals assessed on cognitive and medical traits at 70 years of age[26, 27]. They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland. The sample of 548 men and 543 women had a mean age 69.6 years (SD = 0.8). A full description of participant recruitment and testing can be found elsewhere [26]. Ethics permission for the study was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent. Venous blood was collected in 2.7ml Sarstedt tubes and anticoagulated with EDTA. Full blood counts were performed on the same day using a Coulter LH 750 Haematology Analyser (Beckman Coulter Inc, Milton, UK).

LifeLines. The LifeLines Cohort Study is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of The Netherlands[28]. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity. In addition, the LifeLines project comprises a number of cross-sectional sub-studies, which investigate specific age-related conditions. These include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function

and depression, and musculoskeletal conditions. Written informed consent was obtained from every participant. All participants are between 18 and 90 years old at the time of enrolment. Recruitment has been going on since the end of 2006, and until November 2011 over 62,000 participants have been included. Blood was drawn in BD tubes anticoagulated with EDTA. Blood count measurements were performed using a Sysmex XE2100.

LOLIPOP: London Life Sciences Population study. LOLIPOP is a population based cohort study of ~30,000 Indian Asian and European white men and women, aged 35-75 years, recruited from the lists of 58 General Practitioners in West London, UK[29]. Venous blood was anticoagulated with EDTA and transferred in 4 ml BD Vacutainer® Rapid Serum Tubes. Full blood counts were performed using a XE2100 automated haematology analyser.

MDC: Malmo Diet and Cancer study. A random sample of all men and women, born between 1923 and 1950 and living in Malmö, Sweden, were invited to participate in the Malmö Diet and Cancer Study, MDC[30]. Between March 1991 and September 1996, the respondents participated in several clinical examinations at the screening centre, and a self-administered questionnaire. The cohort consisted of 28,449 subjects (11,246 men and 17,203 women) from the eligible population of about 74,000 individuals. The regional ethics committee approved the MDC study. Participants provided informed consent. The cohort has been shown to be representative considering smoking and overweight, but with a higher mortality rate in non-participants.

A self-administered questionnaire was used to obtain information on smoking habits, diabetes, anti-hypertensive medication, marital status, education level and history of myocardial infarction. Smoking was divided into four different categories; smokers, former smokers, non-smokers and missing. Marital status was classified into two groups; unmarried (single, divorced, or widowed) or married (cohabiting). Educational level was divided into low (≤8 years), moderate (9 to 12 years), and high (college/university) levels. Blood pressure was measured using a mercury-column sphygmomanometer after 10 minutes of rest in the supine position. Hypertension was defined as systolic BP ≥140/90 mm Hg or use of anti-hypertensive medication. Body weight, height, and waist circumference were measured. Diabetes mellitus was defined as self-reported diabetes according to the questionnaire, and/or treatment with anti-diabetic medication. Total and differential blood cell count were analysed using a SYSMEX K1000 automatic counter (Sysmex Europe, Norderstedt, Germany). The analyses were performed consecutively at the time of the screening examination, at the central laboratory of Malmö Hospital, using fresh heparinised blood. The subjects included in the GWAS study were selected from the lower 9% of the blood pressure distribution.[31]

MICROS: Microisolates in South Tyrol. The MICROS study is part of the genomic health care program 'GenNova' and was carried out in three villages of the Val Venosta, South Tyrol (Italy), in 2001-2003.[32] Briefly, study participants were volunteers from three isolated villages located in the Italian Alps, in a German-speaking region bordering with Austria and Switzerland. Owing to geographical, historical and political reasons, the entire region experienced a prolonged period of isolation from surrounding populations. Information on the participant's health status was collected through a standardized questionnaire. Laboratory data were obtained from standard blood analyses Blood count measurements.

NESDA: Netherlands Study of Depression and Anxiety. NESDA is a multi-centre study designed to examine the long-term course and consequences of depressive and anxiety disorders (http://www.nesda.nl).[33] NESDA included both individuals with depressive and/or anxiety disorders and controls without psychiatric conditions. Inclusion criteria were age 18-65 years and self-reported western European ancestry, exclusion criteria were not being fluent in Dutch and having a primary diagnosis of another psychiatric condition (psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use disorder). Venous blood samples after overnight fast were obtained from participants at the baseline NESDA

sample, and transported to local laboratories in the three NESDA regions for analysis the same day. Red Blood Cell count was determined using the Cell-Dyn Sapphire (Abbott Diagnostics).

NFBC1966: The North Finland Birth Cohort of 1966. NFBC1966 was designed to study factors affecting preterm birth, low birth weight, and subsequent morbidity and mortality[34]. The longitudinal data collection includes clinical examination and blood sampling at age 31 years, from which data in the current study are drawn. The attendees in the follow-up (71% response rate) were adequately representative of the original cohort as is the final study sample in the present analyses. A total of 4,763 genotyped samples were available from the NFBC1966. Blood count measurements.

NTR: Netherlands Twin Register. Subjects were registered with the Netherlands Twin Register and took part in the NTR-Biobank project, which targeted an unselected group of Dutch families. Blood samples were taken at the respondents' home between 07.00 and 10.00 am. A haematology profile was obtained within 6h of blood collection using a Coulter instrument.[35] Participants gave informed consent and the study was approved by the Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam (IRB-2991 under Federalwide Assurance 3703; IRB/institute code NTR 03-180).

PREVEND: **Prevention of REnal and Vascular ENd stage Disease** study. This is an ongoing prospective study investigating the natural course of increased levels of urinary albumin excretion and its relation to renal and cardiovascular disease. Details of the protocol have been described elsewhere (www.prevend.org).[36] Blood samples were obtained in the morning hours. Red blood cell measurements were performed at the 2nd visit (~4.2 years from baseline). The drawn blood was anticoagulated with disodium-ethylenediamine tetra-acetic acid and tested with a Coulter Counter (Beckman-Coulter, Fullerton, CA). For 3121 subjects and Genome Wide Data were available for this study. Replication was performed in an additional 2939 subjects.

QIMR. FBC were obtained from 2,538 adolescent twins and their siblings from 1,089 Australian families ascertained from the general population. Twins were enlisted through primary schools, media appeals and by word of mouth and tested longitudinally as close as possible to their twelfth, fourteenth and sixteenth birthdays in the context of an on-going study of melanocytic naevi. Participants (and where appropriate their parents or guardians) gave informed consent to participation, and all studies were approved by appropriate ethics committees. The clinical protocol used for blood collection and processing has been described in detail previously [37]. Briefly, venous blood samples from the twins and, where possible, from their parents and siblings, were collected for haematological assessment (twins and sibs only) and DNA genotyping. FBC were obtained within 24h after venesection using a Coulter (Model STKS) instrument. For each trait, outlier observations (6 SD above the mean) at each time of assessment (ages 12, 14 and 16) were excluded from analysis and the average across all available time points was computed.

SardiNIA. We recruited and phenotyped 6,148 individuals, males and females, ages 14–102 yr, from a cluster of four towns in the Ogliastra province of Sardinia. Both, the IRB at NIA and the Italian ethical Committee approved the study protocol and all participants provided a written informed consent. During physical examination, a blood sample was collected from each individual, and divided into two aliquots; one was used for genomic DNA extraction and the second aliquot to characterize several blood phenotypes as previously described[38]. Among the recruited samples, about 13% was carrier of the beta039 mutation, according to estimates for the Sardinian population[39].

SHIP: The Study of Health in Pomerania. SHIP is a longitudinal population-based cohort study conducted in West Pomerania, the north-east area of Germany[40]. Only individuals with

German citizenship and main residency in the study area were included. The baseline net sample comprised 6,265 eligible subjects, aged 20 to 79 years, out of which 4,308 participated at baseline (response 68.8%). In total 3300 persons took part in the follow-up examinations (83.5% of eligible persons), conducted between 2002 and 2006. Follow-up data from 3183 subjects was available for the present analyses. Non-fasting blood samples were taken in the supine position. The blood count was measured within 60 minutes. Samples were analysed with a Sysmex SE-9000 analyser (Sysmex, Hamburg, Germany). The analysers were calibrated and maintained according to the manufacturer's instructions. Quality control was performed internally daily as well as externally by participating in external proficiency testing programmes.

Sorbs. All subjects are part of a sample from an extensively phenotyped self-contained population from Eastern Germany, the Sorbs[41]. Sampling comprised unrelated subjects as well as families. Extensive phenotyping included standardised questionnaires for past medical history and family history, collection of anthropometric data and a 75g oral glucose tolerance test. The study was approved by the ethics committee of the University of Leipzig and all subjects gave written informed consent before taking part in the study. Venous EDTA blood samples were analysed by use of the haematology automated analyser Sysmex XE-2100.

TwinsUK. The TwinsUK cohort is an adult twin British registry shown to be representative of singleton populations and the United Kingdom population[42]. A total of 1,763 twins (100 % females) were available with FBCs. Ethics approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant to the study. Venous blood was anticoagulated with EDTA and FBCs were performed using either an ADVIA 2120 Haematology System (Siemens Healthcare Diagnostics, Deerfield, IL, US) or a XE2100 automated haematology analyser (Sysmex, Kobe, Japan) (on average within 24 hours from venesection (range 20 - 30 hrs). The two instruments displayed differences in measurements range, with means (SD) of 9.69 (0.96) and 11.17 (1.03) respectively. Hence association analyses were adjusted for instrument type.

UKBS-CC: UK Blood Services Common Controls. The UKBS collection is a national control collection of shared controls for GWAS and was established as part of the Wellcome Trust Case Control Consortium. Full blood counts were measured on a Beckman-Coulter instrument. Measurements were performed between 16-24 hours after phlebotomy.

Replication and population variation cohort methods

CBR: Cambridge BioResource. CBR is a collection of pseudo-anonymised DNA samples from 8,000 healthy blood donors that has been established in 2008 and 2010 by the NIHR funded Cambridge Biomedical Research Centre in collaboration with NHS Blood and Transplant for use in genotype-phenotype association studies [43]. Four thousand donors each were enrolled during 2007 and 2009. Full blood counts (FBCs) were obtained from EDTA anticoagulated samples of blood drawn from the pouches of the donation collection sets. FBCs performed on an ABX Pentra 60 automated haematology analyser (ABX Diagnostics, Montpellier, France) or on a Sysmex XE-2100. For the purpose of calibration measurements, 500 blood samples were performed on both the Beckman-Coulter and Sysmex instruments. Measurements were performed between 16-24 hours after phlebotomy.

deCODE: Red blood parameters were measured in samples from Icelanders at the Landspitali University Hospital Laboratory or at the Icelandic Medical Center (Laeknasetrid) Laboratory in Mjodd (RAM), between the years 1990 and 2010. The measurements were normalized to a standard normal distribution using quantile-quantile normalization and then adjusted for sex, year of birth and age at measurement. For individuals for which more than one measurement was available we used the average of the normalized value.

LURIC: Ludwigshafen Risk and Cardiovascular Health Study. The LURIC Study is a prospective cohort study among 3,316 study participants who were routinely referred to a tertiary care medical centre in south-west Germany between 1997 and 2000[44]. Inclusion criteria were the availability of a coronary angiogram, German ancestry and clinical stability with the exception of acute coronary syndromes (ACS). Exclusion criteria were any acute illness other than ACS, any chronic disease where non-cardiac disease predominated and a history of malignancy within the past five years. Patients were continuously followed up with respect to fatal events. Genotyping was carried out using Affymetrix 500K or 6.0 arrays, excluding SNPs or samples with call rates <90%, gender discrepancy or relatedness. Imputation of missing HapMap2 genotypes was done using MACH.

OGP: Ogliastra Genetic Park. OGP is a population-based epidemiologic survey carried out in 10 Ogliastra villages (Baunei, Escalaplano, Loceri, Perdasdefogu, Seui, Seulo, Talana, Triei, Urzulei and Ussassai) between 2002 and 2008[45]. Mitochondrial analysis traced the original population back to the Neolithic era and showed that Ogliastra inhabitants rank among the most genetically homogenous European population and that they have the lowest values of mtDNA gene diversity with respect to other Sardinia areas. People living in the villages were invited to take part in the study by means of information campaigns and letters sent to residents. Blood count measurements were performed using Coulter LH Haematology analyser (Beckman-Coulter, Brea, CA). For each inhabitant we collected genealogical information dating back to the seventeenth century, medical and pharmacology history data and family history of many disease. Written informed consent was obtained from every participant in the study. DNA for wet-lab genotyping to replicate discovery results of the current study was available in a total of 9,704 OGP participants.

SMART: The Secondary Manifestations of ARTerial disease study. SMART is a prospective outpatient cohort study among patients aged 18-74 years newly referred to the University Medical Center Utrecht, The Netherlands, because of atherosclerotic vascular disease or for treatment of atherosclerotic risk factors [46, 47]. The objective of SMART is to determine the prevalence of concomitant asymptomatic arterial disease and risk factors in patients presenting with a manifestation of arterial disease or risk factor, and to study the incidence of future cardiovascular events and their predictors in these high-risk patients. DNA for wet-lab genotyping to replicate discovery results of the current study was available in a total of 8,361 SMART participants. Wet-lab genotyping for single nucleotide polymorphism (SNP) analysis was carried out by KBiosciences, Hertfordshire, UK. (www.kbioscience.co.uk), whose personnel were blinded to patient status, using their proprietary KASPar PCR technique and Taqman Genotype calling was carried out using an automated system, the results of which were checked manually by study personnel using SNPviewer software.

Young Finns. The Young Finns cohort is a Finnish longitudinal population study sample on the evolution of cardiovascular risk factors from childhood to adulthood[48]. The first cross-sectional study was conducted in the year 1980 in five different centres. It included 3,596 participants in the age groups of 3, 6, 9, 12, 15, and 18, who were randomly chosen from the national population register. After the baseline in 1980 these subjects have been re-examined in 1983 and 1986 as young individuals, and in 2001, 2007 and 2011 as older individuals. For the current analysis a subsample from the latest (2011) follow-up was used from Tampere and Turku (N=616, aged 33-48) where the RBC measurements were available. This study was carried out in accordance with the recommendations of the Declaration of Helsinki. All participants provided written informed consent and the study protocol was approved by the Ethics Committee. Venous blood samples were obtained and anticoagulated with EDTA. RBC parameters were measured by flow cytometric particle counting (cells) and photometry (Hb) using Sysmex XE-5000 and XT-2000i analysers (Sysmex Corporation, Kobe, Japan) with reagents provided by the manufacturer (Cellpack and Sulfolyser). The analysers were accredited by Finnish Accreditation Service (FINAS).

Acknowledgements

ALSPAC. We are extremely grateful to all the individuals who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank 23andMe, the Wellcome Trust Sanger Institute, and the Laboratory Corporation of America, for generating the ALSPAC GWA data. The UK Medical Research Council (grant 74882), the Wellcome Trust (grant 076467) and the University of Bristol provide core support for ALSPAC. DME is supported by a Medical Research Council New Investigator Award (MRC G0800582 to DME). JPK was funded by a Wellcome Trust 4-year PhD studentship in molecular, genetic, and life course epidemiology (WT083431MA).

Amish. We thank our Amish research volunteers for their long-standing partnership in research, and the research staff at the Amish Research Clinic for their hard work and dedication. We are supported by grants and contracts from the NIH including R01 AG18728 (Amish Longevity Study), R01 HL088119 (Amish Calcification Study), U01 GM074518-04 (PAPI Study), U01 HL072515-06 (HAPI Study), U01 HL084756 and NIH K12 RR023250 (University of Maryland MCRDP), the University of Maryland General Clinical Research Center, grant M01 RR 16500, National Research Initiative Competitive Grant no. 2007-35205-17883 from the USDA National Institute of Food and Agriculture, the Baltimore Veterans Administration Medical Center Geriatrics Research and Education Clinical Center and the Paul Beeson Physician Faculty Scholars in Aging Program.

ARIC. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

CBR. The project made use of NHSBT donors from the Cambridge BioResource (http://www.cambridgebioresource.org.uk/). This local resource for genotype-phenotype association studies is supported by a grant from the National Institute for Health Research (NIHR) to the Cambridge Biomedical Research Centre (which supported JS and SM), and by a NIHR programme grant to NHSBT (RP-PG-0310-1002, which supported AA, AR and WHO).

CoLaus. The CoLaus study was supported by research grants from the Swiss National Science Foundation (grant no: 33CSCO-122661) from GlaxoSmithKline and the Faculty of Biology and Medicine of Lausanne, Switzerland. The authors also express their gratitude to the participants in the Lausanne CoLaus study and to the investigators who have contributed to the recruitment, in particular the research nurses of CoLaus: Yolande Barreau, Anne-Lise Bastian, Binasa Ramic, Martine Moranville, Martine Baumer, Marcy Sagette, Jeanne Ecoffey, Sylvie Mermoud.

deCODE. The research performed at deCODE genetics was partly funded through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007- 201413.

DESIR. This work was funded by the French Government ("*Agence Nationale de la Recherche*"), the French Region of "*Nord Pas De Calais*" ("*Contrat de Projets État-Région*"), and the charities: "*Association Française des Diabétiques*", "*Programme National de Recherche*

sur le Diabète" and "*Association de Langue Française pour l'Etude du Diabète et des Maladies Métaboliques*"; as well as by Canadian funding agencies (*"Genome Canada"* and *"Génome Québec"*, and *"Fonds de recherche du Québec – Santé"*). The DESIR study has been supported by contracts with CNAMTS, Lilly, Novartis Pharma and Sanofi-Aventis, by INSERM ("*Réseaux en Santé Publique, Interactions entre les déterminants de la santé*", *"Cohortes Santé TGIR 2008"*), by the "*Association Diabète Risque Vasculaire*", "*Fédération Française de Cardiologie*", "*Fondation de France*", ALFEDIAM, ONIVINS, *"Société Francophone du Diabète",* and by Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche, Topcon. The D.E.S.I.R. Study Group: INSERM U1018: B. Balkau, P. Ducimetière, E. Eschwège; INSERM U367: F. Alhenc-Gelas; CHU D'Angers: Y. Gallois, A. Girault; Bichat Hospital: F. Fumeron, M. Marre, R. Roussel; Centre Hospitalier Universitaire de Rennes: F. Bonnet; Centre National de la Recherche Scientifique UMR 8090, Lille: S. Cauchi, P. Froguel; Medical Examination Services: Alençon, Angers, Blois, Caen, Chartres, Chateauroux, Cholet, Le Mans, Orléans and Tours; Research Institute for General Medicine: J. Cogneau; General practitioners of the region; Inter-Regional Institute for Health (IRSA): C. Born, E. Caces, M. Cailleau, N Copin, J. G. Moreau, F. Rakotozafy, J. Tichet, S. Vol. We are sincerely indebted to all subjects who participated in this study.

EGCUT. EGCUT received financing by (ENGAGE, OPENGENE), targeted financing from Estonian Government SF0180142s08, Estonian Research Roadmap through Estonian Ministry of Education and Research (3.2.0304.11-0312), Center of Excellence in Genomics (EXCEGEN) and Development Fund of University of Tartu (SP1GVARENG). A. Salumets was supported by the Estonian Ministry of Education and Science (Targeted project no. SF0180044s09) and Enterprise Estonia (Grant no. EU30020). We acknowledge EGCUT personnel, especially Mr V. Soo. Data analyzes were carried out in part in the High Performance Computing Center of University of Tartu.

EPIC. The EPIC-Norfolk study is supported by research programme grant funding from Cancer Research UK and the Medical Research Council**.**

GeneBank. The Cleveland Clinic GeneBank study is supported by NIH grants P01HL076491, P01HL098055, R01HL103866, R01HL103931, P20HL113452, and a gift from the Leonard Krieger Fund.

INGI-CARL. We thank Laura Esposito, Angela D'Eustacchio and Emmanouil Athanasakis for technical support. We are very grateful to the municipal administrators for their collaboration on the project and for logistic support. We would like to thank all participants to this study.

INGI-Cilento. We would like to address a special thank to the Cilento populations. This work was supported by grants from the Italian Ministry of Universities (FIRB -RBIN064YAT, RBNE08NKH7), the Assessorato Ricerca Regione Campania, the Ente Parco Nazionale del Cilento e Vallo di Diano, the Fondazione Banco di Napoli to MC.

INGI-FVG. We thank Laura Esposito, Angela D'Eustacchio and Emmanouil Athanasakis for technical support. We are very grateful to the municipal administrators for their collaboration on the project and for logistic support. We would like to thank all participants to this study. Project granted by FVG – Regional Government, and Fondo Trieste.

INGI-Val Borbera. We thank the inhabitants of the Val Borbera that made this study possible, the local administrations, the Tortona and Genova archdiocese and the ASL-Al, Novi Ligure (Al) for support. We particularly thank Clara Camaschella (Vita e Salute University, Milano) who supervised data collection and phenotype analysis, Federico Caligaris-Cappio (Oncology-Hsr) and the internal medicine MDs of the San Raffaele Hospital, Maria Rosa Biglieri, Diego Sabbi and Silvia Bione (IGM,CNR) for their contribution to clinical data collection. We thank Massimo

Ginestri (Asl-Al) for help with laboratory data analysis and Fiammetta Viganò for technical support. The research was supported by funds from Compagnia di San Paolo, Torino, Italy; Fondazione Cariplo, Italy and Ministry of Health, Ricerca Finalizzata 2008 to DT.

KORA F3 and **KORA F4**. The KORA research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823). KORA research was also supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ.

LBC1921 and **LBC1936**. We thank the LBC1936 and LBC1921 participants and research team members. We thank the nurses and staff at the Wellcome Trust Clinical Research Facility, where subjects were tested and the genotyping was performed. We also thank the staff at the Department of Haematology, Western General Hospital for the hematology measurements. We thank the staff at the Lothian Health Board, and the staff at the SCRE Centre, University of Glasgow. The whole genome association study was funded by the Biotechnology and Biological Sciences Research Council (BBSRC). The LBC1936 research was supported by a programme grant from Research Into Ageing and continues with programme grants from Help the Aged/Research Into Ageing (Disconnected Mind). The LBC1921 data collection was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC), Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC) and Medical Research Council (MRC) is gratefully acknowledged. Lorna M. Lopez is the beneficiary of a post-doctoral grant from the AXA Research Fund.

LifeLines. The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. We thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Pim van der Harst, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and Medical Biobank Northern Netherlands, and the participating general practitioners and pharmacists. LifeLines Scientific Protocol Preparation: Rudolf de Boer, Hans Hillege, Melanie van der Klauw, Gerjan Navis, Hans Ormel, Dirkje Postma, Judith Rosmalen, Joris Slaets, Ronald Stolk, Bruce Wolffenbuttel; LifeLines GWAS Working Group: Behrooz Alizadeh, Marike Boezen, Marcel Bruinenberg, Noortje Festen, Lude Franke, Pim van der Harst, Gerjan Navis, Dirkje Postma, Harold Snieder, Cisca Wijmenga, Bruce Wolffenbuttel.

LOLIPOP. The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the NIHR Cardiovascular Biomedical Research Unit of Royal Brompton and Harefield NHS Foundation Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z) the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51). We thank the participants and research staff who made the study possible.

LURIC. LURIC received funding through the 6th Framework Program (integrated project Bloodomics, grant LSHM-CT-2004-503485) and 7th of Framework Program (integrated project AtheroRemo, Grant Agreement number 201668) of the European Union. The authors extend appreciation to the participants of the LURIC study. We thank the LURIC study team either temporarily or permanently involved in patient recruitment, sample and data handling, and the laboratory staff at the Ludwigshafen General Hospital and the Universities of Freiburg and Ulm, Germany.

MDC. O Melander was supported by grants from the European Research Council (StG-282255), Swedish Medical Research Council, the Swedish Heart and Lung Foundation, the Medical Faculty of Lund University, Malmö University Hospital, the Albert Påhlsson Research Foundation, the Crafoord Foundation, the Ernhold Lundströms Research Foundation, the Region Skane, the Hulda and Conrad Mossfelt Foundation, the King Gustaf V and Queen Victoria Foundation, the Lennart Hanssons Memorial Fund, and the Wallenberg Foundation. G Engström was supported by grants from the Swedish Medical Research Council (521-2011-3891), the Swedish Heart and Lung Foundation (20100288), funds from Malmö University Hospital, the Ernhold Lundströms Research Foundation, the Region Skane (LU-ALF). B Hedblad was supported by grants from the Swedish Medical Research Council, the Swedish Heart and Lung Foundation, funds from the Skåne University Hospital, the Region of Skåne and the Ernhold Lundströms Research Foundation. The GWAS study of the MDC cohort was supported by the following grants: British Heart Foundation (BHF) Chair CH/98001 to AFD; BHF Programme RG/07/005/23633 to AFD, SP; BHF Special Project Grant SP/08/005/25115 to AFD, SP; SP was supported by an intermediate research fellowship and a travel fellowship from the BHF (FS/05/095/19937; FS/10/016/28162).

MICROS. We thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano, and the South Tyrolean Sparkasse Foundation.

NESDA. NESDA was supported by the Geestkracht program of ZonMW [grant 10-000-1002]; matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ 25 Lentis, GGZ Friesland, GGZ Drenthe). Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis were supported by grants from GAIN and the NIMH (MH081802), the Neuroscience Campus Amsterdam, and the Center for Molecular and Systems Biology (CMSB). Genotype data were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1.

NFBC1966. We acknowledge the support of US National Heart, Lung, and Blood Institute grant HL087679 through the STAMPEED program, grants MH083268, GM053275-14 and U54 RR020278 from the US National Institutes of Health, grant DMS-0239427 from the National Science Foundation, the Medical Research Council of the UK, EURO-BLCS, QLG1-CT-2000-01643 and the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project and grant agreement HEALTH-F4-2007-201413. The authors would like to thank the Center of Excellence in Common Disease Genetics of the Academy of Finland and Nordic Center of Excellence in Disease Genetics, the Sydantautisaatio (Finnish Foundation of Heart Diseases), the Broad Genotyping Center, D. Mirel, H. Hobbs, J. DeYoung, P. Rantakallio, M. Koiranen and M. Isohanni for advice and assistance.

NTR. Funding was obtained from the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW): Genotype/phenotype database for genetic epidemiological studies (NWO 911-09-032);Spinozapremie (SPI 56-464-14192); VENI-016.115.035; NBIC/BioAssist/RK/2008.024); BBMRI –NL: Biobanking and Biomolecular Resources Research Infrastructure (184.021.007); the Institute for Health and Care Research (EMGO+); Neuroscience Campus Amsterdam (NCA); European Science Foundation (ESF): Genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); European Community's Seventh Framework Program (FP7/2007-2013): ENGAGE (HEALTH-F4-2007-201413); the European Science Council : Genetics of Mental Illness (ERC- 230374); Rutgers University Cell and DNA Repository cooperative agreement (NIMH U24 MH068457-06); Collaborative study of the genetics of DZ twinning (NIH R01D0042157-01A); the Genetic Association Information Network, a public–private partnership between the NIH and Pfizer Inc., Affymetrix Inc. and Abbott Laboratories; and the National Institutes of Mental Health (NIMH): GODOT: Integration of Genomics & Transcriptomics in Normal Twins & Major Depression (1RC2 MH089951-01) and DETECT: Developmental TrajECtories in Twins (1RC2MH089995-01).

OGP. We thank the Ogliastra population and all the individuals who participated in this study. We are very grateful to the municipal administrations for their collaboration to the project and for economic and logistic support. This work was supported by grants from the Italian Ministry of Education, University and Research (MIUR) no.5571/DSPAR/2002 and (FIRB) D. M. no. 718/Ric/2005.

PREVEND. PREVEND genetics is supported by the Dutch Kidney Foundation (Grant E033), the National Institutes of Health (grant LM010098), The Netherlands organisation for health research and development (NWO VENI grant 916.761.70 and 916.111.05), the Dutch Interuniversity Cardiology Institute Netherlands (ICIN), and the Landsteiner Foundation for Blood Transfusion Research (LSBR, grant 1133).

QIMR. We thank the Brisbane twins and their families for their participation; Dixie Statham, Ann Eldridge, Marlene Grace (sample collection); Anjali Henders, Leanne Wallace, Lisa Bowdler, Steven Crooks (sample processing); David Smyth, Harry Beeby (IT support). Funding: Australian National Health and Medical Research Council (241944, 339462, 339446, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496739, 552485, 552498, 613627, 619667); Australian Research Council (A7960034, A79906588, A79801419, DP0212016, DP0343921).

SardiNIA. The SardiNIA team was supported by Contracts NO1-AG-1-2109 and HHSN271201100005C from the NIA, and in part by the Intramural Research Program of the National Institute on Aging (NIA), National Institutes of Health (NIH). The SardiNIA authors are grateful to the many volunteers who generously participate in the study, to Monsignore Piseddu, Bishop of Ogliastra, to the Mayors and administrations of the four towns involved and the Head of Public Health Unit ASL4 in Ogliastra.

SHIP. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH.

Sorbs. Financial support was received from the German Research Council (KFO-152), IZKF (B27) and the German Diabetes Association. We would like to thank Knut Krohn (Microarray Core Facility of the Interdisciplinary Centre for Clinical Research, University of Leipzig) for the genotyping/analytical support and Joachim Thiery (Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig) for clinical chemistry services. We thank Nigel W. Rayner (WTCHG, University of Oxford, UK) for the excellent bioinformatics support. This research of Dr. Inga Prokopenko and Dr. Vasiliki Lagou was part funded through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007-201413.

SMART. This research was financially supported by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). Folkert W. Asselbergs is supported by a clinical fellowship from the Netherlands Organisation for Health Research and Development (ZonMw grant 90700342).

TwinsUK. The TwinsUK study is grateful to the volunteer twins who made available their time. We thank the staff from the genotyping facilities at the Wellcome Trust Sanger Institute and the Center for Inherited Disease Research as part of a National Eye Institute/National Institutes of Health project grant. This study received funding from the Wellcome Trust; the European Community's Seventh Framework Program grant agreement (FP7/2007-2013); ENGAGE project grant agreement (HEALTH-F4-2007-201413); the Department of Health via the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. T.D.S. is an NIHR senior investigator. NIcole Soranzo's team is funded by the Wellcome Trust (Sanger Institute, Grant Code 098051), the EU FP7 (EPIGENESYS Grant Code 257082) and FP7 (BLUEPRINT Grant Code HEALTH-F5-2011-282510).

UKBS-CC. The UK Blood Services collection of Common Controls (UKBS-CC collection), funded by the Wellcome Trust grants 076113/C/04/Z and 084183/Z/07/Z and by the Juvenile Diabetes Research Foundation grant WT061858, was established as part of the Wellcome Trust Case-Control Consortium, with the support of the blood services of England, Scotland and Wales. We gratefully acknowledge the generosity of the blood donors in England, Scotland and Wales who have made their DNA samples available to this project as a "free gift".

YoungFinns. The Young Finns Study has been financially supported by the Academy of Finland: grants 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi), the Social Insurance Institution of Finland, Kuopio, Tampere and Turku University Hospital Medical Funds (grant 9M048 and 9N035 for TeLeht), Juho Vainio Foundation, Paavo Nurmi Foundation, Finnish Foundation of Cardiovascular Research and Finnish Cultural Foundation, Tampere Tuberculosis Foundation and Emil Aaltonen Foundation (T.L). The expert technical assistance in the statistical analyses by Ville Aalto and Irina Lisinen is gratefully acknowledged.

Drosophila studies. Ulrich Elling was supported by a grant from the FWF (Project I434-B13). Gerald Wirnsberger was funded from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n 223151 - INFLA-CARE. Maria Novatchkova and Urszula Puc were funded by IMBA. Josef Penninger has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 233407 - COMBINE and was supported by grants from IMBA, the Austrian Ministry of Sciences, and the Austrian Academy of Sciences, and GEN-AU (Austro-Mouse).

Expression QTL studies. LF received a Horizon Breakthrough grant from the Netherlands Genomics Initiative (93519031), a VENI grant from NWO (ZonMW grant 916.10.135), a SIA-

raakPRO subsidy for project BioCOMP and a grant from the European Community's Health Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 259867.

Formaldehyde-assisted isolation of regulatory elements (FAIRE) studies. We thank S.F. Garner (Department of Haematology, University of Cambridge & NHS Blood and Transplant, Cambridge, UK), D.M. Bloxham (Department of Haematology, University of Cambridge & Addenbrooke's Hospital, Cambridge) and all donors and staff at the NHS Blood and Transplant, Cambridge, UK. We thank the core informatics, library-making and sequencing teams at the Wellcome Trust Sanger Institute. D.S. Paul and K. Voss are supported by the Marie-Curie Initial Training Network NetSim (EC-215820). C.A. Albers and A. Rendon are funded by the British Heart Foundation Program Grant RG/09/12/28096. J. Stephens and W.H. Ouwehand are supported by a grant from the National Institutes for Health Research (RP-PG-0310-1002). N. Soranzo and P. Deloukas are supported by the Wellcome Trust (098051).

Supplementary Table 1. Characteristics of participants in the genome-wide and replication cohorts. Results are provided as mean (SD) or %.

Cohort	Country	Design	N	Female (%)	Age (yrs)	Hb (g/dl)	MCH (pg)	MCHC (g/dl)	MCV (fl)	PCV (%)	RBC (10^{12}/l)
European GWA											
ALSPAC	UK	Population	2526	47.7	7.5 (0.2)	12.45 (0.72)	-	-	-	-	-
AMISH	USA	Population	1578	50.4	48.4 (16.2)	13.8 (1.2)	30.9 (1.35)	34.2 (0.69)	90.4 (3.49)	40.5 (0.08)	4.47 (0.39)
COLAUS	Switzerland	Population	854	56.7	59.3 (10.7)	14.34 (1.21)	30.04 (1.54)	33.37 (0.87)	90.11 (4.14)	43.02 (3.20)	4.78 (0.42)
DESIR	France	Population	716	24.9	50.4 (8.1)	13.86 (1.14)	29.93 (1.61)	30.47 (0.91)	91.12 (4.33)	-	4.64 (0.37)
EGCUT	Estonia	Population	893	50.7	37.5 (15.6)	14.3 (1.44)	29.7 (1.71)	33.7 (1.05)	88.0 (4.32)	42.3 (3.7)	4.81 (0.45)
EPIC-case	UK	Population	844	56.9	59.3 (8.8)	14.01 (1.30)	30.26 (1.56)	33.99 (1.15)	89.11 (3.93)	41.21 (3.77)	4.64 (0.43)
EPIC-cohort	UK	Population	1847	53.2	59.3 (9.0)	13.86 (1.23)	30.50 (1.58)	34.16 (1.09)	89.26 (3.97)	40.56 (3.61)	4.55 (0.41)
GeneBank	USA	Case-cohort	2671	28.3	63.0 (11.2)	13.27 (1.48)	30.71 (1.62)	34.51 (0.96	88.14 (5.57)	38.90 (4.08)	4.34 (0.49)
INGI CARL	Italy	Population	520	59.8	50.4 (16.3)	13.83 (1.40)	29.45 (2.35)	33.36 (1.50)	88.14 (5.57)	41.49 (3.91)	4.73 (0.44)
INGI CILENTO	Italy	Population	855	55.4	52.5 (19.4)	13.90 (1.47)	29.31 (2.75)	32.95 (1.22)	88.85 (7.1)	42.17 (4.19)	4.74 (0.48)
INGI FVG	Italy	Population	1205	59.8	51.7 (16.4)	14.01 (1.35)	29.88 (1.50)	33.00 (1.00)	90.18 (4.00)	42.36 (3.36)	4.73 (0.42)
INGI Val Borbera	Italy	Population	1662	56.0	54.7 (18.3)	14.5 (1.3)	30.1 (1.6)	33.0 (1.1)	91.0 (4.3)	43.7 (3.6)	4.8 (0.4)
KORA-F3	Germany	Population	1642	50.6	62.5 (10.1)	14.18 (1.14)	30.89 (1.44)	33.71 (0.68)	91.55 (3.79)	42.11 (3.35)	4.6 (0.39)
KORA-F4	Germany	Population	1813	51.2	60.9 (8.9)	14.11 (1.15)	31.2 (1.53)	34.1 (0.73)	91.44 (4.0)	41.42 (3.2)	4.53 (0.37)
LBC1921	UK	Population	496	57.9	79.1 (0.6)	13.46 (1.31)	-	-	90.78 (4.50)	39.46 (3.60)	4.35 (0.41)
LBC1936	UK	Population	987	49.2	69.6 (0.8)	14.55 (1.26)	-	-	91.06 (4.04)	42.06 (3.57)	4.62 (0.41)
LIFELINES	Netherlands	Population	8080	52.1	47.9 (11.2)	14.04 (1.26)	29.92 (1.67)	33.39 (1.07)	89.60 (4.43)	42.02 (3.18)	4.70 (0.39)
LOLIPOP-EW610	UK	Population	927	26.9	56.0 (9.8)	14.66 (1.28)	30.77 (1.89)	34.01 (0.92)	90.48 (5.08)	43.13 (3.88)	4.78 (0.46)
LOLIPOP-EW-A	UK	Population	589	12.9	54.3 (10.4)	14.65 (1.22)	30.66 (1.99)	33.97 (0.74)	90.32 (5.18)	43.14 (3.62)	4.78 (0.43)
LOLIPOP-EW-P	UK	Case-control	652	0.0	55.7 (9.1)	14.88 (1.03)	30.75 (1.63)	33.90 (0.78)	90.72 (4.22)	43.94 (3.13)	4.86 (0.40)
MDC	Sweden	Population	1699	58.0	57.4 (5.9)	13.9 (1.04)	30.3 (1.46)	33.8 (0.98)	89.6 (3.5)	41.2 (2.9)	4.6 (0.37)
MICROS	Italy	Population	1213	56.5	45.8 (16.3)	14.93 (1.41)	30.67 (1.89)	34.48 (0.77)	88.91 (4.80)	-	4.87 (0.45)
NESDA	Netherlands	Case-cohort	1739	67.8	42.4 (12.5)	13.96 (1.22)	31.40 (1.71)	33.89 (0.93)	89.34 (4.09)	41.22 (3.35)	4.62 (0.40)
NFBC	Finland	Population	4761	52.0	31	14.17 (1.35)	30.30 (1.69)	33.84 (0.98)	89.88 (4.71)	41.63 (3.87)	4.64 (0.45)
NTR	Netherlands	Twins	1740	64.0	45.3 (14.1)	14.13 (1.21)	30.75 (1.40)	33.39 (0.65)	92.07 (3.85)	-	4.62 (0.41)
PREVEND	Netherlands	Population	3121	48.8	49.3 (12.0)	13.77 (1.16)	-	33.67 (0.67)	90.66 (4.08)	40.88 (3.39)	-
QIMR	Australian	Twins	2538	51.2	15.0 (2.9)	13.8 (0.96)	29.0 (1.3)	33.1 (0.82)	87.4 (3.7)	-	4.8 (0.34)
SardiNIA	Italy	Population	4694	56.3	43.3 (17.6)	13.86 (1.49)	28.88 (3.58)	33.28 (1.18)	86.61 (9.31)	41.62 (4.03)	4.85 (0.57)
SHIP	Germany	Population	3183	52.0	54.5 (15.3)	14.07 (1.30)	30.79 (1.54)	33.93 (0.98)	90.68 (4.00)	41.45 (3.60)	4.58 (0.43)
Sorbs	Germany	Population	934	59.1	47.7 (16.3)	14.20 (1.26)	30.07 (1.45)	34.12 (0.95)	88.04 (3.91)	41.59 (3.27)	4.73 (0.39)
TwinsUK	UK	Twins	3419	93.0	52.0 (13.4)	13.43 (1.09)	29.96 (1.85)	32.44 (1.41)	92.45 (5.25)	41.45 (3.31)	4.49 (0.37)
UKBS-CC	UK	Population	2155	50.0	43.7 (12.3)	14.0 (1.13)	29.97 (1.65)	33.92 (0.68)	88.28 (4.30)	-	4.67 (1.08)
South Asian GWA											
LOLIPOP-IA610	UK	Case-control	6557	15.7	55.4 (10.6)	14.41 (1.56)	28.87 (2.50)	33.41 (1.19)	86.67 (6.47)	37.3 (14.9)	4.96 (0.53)
LOLIPOP-IA300	UK	Population	2139	0.0	48.3 (10.5)	14.94 (1.23)	29.05 (2.46)	33.61 (1.09)	86.39 (6.32)	44.5 (3.5)	5.18 (0.50)
LOLIPOP-IA-P	UK	Case-control	612	0.0	51.1 (8.3)	14.77 (1.11)	29.11 (2.32)	33.55 (0.85)	86.73 (5.88)	44.0 (3.6)	5.10 (0.44)

Cohort	Country	Design	N	Female (%)	Age (yrs)	Hb (g/dl)	MCH (pg)	MCHC (g/dl)	MCV (fl)	PCV (%)	RBC (10^{12}/l)
Replication											
deCODE	Iceland	Population	34843	58.6	65.6 (17.0)	12.9 (11.3)	30.3 (2.2)	33.5 (1.1)	90.5 (5.6)	38.4 (4.8)	4.21 (0.64)
CBR	UK	Population	9136	53.4	49.0 (12.9)	14.0 (1.1)	29.87 (1.8)	33.5 (1.0)	89.1 (4.3)	41.8 (3.3)	4.7 (0.4)
OGP	Italy	Population	8192	56.1	49.7(17.9)	14.3 (1.3)	30.0 (2.4)	34.2 (0.8)	87.5 (6.0)	41.9 (3.8)	4.8 (0.7)
PREVEND	Netherlands	Population	2974	57.7	49.5 (12.8	13.7 (1.2)	30.5 (1.7)	33.6 (0.7)	90.6 (4.5)	40.5 (3.6)	4.5 (0.4)
SMART	Netherlands	Case-cohort	8361	32.0	57 (12)	14.3 (1.2)	-	-	-	41.9 (3.5)	-
Population variation											
EGCUT-stage2	Estonia	Population	738	44.9	53.6 (15.8)	14.4 (1.42)	30.0 (1.76)	33.3 (1.74)	90.0 (4.45)	43.3 (3.92)	4.81 (0.45)
Lifelines-stage2	Netherlands	Population	5241	60.0	50.1 (11.8)	14.06 (1.21)	30.0 (1.64)	33.61 (0.96)	89.49 (4.31)	41.81 (3.08)	4.69 (0.38)
LURIC - Graz	Germany	Case-cohort	804	34.0	60.0 (12.0)	13.8 (1.5)	30.4 (1.8)	34.2 (1.3)	89.0 (4.7)	40.0 (4.0)	4.6 (0.5)
LURIC - HD	Germany	Case-cohort	2124	28.8	63.6 (7.0)	13.8 (1.5)	30.4 (1.8)	34.1 (1.1)	89.2 (4.8)	40.0 (4.0)	4.6 (0.5)
NTR-stage2	Netherlands	Twins	3746	62.1	39.5(15.1)	14.1 (1.33)	30.4 (2.00)	20.7 (1.19)	92.3 (4.71)	42.5 (3.95)	4.65 (0.45)
Young Finns	Finland	Population	616	51.9	41.9 (5.1)	14.2 (1.3)	30.2 (1.7)	33.6 (1.0)	89.5 (4.3)	42.1 (3.3)	4.7 (0.4)

Supplementary Table 2. Summary of study genotyping methods in the genome-wide association cohorts.

	Platform (s)	Genotype calling	Sample call rate cut-off	SNP call rate cut-off	MAF cut off	P_{HWE} cut-off	SNPs	Imputation package	NCBI build	GWAS statistics package	Study-specific Covariates
ALSPAC	Illumina 317 / 610	GenCall	0.97	0.97 (317), 0.95 (610)	0.05	5.0E-07	285531	MACH	36	MACH2QTL	PC1-2
AMISH	Affymetrix	BRLMM	0.95	0.95	0.01	1.0E-06	338598	MACH v1.0.15	36; v22	Measured genotype accounting for polygenic component	Adjusted for relatedness
COLAUS	Affimetrix 500K	BRLMM	0.9	0.7	0.01	4.07E-04	390631	IMPUTE v0.2.0	35	Matlab	PC1-2
DESIR	Illumina 300K	BeadStudio	0.95	0.9	0.01	1.0E-03	291609	IMPUTE	36	SNPTEST v2	
EGCUT	Illumina Human370CNV / OmniExpress	GenomeStudio	0.95	0.95	0.01	4.07E-04	189000	IMPUTE	36; v22	SNPTEST	PC 1-3
EPIC-case	Affymetrix 500K	BRLMM	0.9	0.9	0.01	1.0E-06	382037	IMPUTE v0.3.1	35	SNPTEST	
EPIC-cohort	Affymetrix 500K	BRLMM	0.9	0.9	0.01	1.0E-06	382037	IMPUTE 0.3.1	35	SNPTEST	
GeneBank	Affymetrix 6.0	Birdseed	0.97	0.95	0.01	1.0E-04	562554	MACH v1.0.16	36; v22	PLINK	CAD case-control status, age
INGI CARL	Illumina 370CNV	Beadstudio	0.9	0.9	0.01	1.0E-04	276271	MACH v1.0.16	36; v22	GenABEL/ ProbABEL	Linear Mixed Model
INGO CILENTO	Illumina 370K	GenomeStudio	NA	0.95	NA	NA	305009	MACH v1.0.16a	36; v22	R, linear model, GenABEL and ProbABEL	
INGI FVG	Illumina 370CNV	Beadstudio	0.95	0.9	0.01	1.0E-04	276271	MACH v1.0.16	36; v22	GenABEL/ ProbABEL	Linear Mixed Model
INGI Val Borbera	Illumina 370k	BeadStudio	0.95	0.9	0.01	1.0E-04	324319	MACH	36, v22	GenABEL - ProbABEL	Age, MDS1-3
KORA-F3	Affymetrix 500K	BRLMM	0.93	-	-	-	490033	IMPUTE	35	SNPTEST v2.1.1	Age, sex
KORA-F4	Affymetrix1000K	Birdseed2	0.93	0.93	-	-	909622	IMPUTE	36	SNPTEST v2.1.1	Age, sex
LBC1921	Illumina Human610	Illumina	0.95	0.98	0.01	1.0E-03	535709	MACH	36	MACH2QTL	PC1-4
LBC1936	Illumina Human610	Illumina	0.95	0.98	0.01	1.0E-03	535709	MACH	36	MACH2QTL	PC1-4
Lifelines	Illumina CytoSNP12 V2	Illumina	QC filters	0.95	0.01	1.0E-05	257581	BEAGLE v3.1.0	36, rel 23a	PLINK	
LOLIPOP-EW610	Illumin Human610	Beadstudio	0.95	0.9	0.01	1.0E-06	544620	MACH	36, v22	MACH2QTL	PC1-10
LOLIPOP-EW-A	Affymetrix 500K	BRLMM	0.95	0.9	0.01	1.0E-06	374773	MACH	35, v21	MACH2QTL	PC1-10
LOLIPOP-EW-P	Perlegen Custom	Perlegen	0.95	0.9	0.01	1.0E-06	202544	MACH	35, v21	MACH2QTL	PC1-10
LOLIPOP-	Illumin Human610	Beadstudio	0.95	0.9	0.01	1.0E-	544620	MACH	36,	MACH2QTL	PC1-10

IA610						06			v22		
LOLIPOP-IA300	Illumina HumanHap 300K	BeadStudio	0.95	0.9	0.01	1.0E-06	245892	MACH	35, v21	MACH2QTL	PC1-10
LOLIPOP-IA-P	Perlegen custom	Perlegen	0.95	0.9	0.01	1.0E-06	170055	MACH	35, v21	MACH2QTL	PC1-10
MDC	Illumina 610Quad	BeadStudio	0.95	0.9	0.01	5.0E-07	521220	IMPUTE v2	36, v22	SNPTEST	PC1-3
MICROS	Illumina HumHap300	BeadStudio	0.95	0.98	0.01	1.0E-06	292.917	MACH v1.0.16	36	ProbABEL	Study location (village)
NESDA	Perlegen 600K	Perlegen	0.95	0.95	0.01	-	435291	IMPUTE v0.3.2	36, v22	SNPTEST v2.1.1	PC1-5
NFBC	Illumina Infinium 370 cnvDuo array	Beadstudio	0.95	0.95	0.01	1.0E-04	328 007	IMPUTE v.1.0	35	SNPTEST	PC1-10
NTR	Perlegen 600K	Perlegen	0.95	0.95	0.01	-	435,291	IMPUTE v0.3.2	36, v22	Merlin	Age, time of collection
PREVEND	Illumina CytoSNP12 v2	GenomeStudio	0.95	0.98	0.01	1.0E-5	244868	BEAGLE v3.1.0	36, rel 23a	PLINK	
QIMR	Illumina Human610	BeadStudio	0.95	0.95	0.01	1.0E-06	269,840	MACH	36, I+II	Merlin	Age
SardiNIA	Affymetrix 10K, 500K, 6.0	BRLMM (10K/500K), Birdseed (6.0)	0.95	0.90 (10K, 500K); 0.95 (6.0)	0.05 (10K, 500K); 0.01 (6.0)	1.0E-06	731,209	MACH	36	Merlin	Alphacarrier, betacarrier
SHIP	Affymetrix SNP 6.0	Birdseed V2	0.92	-	-	-	869,224	IMPUTE v0.5.0	36	QUICKTEST 0.95	PC1-10
Sorbs	Affymetrix 500K and 6.0	BRLMM (500K), Birdseed (6.0)	-	0.95	0.01	1.0E-04	378,513	IMPUTE v1.0.0	35	GenABEL, ProbABEL	PC1-3
TwinsUK	Illumina HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M	Illuminus calling algorithm	0.98	0.97 for MAF>0.05, 0.99 for 0.01<MAF<0.05	0.01	1.0E-06	534,633	IMPUTE v2	36	MERLIN	Instrument, PC1-2
UKBS-CC	Affymetrix Genome-Wide Human SNP Array 6.0	Chiamo	0.90	0.90	0.01	1.0E-06	2492005	IMPUTE v2	36	SNPTEST	

Supplementary Table 3. Genomic control inflation factors

	Gender	Hb	MCH	MCHC	MCV	PCV	RBC
Meta-analysis							
- Europeans		1.10	1.13	1.08	1.14	1.10	1.14
- South Asians		1.02	1.04	1.02	1.04	1.03	1.03
- Combined		1.10	1.13	1.08	1.14	1.10	1.14
Individual cohorts							
ALSPAC	F	1.00					
ALSPAC	M	0.99					
Amish	F	1.04	1.05	1.04	1.07	1.03	1.05
Amish	M	1.05	1.03	1.07	1.04	1.04	1.01
CoLaus	F	1.02	1.01	1.00	1.01	1.01	1.01
CoLaus	M	1.00	1.03	0.99	1.02	1.00	1.00
DESIR	F	1.00	1.01	1.02	1.00	1.01	1.01
DESIR	M	1.00	1.01	1.02	1.00	1.01	1.01
EGCUT	F	1.01	1.02	1.02	1.02	1.01	1.04
EGCUT	M	1.03	1.01	1.01	1.02	1.02	1.02
EPIC_case	F	1.01	1.01	1.00	1.01	1.01	1.02
EPIC_case	M	1.01	0.99	1.01	1.02	1.02	1.01
EPIC_cohort	F	1.01	1.02	1.03	1.01	0.99	0.99
EPIC_cohort	M	1.01	1.01	1.00	1.01	1.00	1.02
GeneBank	F	1.01	1.00	1.01	1.00	1.01	1.01
GeneBank	M	0.98	1.05	1.02	1.08	0.99	1.02
INGI CARL	F	1.01	0.99	1.01	1.00	1.00	1.00
INGI CARL	M	0.99	1.00	1.02	0.99	1.02	1.02
INGI CILEN	F	1.01	1.00	0.99	1.00	1.01	0.99
INGI CILEN	M	1.00	0.99	1.00	1.03	1.00	1.00
INGI FVG	F	1.03	1.06	1.06	1.03	1.03	1.01
INGI FVG	M	0.99	1.00	1.03	0.98	0.98	0.98
INGI ValBorbera	F	1.00	1.00	1.01	1.00	1.01	1.00
INGI ValBorbera	M	1.01	1.01	1.01	1.00	1.00	1.00
KORA F3	F	1.02	1.01	1.01	1.01	1.02	1.01
KORA F3	M	1.01	1.02	1.01	1.02	1.01	1.01
KORA F4	F	0.99	1.01	1.00	1.01	0.99	1.00
KORA F4	M	1.01	1.02	1.01	1.01	1.01	1.01
LBC1921	F	1.01			0.99	1.01	1.00
LBC1921	M	0.99			1.00	0.99	1.00
LBC1936	F	1.00			1.00	0.99	1.00
LBC1936	M	1.00			0.99	1.00	0.99
LIFELINES	F	1.01	1.03	1.04	1.03	1.02	1.05
LIFELINES	M	1.02	1.01	1.03	1.02	1.03	1.03
LOLIPOP EW_A	M	1.01	1.00	1.00	1.01	1.01	1.00
LOLIPOP EW_P	M	1.00	1.00	1.00	1.00	1.00	1.00
LOLIPOP EW610	F	1.01	1.00	0.99	1.00	1.00	1.01
LOLIPOP EW610	M	1.00	1.01	1.01	1.00	1.00	1.00
LOLIPOP IA_P	M	0.99	1.01	1.02	1.02	0.98	1.02
LOLIPOP IA300	M	1.01	1.02	1.02	1.02	1.00	1.01
LOLIPOP IA610	F	1.01	1.01	1.01	0.99	1.01	1.03
LOLIPOP IA610	M	1.03	1.05	1.02	1.05	1.02	1.02
MDC	F	1.01	1.01	1.00	1.01	1.02	1.00
MDC	M	1.01	1.01	1.00	1.01	1.01	1.00
MICROS	F	0.99	1.01	1.00	1.01	0.99	1.00
MICROS	M	0.99	1.01	1.01	1.01	1.00	1.00
NESDA	F	1.02	1.01	1.00	1.02	1.02	1.02
NESDA	M	1.02	1.01	1.02	1.03	1.03	1.02
NFBC	F	1.02	1.01	1.00	1.01	1.03	1.02

NFBC	M	1.02	1.01	1.02	1.00	1.02	1.01
Prevend	F	1.02		1.00	1.03	1.02	
Prevend	M	1.03		1.00	1.02	1.02	
QIMR-NTR	F	1.00	1.00	1.00	1.00		1.00
QIMR-NTR	M	1.05	1.00	1.05	1.00		1.00
SardiNIA	F	1.08	1.08	1.14	1.13	1.07	1.09
SardiNIA	M	1.01	1.10	1.14	1.12	1.03	1.05
SHIP	F	0.99	1.00	0.99	1.01	1.00	1.01
SHIP	M	1.01	1.01	1.02	1.01	1.00	0.99
Sorbs	F	0.98	1.00	1.01	0.98	0.98	1.00
Sorbs	M	1.00	0.99	1.00	0.99	0.99	0.99
TwinsUK	F	1.00	1.00	1.01	1.01	1.02	1.00
TwinsUK	M	1.02	1.01	1.02	1.02	1.02	1.02
UKBS-CC	F	1.00	1.00	1.00	1.01	1.01	1.01
UKBS-CC	M	1.02	1.01	1.01	1.00	1.01	1.01

Supplementary Table 4. Excel spreadsheet (online) providing association test results for all SNPs reaching P<10-6 in the red blood cell genome-wide association study.

Supplementary Table 5. Excel spreadsheet (online) providing genome-wide association and replication test results for the 75 sentinel SNPs.

Supplementary Table 6. Association in the European GWA samples (current study) for SNPs previously reported to be associated with red blood cell phenotypes.[2-6, 49] SNPs that are not highlighted reach the conventional threshold for genome-wide significance ($P<5x10^{-8}$), SNPs highlighted yellow replicate at $P<0.05$, SNPs highlighted in green do not show replication.

				Discovery GWA		Current study		
No	**Region**	**SNP**	**Position**	**Pheno**	**P**	**N**	**P**	**Authors**
1	1q22	rs6684514	154522080	MCHC	3.0E-09	48976	4.8E-01	Kamatani Y
2	1q23	rs857721	156879172	MCHC	1.0E-10	56429	1.6E-15	Ganesh SK
3	1q31	rs12127588	196862129	MCH	7.0E-10	35962	7.0E-02	Kamatani Y
4	1q44	rs11204538	246112895	MCV	2.0E-08	30151	1.8E-02	Kamatani Y
5	2p21	rs10495928	46206670	HB	7.0E-13	52101	1.0E-13	Ganesh SK
	2p21	rs10495928	46206670	RBC	4.0E-08	44613	6.4E-10	Kamatani Y
	2p21	rs10168349	46214411	MCV	4.0E-15	53027	2.6E-16	Ganesh SK
6	2p16	rs2540917	60462263	MCV	1.0E-14	57790	1.2E-12	Ganesh SK
7	3p24	rs9310736	24325815	MCH	4.0E-10	51406	3.4E-15	Kamatani Y
	3p24	rs9310736	24325815	MCV	3.0E-08	57810	6.0E-16	Kamatani Y, Ding K
8	3q29	rs9859260	197284944	MCV	8.0E-14	51918	2.3E-10	Ganesh SK
	3q29	rs11915082	197293536	MCH	8.0E-13	46748	2.9E-13	Ganesh SK
9	4q12	rs218237	55088929	MCH	3.0E-25	43231	1.0E-18	Kamatani Y
	4q12	rs218237	55088929	RBC	2.0E-17	45064	2.3E-32	Kamatani Y
	4q12	rs172629	55102519	MCV	1.0E-15	51865	1.9E-25	Ganesh SK
10	5p15	rs4580814	1166244	MCHC	5.0E-10	33137	1.2E-04	Kamatani Y
	5p15	rs2736100	1339516	RBC	3.0E-08	31725	2.2E-03	Kamatani Y
11	6p22	rs17342717	25929749	MCH	5.0E-08	37058	9.6E-40	Kullo IJ
	6p22	rs1408272	25950930	MCH	1.0E-11	36605	4.8E-67	Ferreira MA
	6p22	rs1408272	25950930	MCH	4.0E-39	36605	4.8E-67	Ganesh SK
	6p22	rs1800562	26201120	HB	6.0E-19	34188	5.6E-13	Ganesh SK
	6p22	rs1800562	26201120	MCH	3.0E-09	29711	6.5E-56	Kullo IJ
	6p22	rs1800562	26201120	MCH	3.0E-09	29711	6.5E-56	Kullo IJ
	6p22	rs1800562	26201120	MCV	2.0E-08	31217	1.2E-42	Benyamin B
	6p22	rs1800562	26201120	MCV	1.0E-46	31217	1.2E-42	Ganesh SK
	6p22	rs1800562	26201120	MCV	1.0E-23	31217	1.2E-42	Soranzo N
	6p22	rs1800562	26201120	PCV	2.0E-09	32016	4.3E-05	Ganesh SK
	6p22	rs198846	26215442	HB	1.0E-08	60869	1.4E-30	Chambers JC
12	6p21	rs3218097	42013253	RBC	1.0E-10	49292	2.8E-16	Kamatani Y
	6p21	rs9349205	42033137	MCH	8.0E-20	43669	1.0E-21	Ganesh SK
	6p21	rs9349205	42033137	MCV	1.0E-31	46928	5.5E-30	Ganesh SK
	6p21	rs11970772	42033268	MCV	7.0E-19	58026	4.9E-32	Soranzo N
13	6q21	rs9374080	109723113	MCV	4.0E-10	46686	2.3E-18	Ganesh SK
	6q21	rs11966072	109741521	MCH	1.0E-08	5054	1.9E-02	Kamatani Y
	6q21	rs11966072	109741521	RBC	7.0E-09	5093	1.0E-02	Kamatani Y
14	6q23	rs7775698	135460328	MCH	5.0E-13	39576	2.4E-73	Ferreira MA
	6q23	rs7775698	135460328	MCH	3.0E-66	39576	2.4E-73	Kamatani Y
	6q23	rs7775698	135460328	MCH	1.0E-15	39576	2.4E-73	Kullo IJ
	6q23	rs7775698	135460328	MCV	8.0E-18	45966	1.5E-78	Ferreira MA
	6q23	rs7775698	135460328	MCV	3.0E-56	45966	1.5E-78	Kamatani Y
	6q23	rs7775698	135460328	RBC	1.0E-14	41602	2.6E-70	Kullo IJ
	6q23	rs7776054	135460609	MCH	7.0E-69	56217	6.6E-108	Ganesh SK
	6q23	rs9373124	135464902	MCHC	7.0E-14	56151	1.3E-13	Ganesh SK
	6q23	rs4895441	135468266	MCV	7.0E-86	57837	1.0E-107	Ganesh SK
	6q23	rs9402686	135469510	MCV	7.0E-42	57853	1.7E-106	Soranzo N

	6q23	rs9494145	135474245	MCV	3.0E-15	53763	1.1E-89	Kullo IJ
	6q23	rs9483788	135477194	PCV	3.0E-15	52759	3.2E-17	Ganesh SK
	6q23	rs9483788	135477194	RBC	1.0E-47	53332	5.8E-75	Ganesh SK
	6q23	rs6569992	135493845	MCH	1.0E-08	45630	6.3E-55	Kullo IJ
	6q23	rs6569992	135493845	MCV	3.0E-08	47188	3.9E-48	Kullo IJ
	6q23	rs6569992	135493845	RBC	6.0E-09	44394	1.4E-50	Kullo IJ
15	6q24	rs628751	139880112	MCH	1.0E-17	50565	2.8E-26	Ganesh SK
	6q24	rs643381	139881116	MCV	5.0E-25	41038	7.8E-27	Ganesh SK
16	7p12	rs12718597	50395922	MCV	5.0E-13	36413	2.2E-11	Ganesh SK
17	7q22	rs7786877	100051951	MCV	3.0E-11	45829	8.0E-10	Ganesh SK
	7q22	rs7385804	100073906	PCV	4.0E-10	41886	7.9E-04	Ganesh SK
	7q22	rs7385804	100073906	RBC	5.0E-10	45574	1.6E-17	Soranzo N
	7q22	rs2075671	100183042	RBC	1.0E-09	48904	1.9E-07	Ganesh SK
18	7q36	rs10224002	151045974	HB	3.0E-15	49958	1.2E-13	Ganesh SK
	7q36	rs10224002	151045974	PCV	6.0E-15	41868	1.3E-11	Ganesh SK
19	8p21	rs7843479	21876759	MCV	3.0E-08	45518	2.2E-01	Kamatani Y
20	9p24	rs10758658	4846877	MCH	2.0E-14	56189	1.4E-14	Ganesh SK
	9p24	rs10758658	4846877	MCV	3.0E-20	57827	5.2E-19	Ganesh SK
21	9q34	rs8176746	135121143	MCHC	4.0E-08	44984	8.7E-05	Kamatani Y
	9q34	rs495828	135144688	HB	1.0E-11	52848	1.6E-15	Kamatani Y
	9q34	rs495828	135144688	PCV	6.0E-10	44764	1.5E-13	Kamatani Y
	9q34	rs495828	135144688	RBC	3.0E-12	45331	1.4E-15	Kamatani Y
22	10q11	rs2279434	45275070	MCH	4.0E-12	47419	2.9E-07	Kamatani Y
	10q11	rs11239550	45344735	MCV	1.0E-10	57985	4.6E-15	Ganesh SK
23	10q11	rs7085433	51263360	MCH	6.0E-10	25731	4.9E-02	Kamatani Y
	10q11	rs7085433	51263360	MCV	7.0E-09	27222	3.9E-02	Kamatani Y
24	10q21	rs16926246	70763398	HB	2.0E-11	40486	6.0E-19	Ganesh SK
	10q21	rs16926246	70763398	PCV	1.0E-13	32401	1.5E-15	Ganesh SK
25	12p13	rs11611647	4204180	RBC	6.0E-09	44531	1.2E-08	Kamatani Y
26	12q24	rs11065987	110556807	HB	1.0E-11	56136	1.7E-13	Ganesh SK
	12q24	rs11065987	110556807	PCV	1.0E-12	48077	1.5E-10	Ganesh SK
27	14q23	rs4466998	64545293	MCV	5.0E-08	43114	8.4E-07	Ganesh SK
28	15q22	rs6494537	63838399	MCH	3.0E-09	42140	5.0E-06	Kamatani Y
	15q22	rs8035639	63731602	MCH	8.0E-09	34982	1.3E-05	Ding K
29	16p13	rs7189020	244804	MCV	2.0E-12	48237	1.3E-10	Ganesh SK
	16p13	rs1122794	249156	MCH	3.0E-10	31659	3.3E-09	Ganesh SK
30	16q24	rs9937239	85664621	MCHC	9.0E-08	40048	0.24	Ding K
	16q24	rs837763	87381230	MCHC	5.0E-10	37768	1.9E-22	Ding K
	16q24	rs837763	87381230	MCHC	4.0E-13	37768	1.9E-22	Kamatani Y
31	19p13	rs7255045	12793269	MCV	2.0E-12	53098	4.0E-05	Ganesh SK
32	19p13	rs11085824	12862547	MCH	1.0E-11	45954	1.6E-19	Ganesh SK
33	20q13	rs6013509	50751758	HB	1.0E-10	46778	2.9E-02	Ganesh SK
34	20q13	rs6092477	55425101	MCV	1.0E-08	45588	1.5E-11	Kamatani Y
35	22q11	rs4821112	20294761	MCV	1.0E-08	57847	9.3E-09	Kamatani Y
36	22q12	rs9609565	31197528	MCV	4.0E-10	57799	7.9E-09	Soranzo N
37	22q12	rs855791	35792882	HB	2.0E-13	46184	4.6E-40	Chambers JC
	22q12	rs855791	35792882	HB	3.0E-25	46184	4.6E-40	Ganesh SK
	22q12	rs855791	35792882	MCH	1.0E-12	41609	1.0E-69	Kullo IJ
	22q12	rs855791	35792882	MCV	1.0E-10	43197	2.4E-54	Benyamin B
	22q12	rs855791	35792882	MCV	5.0E-09	43197	2.4E-54	Kullo IJ
	22q12	rs5756506	35797338	MCH	1.0E-09	56070	6.3E-33	Soranzo N
	22q12	rs4820268	35799537	MCH	3.0E-10	39276	2.7E-51	Ferreira MA

	22q12	rs4820268	35799537	MCHC	1.0E-12	39276	5.6E-12	Kullo IJ
	22q12	rs4820268	35799537	MCV	4.0E-12	40974	3.4E-52	Ferreira MA
	22q12	rs2413450	35800170	MCH	9.0E-34	44912	1.2E-55	Ganesh SK
	22q12	rs2413450	35800170	MCV	3.0E-41	46593	1.3E-50	Ganesh SK
	22q12	rs2413450	35800170	PCV	2.0E-13	41533	2.5E-15	Ganesh SK
38	22q13	rs470119	49313780	MCH	4.0E-08	44146	2.2E-12	Kamatani Y

Supplementary Table 7. Comprehensive list of locus-phenotype associations identified. Locus sentinel: 1 = the discovery association for the locus (SNP with lowest P value against any red blood cell phenotype); 0 = secondary phenotype(s) associated with the locus at $P<1x10^{-8}$

Region	SNP	BP	Pheno	Replication testing	Total N	P=	Locus sentinel	Novel
1p36	rs1175550	3681388	MCHC	0	50425	8.6E-15	1	1
1p34	rs3916164	39842526	MCH	1	91874	3.1E-10	1	1
1p32	rs741959	47448820	MCV	0	58002	6.0E-10	1	1
1q23	rs857684	156842353	MCHC	0	56373	3.5E-16	1	0
1q23	rs3737515	156864131	MCV	1	92648	8.3E-09	0	1
1q32	rs7529925	197273831	RBC	1	86337	8.3E-09	1	1
1q32	rs7551442	201921744	MCHC	0	50411	9.7E-12	1	1
1q32	rs4951381	201927461	MCV	1	86736	2.2E-09	0	1
1q32	rs9660992	203516073	MCH	0	51249	7.1E-10	1	1
1q32	rs9660992	203516073	MCV	0	57652	1.0E-09	0	1
1q44	rs3811444	246106074	RBC	0	34323	4.5E-10	1	1
1q44	rs3811444	246106074	MCV	0	35742	2.2E-09	0	0
2p21	rs4953318	46208555	PCV	0	53032	3.1E-19	1	0
2p21	rs4953318	46208555	HB	0	61099	2.3E-16	0	0
2p21	rs4953318	46208555	RBC	0	53605	1.7E-12	0	0
2p16	rs243070	60473790	MCV	0	57740	4.4E-13	1	0
2p16	rs13027161	60461232	MCH	0	51373	7.2E-12	0	1
2p16	rs2540913	60464316	RBC	0	53229	6.0E-13	0	1
2q13	rs10207392	111566130	MCV	1	104493	4.4E-11	1	1
3p24	rs9310736	24325815	MCV	0	57810	6.1E-16	1	0
3p24	rs9310736	24325815	MCH	0	51406	3.4E-15	0	0
3p24	rs9310736	24325815	RBC	0	53356	5.4E-14	0	1
3q22	rs6776003	142749183	MCV	1	101281	3.7E-11	1	1
3q23	rs13061823	143603476	MCV	0	57678	4.7E-13	1	1
3q23	rs7615316	143838616	MCH	1	86235	6.8E-09	0	1
3q29	rs11717368	197318754	MCH	0	51664	6.6E-19	1	0
3q29	rs11717368	197318754	MCV	0	58067	2.0E-14	0	0
4q11	rs218238	55089781	RBC	0	53374	2.8E-39	1	0
4q11	rs218238	55089781	PCV	0	52802	2.9E-12	0	1
4q11	rs218238	55089781	HB	0	60868	1.5E-11	0	1
4q11	rs218264	55103632	MCV	0	56287	3.8E-32	0	0
4q11	rs218264	55103632	MCH	0	49888	7.4E-28	0	0
4q27	rs13152701	122970511	MCV	0	53708	9.0E-10	1	1
6p23	rs6914805	16389166	MCH	0	47195	1.2E-19	1	1
6p23	rs6914805	16389166	MCV	0	51868	2.2E-11	0	1
6p21	rs1408272	25950930	MCH	0	36605	4.8E-67	1	0
6p21	rs1408272	25950930	MCV	0	39868	3.2E-47	0	0
6p21	rs198846	26215442	HB	0	60869	1.4E-30	0	0
6p21	rs198846	26215442	MCHC	0	56189	7.5E-21	0	1
6p21	rs198846	26215442	PCV	0	52802	3.5E-10	0	0
6p22	rs13219787	27969649	MCH	0	42060	5.9E-17	1	1
6p22	rs2097775	30462282	HB	0	61058	1.3E-10	1	1

6p22	rs2097775	30462282	MCH	0	51613	2.3E-11	0	1
6p21	rs9272219	32710247	RBC	0	49302	4.3E-10	1	1
6p21	rs9349204	42022356	MCV	0	53153	2.4E-40	1	0
6p21	rs3218108	42010633	RBC	0	53404	5.2E-19	0	0
6p21	rs11968166	42033282	MCH	0	51388	9.4E-34	0	0
6p12	rs9369427	43919408	HB	0	60855	5.6E-12	1	1
6p12	rs9369427	43919408	PCV	0	52787	1.5E-09	0	1
6q21	rs1008084	109733658	MCH	0	51455	6.4E-26	1	0
6q21	rs6568571	109719945	MCV	0	57762	1.6E-22	0	0
6q21	rs6568571	109719945	RBC	0	53309	2.1E-18	0	0
6q21	rs1341271	109724235	MCHC	1	91057	2.1E-12	0	1
6q23	rs9389269	135468852	MCV	0	57855	2.6E-109	1	0
6q23	rs7776054	135460609	MCH	0	51453	6.6E-108	0	0
6q23	rs7776054	135460609	MCHC	0	56217	1.3E-14	0	0
6q23	rs9373124	135464902	RBC	0	53337	5.1E-97	0	0
6q23	rs9373124	135464902	PCV	0	52764	3.9E-22	0	0
6q23	rs9389269	135468852	HB	0	60896	3.3E-10	0	1
6q24	rs590856	139886122	MCV	0	58041	5.0E-36	1	0
6q24	rs590856	139886122	MCH	0	51638	3.8E-28	0	0
6q26	rs736661	164402826	MCH	0	51397	1.6E-11	1	1
6q26	rs9356181	164397016	MCV	1	92591	3.7E-09	0	1
7p13	rs12718598	50395939	MCV	0	37967	1.6E-13	1	0
7p13	rs7385935	50398365	RBC	0	36708	1.8E-11	0	1
7q22	rs2075672	100078232	RBC	0	41805	1.9E-20	1	0
7q22	rs7385804	100073906	MCV	0	46944	1.8E-15	0	0
7q22	rs2075672	100078232	MCH	0	39779	3.5E-19	0	1
7q22	rs1734910	100147480	MCHC	1	74688	3.5E-09	0	1
7q22	rs1734910	100147480	PCV	1	71217	2.1E-17	0	0
7q36	rs10480300	151036938	HB	0	49771	7.8E-15	1	0
7q36	rs10224210	151044127	RBC	0	45569	7.0E-13	0	1
7q36	rs10224210	151044127	PCV	0	41880	5.7E-12	0	0
8p11	rs4737009	41749562	MCHC	0	54462	4.9E-11	1	1
8p11	rs6987853	42576607	MCHC	0	52954	6.1E-11	1	1
9p24	rs2236496	4834265	MCV	0	53761	1.4E-19	1	0
9p24	rs2236496	4834265	MCH	0	47382	2.3E-16	0	0
9p24	rs10758658	4846877	RBC	0	53374	2.3E-11	0	0
9q34	rs579459	135143989	RBC	0	53362	9.3E-18	1	0
9q34	rs7853989	135121413	HB	0	60898	1.1E-17	0	0
9q34	rs579459	135143989	PCV	0	52789	7.6E-14	0	0
10q11	rs901683	45286428	MCV	0	58051	1.5E-16	1	0
10q11	rs10900128	44713207	RBC	0	53403	2.9E-12	0	1
10q11	rs901683	45286428	MCH	0	51648	1.8E-12	0	0
10q22	rs10159477	70769894	HB	0	45553	4.4E-20	1	0
10q22	rs16926246	70763398	MCV	1	72577	1.0E-09	0	1
10q22	rs10159477	70769894	PCV	0	37078	3.8E-16	0	0
10q24	rs11190134	101272190	MCH	1	86210	1.3E-10	1	1
11p15	rs11042125	8894625	HB	1	95803	1.5E-09	1	1
11p15	rs7936461	9997462	PCV	1	84200	1.0E-09	1	1

11p15	rs10500721	10160186	HB	1	95722	9.7E-09	0	1
11q13	rs2302264	66964002	MCV	0	57841	1.3E-10	1	1
11q13	rs11601325	66983582	MCH	1	82182	9.8E-09	0	1
11q13	rs7125949	72686732	HB	1	83983	2.1E-09	1	1
12p13	rs7312105	2393616	PCV	1	93989	3.2E-09	1	1
12p13	rs10849023	4202739	MCH	0	42647	7.5E-12	1	1
12p13	rs10849023	4202739	MCV	0	45906	1.9E-11	0	1
12p13	rs11611647	4204180	RBC	1	77610	1.4E-11	0	0
12q22	rs11104870	87353425	RBC	1	86405	6.2E-11	1	1
12q24	rs3184504	110368991	HB	0	56784	4.3E-19	1	0
12q24	rs3184504	110368991	PCV	0	48711	7.9E-16	0	0
12q24	rs3184504	110368991	RBC	0	49291	1.8E-12	0	1
12q24	rs3829290	119610821	MCV	1	86709	2.1E-09	1	1
12q24	rs3829290	119610821	MCH	1	82035	9.2E-09	0	1
14q23	rs7155454	64571992	MCH	0	51228	1.8E-12	1	1
14q24	rs11627546	69435677	MCV	1	104591	1.1E-09	1	1
14q32	rs17616316	102892515	MCH	1	90431	8.2E-11	1	1
15q21	rs1532085	56470658	HB	1	81428	6.7E-11	1	1
15q22	rs2572207	63857747	MCV	1	92608	3.4E-09	1	1
15q22	rs2572207	63857747	MCH	1	86206	1.1E-16	0	0
15q24	rs8028632	73108315	MCV	0	53602	6.9E-10	1	1
15q24	rs11072566	74081026	HB	1	115702	3.0E-10	1	1
15q25	rs2867932	76378092	MCHC	1	91040	3.3E-09	1	1
16p11	rs11248850	103598	MCH	0	51345	6.3E-23	1	0
16p11	rs11248850	103598	MCV	0	57749	4.4E-19	0	0
16q22	rs2271294	66459827	RBC	1	86678	1.1E-09	1	1
16q24	rs10445033	87367963	MCHC	0	42050	1.5E-22	1	0
16q24	rs837763	87381230	HB	0	42032	7.1E-11	0	1
17p11	rs888424	19926019	MCH	0	51274	5.4E-20	1	1
17p11	rs17759083	19904197	MCV	0	57707	1.5E-17	0	1
17q11	rs2070265	24099550	MCH	0	51503	5.1E-14	1	1
17q11	rs7215310	24112752	MCV	1	92812	1.5E-09	0	1
17q11	rs7221773	24227014	MCHC	1	90942	4.2E-12	0	1
17q12	rs8182252	34981476	RBC	1	82891	5.9E-09	1	1
17q21	rs2269906	39649863	MCHC	0	56263	2.0E-11	1	1
17q21	rs12150672	41182408	RBC	0	53489	4.7E-12	1	1
17q21	rs12150672	41182408	PCV	0	52920	2.7E-09	0	1
17q21	rs17426106	41184706	HB	1	95789	4.9E-09	0	1
17q25	rs4969184	73905008	HB	1	95722	7.0E-09	1	1
18q21	rs4890633	42087276	MCH	0	51375	1.9E-23	1	1
18q21	rs4890633	42087276	MCV	0	57778	1.1E-17	0	1
19p13	rs2159213	2087102	HB	1	95656	1.9E-09	1	1
19p13	rs2159213	2087102	PCV	1	87602	2.7E-09	0	1
19p13	rs12982593	2126891	RBC	1	81802	2.0E-11	0	1
19p13	rs732716	4317219	MCV	0	58044	1.5E-14	1	1
19p13	rs732716	4317219	MCH	0	51641	6.0E-13	0	1
19p13	rs741702	12885250	MCH	0	45178	8.2E-20	1	0
19p13	rs11085824	12862547	RBC	0	44718	5.9E-12	0	1

19p13	rs741702	12885250	MCV	0	49482	1.7E-17	0	0
19q13	rs3892630	37873324	MCV	1	104479	1.0E-10	1	1
20q13	rs737092	55423811	MCV	0	35156	4.0E-13	1	0
20q13	rs737092	55423811	MCH	0	32108	6.4E-11	0	1
20q13	rs737092	55423811	RBC	0	33925	9.9E-09	0	1
21q22	rs2032314	34276393	PCV	1	111306	7.5E-10	1	1
21q22	rs11910015	34260508	HB	1	96063	7.2E-12	0	1
22q11	rs5754217	20269675	MCV	0	53759	8.6E-10	1	0
22q12	rs5749446	31210585	MCH	0	51609	3.3E-13	1	1
22q12	rs5749446	31210585	MCV	0	58012	4.0E-11	0	0
22q12	rs855791	35792882	MCH	0	38547	1.0E-69	1	0
22q12	rs855791	35792882	MCV	0	43197	2.4E-54	0	0
22q12	rs855791	35792882	HB	0	46184	4.7E-40	0	0
22q12	rs855791	35792882	PCV	0	44036	4.3E-20	0	0
22q12	rs855791	35792882	MCHC	0	41609	3.1E-17	0	0
22q13	rs140522	49318132	MCV	0	44680	4.5E-23	1	1
22q13	rs140522	49318132	MCH	0	38308	8.0E-21	0	0
22q13	rs140522	49318132	RBC	0	39904	3.5E-09	0	1

Supplementary Table 8. Potential secondary SNPs with independent effects on phenotype at the genomic region associated with red blood cell phenotypes. For each genomic region, the SNP most closely associated with phenotype in the discovery GWAS is listed (lead SNP) along with the identity and association test results for SNPs showing independent association with phenotype in conditional analysis.

Region	Phenotype	GWAS Lead SNP	GWAS P	Position	SNPs at P<10-8 in conditional analysis
3q29	MCH	rs11717368	6.6E-19	197318754	rs11717368 (1.9E-15; *TFRC* [n]), rs4916478 (1.8E-10; *TFRC* [e], *ZDHHC19* [n])
	MCV	rs11717368	2.0E-14	197318754	rs4916478 (2.9E-13; *TFRC* [e], *ZDHHC19* [n]), rs3804139 (1.6E-12; *TFRC* [nc])
4q11	RBC	rs218238	2.8E-39	55089781	rs218238 (1.1E-43; *KIT* [n]), rs6824783 (3.9E-09; *KIT* [n])
6p21	HB	rs198846	1.4E-30	26215442	rs198846 (4.5E-34; *HFE* [c], *HIST1H1T* [n]), rs1408272 (2.6E-20; *HFE* [c], *SLC17A3* [n])
	MCH	rs1408272	4.8E-67	25950930	rs198846 (1.6E-72; *HFE* [c], *HIST1H1T* [n]), rs1408272 (6.7E-72; *HFE* [c], *SLC17A3* [n])
	MCV	rs1408272	3.2E-47	25950930	rs198846 (2.9E-50; *HFE* [c], *HIST1H1T* [n]), rs1408272 (2.1E-49; *HFE* [c], *SLC17A3* [n])
6p21	MCH	rs11968166	9.4E-34	42033282	rs11970772 (2.1E-47; *CCND3* [n]), rs6934551 (2.0E-19; *CCND3* [n])
	MCV	rs9349204	2.4E-40	42022356	rs9349204 (3.6E-55; *CCND3* [n]), rs2479720 (2.2E-23; *CCND3* [n])
6q23	MCH	rs7776054	6.6E-108	135460609	rs7776054 (3.8E-77; *HBS1L* [n]), rs2210366 (1.4E-13; *HBS1L* [n])
	MCV	rs9389269	2.6E-109	135468852	rs9389269 (2.2E-73; *HBS1L* [n]), rs2210366 (1.8E-13; *HBS1L* [n])
	RBC	rs9373124	5.1E-97	135464902	rs9373124 (6.0E-48; *HBS1L* [n]), rs1411919 (4.6E-30; *HBS1L* [n]), rs1320962 (2.1E-22; *MYB* [n]), rs10484494 (3.9E-14; *HBS1L* [n])
9q34	HB	rs7853989	1.1E-17	135121413	rs7853989 (7.0E-17; *ABO* [n]), rs579459 (1.9E-14; *ABO* [n])
	RBC	rs579459	9.3E-18	135143989	rs579459 (2.9E-17; *ABO* [n]), rs8176725 (1.2E-12; *ABO* [n])
10q11	MCH	rs901683	1.8E-12	45286428	rs901683 (1.9E-10; *MARCH8* [nce]), rs7909074 (1.6E-09; *RASSF4* [n])
	MCV	rs901683	1.5E-16	45286428	rs901683 (1.4E-13; *MARCH8* [nce]), rs7909074 (4.7E-13; *RASSF4* [n])
16p11	MCH	rs11248850	6.3E-23	103598	rs11248850 (8.8E-17; *NPRL3* [n]), rs11248914 (8.4E-13; *HBM* [e], *ITFG3* [ne]), rs17525396 (1.3E-12; *NPRL3* [n])

Supplementary Table 9. Coding SNPs in transcribed genes in LD at r^2>0.8 (1000G EUR) with sentinel SNPs identified in the red blood cell phenotype genome-wide association study. Results are shown for the sentinel SNPs at each region, and also for secondary SNPs identified through conditional analysis (Supplementary Table 7). AF is frequency of allele A1 in the EUR population. Effect is per unit copy of allele A1, where available. P_{pheno} is for the association of coding SNP with phenotype; P_{hetero} is for comparison of effect size on phenotype between sentinel SNP and coding SNP; r^2 is LD between sentinel and coding SNP.

Region	GWAS SNP	Pheno	Coding SNP	Pos	Alleles (A1/A2)	r^2	AF	Effect	N	P_{pheno}	P_{hetero}	Gene	Amino Acid change	Protein Position
Sentinel SNPs														
1q23	rs857684	MCHC	rs41273491	156783766	C/T	0.81	0.25	NA	NA	NA	NA	*OR6Y1*	VAL,ILE	252
1q23	rs857684	MCHC	rs857685	156843733	A/C	1.00	0.27	0.018 (0.004)	56209	5.60E-16	1.00	*OR10Z1*	ASN,THR	294
1q23	rs857684	MCHC	rs857725	156874559	T/G	0.95	0.28	0.014 (0.004)	52085	1.60E-13	0.51	*SPTA1*	LYS,GLN	1693
1q44	rs3811444	RBC	rs3811444	246106074	C/T	1.00	0.33	0.018 (0.003)	34323	4.50E-10	1.00	*TRIM58*	THR,MET	374
6p21	rs1408272	MCH	rs1800562	26201120	G/A	0.82	0.05	0.425 (0.028)	29711	6.50E-56	0.92	*HFE*	CYS,TYR	282
6p21	rs9272219	RBC	rs1142323	32717170	A/G	0.84	0.28	NA	NA	NA	NA	*HLA-DQA1*	GLU,GLY	63
10q11	rs901683	MCV	rs3764990	45276834	G/A	1.00	0.08	0.350 (0.050)	57857	1.20E-15	0.84	*MARCH8*	PRO,SER	92
11q13	rs2302264	MCV	rs4930427	66957395	C/T	0.99	0.44	0.104 (0.029)	45955	6.75E-06	0.35	*RPS6KB2*	PHE,LEU	269
11q13	rs2302264	MCV	rs13859	66958732	C/T	0.99	0.44	NA	NA	NA	NA	*RPS6KB2*	ALA,VAL	420
11q13	rs7125949	HB	rs3741151	72698494	G/T	0.83	0.1	NA	NA	NA	NA	*ARHGEF17*	ARG,LEU	388
12q24	rs3184504	HB	rs3184504	110368991	T/C	1.00	0.53	0.051 (0.006)	56784	4.30E-19	1.00	*SH2B3*	TRP,ARG	262
12q24	rs3829290	MCV	rs555404	119660367	T/C	0.99	0.47	NA	NA	NA	NA	*ACADS*	LEU,PRO	202
16q22	rs2271294	RBC	rs1134760	66521704	T/C	0.95	0.18	0.011 (0.006)	11607	7.50E-02	0.38	*CTRL*	HIS,ARG	173
16q22	rs2271294	RBC	rs20549	66527431	A/G	0.95	0.18	0.016 (0.003)	53376	3.90E-09	0.88	*PSMB10*	LEU,ILE	107
17q21	rs12150672	RBC	rs114107890	41074926	A/G	0.99	0.23	NA	NA	NA	NA	*C17orf69*	TYR,CYS	132
17q21	rs12150672	RBC	rs16940681	41267940	G/C	1.00	0.23	NA	NA	NA	NA	*CRHR1*	GLU,GLN	280
17q21	rs12150672	RBC	rs62621252	41278722	T/C	1.00	0.23	NA	NA	NA	NA	*SPPL2C*	SER,PRO	224
17q21	rs12150672	RBC	rs62054815	41279046	G/A	1.00	0.23	NA	NA	NA	NA	*SPPL2C*	ALA,THR	332
17q21	rs12150672	RBC	rs12185233	41279434	G/C	1.00	0.23	0.015 (0.003)	53406	1.50E-09	0.53	*SPPL2C*	ARG,PRO	461
17q21	rs12150672	RBC	rs12185268	41279463	A/G	1.00	0.23	0.015 (0.003)	42181	5.00E-08	0.51	*SPPL2C*	ILE,VAL	471
17q21	rs12150672	RBC	rs12373123	41279853	T/C	1.00	0.23	0.016 (0.003)	45376	6.70E-09	0.73	*SPPL2C*	SER,PRO	601
17q21	rs12150672	RBC	rs12373139	41279910	G/A	1.00	0.23	0.014 (0.003)	42186	2.40E-07	0.40	*SPPL2C*	GLY,ARG	620
17q21	rs12150672	RBC	rs12373142	41279980	C/G	0.99	0.23	0.016 (0.003)	53406	3.80E-10	0.62	*SPPL2C*	PRO,ARG	643
17q21	rs12150672	RBC	rs63750417	41416612	C/T	1.00	0.23	NA	NA	NA	NA	*MAPT*	PRO,LEU	202
17q21	rs12150672	RBC	rs62063786	41416860	G/A	1.00	0.23	NA	NA	NA	NA	*MAPT*	ASP,ASN	285

17q21	rs12150672	RBC	rs62063787	41416873	T/C	1.00	0.23	NA	NA	NA	NA	*MAPT*	VAL,ALA	289
17q21	rs12150672	RBC	rs17651549	41417115	C/T	1.00	0.23	0.016 (0.003)	53394	2.70E-10	0.66	*MAPT*	ARG,TRP	370
17q21	rs12150672	RBC	rs10445337	41423237	T/C	1.00	0.23	0.016 (0.003)	45375	1.10E-08	0.68	*MAPT*	SER,PRO	447
17q21	rs12150672	RBC	rs116444268	41429810	T/C	1.00	0.23	NA	NA	NA	NA	*MAPT*	ASN,LYS	197
17q21	rs12150672	RBC	rs62063857	41432502	A/G	1.00	0.23	NA	NA	NA	NA	*MAPT,STH*	GLN,ARG	7
17q21	rs12150672	RBC	rs34579536	41464753	A/G	1.00	0.23	NA	NA	NA	NA	*KANSL1*	ILE,THR	1085
17q21	rs12150672	RBC	rs34043286	41472966	A/G	1.00	0.23	NA	NA	NA	NA	*KANSL1*	SER,PRO	718
19p13	rs732716	MCV	rs1127888	4405083	C/T	0.83	0.25	NA	NA	NA	NA	*UBXD1*	ALA,THR	31
19q13	rs3892630	MCV	rs8108621	37875131	G/A	0.99	0.2	NA	NA	NA	NA	*NUDT19*	ARG,GLN	142
22q11	rs5754217	MCV	rs2298428	20312892	C/T	0.89	0.18	0.172 (0.033)	48734	2.50E-07	0.62	*YDJC*	ALA,THR	263
22q12	rs5749446	MCH	rs11107	31205190	G/A	1.00	0.39	0.055 (0.017)	15759	5.60E-04	0.46	*FBXO7*	MET,ILE	36
22q12	rs855791	MCH	rs855791	35792882	A/G	1.00	0.6	0.193 (0.011)	38547	1.00E-69	1.00	*TMPRSS6*	VAL,ALA	736
Secondary SNPs														
3q29	rs3804139	MCH	rs3817672	197285208	C/T	0.84	0.55	0.061 (0.011)	42699	5.81E-10	0.15	*TFRC*	GLY,SER	142
6p21	rs198846	MCH	rs1799945	26199158	C/G	0.96	0.15	0.217 (0.015)	43454	4.01E-47	0.49	*HFE*	HIS,ASP	63

Supplementary Table 10. Relationship between sentinel SNPs from the GWAS with expression of cis genes (±1MB) in peripheral blood lymphocytes from i. 206 families of European descent (eQTL1)[50]; and ii. 1,469 unrelated individuals from the UK and Netherlands (eQTL2)[51]. Genes identified as eQTLs based on: $P<1x10^{-5}$ for association of sentinel SNP with transcript expression (Tx P1) and $r^2>0.8$ between Sentinel SNP and Transcript SNP (the SNP most closely associated with transcript). Tx P2: association of Transcript SNP with expression; LD between sentinel and peak SNPs ($\mathbf{r^2}$) calculated from HapMap CEU population.

Chr	Band	Sentinel SNP	Position1	Primary Pheno	Gene	Tx P1	Transcript SNP	Position2	Distance	Tx P2	r^2	Dataset
4	4q27	rs13152701	122970511	MCV	*CCNA2*	5.9E-13	rs4833236	122995517	25006	5.4E-13	1.00	eQTL1
6	6p23	rs6914805	16389166	MCH	*GMPR*	3.1E-09	rs9396658	16360832	-28334	2.2E-09	1.00	eQTL2
6	6p21	rs9272219	32710247	RBC	*HLA-DQA1 / HLA-DQA2*	3.2E-134	rs9272219	32710247	0	3.2E-134	1.00	eQTL2
8	8p11	rs6987853	42576607	MCHC	*C8orf40*	3.7E-30	rs2974354	42533100	-43507	5.6E-34	0.80	eQTL1
8	8p11	rs6987853	42576607	MCHC	*C8orf40*	3.7E-41	rs2923427	42504905	-71702	1.1E-43	0.89	eQTL2
10	10q11	rs901683	45286428	MCV	*MARCH8*	6.2E-07	rs2288619	45259824	-26604	2.9E-07	1.00	eQTL2
11	11p15	rs11042125	8894625	HB	*AKIP1 / C11orf16*	1.2E-43	rs10840147	8877280	-17345	5.8E-46	0.83	eQTL2
11	11p15	rs11042125	8894625	HB	*NRIP3*	3.2E-06	rs10840147	8877280	-17345	1.3E-06	0.83	eQTL2
11	11q13	rs2302264	66964002	MCV	*RPS6KB2*	2.5E-17	rs1638588	66956282	-7720	2.5E-17	1.00	eQTL2
11	11q13	rs2302264	66964002	MCV	*PTPRCAP, CORO1B*	2.6E-19	rs7114510	66981170	17168	7.0E-20	0.94	eQTL2
11	11q13	rs7125949	72686732	HB	*ARHGEF17*	4.5E-14	rs1002127	72781483	94751	1.3E-14	0.94	eQTL2
15	15q22	rs2572207	63857747	MCV	*PTPLAD1*	8.6E-17	rs688223	63518793	-338954	1.6E-17	0.90	eQTL1
15	15q25	rs2867932	76378092	MCHC	*DNAJA4*	1.5E-08	rs2867932	76378092	0	1.5E-08	1.00	eQTL1
16	16q22	rs2271294	66459827	RBC	*DUS2L*	4.2E-46	rs6499157	66662975	203148	1.5E-59	0.80	eQTL2
17	17q11	rs2070265	24099550	MCH	*ERAL1*	3.2E-10	rs2242345	24209953	110403	6.8E-11	1.00	eQTL2
17	17q11	rs2070265	24099550	MCH	*TRAF4*	1.3E-15	rs2242345	24209953	110403	2.2E-16	1.00	eQTL1
17	17q12	rs8182252	34981476	RBC	*CDK12*	2.6E-09	rs6503513	34815139	-166337	8.7E-11	1.00	eQTL1
17	17q12	rs12150672	41182408	RBC	*C17orf69*	4.5E-49	rs413844	41085167	-97241	7.9E-50	0.96	eQTL2
17	17q12	rs12150672	41182408	RBC	*ARHGAP27*	7.2E-11	rs10514879	41158754	-23654	2.9E-11	0.95	eQTL1
17	17q12	rs12150672	41182408	RBC	*ARL17B*	2.4E-20	rs199443	42174733	992325	6.0E-24	0.94	eQTL1
17	17q25	rs4969184	73905008	HB	*PGS1*	5.0E-10	rs4969183	73904967	-41	5.0E-10	1.00	eQTL1
18	18q21	rs4890633	42087276	MCH	*C18orf25*	1.5E-12	rs4574015	42056124	-31152	1.2E-12	0.96	eQTL1
19	19p13	rs741702	12885250	MCH	*CALR*	4.7E-10	rs2242517	12863563	-21687	9.8E-11	0.81	eQTL1
19	19p13	rs741702	12885250	MCH	*FARSA*	4.6E-10	rs2293683	12900284	15034	3.1E-10	1.00	eQTL1
22	22q11	rs5754217	20269674	MCV	*UBE2L3*	1.5E-23	rs5754217	20269674	0	1.5E-23	1.00	eQTL1
22	22q13	rs140522	49318132	MCV	*ECGF1*	1.3E-133	rs140522	49318132	0	1.3E-133	1.00	eQTL2
22	22q13	rs140522	49318132	MCV	*ECGF1*	3.0E-12	rs140522	49318132	0	3.0E-12	1.00	eQTL1

Supplementary Table 11. Candidate genes identified by GRAIL using Pubmed 2006 or 2011 datasets. P values are corrected for multiple testing.

			GRAIL 2006		GRAIL 2011	
Region	**SNP**	**Position**	**Gene**	**P**	**Gene**	**P**
1p36	rs1175550	3681388	*KIAA0562*	9.4E-01	*LRRC47*	2.0E-01
1p34	rs3916164	39842526	*BMP8A*	9.3E-01	*LOC728448*	5.4E-01
1p32	rs741959	47448820	*TAL1**	1.1E-02	*TAL1*	6.5E-02
1q23	rs857684	156842353	*SPTA1**	2.4E-02	*SPTA1**	1.5E-02
1q32	rs7529925	197273831	*PTPRC*	1.3E-01	*PTPRC*	1.8E-01
1q32	rs7551442	201921744	*ATP2B4*	1.0E-01	*ATP2B4*	4.4E-02
1q32	rs9660992	203516073	*NUAK2*	6.5E-01	*RBBP5*	3.4E-01
1q44	rs3811444	246106074	*TRIM58*	1.0E+00	*TRIM58*	9.6E-01
2p21	rs4953318	46208555	*PRKCE*	4.5E-01	*PRKCE*	3.2E-01
2p16	rs243070	60473790	*BCL11A*	6.0E-01	*BCL11A*	5.5E-02
2q13	rs10207392	111566130	*ACOXL*	2.2E-01	*ACOXL*	1.9E-01
3p24	rs9310736	24325815	*THRB*	1.4E-01	*THRB*	4.1E-01
3q22	rs6776003	142749183	*ZBTB38*	8.1E-01	*ZBTB38*	9.5E-01
3q23	rs13061823	143603476	*TRPC1*	6.2E-02	*TRPC1*	8.4E-02
3q29	rs11717368	197318754	*TFRC**	3.6E-02	*TFRC**	1.2E-02
4q11	rs218238	55089781	*KIT*	1.7E-01	*KIT*	5.1E-01
4q27	rs13152701	122970511	*TRPC3*	7.9E-02	*TRPC3*	7.3E-02
6p23	rs6914805	16389166	*GMPR*	5.7E-01	*GMPR*	1.5E-01
6p21	rs1408272	25950930	*HFE*	1.1E-01	*HFE**	5.2E-03
6p22	rs13219787	27969649	*HIST1H1B*	9.9E-01	*HIST1H3H*	9.2E-01
6p22	rs2097775	30462282	*RPP21*	9.6E-01	*RPP21*	5.2E-01
6p21	rs9272219	32710247	*HLA-DQA1*	9.9E-01	*HLA-DQA1*	9.3E-01
6p21	rs9349204	42022356	*CCND3**	3.0E-02	*CCND3*	8.1E-02
6p12	rs9369427	43919408	*VEGFA*	6.4E-01	*GTPBP2*	8.9E-01
6q21	rs1008084	109733658	*CD164*	7.7E-01	*ARMC2*	8.8E-01
6q23	rs9389269	135468852	*MYB*	1.4E-01	*HBS1L*	2.6E-01
6q24	rs590856	139886122	*TXLNB*	8.4E-01	*TXLNB*	8.1E-01
6q26	rs736661	164402826	*N/A*	N/A	*N/A*	N/A
7p13	rs12718598	50395939	*IKZF1**	1.8E-02	*IKZF1*	7.6E-02
7q22	rs2075672	100078232	*TFR2*	1.2E-01	*TFR2**	9.7E-03
7q36	rs10480300	151036938	*PRKAG2*	5.7E-02	*PRKAG2*	3.7E-02
8p11	rs4737009	41749562	*ANK1**	2.2E-02	*ANK1**	2.7E-02
8p11	rs6987853	42576607	*VDAC3*	3.0E-01	*SLC20A2*	2.7E-01
9p24	rs2236496	4834265	*RCL1*	8.4E-01	*RCL1*	9.0E-01
9q34	rs579459	135143989	*ABO*	3.1E-01	*ABO*	2.1E-01
10q11	rs901683	45286428	*ANXA8*	7.5E-01	*AGAP4*	3.1E-04
10q22	rs10159477	70769894	*HK1*	6.8E-02	*HK1**	1.1E-02
10q24	rs11190134	101272190	*NKX2-3*	5.5E-01	*NKX2-3*	9.7E-01
11p15	rs11042125	8894625	*AKIP1*	9.2E-01	*AKIP1*	9.1E-01
11p15	rs7936461	9997462	*ADM*	2.5E-01	*ADM*	2.2E-01
11q13	rs2302264	66964002	*CABP4*	1.3E-01	*CABP4*	1.4E-01
11q13	rs7125949	72686732	*P2RY6*	8.2E-01	*P2RY6*	8.5E-01

12p13	rs7312105	2393616	*CACNA1C*	2.1E-01	*CACNA1C*	1.5E-01
12p13	rs10849023	4202739	*CCND2**	1.2E-02	*CCND2**	3.4E-02
12q22	rs11104870	87353425	*TMTC3*	8.0E-01	*TMTC3*	6.0E-01
12q24	rs3184504	110368991	*SH2B3*	1.7E-01	*SH2B3*	2.6E-01
12q24	rs3829290	119610821	*ACADS*	2.6E-01	*ACADS*	1.0E-01
14q23	rs7155454	64571992	*RAB15*	3.4E-01	*FNTB*	2.7E-01
14q24	rs11627546	69435677	*SMOC1*	5.1E-01	*SMOC1*	8.2E-01
14q32	rs17616316	102892515	*CKB*	1.3E-01	*CKB*	1.1E-01
15q21	rs1532085	56470658	*LIPC*	8.2E-01	*LIPC*	7.6E-01
15q22	rs2572207	63857747	*SLC24A1*	4.7E-01	*SLC24A1*	6.3E-01
15q24	rs8028632	73108315	*RPP25*	8.6E-01	*RPP25*	5.7E-02
15q24	rs11072566	74081026	*NRG4*	4.2E-01	*NRG4*	3.5E-01
15q25	rs2867932	76378092	*ACSBG1*	5.5E-01	*ACSBG1*	6.4E-01
16p11	rs11248850	103598	*HBA1*	3.2E-01	*HBA1**	7.2E-04
16q22	rs2271294	66459827	*SLC9A5*	4.5E-01	*CBFB*	9.0E-01
16q24	rs10445033	87367963	*ZFPM1*	4.2E-01	*MVD*	1.3E-01
17p11	rs888424	19926019	*LOC284194*	9.9E-01	*AKAP10*	9.8E-01
17q11	rs2070265	24099550	*PROCA1*	2.8E-01	*PROCA1*	6.3E-01
17q12	rs8182252	34981476	*IKZF3*	1.0E-01	*IKZF3*	5.3E-01
17q21	rs2269906	39649863	*SLC4A1**	4.0E-02	*SLC4A1**	9.0E-03
17q21	rs12150672	41182408	*CRHR1*	9.6E-01	*KANSL1*	9.1E-01
17q25	rs4969184	73905008	*SOCS3*	3.2E-01	*PGS1*	8.8E-02
18q21	rs4890633	42087276	*ATP5A1*	3.5E-01	*ATP5A1*	2.9E-01
19p13	rs2159213	2087102	*DOT1L*	7.6E-01	*PLEKHJ1*	3.1E-01
19p13	rs732716	4317219	*SH3GL1*	7.9E-01	*CHAF1A*	8.8E-01
19p13	rs741702	12885250	*KLF1*	2.3E-01	*KLF1*	2.0E-01
19q13	rs3892630	37873324	*RGS9BP*	8.6E-01	*ANKRD27*	3.0E-01
20q13	rs737092	55423811	*PCK1*	9.6E-01	*SPO11*	8.5E-01
21q22	rs2032314	34276393	*ATP5O*	2.8E-01	*ATP5O*	1.2E-01
22q11	rs5754217	20269675	*HIC2*	1.0E+00	*HIC2*	7.1E-01
22q12	rs5749446	31210585	*FBXO7*	1.3E-02	*FBXO7*	3.4E-02
22q12	rs855791	35792882	*TMPRSS6*	8.1E-01	*TMPRSS6*	1.9E-01
22q13	rs140522	49318132	*CPT1B*	7.7E-01	*LOC440836*	2.0E-01

Supplementary Table 12. Canonical pathways analysis using the Ingenuity Pathway Analysis tool (IPA, Ingenuity Systems, CA, USA). The IPA Knowledge Base was used to explore the functional relationship between proteins encoded by the 121 candidate genes identified at the 75 loci associated with red blood cell phenotypes. Genes were analysed for direct interactions only and networks were generated with a maximum size of 35 molecules.

Candidate genes	Additional genes	P	Top Functions
26 genes: ANK1, ATP2B4, CACNA1C, CALR, CCND2, CCND3, CDK12, FBX07, HK1, HLADQA1, IKZF1, KIT, KITLG, MAPT, MAX, MPND, PRKCE, PTPRCAP, QKI, RPS6KB2, SH2B3, SLC4A1, SPTA1, STH, TAL1, THRB	*Akt, Calmoldulin, estrogen receptor, Histone h3, P38 MAPK, PI3K, PP2A, RNA polymerase II, Spectrin*	10^{-54}	Hematological Disease, Cellular Development, Hematological System Development and Function
15 genes: CCNA2,CITED2, CRHR1, EDC4, HEYL, HIST1H2AG, HIST1H3A, KANSL1, NPRL3, NRIP3, P2RY6, RBM38, SCO2, SPECC1, VEGFA	*CDK1, E2f, Gpcr, GPR55, GPR61, GPR84, GPR144, GPR173, GPR174, GPR176, GPR89A/GPR89B, GPR89C, GPRC6A, GRK5, LGR4, NR3C2, SMAD7, TAAR8, TP53, VN1R2*	10^{-26}	Developmental Disorder, Endocrine System Disorders, Cellular Growth and Proliferation
14 genes: BBS7, BCL11A, C18orf25, EIF5, FNTB, GMPR, HBS1L, NCAPH2, NEUROD2, NKX2-3, NUTF2, PPCDC, TMCC2, TRAF4,	*DBI,EBP, HNF4A, JUN, MEF2BNB-MEF2B, MITF, NR2F2, OSCAR, PMEL, POUF1, PXDN, TBC1D16, TCF4, TDRD7, TMC6, Tpsab1, Tsc22d3, UBE2I, VSX1, ZCCHC8, ZNF146*	10^{-24}	Cellular Development, Embryonic Development, Amino Acid Metabolism
13 genes: ACTL6B, ARHGAP27, ARHGEF17, CORO1B, ERAL1, LIPC, NRG4, ODF3B, PTPLAD1, SCAMP5, TMPRSS6, UBTF, XRN1	*AKT1, ARHGEF4, ARHGEF19, ARHGEF25, ERBB4, Erb4 dimer, ESR1, FN1, INPP5A, MYBPH, MYC, MYO9B, NDRG2, NRG3, OPHN1, PLEKHG2, PRKCA, RAC1, RHOA, RIBIN, SMARCA4, TMEM97*	10^{-22}	Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Tissue Development
13 genes: AKIP1, AP3D1, CTRL, DENND4A, HIST1H2BH/HIST1H2BO, SPPL2C, NEK8, PSMB10, SBF2, SH3GL1, TYMP, UBXN6, WDR61	*Asc2, ATP6AP2, BATF2, DAPK2, DEFB103A/DEFB103B, FIP1L1, FOS, GNL2, GSK3B, ILKAP, MDM2, NR5A2, PDLIM2, RELA, SNN, SOAT1, STAT1, TFGBR1, TREM1, YBX2, YWHAZ, ZNF133*	10^{-22}	Cell Death, Infectious Disease, Gene Expression
11 genes: ACADS, ATP5O, ATXN2, DNAJA4, FARSA, KCTD17, MARCH8, mir-181, MLEC, ST5, UBE2L3	*ATP5G1, ATP6AP1, ATP6V0E2, ATP6V1C1, ATP6V1D*	10^{-17}	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry

Supplementary Table 13. Top biological functions of candidate genes using the IPA software tool.

Biological function	P-value	P-value (corrected)	Candidate genes (N)
Diseases and Disorders			
Hematological Disease	4.99E-08	1.21E-02	23
Developmental Disorder	8.51E-07	1.21E-02	23
Genetic Disorder	8.51E-07	1.21E-02	23
Organismal Injury and Abnormalities	8.51E-07	1.21E-02	15
Metabolic Disease	8.51E-07	1.21E-02	23
Molecular and Cellular Functions			
Cellular Development	1.20E-06	1.21E-02	32
Cell Morphology	4.46E-06	1.15E-02	25
Cellular Growth and Proliferation	5.09E-06	1.21E-02	29
Cellular Movement	2.49E-05	1.21E-02	6
Cell Death	3.36E-05	1.21E-02	19
Physiological System Development and Function			
Hematological System Development and Function	1.41E-08	1.21E-02	26
Hematopoiesis	1.41E-08	1.21E-02	22
Lymphoid Tissue Structure and Development	5.09E-06	1.21E-02	20
Tissue Morpology	6.70E-06	1.09E-02	27
Immune Cell Trafficking	2.49E-05	1.15E-03	7

Supplementary Table 14. Summary of known biology for the 121 candidate genes. Blood cell phenotypes in humans are highlighted (red).

Region	SNP	Gene	Mouse homolog	Mouse pheno	OMIM	Gene summary
1p36	rs1175550	*CCDC27*	*Ccdc27*			Coiled-coil domain containing 27. Function unknown.
1p36	rs1175550	*LRRC48*	*Lrrc48*			Leucine rich repeat containing 48. Function unknown.
1p34	rs3916164	*HEYL*	*Heyl*			Hairy/enhancer-of-split related with YRPW motif-like. Encodes a basic helix-loop-helix-type transcription factor, thought to be involved in Notch signalling and a regulator of cell fate decisions[52]. HEYL may also influence androgen receptor function and be involved in prostate cancer pathogenesis[53].
1p32	rs741959	*TAL1*	*Tal1*	*Yes*	Leukemia-1, T-cell acute lymphocytic	T-cell acute lymphocytic leukemia 1. TAL1 is a basic loop helix transcriptional regulator with a key role in haematopoiesis. TAL1 is required for terminal differentiation and maturation of red blood cells [54], and also involved in oncogenesis in the T-cell lineage [55].
1q23	rs857684	*OR6Y1*	*Olfr220*			Olfactory receptor, family 6, subfamily Y, member 1. Encodes an olfactory receptor.
1q23	rs857684	*OR10Z1*	*Olfr419*			Olfactory receptor, family 10, subfamily Z, member 1. Encodes an olfactory receptor.
1q23	rs857684	*SPTA1*	*Spna1*	*Yes*	Elliptocytosis-2 MIM:130600; Pyropoikilocytosis MIM:266140; Spherocytosis, type 3 MIM:270970	Spectrin, alpha, erythrocytic 1. Spectrin is an actin cross linking and molecular scaffold protein that links the plasma membrane to the actin cytoskeleton, and functions in the determination of cell shape, arrangement of transmembrane proteins, and organization of organelles. Mutations in this gene result in a variety of hereditary red blood cell disorders[56].
1q32	rs7529925	*MIR181A1*	*Mir181a-1*			MicroRNA 181a-1. A non-coding gene, which is involved in post-transcriptional regulation of gene expression. miR-181a regulates p27 mRNA translation during myeloid cell differentiation and may play a role in myelodysplastic syndromes[57]. miRNA-181a has also been implicated in non-haematological malignancies including thyroid papillary[58], oral squamous cell[59], malignant glioma[60] and lung cancer[61].
1q32	rs7551442	*ATP2B4*	*Atp2b4*			ATPase, Ca++ transporting, plasma membrane 4. ATP2B4 belongs to the family of ion transport ATPases and plays a role in calcium homeostasis[62]. A role in red blood cell biology not described.
1q32	rs9660992	*TMCC2*	*Tmcc2*			Transmembrane and coiled-coil domain family 2. TMCC2 is a transmembrane protein, reported to be involved in amyloid production underlying Alzheimers Disease [63].
1q44	rs3811444	*TRIM58*	*Trim58*			Tripartite motif containing 58. Encodes a gene of unknown function that is strongly expressed in bone marrow.
2p21	rs4953318	*PRKCE*	*Prkce*	*Yes*		Protein kinase C, epsilon. PRKCE is a protein kinase involved in calcium and second messenger signalling. PRKCE has been shown to be involved in many

						different cellular functions, including macrophage activation, apoptosis, cardioprotection from ischemia, heat shock response, as well as insulin exocytosis[64].
2p16	rs243070	*BCL11A*	*Bcl11a*	*Yes*		B-cell CLL/lymphoma 11A. BCL11A encodes a zinc-finger protein involved in haematopoiesis. BCL11A influences HbF levels and disease severity in patients with beta-thalassemia[65]. Variants near BCL11A are also associated with type-2 diabetes[66].
2q13	rs10207392	*ACOXL*	*Acoxl*			Acyl-CoA oxidase-like. Function unknown.
3p24	rs9310736	*THRB*	*Thrb*	*Yes*	Thyroid hormone resistance MIM:188570; Thyroid hormone resistance, autosomal recessive MIM:274300; Thyroid hormone resistance, selective pituitary MIM:145650	Thyroid hormone receptor beta. THRB is a nuclear hormone receptor for triiodothyronine. Mutations in this gene are known to cause generalized thyroid hormone resistance[67].
3q22	rs6776003	*RASA2*	*Rasa2*			RAS p21 protein activator 2. RASA2 is a GTPase-activating protein, which acts to suppress RAS function, thereby influencing cellular proliferation and differentiation[68].
3q23	rs13061823	*XRN1*	*Xrn1*			5'-3' exoribonuclease 1. XRN1 localizes to cytoplasmic foci containing a complex of mRNA-degrading enzymes. XRN1 has been implicated in homologous recombination, meiosis, telomere maintenance, and microtubule assembly[69].
3q29	rs11717368	*TFRC*	*Tfrc*	*Yes*		Transferrin receptor. TFRC is the key recepetor for cellular uptake of iron, required for hameoglobin synthesis[70].
4q11	rs218238	*KIT*	*Kit*	*Yes*	Gastrointestinal stromal tumor, somatic MIM: 606764; Germ cell tumors MIM:273300; Leukemia, acute myeloid MIM:601626; Mast cell leukemia; Mastocytosis with associated hematologic disorder; Piebaldism MIM:172800	Kit proto-oncogene. KIT is a tyrosine kinase growth factor receptor expressed at high levels on haemopoietic stem cells and multipotent progenitor cells, which acts as a receptor for growth factors such as Stem Cell Factor[71].
4q27	rs13152701	*BBS7*	*Bbs7*		Bardet-Biedl syndrome 7 MIM:209900	Bardet-Biedl syndrome 7. BBS7 encodes one of seven proteins that form the BBSome complex which is involved in cilia formation and morphogenesis[72].
4q27	rs13152701	*CCNA2*	*Ccna2*	*Yes*		Cyclin A2. CCNA2 belongs to the highly conserved cyclin family, which function as regulators of CDK kinases, and may influence cell cycle transitions[73].
6p23	rs6914805	*GMPR*	*Gmpr*			Guanosine monophosphate reductase. GMPR encodes an enzyme that catalyzes conversion of guanosine monophosphate to inosine monophosphate, and which may be involved in the re-utilization of purine nucleosides[74].
6p21	rs1408272	*HFE*	*Hfe*	*Yes*	Hemochromatosis MIM:235200	Hemochromatosis. HFE is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin. This protein influences iron

6p21	rs1408272	*SLC17A3*	*Slc17a3*		Solute carrier family 17, sodium phosphate member 3. SLC17A3 is a voltage-driven organic anion transporter involved in elimination of various anionic drugs, as well as in the secretion of endogenous substrates such as urate[76].
6p22	rs13219787	*HIST1H2AM*			Histone cluster 1, H2am. HIST1H2AM encodes a member of the histone H2B family and is involved in determination of chromatin structure.
6p22	rs13219787	*HIST1H2BO*			Histone cluster 1. HIST1H2BO encodes a member of the histone H2B family and is involved in determination of chromatin structure.
6p22	rs13219787	*HIST1H3J*	*Hist1h3g*		HIST1H3J histone cluster 1, H3j. HIST1H3J encodes a member of the histone H3 family and is involved in determination of chromatin structure.
6p22	rs2097775	*TRIM39-RPP21*			TRIM39-RPP21 read through. This locus represents naturally occurring read-through transcription between the neighboring TRIM39 (tripartite motif-containing 39) and RPP21 (ribonuclease P/MRP 21kDa subunit) genes on chromosome 6. The read-through transcript encodes a fusion protein that shares sequence identity with each individual gene product[77].
6p21	rs9272219	*HLA-DQA1*			Major histocompatibility complex, class II, DQ alpha 1. HLA-DQA1 belongs to the HLA class II alpha chain paralogues and are expressed in antigen presenting cells. HLA-DQA1 plays a central role in the immune system by presenting peptides derived from extracellular proteins. Genetic variants at HLA-DQA1 are associated with multiple inflammatory disorders such as ulcerative colitis[78], SLE[79] and type-1 diabetes mellitus[80].
6p21	rs9272219	*HLA-DQA2*			Major histocompatibility complex, class II, DQ alpha 2. HLA-DQA2 belongs to the HLA class II alpha chain family, is located in intracellular vesicles and plays a central role in the peptide loading of MHC class II molecules. Class II molecules are expressed in antigen presenting cells and are used to present antigenic peptides on the cell surface to be recognized by CD4 T-cells. Genetic variants at HLA-DQA2 are associated with narcolepsy[81], SLE[82] and rheumatoid arthritis[83].
6p21	rs9349204	*CCND3*	*Ccnd3*	*Yes*	Cyclin D3. CCD3 is a member of the cyclin family, involved in cell cycle progression. CCND3 is thought to be critical for expansion of hematopoietic stem cells[84].
6p12	rs9369427	*VEGFA*	*Vegfa*	*Yes*	Vascular endothelial growth factor A. VEGFA is a member of the PDGF/VEGF growth factor family which binds to and activates a receptor tyrosine kinase involved in the regulation of angiogenesis and vasculogenesis[85]. Elevated levels of this protein is linked to POEMS syndrome, also known as Crow-Fukase syndrome[86]. Mutations in this gene have been associated with proliferative and nonproliferative diabetic retinopathy[87].
6q21	rs1008084	*CCDC162P*	*Ccdc162*		Coiled-coil domain containing 162, pseudogene
6q23	rs9389269	*HBS1L*	*Hbs1l*		HBS1-like (S. cerevisiae). HBS1L encodes a GTP-binding elongation factor expressed in multiple tissues. This genomic region influences erythrocyte,

						platelet, and monocyte counts as well as erythrocyte volume and hemoglobin content, and genetic variants at this locus are associated with fetal hemoglobin levels, pain crises in sickle cell disease, and with severity in beta-thalassemia/Hemoglobin E[88-90].
6q24	rs590856	*CITED2*	*Cited2*	*Yes*		Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2. CITED2 inhibits transactivation of HIF1A-induced genes by competing with binding of HIF1a to p300-CH1 and is an essential regulator of adult hematopoietic stem cells maintenance[91]. Mutations in this gene are a cause of cardiac septal defects[92].
6q26	rs736661	*QKI*	*Qk*	*Yes*		Quaking homolog. QK1 is an RNA-binding protein that regulates splicing, stability, and transport of mRNA[93]. Role in haemopoiesis not described.
7p13	rs12718598	*IKZF1*	*Ikzf1*	*Yes*	Leukemia, acute lymphoblastic	IKAROS family zinc finger 1. IKZF1 is a transcription factor expressed in the fetal and adult hemo-lymphopoietic system, which regulates lymphocyte differentiation[94]. IKZF1 may be involved in the development of acute leukaemia[95].
7q22	rs2075672	*ACTL6B*	*Actl6b*			Actin-like 6B. ACTL6B is an actin-related protein involved in cellular vesicular transport, spindle orientation, nuclear migration and chromatin remodelling[96].
7q22	rs2075672	*TFR2*	*Trfr2*	*Yes*	Hemochromatosis, type 3 MIM:604250	Transferrin receptor 2. TFR2 mediates cellular uptake of transferrin-bound iron, and which is involved in iron metabolism, hepatocyte function and erythrocyte differentiation[97].
7q36	rs10480300	*PRKAG2*	*Prkag2*		Cardiomyopathy, familial hypertrophic 6 MIM:600858; Glycogen storage disease of heart, lethal congenital MIM:261740; Wolff-Parkinson-White syndrome MIM:194200	Protein kinase, AMP-activated, gamma 2 non-catalytic subunit AMP-activated protein kinase. PRKAG2 is a member of the AMPK gamma subunit family which monitors cellular energy status and inactivates key enzymes involved in fatty acid and cholesterol biosynthesis[98].
8p11	rs4737009	*ANK1*	*Ank1*	*Yes*	Spherocytosis, type 1 MIM:182900	Ankyrin 1. ANK1 link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and is expressed in erythrocytes, brain and muscles[99]. Mutations in ANK1 are associated with hereditary spherocytosis[100].
8p11	rs6987853	*C8orf40*	*AI316807*			Hypothetical protein
9p24	rs2236496	*RCL1*	*Rcl1*			RNA terminal phosphate cyclase-like. RCL1 plays a role in 18S-ribosomal-subunit biogenesis in early pre-rRNA processing[101].
9q34	rs579459	*ABO*	*abo*			ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminyltransferase; transferase B, alpha 1-3-galactosyltransferase). ABO encodes a glycosyltransferase which converts the H antigen (a cell surface carbohydrate sequence) into the A or B antigen. ABO variants are strongly associated with several phenotypes, including pancreatic[102] and gastric carcinoma[103] and auto-immune disease[104].
10q11	rs901683	*MARCH8*	*March8*			MARCH8 is a ubiquitin ligase which adds ubiquitin to substrate proteins. MARCH8 plays a role in vesicular transport between membrane compartments and induces the internalization of membrane glycoproteins[105, 106]

10q22	rs10159477	*HK1*	*Hk1*	*Yes*	Hemolytic anemia due to hexokinase deficiency MIM:235700	Hexokinase 1. HK1 encodes a hexokinase present in the outer membrane of mitochondria, which converts glucose to glucose-6-phosphate, the substrate for glycolysis[107]. Mutations in this gene are associated with hemolytic anemia[108].
10q24	rs11190134	*NKX2-3*	*Nkx2-3*	*Yes*		Homeobox protein Nkx-2.3. NKX2 is a transcription factor which plays a role in the correct association of lymphocytes and splenic stromal elements[109]. Common variants at NKX2-3 are associated with Crohns disease and ulcerative colitis[110].
11p15	rs11042125	*AKIP1*	*D930014E17Rik*			AKIP1 A kinase (PRKA) interacting protein 1. AKIP1 is a p65-interacting protein involved in NF-kB signalling[111].
11p15	rs11042125	*C11orf16*	*BC051019*			Hypothetical protein
11p15	rs11042125	*NRIP3*	*Nrip3*			Nuclear receptor interacting protein 3. Function unknown.
11p15	rs11042125	*ST5*	*St5*			Suppression of tumorigenicity 5. ST5 acts as a regulator of MAPK1/ERK2 kinase, and may play a role in tumorigenesis[112]. Disruption of ST5 is associated with mental retardation and multiple congenital anomalies[113].
11p15	rs7936461	*SBF2*	*Sbf2*		Charcot-Marie-Tooth disease, type 4B2 MIM:604563	SET binding factor 2. SBF2 is member of the myotubularin-related protein family and encodes an inactive phosphoinositide 3-phosphatase [114].
11q13	rs2302264	*CORO1B*	*Coro1b*			Coronin, actin binding protein, 1B. CORO1B, are WD repeat-containing actin-binding proteins that regulate cell motility[115].
11q13	rs2302264	*PTPRCAP*	*Ptprcap*	*Yes*		Protein tyrosine phosphatase, receptor type, C-associated protein. PTPRCAP is a transmembrane phosphoprotein associated with tyrosine phosphatase PTPRC/CD45, a regulator of T- and B-lymphocyte activation[116]. Function in red blood cell lineage unclear.
11q13	rs2302264	*RPS6KB2*	*Rps6kb2*			Ribosomal protein S6 kinase, 70kDa, polypeptide 2. RPS6KB2 is a member of the ribosomal S6 kinase family of serine/threonine kinases and is involved in several pathways central to the carcinogenic process, including regulation of cell growth, insulin, and inflammation[117].
11q13	rs7125949	*ARHGEF17*	*Arhgef17*			Rho guanine nucleotide exchange factor (GEF) 17. ARHGEF17 is a guanine nucleotide exchange factors which facilitate the exchange of GDP for GTP to activate Rho- family GTPase. ARHGEF17 is involved in several signalling networks and is predominantly expressed in the heart[118].
11q13	rs7125949	*P2RY6*	*P2ry6*			Pyrimidinergic receptor P2Y, G-protein coupled. P2RY6 is a G-protein coupled receptor activated by extracellular nucleotides. PR2Y6 is expressed in macrophages and aortic endothelial and smooth muscle cells and is involved in both the direct contraction and endothelium-dependent relaxation of the aorta by UDP[119]. P2RY6 plays a role in enhancing vascular inflammation[120].
12p13	rs7312105	*CACNA1C*	*Cacna1c*		Brugada syndrome 3 MIM:611875; Timothy syndrome MIM:601005	Calcium channel, voltage-dependent, L type, alpha 1C subunit. CACNA1C encodes an alpha-1 subunit of a voltage-dependent calcium channel[121].
12p13	rs10849023	*CCND2*	*Ccnd2*	*Yes*		Cyclin D2. CCND2 belongs to the cyclin family, which regulates CDK kinases and plays a role in the cell cycle and tumorigenesis[122].

12q22	rs11104870	*KITLG*	*Kitl*	*Yes*	Hyperpigmentation, familial progressive, 2 MIM:145250;	KIT ligand. KITLG encodes the ligand of the KIT tyrosine-kinase receptor, separately identified as a candidate gene locus influencing red blood cell traits. KITLG is also known as Stem Cell Factor and augments mobilisation and proliferation of haematopoietic progenitor cells, synergistically with other haemopoiteic growth factors[123].
12q24	rs3184504	*ATXN2*	*Atxn2*		Spinocerebellar ataxia 2 MIM:183090	Ataxin 2. ATXN2 is a protein whose function is unknown. Mutations in ATXN2 cause spinocerebellar ataxia type 2 (SCA2) and increases the risk of developing ALS[124, 125].
12q24	rs3184504	*SH2B3*	*Sh2b3*	*Yes*	Erythrocytosis, somatic MIM:133100; Myelofibrosis, somatic MIM:254450; Thrombocythemia, somatic MIM:187950	SH2B adaptor protein 3. SH2B3 is a member of the SH2B adaptor family of proteins, and is a regulator of cytokine signalling by growth factors involved in haematopoiesis[126].
12q24	rs3829290	*ACADS*	*Acads*		Short-chain acyl-CoA dehydrogenase deficiency MIM: 606885	Encodes a mitochondrial flavoprotein, which catalyzes the initial step of the mitochondrial fatty acid beta-oxidation pathway. Mutations in this gene Underlie Short Chain Acyl-CoA Dehydrogenase Deficiency.
12q24	rs3829290	*MLEC*	*Mlec*			Malectin. MLEC is a carbohydrate-binding protein which plays a role in the genesis, processing and secretion of N-glycosylated proteins[127].
14q23	rs7155454	*FNTB*	*Fntb*	*Yes*		Farnesyltransferase, CAAX box, beta. FNTP catalyzes the transfer of a farnesyl moiety from farnesyl pyrophosphate to a cysteine residues as post-translational modification[128].
14q23	rs7155454	*MAX*	*Max*			MYC associated factor X. MAX is a transcription regulator involved in cell proliferation, differentiation and apoptosis[129]. Function in red blood cell lineage unknown.
14q24	rs11627546	*SMOC1*	*Smoc1*		Microphthalmia with limb anomalies MIM:206920	SPARC related modular calcium binding 1. SMOC1 is a calcium binding glycoprotein found in the basement membrane that may have a role in ocular and limb development[130, 131].
14q32	rs17616316	*EIF5*	*Eif5*			Eukaryotic translation initiation factor-5. EIF5 interacts with the 40S initiation complex to promote hydrolysis of bound GTP. The resulting functional 80S ribosomal initiation complex is then active[132].
15q21	rs1532085	*LIPC*	*Lipc*		Hepatic lipase deficiency MIM:614025;	Lipase, hepatic. LIPC encodes hepatic triglyceride lipase, which is expressed in liver. LIPC has the dual functions of triglyceride hydrolase and ligand/bridging factor for receptor-mediated lipoprotein uptake[133]. Variants in LIPC are associated with intermediate and large drusen, as well as advanced age-related macular degeneration[134].
15q22	rs2572207	*DENND4A*	*Dennd4a*			DENN/MADD domain containing 4A. DENND4A is a transcription factor involved in interferon signalling. DENND4A is ubiquitously expressed with highest levels in bone marrow[135]. Function in red blood cell lineage not described.
15q22	rs2572207	*PTPLAD1*	*Ptplad1*			Protein tyrosine phosphatase-like A domain containing 1. PTPLAD1 is a transmembrane protein involved in very long chain fatty acid synthesis and which may also influence gene expression[136].

15q24	rs8028632	*PPCDC*	*Ppcdc*			Phosphopantothenoylcysteine decarboxylase. PPCDC is an enzyme involved in biosynthesis of coenzyme A, an essential component of fatty acid synthesis and oxidation, and pyruvate oxidation[137].
15q24	rs8028632	*SCAMP5*	*Scamp5*			Secretory carrier membrane protein 5. SCAMP5 is a membrane protein required for the calcium-dependent exocytosis of signal sequence-containing cytokines[138]. May play a role in accumulation of expanded polyglutamine protein huntingtin in case of endoplasmic reticulum stress by inhibiting the endocytosis pathway[139].
15q24	rs11072566	*NRG4*	*Nrg4*			Neuregulin 4. NRG4 is a ligand for tyrosine-kinase receptors involved in growth factor signalling epidermal growth factor receptors[140].
15q25	rs2867932	*DNAJA4*	*Dnaja4*			DNAJ (Hsp40) homolog, subfamily A, member 4 . DNAJA4 is a SREBP-responsive gene which has been reported to be involved in cholesterol biosynthesis[141].
15q25	rs2867932	*WDR61*	*Wdr61*			WD repeat domain 61. WDR61 may be involved in transcriptional regulation[142]. Function in red blood cell lineage not described.
16p11	rs11248850	*NPRL3*	*nprl3*			Nitrogen permease regulator-like 3. NPRL3 is a protein that forms a heterodimer with NPRL2[143]. Function unknown.
16q22	rs2271294	*CTRL*	*Ctrl*			Chymotrypsin-like. CTRL is a protease related to Chymotrypsin and synthesized primarily in pancreas[144]. Function uncertain.
16q22	rs2271294	*EDC4*	*Edc4*			Enhancer of mRNA decapping 4. EDC4 is a component of a complex containing DCP2 and DCP1A that plays a role in mRNA decapping during the process of mRNA degradation[145]. Red blood cell function unknown.
16q22	rs2271294	*NUTF2*	*Nutf2-ps1*			Nuclear transport factor 2. NUTF2 is a cytosolic factor that facilitates protein transport into the nucleus via interaction with the nuclear pore complex glycoprotein[146].
16q22	rs2271294	*PSMB10*	*Psmb10*	*Yes*		Proteasome (prosome, macropain) subunit, beta type, 10. The proteasome is a multicatalytic proteinase complex. This subunit is involved in antigen processing to generate class I binding peptides.
16q24	rs10445033	*PIEZO1*	*Fam38a*			Piezo-type mechanosensitive ion channel component 1. PIEZO1 is a component of mechanosensitive channel required for the mechanosensitive currents. Plays a role in epithelial cell adhesion by maintaining integrin activation[147].
17p11	rs888424	*SPECC1*	*Specc1*			Sperm antigen with calponin homology and coiled-coil domains 1. SPECC1 belongs to the cytospin-A family and is localized in the nucleus. SPECC1 highly expressed in testis and some cancer cell lines[148]. A chromosomal translocation involving this gene and platelet-derived growth factor receptor, beta gene may be a cause of juvenile myelomonocytic leukemia[149].
17q11	rs2070265	*C17orf63*	*BC017647*			Hypothetical protein
17q11	rs2070265	*ERAL1*	*Eral1*			Era G-protein-like 1 (E. coli). ERAL1 is a GTPase, which may play a role in mitochondrial ribosomal small subunit assembly[150].
17q11	rs2070265	*NEK8*	*Nek8*		Nephronophthisis 9 MIM:613824	NIMA (never in mitosis gene a)- related kinase 8. NEK8 a member of the serine/threionine protein kinase family related to NIMA of Aspergillus nidulans[151].

						NEK8 is required for renal tubular integrity and may play a role in cell cycle progression[152, 153].
17q11	rs2070265	*TRAF4*	*Traf4*			TNF receptor-associated factor 4. TRAF4 is an adapter protein and signal transducer that links members of the tumor necrosis factor receptor family to different signaling pathways[154]. Plays a role in the activation of NF-kappa-B and JNK, and in the regulation of cell survival and apoptosis[155].
17q12	rs8182252	*CDK12*	*Cdk12*			Cyclin-dependent kinase 12. CDK12 is a cyclin-dependent kinase required for RNA splicing[156]. CDK12 may be involved in regulation of transcription elongation.[157]
17q12	rs8182252	*NEUROD2*	*Neurod2*			Neurogenic differentiation 2. NEUROD2 is a neurogenic basic helix-loop-helix protein found mainly in endovascular invasive cells. NEUROD2 can induce transcription from neuron-specific promoters[158] NEUROD2 is thought to play a role in the determination and maintenance of neuronal cell fates[159].
17q21	rs2269906	*SLC4A1*	*Slc4a1*	*Yes*	Ovalocytosis MIM:166900 ; Renal tubular acidosis, distal, AD MIM:179800; Renal tubular acidosis, distal, AR MIM:611590; Spherocytosis, type 4 MIM:612653;	Solute carrier family 4, anion exchanger, member 1. SLC4A1 is an anion exchanger expressed in the erythrocyte plasma membrane involved in carbon dioxide transport from tissues to lungs[160]. SLC4A1 is associated with the red blood cell membrane protein glycophorin A[161]. Mutations in SLC4A1 can result in hereditary spherocytosis, ovalocytosis and anaemia[162].
17q21	rs2269906	*UBTF*	*Ubtfl1*			Upstream binding transcription factor, RNA polymerase. UBTF is a transcription factor involved in expression of ribosomal RNA[163].
17q21	rs12150672	*ARHGAP27*	*Arhgap27*			Rho GTPase activating protein 27. ARHGAP27 is a GTPase-activating protein which may be involved in clathrin-mediated endocytosis, and internalization of transferrin receptors[164].
17q21	rs12150672	*ARL17B*				ADP-ribosylation factor-like 17B. ARL17B is a GTP-binding protein which may be involved in protein trafficking and vesicle transport[165].
17q21	rs12150672	*C17orf69*				Hypothetical protein
17q21	rs12150672	*CRHR1*	*Crhr1*			Corticotropin releasing hormone receptor 1. CRHR1 is a G-protein coupled receptor found in the central nervous system, which may influence ACTH release and physiological processes such as reproduction, immune response and obesity[166].
17q21	rs12150672	*SPPL2C*	*4933407P14Rik*			Signal peptide peptidase like 2C. SPPL2C is an enzyme that may act as intramembrane protease[167]. Function unknown.
17q21	rs12150672	*KANSL1*	*1700081L11Rik*			KAT8 regulatory NSL complex subunit 1. Encodes a protein that is a component of the MLL1/MLL complex[168], a transcriptional coactivator that plays an role in regulating gene expression during early development and hematopoiesis[169].
17q21	rs12150672	*MAPT*	*Mapt*	*Yes*	Dementia, frontotemporal, with or without parkinsonism MIM:600274; Pick disease MIM:172700; Supranuclear	Microtubule-associated protein tau. MAPT promotes microtubule assembly and stability[170], and might be involved in the establishment and maintenance of neuronal polarity[171]. Mutations in MAPT have been linked to several neurological disorders[172, 173].

					palsy, progressive MIM:601104; Supranuclear palsy, progressive atypical MIM:260540; Tauopathy and respiratory failure MIM:601104 ;	
17q21	rs12150672	*STH*				Saitohin. STH is a gene located in the human TAU gene, and which be involved in the pathogenesis of neurodegenerative disease[174]. Function in red blood cell lineage not described.
17q25	rs4969184	*PGS1*	*Pgs1*			Phosphatidylglycerophosphate synthase 1. PGS1 is an enzyme involved in the biosynthesis of phospholipids and cardiolipin[175]. Function in red blood cell lineage not described.
18q21	rs4890633	*C18orf25*	*8030462N17Rik*			Hypothetical protein
19p13	rs2159213	*AP3D1*	*Ap3d1*	*Yes*		Adaptor-related protein complex 3, delta 1 subunit. AP3D1 is part of the AP-3 complex, an adaptor-related complex which may play a role in vesicle formation and lysosomal trafficking[176].
19p13	rs732716	*MPND*	*Mpnd*			MPN domain-containing protein. MPND is likely to be a protease. Function unknown.
19p13	rs732716	*SH3GL1*	*Sh3gl1*		Leukemia, acute myeloid MIM:601626	SH3-domain GRB2-like. SH3GL1 is a member of the endophilin family of Src homology 3 domain-containing proteins, and is involved in endocytosis and regulation of the cell cycle[177]. SH3GL1 is implicated in acute myeloid leukemia as a fusion partner of the myeloid-lymphoid leukemia protein[178].
19p13	rs732716	*UBXN6*	*Ubxn6*			UBX domainprotein 6. UBXN6 is a VCP-interacting protein that is involved in endoplasmic reticulum-associated degradation[179].
19p13	rs741702	*CALR*	*Calr*			Calreticulin. CALR is a multifunctional Ca(2+) binding chaperone in the endoplasmic reticulum and expression of the protein is tightly regulated at the transcriptional level. CALR is critical for cardiac development and expression of the protein must be tightly regulated during cardiogenesis. Differential expression of calreticulin has been associated with several diseases, including neurodegenerative problems, cancers, autoimmune diseases and wound healing[180].
19p13	rs741702	*FARSA*	*Farsa*			Phenylalanyl-tRNA synthetase, alpha subunit. FARSA belongs to a class of enzymes that charge tRNAs with their cognate amino acids[181]. FARSA is expressed in a tumor-selective and cycle stale stage- and differentiation-dependent manner[182].
19p13	rs741702	*SYCE2*	*Syce2*			Synaptonemal complex central element protein 2. SYCE2 is a major component of the transverse central element of synaptonemal complexes (SCS), formed between homologous chromosomes during meiotic prophase[183]. May have a role in the synaptonemal complexes assembly, stabilization and recombination[184].

19q13	rs3892630	*NUDT19*	*Nudt19*			Nudix (nucleoside diphosphate linked moiety X)-type motif 19. NUDT19 is a coenzyme A diphosphatase that mediates the hydrolysis of a wide range of CoA esters[185].
20q13	rs737092	*RBM38*	*Rbm38*			RBM38 RNA binding motif protein 38. RBM38 acts as a mediator of the p53/TP53 family to regulate CDKN1A to induce cell cycle arrest[186].
21q22	rs2032314	*ATP5O*	*Atp5o*			ATP5O ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit. ATP5O is a component of the F-type ATPase found in the mitochondrial matrix which produces ATP from ADP[187].
22q11	rs5754217	*UBE2L3*	*Ube2l3*			Ubiquitin-conjugating enzyme E2L 3. UBE2L3 is a member of the E2 ubiquitin-conjugating enzyme family[188].
22q11	rs5754217	*YDJC*	*Ydjc*			YdjC homolog; function unknown.
22q12	rs5749446	*FBXO7*	*Fbxo7*		Parkinson disease 15, autosomal recessive MIM:260300	F-box protein 7. FBOX7 is a member of the F-box protein family which function in phosphorylation-dependent ubiquitination[189]. FBOX7 may play a role in regulation of haematopoiesis[190].
22q12	rs855791	*KCTD17*	*Kctd17*			Potassium channel tetramerisation domain containing 17. Function unknown.
22q12	rs855791	*TMPRSS6*	*Tmprss6*	*Yes*	Iron-refractory iron deficiency anemia MIM:206200	Transmembrane protease, serine 6. TMPRSS6 is a transmembrane serine proteinase involved in iron metabolism[191]. Mutations in TMPRSS6 cause iron-refractory iron deficiency anaemia[192].
22q13	rs140522	*TYMP*	*Tymp*		Mitochondrial DNA depletion syndrome 1 (MNGIE type) MIM:603041	Thymidine Phosphorylase. TYPM encodes an angiogenic factor which promotes angiogenesis in vivo and stimulates the in vitro growth of a variety of endothelial cells. It has a highly restricted target cell specificity acting only on endothelial cells[193].
22q13	rs140522	*NCAPH2*	*Ncaph2*	*Yes*		Non-SMC condensin II complex, subunit H2. NCAPH2 is a regulatory subunit of the condensin-2 complex, a complex that plays a role in mitotic chromosome assembly[194]. May play a lineage-specific role in T-cell development[195].
22q13	rs140522	*ODF3B*	*Odf3b*			Outer dense fiber of sperm tails 3B. Function unknown
22q13	rs140522	*SCO2*	*Sco2*		Cardioencephalomyopathy, fatal infantile, due to cytochrome c oxidase deficiency MIM:604377	SCO cytochrome oxidase deficient homolog 2 (yeast). SCO2 plays a role in ATP production[196]. Mutations in this gene are associated with cardiac and neurologic disease[197].

Supplementary Table 15. Definitions of cell types in the Differentiation Map of Haematology (http://www.broadinstitute.org/dmap/home) used to investigate expression of the 121 candidate genes in haematologic precursors.

Abbreviation	Description
BASO	Basophils
TCELL1	CD4+ Central Memory
TCELL2	CD4+ Effector Memory
TCELL3	CD8+ Central Memory
TCELL4	CD8+ Effector Memory
TCELL5	CD8+ Effector Memory RA
GRAN1	Colony Forming Unit-Granulocyte
MK1	Colony Forming Unit-Megakaryocytic
MONO1	Colony Forming Unit-Monocyte
CMP	Common myeloid progenitor
PreBCELL2	Early B-cell
EOS	Eosinophill
EB1	Erythroid_CD34+ CD71+ GlyA-
EB2	Erythroid_CD34- CD71+ GlyA-
EB3	Erythroid_CD34- CD71+ GlyA+
EB4	Erythroid_CD34- CD71lo GlyA+
EB5	Erythroid_CD34- CD71- GlyA+
GMP	Granulocyte/monocyte progenitor
GRAN3	Granulocyte (Neutrophil)
GRAN2	Granulocyte (Neutrophilic Metamyelocyte)
HSC1	Hematopoietic stem cell_CD133+ CD34dim
HSC2	Hematopoietic stem cell_CD38- CD34+
BCELL2	Mature B-cell class able to switch
BCELL3	Mature B-cell class switched
BCELL4	Mature B-cells
NK1	Mature NK cell_CD56- CD16- CD3-
NK2	Mature NK cell_CD56- CD16+ CD3-
NK3	Mature NK cell_CD56+ CD16+ CD3-
MK2	Megakaryocyte
MEP	Megakaryocyte/ erythroid progenitor
MONO2	Monocyte
DEND1	Myeloid Dendritic Cell
BCELL1	NaÃ¯ve B-cells
TCELL6	Naive CD4+ T-cell
TCELL7	Naive CD8+ T-cell
NKT	NKT
DEND2	Plasmacytoid Dendritic Cell
PreBCELL1	Pro B-cell

Supplementary Table 16. Nucleosome deplete regions identified by FAIRE-seq, containing either a sentinel SNP from the RBC GWAS or a SNP in high LD (r^2>0.8). Cell type: EB - erythroblasts, MO - monocytes, MK - megakaryocytes. Distance: the distance between the Sentinel SNP and the NDR SNP.

GWAS region						Nucleosome-depleted region							
#	Region	Sentinel SNP	Chr	Position	Candidate gene(s)	Cell type	Start position	End position	NDR SNP	Position	Allele freq	r^2	Distance (bp)
1	1p36	rs1175550	1	3,691,528	*CCDC27, LRRC48*	EB	3,691,366	3,691,660	rs1175550	3,691,528	0.206	1.000	0
2	1p32	rs741959	1	47,676,233	*TAL1*	EB	47,679,193	47,679,296	rs4926524	47,679,258	0.561	0.850	3,025
3	1q23	rs857684	1	158,575,729	*OR6Y1, OR10Z1, SPTA1*	MO	158,596,370	158,596,475	rs2482963	158,596,438	0.278	0.941	20,709
4	1q32	rs7529925	1	199,007,208	*MIR181A1*	EB	199,010,573	199,011,066	rs1434282	199,010,721	0.724	0.929	3,513
						MK	199,010,612	199,011,128	rs1434282	199,010,721	0.724	0.929	3,513
5	2q13	rs10207392	2	111,849,659	*ACOXL*	MO	111,843,101	111,843,381	rs2880112	111,843,166	0.435	0.842	6,493
6	3q22	rs6776003	3	141,266,493	*RASA2*	EB	141,217,874	141,218,028	rs6808837	141,217,954	0.384	0.857	48,539
						MO	141,217,713	141,218,095	rs6808837	141,217,954	0.384	0.857	48,539
7	3q23	rs13061823	3	142,120,786	*XRN1*	MK	142,233,859	142,234,023	rs6791816	142,233,990	0.589	0.824	113,204
8	4q11	rs218238	4	55,395,024	*KIT*	EB	55,408,759	55,409,035	rs218264	55,408,875	0.258	0.834	13,851
9	4q27	rs13152701	4	122,751,061	*BBS7, CCNA2*	EB	122,744,961	122,745,400	rs769236	122,745,038	0.368	1.000	6,023
						MK	122,750,001	122,750,254	rs13145213	122,750,079	0.369	0.994	982
						EB	122,791,573	122,791,906	rs2271176	122,791,601	0.368	1.000	40,540
10	6p21	rs9349204	6	41,914,378	*CCND3*	EB	41,924,850	41,925,202	rs9349205	41,925,159	0.234	0.850	10,781
11	6q21	rs1008084	6	109,626,965	*CCDC162*	EB	109,625,663	109,626,087	rs1546723	109,625,879	0.422	0.989	1,086
						MK	109,625,692	109,625,981	rs1546723	109,625,879	0.422	0.989	1,086
12	6q23	rs9389269	6	135,427,159	*HBS1L*	EB	135,419,430	135,419,712	rs9389268	135,419,631	0.271	0.911	7,528
						EB	135,419,430	135,419,712	rs9376091	135,419,636	0.271	0.911	7,523
						EB	135,419,430	135,419,712	rs9402685	135,419,688	0.271	0.911	7,471
						EB	135,431,304	135,431,674	rs6920211	135,431,318	0.258	0.850	4,159
						EB	135,431,304	135,431,674	rs9494142	135,431,640	0.255	0.863	4,481
13	6q24	rs590856	6	139,844,429	*CITED2*	EB	139,839,765	139,840,281	rs589235	139,839,960	0.499	0.904	4,469
14	6q26	rs736661	6	164,482,836	*QKI*	EB	164,463,287	164,463,672	rs4709819	164,463,355	0.446	1.000	19,481
						MK	164,463,305	164,463,683	rs4709819	164,463,355	0.446	1.000	19,481
						EB	164,463,287	164,463,672	rs4709820	164,463,572	0.446	1.000	19,264
						MK	164,463,305	164,463,683	rs4709820	164,463,572	0.446	1.000	19,264
15	8p11	rs4737009	8	41,630,405	*ANK1*	EB	41,630,153	41,630,603	rs4737009	41,630,405	0.261	1.000	0
						MK	41,630,356	41,630,449	rs4737009	41,630,405	0.261	1.000	0
						EB	41,630,153	41,630,603	rs4737010	41,630,447	0.250	0.946	42
						MK	41,630,356	41,630,449	rs4737010	41,630,447	0.250	0.946	42
16	9p24	rs2236496	9	4,844,265	*RCL1*	EB	4,852,346	4,852,777	rs10758656	4,852,599	0.193	0.950	8,334
17	10q11	rs901683	10	45,966,422	*MARCH8*	EB	45,966,011	45,966,515	rs901683	45,966,422	0.079	1.000	0
						EB	46,039,726	46,040,064	rs75595592	46,039,930	0.079	1.000	73,508
						MK	46,052,986	46,053,213	rs9422657	46,053,061	0.079	1.000	86,639

18	11p15	rs11042125	11	8,938,049	*AKIP1, C11orf16, NRIP3, ST5*	MO	9,023,307	9,023,742	rs7479407	9,023,421	0.413	0.873	85,372
19	11q13	rs7125949	11	73,009,084	*ARHGEF17, P2RY6*	MO	73,115,131	73,115,357	rs7114009	73,115,314	0.114	0.827	106,230
20	14q23	rs7155454	14	65,502,239	*FNTB, MAX*	MK	65,499,871	65,500,238	rs12435835	65,499,909	0.481	0.989	2,330
						EB	65,509,783	65,510,185	rs11628273	65,509,878	0.481	0.989	7,639
						MK	65,509,766	65,510,128	rs11628273	65,509,878	0.481	0.989	7,639
21	15q22	rs2572207	15	66,070,693	*DENND4A, PTPLAD1*	EB	66,070,410	66,070,913	rs2572207	66,070,693	0.209	1.000	0
22	15q24	rs8028632	15	75,321,262	*PPCDC, SCAMP5*	EB	75,315,650	75,316,026	rs2304903	75,315,778	0.225	1.000	5,484
						MK	75,315,722	75,315,936	rs2304903	75,315,778	0.225	1.000	5,484
						MK	75,322,112	75,322,277	rs35911108	75,322,179	0.225	1.000	917
						MO	75,354,500	75,354,655	rs35577967	75,354,621	0.212	0.881	33,359
23	16p11	rs11248850	16	163,598	*NPRL3*	EB	163,293	163,915	rs11248850	163,598	0.487	1.000	0
						MK	163,466	163,821	rs11248850	163,598	0.487	1.000	0
						EB	163,293	163,915	rs11865131	163,667	0.487	1.000	69
						MK	163,466	163,821	rs11865131	163,667	0.487	1.000	69
						EB	169,902	170,268	rs11866877	170,044	0.460	0.850	6,446
24	16q22	rs2271294	16	67,902,326	*CTRL, EDC4, NUTF2, PSMB10*	EB	67,926,933	67,927,181	rs7196789	67,927,124	0.172	0.991	24,798
						MK	67,926,943	67,927,172	rs7196789	67,927,124	0.172	0.991	24,798
25	16q24	rs10445033	16	88,840,462	*PIEZO1*	EB	88,840,369	88,840,702	rs10445033	88,840,462	0.634	1.000	0
						MK	88,840,437	88,840,569	rs10445033	88,840,462	0.634	1.000	0
26	17q21	rs2269906	17	42,294,337	*SLC4A1, UBTF*	EB	42,323,033	42,323,732	rs7209801	42,323,376	0.272	0.804	29,039
27	17q21	rs12150672	17	43,826,637	*ARHGAP27, ARL17B, C17orf69, CRHR1, SPPL2C, KANSL1, MAPT, STH*	MK	44,217,038	44,217,183	rs2532314	44,217,112	0.230	1.000	390,475
						EB	44,253,293	44,253,384	rs2532259	44,253,364	0.230	1.000	426,727
						EB	44,271,327	44,271,746	rs2532236	44,271,430	0.230	1.000	444,793
						MK	44,271,426	44,271,848	rs2532236	44,271,430	0.230	1.000	444,793
						MO	44,271,954	44,272,323	rs2532235	44,272,000	0.230	1.000	445,363
						MO	44,271,954	44,272,323	rs2532234	44,272,266	0.230	1.000	445,629
						MK	44,272,445	44,272,713	rs17663792	44,272,552	0.231	0.993	445,915
						MO	44,272,408	44,272,673	rs17663792	44,272,552	0.231	0.993	445,915
						MK	44,272,445	44,272,713	rs2732660	44,272,679	0.230	1.000	446,042
						MK	44,276,225	44,276,411	rs1918785	44,276,330	0.230	1.000	449,693
						MK	44,280,017	44,280,226	rs2732675	44,280,188	0.230	1.000	453,551
						MO	44,289,093	44,289,164	rs2732629	44,289,101	0.230	0.985	462,464
						MO	44,289,093	44,289,164	rs2732630	44,289,150	0.230	0.985	462,513
28	18q21	rs4890633	18	43,833,278	*C18orf25*	EB	43,802,643	43,803,150	rs12607898	43,802,778	0.733	0.980	30,500
29	19p13	rs732716	19	4,366,219	*MPND, SH3GL1, UBXD1*	EB	4,457,808	4,458,132	rs11670503	4,458,063	0.254	0.839	91,844
30	19p13	rs741702	19	13,024,250	*CALR, FARSA, SYCE2*	EB	13,001,484	13,002,038	rs11085824	13,001,547	0.310	0.840	22,703
						MK	13,001,466	13,002,026	rs11085824	13,001,547	0.310	0.840	22,703
						MO	13,001,382	13,002,002	rs11085824	13,001,547	0.310	0.840	22,703
						EB	13,030,047	13,030,376	rs8113575	13,030,280	0.701	0.921	6,030
						MO	13,044,538	13,044,646	rs2974750	13,044,544	0.700	0.903	20,294
31	20q13	rs737092	20	55,990,405	*RBM38*	EB	55,990,051	55,990,746	rs737092	55,990,405	0.496	1.000	0

32	22q12	rs5749446	22	32,880,585	*FBXO7*	EB	32,887,371	32,887,591	rs6518786	32,887,498	0.386	1.000	6,913
						EB	32,887,371	32,887,591	rs5754113	32,887,566	0.386	1.000	6,981
NDRs overlapping secondary SNPs at loci														
1	6p21	rs2479720	6	41915704	*CCND3*	EB	41,924,851	41,925,202	rs16895130	41924931	0.3042328	0.933	9,227

Supplementary Table 17. Number and tissue distribution of NDRs identified by FAIRE-seq, using either stringent or relaxed criteria for NDR calling. The numbers of SNPs in NDR regions that are potential causal variants (defined as being either i. a sentinel SNP from the GWAS or ii. a SNP in high LG (r^2>0.8) with a sentinel SNP are shown, along with enrichment compared to expected under null hypothesis.

Peak calling threshold[a]	Tissue	NDRs	Potential causal variants		Fold Enrichment	Binomial P (two sided)*	Bonferroni corrected P°
			Observed	Expected			
Stringent - T 8	EB	23570	25	13.7	1.8	0.001	0.007
	EB/MK	9377	13	5.4	2.4	0.002	0.017
	EB/MK/MO	3623	1	2.1	0.5	0.725	n.s.
	EB/MO	1352	1	0.8	1.3	0.546	n.s.
	MK	35621	9	20.7	0.4	0.001	0.007
	MK/MO	1578	1	0.9	1.1	0.603	n.s.
	MO	28187	10	16.4	0.6	0.081	n.s.
	All	103308	60				
Relaxed - T 6	EB	59698	47	26.0	1.8	2.27E-5	1.13E-4
	EB/MK	27527	18	12.0	1.5	0.09	n.s.
	EB/MK/MO	6913	3	3.0	1.0	1.00	n.s.
	EB/MO	4231	2	1.8	1.1	0.71	n.s.
	MK	110733	20	48.2	0.4	8.51E-8	4.26E-7
	MK/MO	5527	3	2.4	1.2	0.52	n.s.
	MO	86252	38	37.6	1.0	0.92	n.s.
	All	300881	131				

*Two tailed binomial tests
[a]The T parameter from Fseq makes reference to the number of standard deviations above the background to call a peak
°Bonferroni corrected for 7 tests.

Supplementary Table 18. Genes closest to an NDR containing a sentinel SNP or a SNP in high LD in erythroblasts only (see methods). For each gene, we determined the relationship of gene expression with time during erythropoiesis using linear regression, and calculated the t-statistic for the difference in beta from zero. Distance refers to the distance from the SNP to the transcription start site (TSS) of the transcript.

SNP in NDR	Chr	Position	Gene	Gene biotype	Distance to TSS	t statistic
rs1175550	1	3691528	*RP1-286D6.2*	protein_coding	1901	8.11
rs4926524	1	47679258	*TAL1*	protein_coding	-10490	5.20
rs218264	4	55408875	*AC006552.1*	lincRNA	-64423	No probe
rs769236	4	122745038	*CCNA2*	protein_coding	-49	2.03
rs2271176	4	122791601	*BBS7*	protein_coding	-4	2.86
rs9349205	6	41925159	*CCND3*	protein_coding	15573	4.20
rs9389268	6	135419631	*HBS1L*	protein_coding	-4563	1.76
rs9376091	6	135419636	*HBS1L*	protein_coding	-4558	1.76
rs9402685	6	135419688	*HBS1L*	protein_coding	-4506	1.76
rs6920211	6	135431318	*HBS1L*	protein_coding	7124	1.76
rs9494142	6	135431640	*HBS1L*	protein_coding	7446	1.76
rs589235	6	139839960	*RP11-12A2.3*	lincRNA	45768	No probe
rs10758656	9	4852599	*RP11-125K10.5*	processed_transcript	2226	No probe
rs901683	10	45966422	*MARCH8*	protein_coding	7089	4.35
rs75595592	10	46039930	*MARCH8*	protein_coding	9111	4.35
rs2572207	15	66070693	*DENND4A*	protein_coding	-13769	7.83
rs11866877	16	170044	*NPRL3*	protein_coding	640	15.04
rs7209801	17	42323376	*AC003102.1*	protein_coding	-3860	No probe
rs2532259	17	44253364	*KANSL1*	protein_coding	3770	-0.06
rs12607898	18	43802778	*C18orf25*	protein_coding	48790	3.27
rs11670503	19	4458063	*UBXN6*	protein_coding	272	4.67
rs8113575	19	13030280	*SYCE2*	protein_coding	204	-2.32
rs737092	20	55990405	*RBM38*	protein_coding	23119	4.83
rs6518786	22	32887498	*FBXO7*	protein_coding	4223	11.34
rs5754113	22	32887566	*FBXO7*	protein_coding	4291	11.34

Supplementary Table 19. Bioinformatic approaches used for identification of candidate genes and overlap with blood cell phenotypes in model organisms.

			Bioinformatic strategy				Phenotype in model organism	
Region	**Sentinel SNP**	**Gene**	**Proximity**	**eQTL**	**Coding**	**GRAIL**	**Mouse**	**Drosophila**
1p36	rs1175550	*CCDC27*	+					
1p36	rs1175550	*LRRC48*	+					
1p34	rs3916164	*HEYL*	+					
1p32	rs741959	*TAL1*	+			+	+	
1q23	rs857684	*OR10Z1*	+		+			
1q23	rs857684	*OR6Y1*			+			
1q23	rs857684	*SPTA1*	+		+	+	+	
1q32	rs7529925	*MIR181A1*	+					
1q32	rs7551442	*ATP2B4*	+					+
1q32	rs9660992	*TMCC2*	+					
1q44	rs3811444	*TRIM58*	+		+			
2p21	rs4953318	*PRKCE*	+				+	
2p16	rs243070	*BCL11A*	+				+	
2q13	rs10207392	*ACOXL*	+					
3p24	rs9310736	*THRB*	+				+	
3q22	rs6776003	*RASA2*	+					
3q23	rs13061823	*XRN1*	+					
3q29	rs11717368	*TFRC*	+			+	+	
4q11	rs218238	*KIT*	+				+	+
4q27	rs13152701	*BBS7*	+					
4q27	rs13152701	*CCNA2*	+	+			+	+
6p23	rs6914805	*GMPR*	+	+				
6p21	rs1408272	*HFE*			+		+	
6p21	rs1408272	*SLC17A3*	+					
6p22	rs13219787	*HIST1H2AM*	+					
6p22	rs13219787	*HIST1H2BO*	+					
6p22	rs13219787	*HIST1H3J*	+					
6p22	rs2097775	*TRIM39-RPP21*	+					
6p21	rs9272219	*HLA-DQA1*	+	+	+			
6p21	rs9272219	*HLA-DQA2*		+				
6p21	rs9349204	*CCND3*	+			+	+	
6p12	rs9369427	*VEGFA*	+				+	
6q21	rs1008084	*CCDC162P*	+					
6q23	rs9389269	*HBS1L*	+					+
6q24	rs590856	*CITED2*	+				+	
6q26	rs736661	*QKI*	+				+	
7p13	rs12718598	*IKZF1*	+			+	+	
7q22	rs2075672	*ACTL6B*	+					+
7q22	rs2075672	*TFR2*	+				+	
7q36	rs10480300	*PRKAG2*	+					
8p11	rs4737009	*ANK1*	+			+	+	
8p11	rs6987853	*C8orf40*	+	+				
9p24	rs2236496	*RCL1*	+					
9q34	rs579459	*ABO*	+					
10q11	rs901683	*MARCH8*	+	+	+			
10q22	rs10159477	*HK1*	+				+	+
10q24	rs11190134	*NKX2-3*	+				+	

Region	Sentinel SNP	*Gene*	Proximity	eQTL	Coding	GRAIL	Mouse	Drosophila
11p15	rs11042125	*AKIP1*	+	+				
11p15	rs11042125	*C11orf16*	+	+				
11p15	rs11042125	*NRIP3*		+				
11p15	rs11042125	*ST5*	+					
11p15	rs7936461	*SBF2*	+					
11q13	rs2302264	*CORO1B*	+	+				+
11q13	rs2302264	*PTPRCAP*	+	+			+	
11q13	rs2302264	*RPS6KB2*	+	+	+			+
11q13	rs7125949	*ARHGEF17*		+	+			
11q13	rs7125949	*P2RY6*	+					
12p13	rs7312105	*CACNA1C*	+					
12p13	rs10849023	*CCND2*	+			+	+	
12q22	rs11104870	*KITLG*	+				+	
12q24	rs3184504	*ATXN2*	+					
12q24	rs3184504	*SH2B3*	+		+		+	
12q24	rs3829290	*ACADS*			+			
12q24	rs3829290	*MLEC*	+					
14q23	rs7155454	*FNTB*	+				+	
14q23	rs7155454	*MAX*	+					
14q24	rs11627546	*SMOC1*	+					
14q32	rs17616316	*EIF5*	+					+
15q21	rs1532085	*LIPC*	+					
15q22	rs2572207	*DENND4A*	+					
15q22	rs2572207	*PTPLAD1*		+				
15q24	rs8028632	*PPCDC*	+					
15q24	rs8028632	*SCAMP5*	+					
15q24	rs11072566	*NRG4*	+					
15q25	rs2867932	*DNAJA4*		+				+
15q25	rs2867932	*WDR61*	+					+
16p11	rs11248850	*NPRL3*	+					
16q22	rs2271294	*CTRL*			+			
16q22	rs2271294	*EDC4*	+					+
16q22	rs2271294	*NUTF2*	+					
16q22	rs2271294	*PSMB10*			+		+	+
16q24	rs10445033	*PIEZO1*	+					
17p11	rs888424	*SPECC1*	+					
17q11	rs2070265	*C17orf63*	+					
17q11	rs2070265	*ERAL1*		+				
17q11	rs2070265	*NEK8*	+					
17q11	rs2070265	*TRAF4*	+	+				
17q12	rs8182252	*CDK12*		+				
17q12	rs8182252	*NEUROD2*	+					
17q21	rs2269906	*SLC4A1*				+	+	
17q21	rs2269906	*UBTF*	+					
17q12	rs12150672	*ARHGAP27*		+				
17q12	rs12150672	*ARL17B*		+				
17q12	rs12150672	*C17orf69*		+	+			
17q12	rs12150672	*CRHR1*	+		+			+
17q12	rs12150672	*SPPL2C*			+			
17q12	rs12150672	*KANSL1*			+			
17q12	rs12150672	*MAPT*			+		+	
17q12	rs12150672	*STH*			+			
17q25	rs4969184	*PGS1*	+	+				
18q21	rs4890633	*C18orf25*	+	+				

Region	Sentinel SNP	*Gene*	Proximity	eQTL	Coding	GRAIL	Mouse	Drosophila
19p13	rs2159213	*AP3D1*	+				+	+
19p13	rs732716	*MPND*	+					
19p13	rs732716	*SH3GL1*	+					
19p13	rs732716	*UBXD1*			+			
19p13	rs741702	*CALR*		+				
19p13	rs741702	*FARSA*	+	+				
19p13	rs741702	*SYCE2*	+					
19q13	rs3892630	*NUDT19*	+		+			
20q13	rs737092	*RBM38*	+					
21q22	rs2032314	*ATP5O*	+					+
22q11	rs5754217	*UBE2L3*	+	+				+
22q11	rs5754217	*YDJC*			+			
22q12	rs5749446	*FBXO7*	+		+	+		
22q12	rs855791	*KCTD17*	+					
22q12	rs855791	*TMPRSS6*	+		+		+	
22q13	rs140522	*ECGF1*	+	+				
22q13	rs140522	*NCAPH2*	+				+	
22q13	rs140522	*ODF3B*	+					+
22q13	rs140522	*SCO2*	+					+

Supplementary Table 20. Blood cell phenotypic abnormalities identified in *D. melanogaster* with RNAi silencing of orthologs for the red blood cell candidate genes. Results are provided for the ortholog with strongest phenotype: reduced cell numbers (-) or raised cell numbers (+). Drosophila orthologs are listed as: Flybase ID (gene name)

Gene	Region	*D.melanogaster* ortholog	Phenotype for RNAi knockdown of D. melanogaster ortholog			Mouse phenotype
			Crystal Cells early larvae	Crystal Cell late larvae	Plasmatocytes	
ACTL6B	7q22	CG6546 (*Bap55*)	++	+++	--	
AP3D1	19p13	CG10986 (*garnet*)	-	--	0	+
ATP2B4	1q32	CG42314 (*PMCA*)	-	++	++	
ATP5O	21q22	CG4307 (*Oscp*)	+	+++	+	
CCNA2	4q27	CG5940 (*CycA*)	++	+	0	+
CORO1B	11q13	CG9446 (*coro*)	--	-	-	
CRHR1	17q12	CG13758 (*Pdfr*)	+	+++	++	
DNAJA4	15q25	CG8863 (*Droj2*)	+	++	+++	
EDC4	16q22	CG6181 (*Ge-1*)	0	-	---	
EIF5	14q32	CG9177 (*eIF5*)	0	0	---	
HBS1L	6q23	CG1898 (*HBS1*)	++	+	0	
HK1	10q22	CG3001 (*Hex-A*)	++	+++	0	+
KIT	4q11	CG8222 (*Pvr*)	0	-	---	+
ODF3B	22q13	CG8086 (-)	-	+	---	
PSMB10	16q22	CG12161 (*Prosβ2R2*)	++	++	0	+
RPS6KB2	11q13	CG10539 (*S6K*)	0	+	--	
SCO2	22q13	CG8885 (*Scox*)	+	+++	0	
UBE2L3	22q11	CG7425 (*eff*)	--	0	++	
WDR61	15q25	CG3909 (-)	---	-	-	

Supplementary Table 21. Phenotypic variance explained by sentinel SNPs.

	GWA cohorts	Non-GWA cohorts	Combined
Sample size	11,898	13,264	25,156
Model 1 - Phenotype specific SNPs			
HB	0.027	0.020	0.023
MCH	0.077	0.067	0.071
MCHC	0.027	0.024	0.025
MCV	0.069	0.072	0.071
PCV	0.026	0.017	0.021
RBC	0.046	0.038	0.042
Model 2 - All Sentinel SNPs			
HB	0.044	0.042	0.043
MCH	0.096	0.084	0.090
MCHC	0.046	0.056	0.051
MCV	0.084	0.089	0.087
PCV	0.045	0.039	0.042
RBC	0.070	0.068	0.069

Supplementary Table 22. Effect sizes at the 75 loci associated with red blood cell phenotypes in i. 460 β-thalassaemia heterozygotes, and ii. 3786 controls with normal genotype from the SardiNIA study. P_{hetero} is for comparison of effect size between heterozygotes and controls. Genetic loci showing heterogeneity of effect are highlighted (green: P<0.05, yellow: P<7x10^{-4}).

						β-thalassaemia heterozygotes		Normal Controls		
Region	**SNP**	**r^{2*}**	**Position**	**Pheno**	**Gene**	**Effect**	**P**	**Effect**	**P**	**P_{hetero}**
1p36	rs1175549	0.82	3681587	MCHC	*CCDC27,LRRC48*	0.00 (0.07)	1.00	-0.08 (0.03)	0.02	0.33
1p34	rs3916164		39842526	MCH	*HEYL*	-0.09 (0.11)	0.42	0.01 (0.07)	0.93	0.47
1p32	rs741959		47448820	MCV	*TAL1*	0.24 (0.32)	0.45	0.27 (0.17)	0.10	0.92
1p23	rs857684		156842353	MCHC	*OR6Y1, OR10Z1, SPTA1*	-0.05 (0.08)	0.47	0.10 (0.03)	9.6E-04	0.05
1q32	rs7529925		197273831	RBC	*MIR181A1*	-0.01 (0.05)	0.88	-0.02 (0.02)	0.20	0.83
1q32	rs7551442		201921744	MCHC	*ATP2B4*	0.06 (0.09)	0.53	-0.01 (0.04)	0.87	0.53
1q32	rs9660992		203516073	MCH	*TMCC2*	-0.21 (0.11)	0.04	-0.10 (0.06)	0.11	0.34
1q44	rs12404125	0.252	246111082	RBC	*TRIM58*	-0.01 (0.05)	0.81	0.00 (0.01)	0.85	0.79
2p21	rs4953318		46208555	PCV	*PRKCE*	-0.17 (0.30)	0.57	0.14 (0.12)	0.22	0.33
2p16	rs243070		60473790	MCV	*BCL11A*	-0.69 (0.37)	0.06	-0.29 (0.20)	0.14	0.34
2q13	rs10207392		111566130	MCV	*ACOXL*	0.05 (0.31)	0.88	0.50 (0.17)	2.8E-03	0.19
3p24	rs9310736		24325815	MCV	*THRB*	0.03 (0.32)	0.93	0.18 (0.16)	0.26	0.67
3q22	rs6776003		142749183	MCV	*RASA2*	0.04 (0.33)	0.91	0.18 (0.16)	0.27	0.69
3q23	rs13061823		143603476	MCV	*XRN1*	-0.51 (0.31)	0.11	-0.26 (0.16)	0.12	0.48
3q29	rs11717368		197318754	MCH	*TFRC*	-0.02 (0.10)	0.82	0.00 (0.06)	0.96	0.87
4q11	rs218238		55089781	RBC	*KIT*	-0.08 (0.06)	0.18	0.02 (0.02)	0.15	0.09
4q27	rs13152701		122970511	MCV	*BBS7*, *CCNA2*	-0.36 (0.32)	0.27	0.03 (0.17)	0.89	0.30
6p23	rs6914805		16389166	MCH	*GMPR*	0.11 (0.13)	0.42	0.11 (0.08)	0.18	0.98
6q21	rs1800562	0.793	26201120	MCH	*HFE*, *SLC17A3*	-0.10 (0.27)	0.72	0.04 (0.16)	0.79	0.66
6p22	rs13214703	1	28049366	MCH	*HIST1H2AM, HIST1H2BO, HIST1H3J*	0.45 (0.24)	0.05	0.34 (0.14)	0.02	0.66
6p22	rs2097775		30462282	HB	*TRIM39-RPP21*	0.15 (0.23)	0.52	-0.07 (0.08)	0.36	0.36
6p21	rs9272219		32710247	RBC	*HLA-DQA1*, *HLA-DQA2*	0.01 (0.05)	0.81	0.03 (0.01)	0.02	0.70
6p21	rs17318575	0.379	25601217	MCV	*CCND3*	1.36 (0.87)	0.12	0.12 (0.41)	0.77	0.20

6p12	rs9369427		43919408	HB	*VEGFA*	0.04 (0.10)	0.69	0.09 (0.04)	0.03	0.65
6q21	rs1008084		109733658	MCH	*CCDC162P*	-0.15 (0.10)	0.14	-0.08 (0.06)	0.18	0.55
6q23	rs9389269		135468852	MCV	*HBS1L*	-0.47 (0.35)	0.18	-0.28 (0.19)	0.14	0.64
6q24	rs590856		139886122	MCV	*CITED2*	-0.02 (0.31)	0.94	-0.63 (0.17)	1.4E-04	0.09
6q26	rs736661		164402826	MCH	*QKI*	0.05 (0.10)	0.60	0.01 (0.06)	0.88	0.71
7p13	rs12669559	0.61	50403271	MCV	*IKZF1*	0.29 (0.41)	0.47	0.38 (0.22)	0.09	0.85
7q22	rs2075672		100078232	RBC	*ACTL6B*, *TFR2*	-0.08 (0.05)	0.07	-0.02 (0.01)	0.10	0.20
7q36	rs10480300		151036938	HB	*PRKAG2*	0.27 (0.12)	0.03	0.10 (0.05)	0.02	0.20
8p11	rs4737009		41749562	MCHC	*ANK1*	-0.08 (0.07)	0.30	0.05 (0.04)	0.12	0.11
8p11	rs6987853		42576607	MCHC	*C8orf40*	-0.02 (0.07)	0.74	0.04 (0.03)	0.17	0.39
9p24	rs2236496		4834265	MCV	*RCL1*	0.05 (0.45)	0.90	0.13 (0.21)	0.56	0.88
9q34	rs579459		135143989	RBC	*ABO*	0.16 (0.05)	2.1E-03	0.03 (0.02)	0.06	0.02
10q11	rs901683		45286428	MCV	*MARCH8*	0.54 (0.54)	0.32	-0.96 (0.28)	5.9E-04	0.01
10q22	rs10159477		70769894	HB	*HK1*	-0.03 (0.15)	0.84	-0.04 (0.05)	0.48	0.95
10q22	rs11190134		101272190	MCH	*NKX2-3*	-0.09 (0.11)	0.40	0.05 (0.06)	0.45	0.26
11p15	rs11042125		8894625	HB	*AKIP1*, *C11orf16*, *NRIP3*, *ST5*	0.05 (0.12)	0.66	0.11 (0.05)	0.02	0.64
11p15	rs7936461		9997462	PCV	*SBF2*	0.84 (0.39)	0.03	0.17 (0.15)	0.24	0.11
11q13	rs2302264		66964002	MCV	*CORO1B*, *PTPRCAP*, *RPS6KB2*	-0.28 (0.32)	0.39	0.41 (0.17)	0.02	0.06
11q13	rs7125949		72686732	HB	*ARHGEF17*, *P2RY6*	-0.22 (0.17)	0.22	-0.26 (0.08)	4.5E-04	0.80
12p13	rs7312107	1	2393631	PCV	*CACNA1C*	-0.03 (0.37)	0.93	-0.08 (0.13)	0.56	0.91
12p13	rs10849023		4202739	MCH	*CCND2*	0.06 (0.11)	0.56	-0.11 (0.06)	0.09	0.17
12q22	rs11104870		87353425	RBC	*KITLG*	0.02 (0.05)	0.66	-0.01 (0.02)	0.52	0.53
12q24	rs3184504		110368991	HB	*ATXN2*, *SH2B3*	-0.06 (0.09)	0.55	-0.09 (0.04)	0.03	0.74
12q24	rs3829290		119610821	MCV	*ACADS, MLEC*	-0.07 (0.30)	0.82	0.15 (0.16)	0.37	0.53
14q24	rs11627546		69435677	MCV	*SMOC1*	0.50 (0.36)	0.17	0.25 (0.19)	0.19	0.55
14q32	rs7155454		64571992	MCH	*FNTB*, *MAX*	-0.09 (0.10)	0.37	-0.04 (0.06)	0.51	0.68
14q32	rs17616316		102892515	MCH	*EIF5*	-1.37 (0.29)	3.0E-06	-0.12 (0.21)	0.56	5.3E-04
15q21	rs1532085		56470658	HB	*LIPC*	0.02 (0.10)	0.82	0.03 (0.04)	0.37	0.91
15q22	rs2572207		63857747	MCV	*DENND4A*, *PTPLAD1*	0.30 (0.33)	0.36	0.18 (0.17)	0.30	0.74
15q24	rs8028632		73108315	MCV	*PPCDC*, *SCAMP5*	0.44 (0.32)	0.17	0.12 (0.18)	0.50	0.38
15q24	rs2867932		76378092	MCHC	*DNAJA4*, *WDR61*	-0.04 (0.07)	0.62	-0.06 (0.03)	0.05	0.72

15q25	rs11072566		74081026	HB	*NRG4*	-0.05 (0.09)	0.59	0.04 (0.04)	0.33	0.38
16p11	rs11248850		103598	MCH	*NPRL3*	0.14 (0.11)	0.20	-0.19 (0.06)	2.5E-03	0.01
16q22	rs2271294		66459827	RBC	*CTRL*, *EDC4*, *NUTF2, PSMB10*	-0.11 (0.05)	0.05	0.00 (0.02)	0.87	0.05
16q24	rs9933309	0.711	87372433	MCHC	*PIEZO1*	-0.13 (0.08)	0.11	-0.08 (0.04)	0.03	0.57
17p11	rs888424		19926019	MCH	*SPECC1*	-0.02 (0.10)	0.86	-0.07 (0.06)	0.26	0.67
17q11	rs2070265		24099550	MCH	*C17orf63*, *ERAL1*, *NEK8*, *NEK8, TRAF4*	-0.33 (0.18)	0.06	-0.19 (0.10)	0.05	0.46
17q12	rs8182252		34981476	RBC	*CDK12*, *NEUROD2*	-0.08 (0.05)	0.07	-0.01 (0.01)	0.33	0.15
17q21	rs2269906		39649863	MCHC	*SLC4A1*, *UBTF*	-0.11 (0.08)	0.14	-0.04 (0.03)	0.24	0.38
17q21	rs12150672		41182408	RBC	*ARHGAP27*, *ARL17B*, *C17orf69*, *CRHR1*, *SPPL2C, KANSL1*, *MAPT*, *STH*	-0.03 (0.05)	0.53	0.00 (0.01)	0.87	0.58
17q25	rs4969184		73905008	HB	*PGS1*	-0.01 (0.09)	0.89	0.15 (0.04)	1.6E-04	0.11
18q21	rs4890633		42087276	MCH	*C18orf25*	-0.12 (0.13)	0.36	-0.20 (0.08)	0.01	0.60
19p13	rs2159213		2087102	HB	*AP3D1*	0.13 (0.09)	0.14	-0.03 (0.04)	0.43	0.09
19p13	rs732716		4317219	MCV	*MPND, SH3GL1, UBXN6*	0.42 (0.31)	0.17	0.35 (0.17)	0.03	0.84
19p13	rs741702		12885250	MCH	*CALR, FARSA, SYCE2*	-0.01 (0.11)	0.95	-0.13 (0.07)	0.08	0.37
19q13	rs3892630		37873324	MCV	*NUDT19*	-0.12 (0.39)	0.75	-0.50 (0.22)	0.03	0.41
20q13	rs99595	0.561	55423214	MCV	*RBM38*	0.69 (0.29)	0.02	0.30 (0.16)	0.06	0.25
21q22	rs2032314		34276393	PCV	*ATP5O*	-0.18 (0.42)	0.68	-0.40 (0.16)	0.01	0.62
22q11	rs5754217		20269675	MCV	*UBE2L3, YDJC*	0.04 (0.43)	0.93	-0.06 (0.23)	0.79	0.84
22q12	rs5749446		31210585	MCH	*FBXO7*	0.08 (0.11)	0.45	0.17 (0.07)	0.01	0.52
22q12	rs2413450[§]	0.867[*]	35800170	MCH	*KCTD17, TMPRSS6*	0.02 (0.13)	0.87	-0.18 (0.09)	0.04	0.20
22q13	rs470119[§]	0.669[*]	49313780	MCV	T*YMP*, *NCAPH2, ODF3B, SCO2*	0.39 (0.42)	0.35	0.31 (0.23)	0.17	0.14

Supplementary Table 23. Effect sizes in South Asians (SA) at the novel loci associated with red blood cell phenotypes amongst Europeans (EW) in the current GWA study. P_{hetero} is the P values for comparison of effect size between Europeans and South Asians. Dir: is direction of effect between Europeans and South Asians (+ is concordant, - is discordant). Blank cells represent genotypes not available in South Asians.

			Alleles		EAF		Europeans			South Asians				
Region	**SNP**	**Pheno**	**Effect**	**Alt**	**EW**	**SA**	**N**	**Effect**	**P**	**N**	**Effect**	**P**	**P_{hetero}**	**Dir**
1p36	rs1175550	MCHC	G	A	0.22		50425	0.008 (0.013)	8.6E-15					
1p32	rs741959	MCV	G	A	0.57		58002	0.157 (0.025)	6.0E-10					
1q23	rs857684	MCHC	C	T	0.74	0.81	56373	-0.006 (0.011)	3.5E-16	7953	0.019 (0.021)	4.1E-01	3.1E-01	-
1q32	rs7529925	RBC	C	T	0.28	0.28	53258	0.014 (0.002)	8.3E-09	7912	-0.003 (0.008)	8.2E-01	5.0E-02	-
1q32	rs7551442	MCHC	A	G	0.09		50411	-0.023 (0.017)	9.7E-12					
1q32	rs9660992	MCH	G	A	0.42	0.17	51249	0.007 (0.004)	7.1E-10	8126	0.012 (0.043)	7.8E-01	8.9E-01	+
1q44	rs3811444	RBC	T	C	0.35	0.34	34323	0.018 (0.003)	4.5E-10	7365	-0.001 (0.008)	9.8E-01	3.1E-02	-
2p21	rs4953318	PCV	A	C	0.62	0.63	53032	0.152 (0.018)	3.1E-19	7941	0.287 (0.058)	6.7E-07	2.4E-02	+
2p16	rs243070	MCV	T	A	0.72		57740	-0.181 (0.027)	4.4E-13					
2q13	rs10207392	MCV	G	A	0.44	0.48	57750	-0.132 (0.025)	2.0E-08	8485	-0.182 (0.082)	2.6E-02	5.6E-01	+
3p24	rs9310736	MCV	A	G	0.35	0.33	57810	-0.210 (0.026)	6.1E-16	8485	-0.248 (0.089)	5.2E-03	6.8E-01	+
3q22	rs6776003	MCV	A	G	0.44		54586	-0.138 (0.026)	7.1E-08					
3q23	rs13061823	MCV	T	C	0.56	0.39	57678	-0.168 (0.025)	4.7E-13	9081	-0.081 (0.080)	3.2E-01	3.0E-01	+
3q29	rs11717368	MCH	C	G	0.52	0.50	51664	0.008 (0.004)	6.6E-19	7533	0.036 (0.034)	2.9E-01	4.2E-01	+
4q11	rs218238	RBC	A	T	0.78	0.66	53374	0.033 (0.003)	2.8E-39	7912	0.042 (0.008)	1.1E-07	2.9E-01	+
4q27	rs13152701	MCV	A	G	0.37	0.33	53708	0.150 (0.026)	9.0E-10	9081	0.066 (0.082)	4.2E-01	3.3E-01	+
6p23	rs6914805	MCH	C	T	0.75	0.59	47195	0.012 (0.004)	1.2E-19	8126	0.053 (0.033)	1.2E-01	2.2E-01	+
6p21	rs1408272	MCH	G	T	0.07		36605	0.033 (0.009)	4.8E-67					
6p22	rs13219787	MCH	A	G	0.09		42060	0.023 (0.007)	5.9E-17					
6p22	rs2097775	HB	A	T	0.15	0.07	61058	0.055 (0.008)	1.3E-10	9213	-0.047 (0.037)	2.6E-01	7.2E-03	-
6p21	rs9272219	RBC	G	T	0.72	0.70	49302	0.015 (0.002)	4.3E-10	7365	-0.005 (0.008)	6.0E-01	2.8E-02	-
6p21	rs9349204	MCV	G	A	0.27	0.19	53153	-0.367 (0.028)	2.4E-40	9081	-0.291 (0.103)	4.3E-03	4.7E-01	+
6p12	rs9369427	HB	A	C	0.68	0.78	60855	0.042 (0.006)	5.6E-12	8605	0.043 (0.024)	6.0E-02	9.7E-01	+
6q21	rs1008084	MCH	G	A	0.56	0.73	51455	-0.010 (0.003)	6.4E-26	8126	-0.064 (0.037)	1.1E-01	1.5E-01	+
6q23	rs9389269	MCV	T	C	0.72	0.87	57855	-0.600 (0.028)	2.6E-109	9081	-0.662 (0.117)	2.8E-08	6.1E-01	+
6q24	rs590856	MCV	G	A	0.43	0.34	58041	0.313 (0.026)	5.0E-36	9081	0.399 (0.088)	7.9E-06	3.5E-01	+

6q26	rs736661	MCH	A	G	0.62	0.59	51397	0.007 (0.004)	1.6E-11	8126	0.065 (0.033)	6.6E-02	8.1E-02	+
7p13	rs12718598	MCV	T	C	0.51	0.51	37967	-0.204 (0.030)	1.6E-13	8485	-0.237 (0.082)	4.4E-03	7.1E-01	+
7q22	rs2075672	RBC	A	G	0.39		41805	0.022 (0.003)	1.9E-20					
7q36	rs10480300	HB	C	T	0.72		49771	0.052 (0.007)	7.8E-15					
8p11	rs4737009	MCHC	G	A	0.74	0.79	54462	-0.014 (0.013)	4.9E-11	7409	0.042 (0.026)	1.2E-01	5.6E-02	-
8p11	rs6987853	MCHC	C	T	0.62	0.71	52954	-0.002 (0.010)	6.1E-11	5376	0.043 (0.024)	7.8E-02	8.5E-02	-
9p24	rs2236496	MCV	C	T	0.22	0.19	53761	-0.279 (0.031)	1.4E-19	9081	-0.225 (0.100)	2.6E-02	6.0E-01	+
9q34	rs579459	RBC	T	C	0.80	0.83	53362	0.021 (0.003)	9.3E-18	7365	0.036 (0.010)	6.2E-04	1.5E-01	+
10q11	rs901683	MCV	A	G	0.08	0.03	58051	0.364 (0.050)	1.5E-16	9081	0.408 (0.256)	1.1E-01	8.7E-01	+
10q22	rs10159477	HB	A	G	0.16		45553	0.087 (0.010)	4.4E-20					
10q24	rs11190134	MCH	G	A	0.60	0.63	51412	-0.011 (0.004)	7.9E-08	8126	-0.003 (0.034)	9.3E-01	8.2E-01	+
11p15	rs11042125	HB	A	T	0.60	0.73	60973	0.032 (0.006)	1.5E-09	9213	0.020 (0.021)	3.6E-01	5.6E-01	+
11p15	rs7936461	PCV	C	T	0.75		49357	0.121 (0.021)	1.0E-09					
11q13	rs2302264	MCV	G	A	0.58	0.71	57841	0.140 (0.025)	1.3E-10	9081	0.010 (0.087)	8.8E-01	1.5E-01	+
11q13	rs7125949	HB	A	G	0.11		49153	0.053 (0.010)	2.1E-09					
12p13	rs7312105	PCV	G	A	0.36	0.38	48278	0.104 (0.019)	3.2E-08	7941	0.107 (0.057)	6.3E-02	9.6E-01	+
12p13	rs10849023	MCH	C	T	0.79	0.86	42647	-0.008 (0.005)	7.5E-12	7533	-0.043 (0.049)	3.8E-01	4.7E-01	+
12q22	rs11104870	RBC	C	T	0.30	0.29	53326	0.013 (0.002)	1.7E-08	7912	0.002 (0.008)	8.2E-01	2.2E-01	+
12q24	rs3184504	HB	T	C	0.48		56784	0.051 (0.006)	4.3E-19					
12q24	rs3829290	MCV	C	T	0.44	0.65	51911	-0.153 (0.026)	2.1E-09	8485	-0.247 (0.089)	5.5E-03	3.1E-01	+
14q23	rs7155454	MCH	A	G	0.51	0.47	51228	0.002 (0.004)	1.8E-12	8126	-0.001 (0.033)	9.7E-01	9.4E-01	-
14q24	rs11627546	MCV	C	A	0.84	0.79	57833	0.162 (0.032)	3.6E-08	9081	0.022 (0.098)	8.2E-01	1.8E-01	+
15q22	rs2572207	MCV	C	T	0.74	0.60	57810	0.153 (0.029)	3.4E-09	9081	0.080 (0.081)	3.1E-01	4.0E-01	+
15q24	rs8028632	MCV	T	C	0.80	0.45	53602	0.188 (0.032)	6.9E-10	9081	0.063 (0.079)	4.4E-01	1.4E-01	+
15q24	rs11072566	HB	A	G	0.48	0.33	60792	0.028 (0.006)	8.3E-08	9213	0.000 (0.019)	9.9E-01	1.5E-01	+
15q25	rs2867932	MCHC	G	A	0.61	0.54	56211	-0.021 (0.010)	3.3E-09	7953	-0.008 (0.017)	5.9E-01	5.0E-01	+
16p11	rs11248850	MCH	G	A	0.50	0.60	51345	0.007 (0.004)	6.3E-23	8126	0.023 (0.033)	5.0E-01	6.4E-01	+
16q22	rs2271294	RBC	T	A	0.15	0.23	53599	0.017 (0.003)	1.1E-09	7912	0.005 (0.008)	5.5E-01	2.0E-01	+
16q24	rs10445033	MCHC	G	A	0.37	0.36	42050	0.020 (0.012)	1.5E-22	7409	0.053 (0.019)	5.1E-03	1.4E-01	+
17p11	rs888424	MCH	A	G	0.43	0.31	51274	0.006 (0.004)	5.4E-20	8126	0.108 (0.034)	1.4E-03	3.0E-03	+
17q11	rs2070265	MCH	T	C	0.20	0.30	51503	0.013 (0.004)	5.1E-14	8126	0.006 (0.034)	8.2E-01	8.5E-01	+
17q12	rs8182252	RBC	C	T	0.18	0.10	49812	0.016 (0.003)	5.9E-09	5340	0.009 (0.015)	5.5E-01	6.6E-01	+

17q21	rs2269906	MCHC	C	A	0.36	0.30	56263	0.027 (0.010)	2.0E-11	7409	0.027 (0.020)	1.8E-01	9.9E-01	+
17q21	rs12150672	RBC	A	G	0.23		53489	0.017 (0.003)	4.7E-12					
17q25	rs4969184	HB	G	A	0.53	0.58	60892	0.031 (0.006)	7.0E-09	9213	0.029 (0.018)	1.3E-01	9.4E-01	+
18q21	rs4890633	MCH	G	A	0.27	0.33	51375	0.005 (0.004)	1.9E-23	8126	0.133 (0.034)	7.6E-05	1.7E-04	+
19p13	rs2159213	HB	C	T	0.50	0.41	60826	0.032 (0.006)	1.9E-09	9213	0.049 (0.019)	9.3E-03	3.8E-01	+
19p13	rs732716	MCV	A	G	0.71	0.73	58044	0.201 (0.028)	1.5E-14	8485	0.200 (0.106)	6.5E-02	9.9E-01	+
19p13	rs741702	MCH	A	C	0.35	0.29	45178	0.006 (0.004)	8.2E-20	8126	0.077 (0.035)	2.4E-02	4.1E-02	+
19q13	rs3892630	MCV	T	C	0.18	0.14	57699	0.176 (0.034)	8.8E-08	9081	-0.007 (0.122)	9.6E-01	1.5E-01	-
20q13	rs737092	MCV	C	T	0.49	0.34	35156	0.216 (0.033)	4.0E-13	8485	0.265 (0.100)	8.5E-03	6.4E-01	+
22q11	rs5754217	MCV	G	T	0.83	0.63	53759	0.194 (0.031)	8.6E-10	9081	0.223 (0.081)	7.8E-03	7.4E-01	+
22q12	rs5749446	MCH	T	C	0.62	0.55	51609	0.007 (0.004)	3.3E-13	8126	0.046 (0.032)	1.5E-01	2.2E-01	+
22q12	rs855791	MCH	G	A	0.57	0.47	38547	0.012 (0.004)	1.0E-69	8126	0.230 (0.032)	5.4E-13	1.3E-11	+
22q13	rs140522	MCV	C	T	0.67	0.70	44680	0.287 (0.030)	4.5E-23	9081	0.179 (0.086)	4.0E-02	2.4E-01	+

Supplementary Table 24. Pearson correlation coefficients between phenotypic traits (amongst LOLIPOP EW sample, shaded orange) and SNP associations (-log10[P] in the European analysis, shaded lilac)

	Hb	MCH	MCHC	MCV	PCV	RBC
Hb		0.23	0.08	0.22	0.96	0.75
MCH	0.14		0.46	0.91	0.09	0.47
MCHC	0.14	0.32		0.07	0.21	0.25
MCV	0.16	0.79	0.22		0.20	0.42
PCV	0.72	0.25	0.19	0.23		0.80
RBC	0.37	0.54	0.20	0.51	0.55	

Supplementary Table 25. Results of permutation testing to determine the effective number phenotypes studied. The genome-wide association study for association of SNPs with red blood cell traits was run repeatedly with the relationship of genotype and phenotype data randomised to simulate expectations under the null hypothesis. On each run, the minimum P value (P_{min}) for association with any phenotype was determined for each SNP, and the number of SNPs with P_{min} reaching suggestive statistical significance determined ($P<10^{-6}$, $P<10^{-7}$, $P<5x10^{-8}$). First the GWAS was run 100 times, with the phenotype data structure intact, to assess the number of associations expected under the null hypothesis for 6 related phenotypes ("Related"). The GWAS was then run a further 100 times, with the phenotype data for the six red blood cell traits now also randomised, to assess the number of associations expected under the null hypothesis for 6 unrelated phenotypes ("Unrelated"). Effective number of phenotypes was calculated as the ratio of SNPs reaching statistical significance under the null hypothesis for related vs unrelated phenotypes multiplied by the actual number of phenotypes studied (6). Permutation testing was carried out amongst both Europeans and South Asians separately.

	Number of SNPs			
P value threshold	**Related phenotypes***	**Unrelated phenotypes †**	**Related / Unrelated**	**Effective number of phenotypes**
Europeans				
P<1.0E-06	1705	1983	0.86	5.16
P<1.0E-07	235	255	0.92	5.53
P<5.0E-08	128	152	0.84	5.05
South Asians				
P<1.0E-06	1217	1347	0.90	5.42
P<1.0E-07	104	129	0.81	4.84
P<5.0E-08	62	75	0.83	4.96

* Related: GWAS for the six red cell phenotypes with correlation matrix intact
† Unrelated: GWAS for the six red cell phenotypes rendered unrelated by randomisation

Supplementary Table 26. Results of replication testing. SNPs that do not show replication (P>0.05 in the replication sample, or P>1x10^{-8} in combined analysis) are shaded grey.

				GWA		Replication		Combined	
Region	**SNP**	**Position**	**Pheno**	**N**	**P**	**N**	**P**	**N**	**P**
Sentinel SNPs at loci identified in GWA of Europeans									
1p36	rs6656196	26631360	MCV	57813	2.9E-08	34798	2.0E-01	92611	2.5E-07
2q13	rs10207392	111566130	MCV	57750	2.0E-08	46743	3.1E-04	104493	4.4E-11
3p25	rs17040409	14879731	HB	60821	4.3E-08	34830	5.8E-01	95651	5.5E-05
3p24	rs6770091	23529811	HB	60884	4.7E-08	63147	1.25E-01	124031	1.9E-06
3q22	rs6776003	142749183	MCV	54586	7.1E-08	46695	9.0E-05	101281	3.7E-11
3q25	rs919520	158019813	RBC	52971	3.2E-08	53201	3.85E-02	106172	6.6E-03
4q21	rs236996	88224227	HB	56729	1.5E-08	55004	1.7E-02	111733	1.1E-08
6p21	rs9380238	31375597	MCH	51449	2.8E-08	34798	3.4E-01	86247	9.9E-07
10q24	rs11190134	101272190	MCH	51412	7.9E-08	34798	3.4E-04	86210	1.3E-10
12p13	rs7312105	2393616	PCV	48278	3.2E-08	45711	5.0E-03	93989	3.2E-09
12q22	rs11104870	87353425	RBC	53326	1.7E-08	33079	6.5E-04	86405	6.2E-11
14q23	rs1256061	63773346	PCV	43992	9.0E-08	54922	3.4E-02	98914	2.7E-07
14q24	rs11627546	69435677	MCV	57833	3.6E-08	46758	2.9E-03	104591	1.1E-09
15q24	rs11072566	74081026	HB	60792	8.3E-08	54910	4.6E-04	115702	3.0E-10
16p11	rs13708	30908310	RBC	53602	2.2E-08	33079	4.3E-01	86681	1.0E-06
19q13	rs3892630	37873324	MCV	57699	8.8E-08	46780	2.0E-04	104479	1.0E-10
22q12	rs695267	27229748	RBC	48988	8.4E-08	33079	6.9E-01	82067	1.1E-05
Sentinel SNPs at loci identified in GWA of Europeans and South Asians									
1p34	rs3916164	39842526	MCH	57076	7.4E-08	34798	8.5E-04	91874	3.1E-10
7q33	rs12530845	134980518	MCH	57204	6.6E-08	37687	3.0E-02	94891	2.7E-08
14q32	rs17616316	102892515	MCH	55633	1.6E-08	34798	8.7E-04	90431	8.2E-11
15q21	rs1532085	56470658	HB	46598	2.9E-09	34830	1.8E-03	81428	6.7E-11
21q22	rs2032314	34276393	PCV	56362	5.4E-08	54944	1.1E-03	111306	7.5E-10
SNPs tested for association with secondary phenotypes									
4q21	rs6840258	88192692	MCH	51387	6.5E-08	46776	1.9E-03	98163	1.4E-09
6p23	rs6914805	16389166	MCHC	50304	5.8E-08	34800	7.4E-08	85104	2.7E-14
6p22	rs13219787	27969649	HB	51167	1.1E-08	34800	1.3E-06	85967	6.9E-14
6p21	rs2853925	31372901	RBC	49019	3.0E-08	33079	2.4E-05	82098	3.4E-12
6p21	rs9272219	32710247	HB	55108	4.1E-08	34800	2.5E-02	89908	1.2E-08
6q21	rs1341271	109724235	MCHC	56228	8.9E-08	34829	5.0E-06	91057	2.1E-12
7p13	rs12718598	50395939	MCH	34731	7.2E-08	34800	8.3E-08	69531	3.0E-14
7q22	rs1734910	100147480	PCV	36374	5.4E-08	34843	4.8E-11	71217	2.1E-17
10q22	rs10159477	70769894	RBC	40764	2.8E-08	0	**	40764	2.8E-08
12p13	rs2239063	2382092	HB	61102	5.8E-08	34830	4.2E-01	95932	1.4E-06
12p13	rs11611647	4204180	RBC	44531	1.2E-08	33079	1.9E-04	77610	1.4E-11
12q24	rs7976497	119619850	RBC	49164	8.1E-08	33079	8.8E-02	82243	1.7E-07
14q23	rs7155454	64571992	MCV	57632	1.9E-08	34800	3.8E-07	92432	4.2E-14
15q22	rs2572207	63857747	MCH	51408	1.2E-08	34798	1.0E-09	86206	1.1E-16
16p11	rs743725	76888	RBC	47164	1.8E-08	33079	2.7E-01	80243	5.3E-07

16p11	rs11248914	233563	MCHC	52056	3.1E-08	34800	3.3E-12	86856	3.5E-18
17q11	rs7221773	24227014	MCHC	56113	3.8E-08	34829	2.5E-05	90942	4.2E-12
19p13	rs12982593	2126891	RBC	48723	2.5E-08	33079	1.5E-04	81802	2.0E-11
21q22	rs11910015	34260508	HB	61233	6.3E-08	34830	2.6E-05	96063	7.2E-12
SNPs tested as secondary signals with primary phenotype									
2p21	rs10184620	46212039	PCV	37432	9.9E-08	34843	3.7E-11	72275	3.5E-17
3q29	rs7625441	197338676	MCH	34728	4.2E-08	46305	4.7E-06	81033	1.8E-12
4q11	rs17084315	54981597	MCV	57857	2.9E-08	34798	9.7E-10	92655	4.3E-16
6p21	rs1034050	25600343	MCV	53730	5.5E-08	46788	5.4E-03	100518	4.3E-09
6p21	rs13203202	25690750	HB	48931	2.0E-08	34830	9.9E-06	83761	9.5E-13
6p21	rs10946795	25822718	MCH	51345	5.1E-08	46770	6.5E-05	98115	2.1E-11
6p21	rs501220	25981004	MCH	51453	3.2E-08	34798	8.7E-06	86251	1.3E-12
6p21	rs10484440	26561721	MCH	47036	1.8E-08	34798	2.4E-05	81834	2.2E-12
6p21	rs2395033	31566533	RBC	49163	5.4E-08	33079	1.6E-01	82242	3.5E-07
6p21	rs9368716	32414068	RBC	53502	3.4E-08	45000	4.9E-03	98502	2.4E-09
6q21	rs12214121	109429247	MCH	51003	8.8E-08	34798	3.7E-06	85801	1.5E-12
6q21	rs12528712	109767577	MCH	51633	7.4E-08	46772	3.8E-03	98405	3.8E-09
6q21	rs12181780	109816229	MCH	51456	5.5E-08	37671	5.7E-04	89127	1.9E-10
6q21	rs12206574	110174466	RBC	49292	1.1E-08	33079	8.6E-07	82371	4.7E-14
6q23	rs1074849	135465105	PCV	52677	6.2E-08	34843	6.2E-06	87520	1.8E-12
16q24	rs750739	87335109	MCHC	41399	2.1E-08	46869	7.0E-03	88268	6.5E-09
22q12	rs7291067	31203496	MCH	47384	5.0E-08	34798	6.1E-10	82182	3.2E-16
22q12	rs1421312	35817756	MCHC	46284	8.5E-08	34800	7.2E-17	81084	1.9E-21
22q12	rs733655	35824997	PCV	48127	3.7E-08	34800	3.8E-02	82927	3.1E-08

Supplementary Table 27. Results of the Drosophila RNAi screen. Gene: candidate gene from GWAS; Flybase Gene ID: unique ID in Flybase for the D. melanogaster ortholog; VDRC stock no: unique ID in the Vienna Drosophila RNAi Center (VDRC) for the D. melanogaster ortholog stock; Construct ID: unique ID in VDRC for the RNAi construct; Specificity score (S19): measure of RNAi specificity for ortholog gene sequence; Cell counts: score for cell count number (mean of two experiments, 15 larvae per experiment, scale -3 to +3); Highest Abs(score): absolute value of greatest departure from zero for any cell type. Fly lines with Highest Abs(score)≥2.0 are highlighted in grey.

Gene	Flybase Gene ID	VDRC stock no	Construct ID	Specificity score (S19)	Crystal cell early L3	Crystal cell late L3	Plasmatocytes	Highest Abs(score)
ACADS	CG4860	34899	11451	1.00	0.0	-1.0	0.0	1.0
ACADS	CG4860	34900	11451	1.00	0.5	0.0	0.0	0.5
ACADS	CG4703	110335	100761	1.00	0.0	0.0	0.0	0.0
ACTL6B	CG6546	24703	11955	1.00	2.0	2.5	-2.0	2.5
ACTL6B	CG6546	24704	11955	1.00	0.5	1.5	0.0	1.5
ANK1	CG32373	101586	103913	1.00	0.0	0.0	0.0	0.0
ANK1	CG32373	5115	2354	1.00	0.0	0.0	0.5	0.5
ANK1	CG42734	46225	16285	0.99	-1.0	0.0	0.0	1.0
ANK1	CG42734	46224	16285	0.99	0.5	0.0	0.0	0.5
ANK1	CG42734	40638	12247	0.99	0.0	0.0	0.0	0.0
ANK1	CG42734	107238	104937	1.00	0.0	0.0	0.0	0.0
ANK1	CG42734	26122	10869	1.00	0.0	0.0	0.0	0.0
ANK1	CG42734	26121	10869	1.00	0.0	0.0	0.0	0.0
ANK1	CG42734	104833	113612	1.00	0.0	0.0	0.0	0.0
ANK1	CG42734	107369	106729	1.00	0.0	0.0	0.0	0.0
ANK1	CG1651	25946	10431	1.00	0.0	0.0	0.0	0.0
ANK1	CG1651	25945	10431	1.00	0.0	0.0	0.0	0.0
AP3D1	CG10986	39766	7158	1.00	-0.5	-2.0	0.0	2.0
AP3D1	CG10986	41369	6405	1.00	0.0	0.0	-0.5	0.5
AP3D1	CG10986	31390	7158	1.00	0.0	0.5	0.0	0.5
AP3D1	CG10986	31391	7158	1.00	0.0	0.0	0.0	0.0
AP3D1	CG10986	41368	6405	1.00	0.0	0.0	0.0	0.0
ARHGEF17	CG43102	110150	102017	1.00	0.0	1.5	0.0	1.5
ARHGEF17	CG43102	49053	17115	1.00	1.0	0.0	0.0	1.0
ARHGEF17	CG43102	13094	5178	1.00	0.0	0.0	0.0	0.0
ARHGEF17	CG43102	13429	5178	1.00	0.0	0.0	0.0	0.0
ARL17B	CG2219	41691	9755	1.00	0.5	0.0	0.5	0.5
ARL17B	CG2219	107995	100576	1.00	0.0	0.0	0.0	0.0
ARL17B	CG2219	41690	9755	1.00	0.0	0.0	0.0	0.0
ATP2B4	CG42314	101743	108105	1.00	-1.0	2.0	2.0	2.0
ATP2B4	CG42314	109188	116060	1.00	-2.0	-1.0	0.0	2.0
ATP2B4	CG42314	30203	3152	1.00	0.0	0.0	0.0	0.0
ATP5O	CG4307	12792	4768	1.00	0.5	2.5	0.5	2.5
ATP5O	CG4307	12792	4768	1.00	0.0	2.5	1.5	2.5
ATP5O	CG4307	106753	107798	1.00	1.0	1.5	0.0	1.5
ATP5O	CG4307	12794	4768	1.00	1.0	1.0	-0.5	1.0
ATP5O	CG4307	12794	4768	1.00	0.5	2.5	1.0	2.5
ATP5O	CG4307	106753	107798	1.00	0.0	1.5	-1.0	1.5
ATXN2	CG5166	34955	11562	0.98	1.5	1.0	0.0	1.5
ATXN2	CG5166	34956	11562	0.98	0.0	0.0	0.0	0.0
BCL11A	CG9650	104402	108364	1.00	0.0	0.0	0.0	0.0
CACNA1C	CG43368	48092	15820	1.00	0.0	0.0	-0.5	0.5
CACNA1C	CG4894	52644	1737	1.00	-0.5	0.5	0.0	0.5
CACNA1C	CG4894	51490	1737	1.00	0.0	-0.5	0.0	0.5
CACNA1C	CG43368	104168	101478	1.00	0.5	0.0	0.0	0.5
CACNA1C	CG43368	5551	3326	0.99	0.0	0.0	0.0	0.0
CACNA1C	CG43368	43368	15341	1.00	0.0	0.0	0.0	0.0
CACNA1C	CG43368	48093	15820	1.00	0.0	0.0	0.0	0.0
CACNA1C	CG4894	51491	1737	1.00	0.0	0.0	0.0	0.0
CALR	CG9429	51272	4328	1.00	0.0	1.0	0.0	1.0
CCNA2	CG5940	32421	8653	1.00	0.0	0.0	0.0	0.0
CCNA2	CG5940	103595	101548	1.00	2.0	1.0	0.0	2.0
CCND2, CCND3	CG9096	105361	108447	1.00	0.0	0.0	-1.0	1.0
CORO1B	CG9446	109644	101987	1.00	-2.0	-1.0	-0.5	2.0
CRHR1	CG8422	110708	108591	1.00	-1.0	0.5	-1.5	1.5

Gene	Flybase Gene ID	VDRC stock no	Construct ID	Specificity score (S19)	Crystal cell early L3	Crystal cell late L3	Plasmatocytes	Highest Abs(score)
CRHR1	CG13758	106381	110677	1.00	0.5	2.5	2.0	2.5
CRHR1	CG13758	42724	712	1.00	0.0	0.0	0.0	0.0
CRHR1	CG12370	43314	15731	1.00	0.0	-1.0	0.0	1.0
CRHR1	CG12370	109558	115681	1.00	1.0	1.0	0.0	1.0
CRHR1	CG12370	102292	111461	1.00	0.0	0.0	0.0	0.0
CDK12	CG7597	25508	9926	0.97	0.0	0.0	0.0	0.0
CTRL	CG15002	32263	6002	0.99	0.0	1.0	1.5	1.5
CTRL	CG15002	103955	103120	1.00	0.0	0.0	0.0	0.0
DENND4A	CG12737	24602	7816	0.97	0.0	0.0	0.0	0.0
DENND4A	CG12737	24604	7816	0.97	0.0	0.0	0.0	0.0
DNAJA4	CG8863	23638	14050	1.00	0.0	0.0	2.0	2.0
DNAJA4	CG8863	104880	107834	1.00	1.0	2.0	2.5	2.5
DUS2L	CG1434	30996	6424	1.00	0.0	1.0	0.0	1.0
DUS2L	CG1434	30994	6424	1.00	0.0	0.0	-1.0	1.0
DUS2L	CG1434	104876	107816	1.00	0.0	0.0	-0.5	0.5
EDC4	CG6181	37945	5260	0.92	0.0	-0.5	-2.5	2.5
EDC4	CG6181	37946	5260	0.92	0.0	1.0	0.0	1.0
EDC4	CG6181	106687	102275	1.00	1.0	0.0	0.0	1.0
EIF5	CG9177	105992	102299	0.98	0.0	1.0	2.0	2.0
EIF5	CG9177	29070	14146	0.98	0.0	0.0	-2.5	2.5
EIF5	CG9177	29070	14146	0.98	0.0	0.0	-1.0	1.0
EIF5	CG9177	29071	14146	0.98	0.0	1.0	0.5	1.0
EIF5	CG9177	105992	102299	0.98	-1.0	1.0	0.5	1.0
ERAL1	CG7488	106677	101760	1.00	0.0	0.0	0.0	0.0
PIEZO1	CG8486	2796	993	1.00	0.0	0.0	0.0	0.0
PIEZO1	CG8486	105132	101815	1.00	0.0	0.0	0.0	0.0
PIEZO1	CG40188	109995	115662	1.00	0.0	0.0	0.0	0.0
FARSA	CG2263	33515	9769	0.98	0.0	0.0	0.0	0.0
FARSA	CG2263	33514	9769	0.98	0.0	0.0	0.0	0.0
FNTB	CG17565	32952	9415	1.00	0.0	-1.0	0.0	1.0
FNTB	CG17565	32951	9415	1.00	0.0	0.5	0.5	0.5
FNTB	CG17565	110646	108342	1.00	0.0	0.0	0.0	0.0
HBS1L	CG1898	33419	9671	0.99	0.0	-1.0	-1.5	1.5
HBS1L	CG1898	104327	107532	1.00	2.0	1.0	0.0	2.0
HBS1L	CG1898	33420	9671	0.99	0.0	1.0	-0.5	1.0
HEYL	CG11194	103570	101381	1.00	0.5	1.0	0.0	1.0
HK1	CG3001	104680	100831	1.00	2.0	2.5	0.0	2.5
HK1	CG3001	21054	9964	0.99	-0.5	-1.0	0.5	1.0
HK1	CG32849	47331	11656	0.99	0.5	0.5	0.0	0.5
HK1	CG33102	46574	11655	1.00	0.0	-0.5	0.0	0.5
HK1	CG8094	35337	12378	1.00	0.0	0.0	0.5	0.5
HK1	CG33102	46573	11655	1.00	-0.5	0.0	0.0	0.5
HK1	CG8094	35338	12378	1.00	0.0	0.0	-0.5	0.5
KCTD17	CG32810	108816	108908	0.99	0.0	0.5	0.0	0.5
KCTD17	CG32810	18225	7621	1.00	0.0	0.0	0.0	0.0
KIT	CG8222	43461	14375	1.00	0.0	-1.0	-3.0	3.0
KIT	CG8222	13502	2571	1.00	0.0	0.0	-3.0	3.0
KIT	CG8222	43459	14375	1.00	0.0	0.0	-2.0	2.0
KIT	CG8222	13503	2571	1.00	1.0	0.0	-1.0	1.0
KIT	CG8222	43461	14375	1.00	0.0	-0.5	-2.0	2.0
KIT	CG8222	105353	101575	1.00	0.0	1.0	-1.0	1.0
KIT	CG8222	977	75	1.00	0.0	-0.5	-1.5	1.5
KIT	CG8222	105353	101575	1.00	0.0	1.0	0.0	1.0
KIT	CG8222	43459	14375	1.00	1.0	0.0	-0.5	1.0
KIT	CG8222	13503	2571	1.00	0.5	0.5	0.0	0.5
KIT	CG8222	977	75	1.00	0.0	0.0	0.0	0.0
LRRC48	CG40440	109257	116234	1.00	1.5	0.5	0.0	1.5
MAPT	CG31057	25023	8682	1.00	0.0	0.5	0.0	0.5
MAPT	CG31057	101386	109359	1.00	0.0	0.0	0.0	0.0
MAPT	CG31057	25024	8682	1.00	0.0	0.0	0.0	0.0
MARCH8	CG4080	9026	2389	1.00	0.0	-1.0	0.0	1.0
MARCH8	CG4080	101426	109126	1.00	0.0	0.0	0.0	0.0
MAX	CG9648	110332	100742	1.00	0.0	0.0	0.0	0.0
MLEC	CG9257	103425	100869	1.00	0.0	1.0	0.0	1.0
MLEC	CG9257	6406	1795	1.00	0.5	0.0	0.0	0.5
MPND	CG4751	45530	11417	1.00	0.0	1.0	0.0	1.0
MPND	CG4751	26623	11417	1.00	0.0	0.5	-0.5	0.5

Gene	Flybase Gene ID	VDRC stock no	Construct ID	Specificity score (S19)	Crystal cell early L3	Crystal cell late L3	Plasmatocytes	Highest Abs(score)
NEK8	CG10951	100823	102962	1.00	0.0	0.0	0.0	0.0
NEK8	CG10951	16121	7142	1.00	0.0	0.0	0.0	0.0
NEK8	CG10951	16120	7142	1.00	0.0	0.0	0.0	0.0
NEUROD2	CG7508	48675	16434	1.00	0.5	0.0	0.0	0.5
NEUROD2	CG7508	2924	1379	1.00	0.5	0.5	0.0	0.5
NKX2-3	CG7895	12656	4155	1.00	0.0	0.0	-1.5	1.5
NKX2-3	CG7895	12655	4155	1.00	0.0	0.0	0.0	0.0
NKX2-3	CG7895	101825	109830	0.97	1.0	0.0	1.0	1.0
NKX2-3	CG7895	32510	4155	1.00	0.0	0.0	0.0	0.0
NPRL3	CG8783	40720	14027	1.00	1.0	0.0	0.0	1.0
NPRL3	CG8783	40721	14027	1.00	0.0	1.0	0.0	1.0
NUDT19	CG10194	41171	4939	1.00	1.0	-0.5	0.0	1.0
NUDT19	CG10194	107721	102407	1.00	0.5	0.0	0.0	0.5
NUDT19	CG10195	109468	104676	1.00	0.0	0.5	0.0	0.5
NUDT19	CG10194	41170	4939	1.00	0.5	0.0	0.0	0.5
NUDT19	CG18094	100138	104998	1.00	0.0	0.0	0.0	0.0
NUDT19	CG10195	26771	12581	1.00	0.0	0.0	0.0	0.0
NUDT19	CG18094	40126	9510	1.00	0.0	0.0	0.0	0.0
NUTF2	CG10174	109227	116168	1.00	0.0	0.0	0.0	0.0
NUTF2	CG1740	110108	114962	1.00	0.0	0.0	0.0	0.0
ODF3B	CG8086	23028	12374	1.00	-0.5	1.0	-2.0	2.0
ODF3B	CG8086	23027	12374	1.00	0.0	1.0	-2.0	2.0
ODF3B	CG8086	24225	13829	1.00	-1.0	0.0	-1.5	1.5
ODF3B	CG8086	24226	13829	1.00	0.0	0.0	0.0	0.0
PGS1	CG7718	25532	9946	1.00	-1.0	-1.5	0.0	1.5
PGS1	CG7718	109405	100360	1.00	0.0	1.0	0.0	1.0
PPCDC	CG30290	41564	8674	0.94	0.0	1.0	-0.5	1.0
PPCDC	CG30290	41565	8674	0.94	0.0	0.0	-0.5	0.5
PPCDC	CG30290	104495	109377	1.00	1.5	1.0	-1.5	1.5
PPCDC	CG30290	49962	16798	1.00	0.5	0.5	0.0	0.5
PRKCE	CG1954	108151	107658	0.99	0.0	1.0	0.0	1.0
PRKCE	CG1954	33434	9685	0.99	0.0	1.0	0.5	1.0
PSMB10	CG18341	52475	9563	1.00	0.0	0.0	1.0	1.0
PSMB10	CG12161	103323	112887	1.00	1.5	2.0	0.0	2.0
PSMB10	CG18341	108141	106009	1.00	1.0	0.5	0.0	1.0
PSMB10	CG12161	31669	7509	1.00	0.0	0.0	0.5	0.5
PSMB10	CG3329	24749	10938	1.00	0.0	0.0	0.0	0.0
PSMB10	CG3329	103575	101430	1.00	0.0	0.0	0.0	0.0
PTPLAD1	CG9267	101546	109012	1.00	0.5	0.0	0.0	0.5
PTPLAD1	CG6746	46513	1093	1.00	0.0	0.0	0.0	0.0
QKI	CG10293	100775	108558	1.00	0.0	0.0	0.0	0.0
RASA2	CG6721	23016	12823	0.99	0.0	-1.0	0.0	1.0
RASA2	CG6721	23017	12823	0.99	0.0	0.0	0.0	0.0
RASA2	CG6721	105383	100364	1.00	0.0	0.0	0.0	0.0
RPS6KB2	CG10539	104369	107986	0.99	0.0	0.5	-2.0	2.0
SBF2	CG3632	26254	11033	1.00	0.0	0.5	0.0	0.5
SBF2	CG3632	110167	102173	1.00	0.0	0.0	0.0	0.0
SBF2	CG6939	22317	12081	1.00	0.0	0.0	1.0	1.0
SCAMP5	CG9195	9130	3371	1.00	0.0	0.5	0.0	0.5
SCO2	CG8885	100005	102902	1.00	0.0	0.0	0.0	0.0
SCO2	CG8885	7861	898	1.00	0.0	2.0	-1.0	2.0
SCO2	CG8885	7860	898	1.00	1.0	2.5	0.0	2.5
SH2B3	CG17367	103646	105731	1.00	1.0	1.0	0.0	1.0
SH2B3	CG17367	32892	9362	1.00	0.0	0.0	0.0	0.0
SLC4A1	CG8177	109594	100095	1.00	1.0	1.0	0.0	1.0
SMOC1	CG2264	106494	108507	1.00	1.0	0.0	0.0	1.0
SPECC1	CG13366	108092	101059	1.00	-1.0	-1.0	-1.0	1.0
SPECC1	CG13366	29606	14994	1.00	0.0	0.0	-1.0	1.0
SPTA1	CG1977	110417	101541	1.00	0.0	0.0	0.0	0.0
ST5	CG18659	104988	107808	1.00	0.0	0.0	0.5	0.5
TAL1	CG2655	30564	4356	1.00	1.0	0.5	0.5	1.0
TAL1	CG2655	104381	108076	1.00	0.0	0.0	0.0	0.0
TMCC2	CG1021	109620	101382	1.00	1.0	0.5	0.0	1.0
TMCC2	CG1021	37336	2834	1.00	0.0	0.0	0.0	0.0
TRAF4	CG3048	110766	107398	1.00	0.0	0.0	0.5	0.5
UBE2L3	CG7425	110767	107438	0.97	-2.0	0.0	-2.0	2.0
UBE2L3	CG7425	110767	107438	0.97	-1.0	-2.0	-1.5	2.0

Gene	Flybase Gene ID	VDRC stock no	Construct ID	Specificity score (S19)	Crystal cell early L3	Crystal cell late L3	Plasmatocytes	Highest Abs(score)
UBE2L3	CG7425	105731	107993	1.00	-1.0	1.0	-1.0	1.0
UBE2L3	CG7425	26011	10600	1.00	0.0	0.0	0.0	0.0
UBE2L3	CG7425	26011	10600	1.00	0.0	-0.5	-2.0	2.0
UBE2L3	CG5788	100570	108273	1.00	0.0	1.0	0.0	1.0
UBE2L3	CG7425	26012	10600	1.00	0.0	0.0	0.5	0.5
UBE2L3	CG5788	48146	16723	0.97	1.5	1.0	0.0	1.5
UBE2L3	CG5788	48145	16723	0.97	1.0	0.0	0.0	1.0
UBE2L3	CG12799	20260	7832	0.98	0.0	-1.0	0.0	1.0
UBE2L3	CG5788	27515	11768	0.98	0.0	1.0	0.0	1.0
UBE2L3	CG5788	48145	16723	0.97	0.5	0.5	0.0	0.5
UBE2L3	CG12799	106363	110243	1.00	0.0	0.5	0.0	0.5
UBE2L3	CG5788	27515	11768	0.98	0.0	0.0	0.0	0.0
UBE2L3	CG12799	20260	7832	0.98	0.0	0.0	0.0	0.0
UBE2L3	CG5788	100570	108273	1.00	0.0	0.0	0.0	0.0
UBE2L3	CG7425	26012	10600	1.00	-1.0	1.0	0.0	1.0
UBE2L3	CG7425	105731	107993	1.00	0.0	0.0	-1.0	1.0
UBE2L3	CG17030	108804	107234	1.00	0.0	0.0	1.0	1.0
UBE2L3	CG2574	40173	9792	1.00	0.0	1.0	0.0	1.0
UBE2L3	CG10862	31372	7110	1.00	1.0	1.0	0.0	1.0
UBE2L3	CG17030	32827	9266	1.00	0.0	-0.5	0.0	0.5
UBE2L3	CG17030	108804	107234	1.00	0.0	-0.5	0.0	0.5
UBE2L3	CG10862	101113	106941	1.00	0.5	0.0	0.0	0.5
UBE2L3	CG2574	105725	107231	0.99	0.0	0.0	0.0	0.0
UBE2L3	CG2574	105725	107231	0.99	0.0	0.0	0.0	0.0
UBE2L3	CG10862	101113	106941	1.00	0.0	0.0	0.0	0.0
UBE2L3	CG17030	32827	9266	1.00	0.0	0.0	0.0	0.0
UBE2L3	CG10862	31372	7110	1.00	0.0	0.0	0.0	0.0
UBE2L3	CG2574	40173	9792	1.00	0.0	0.0	0.0	0.0
UBXD1	CG5469	105104	100852	1.00	0.0	0.0	0.0	0.0
UBXD1	CG5469	39000	11667	1.00	0.0	0.0	0.0	0.0
UBXD1	CG5469	38998	11667	1.00	0.0	0.0	0.0	0.0
WDR61	CG3909	12758	4738	0.99	-2.5	-1.0	-1.0	2.5
WDR61	CG3909	104387	108130	1.00	0.0	0.0	0.0	0.0
XRN1	CG3291	21677	10926	1.00	-1.0	0.0	0.0	1.0
XRN1	CG3291	105739	108511	1.00	0.0	0.0	0.0	0.0

Supplementary references

1. Nyholt,D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765-769 (2004).
2. Chambers,J.C. *et al.* Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat. Genet.* **41**, 1170-1172 (2009).
3. Ganesh,S.K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* **41**, 1191-1198 (2009).
4. Soranzo,N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41**, 1182-1190 (2009).
5. Kamatani,Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet* **42**, 210-215 (2010).
6. Ding,K. *et al.* Genetic Loci implicated in erythroid differentiation and cell cycle regulation are associated with red blood cell traits. *Mayo Clin. Proc.* **87**, 461-474 (2012).
7. Olivieri,N.F. The beta-thalassemias. *N. Engl. J Med.* **341**, 99-109 (1999).
8. Thein,S.L. Genetic modifiers of beta-thalassemia. *Haematologica* **90**, 649-660 (2005).
9. Galanello,R. *et al.* Amelioration of Sardinian beta0 thalassemia by genetic modifiers. *Blood* **114**, 3935-3937 (2009).
10. Macdonald-Wallis,C., Tilling,K., Fraser,A., Nelson,S.M., & Lawlor,D.A. Established preeclampsia risk factors are related to patterns of blood pressure change in normal term pregnancy: findings from the Avon Longitudinal Study of Parents and Children. *J Hypertens.* **29**, 1703-1711 (2011).
11. Pollin,T.I. *et al.* A genome-wide scan of serum lipid levels in the Old Order Amish. *Atherosclerosis* **173**, 89-96 (2004).
12. Mitchell,B.D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *Am. Heart J.* **155**, 823-828 (2008).
13. Firmann,M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC. Cardiovasc. Disord.* **8**, 6 (2008).
14. Vaxillaire,M. *et al.* Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study. *Diabetes* **57**, 244-254 (2008).
15. Nelis,M. *et al.* Genetic structure of Europeans: a view from the North-East. *PLoS. One.* **4**, e5472 (2009).
16. Day,N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br. J Cancer* **80 Suppl 1**, 95-103 (1999).
17. Bhattacharyya,T. *et al.* Relationship of paraoxonase 1 (PON1) gene polymorphisms and functional activity with systemic oxidative stress and cardiovascular risk. *JAMA* **299**, 1265-1276 (2008).
18. Nalls,M.A. *et al.* Multiple loci are associated with white blood cell phenotypes. *PLoS. Genet* **7**, e1002113 (2011).
19. Colonna,V. *et al.* Comparing population structure as inferred from genealogical versus genetic information. *Eur. J Hum Genet* **17**, 1635-1641 (2009).
20. Valent,F., Schiava,F., Gnesutta,D., Brianti,G., & Barbone,F. [Evaluation of the effectiveness of influenza vaccination among the elderly in Friuli-Venezia Giulia (Italy)]. *Epidemiol. Prev.* **29**, 195-203 (2005).
21. Traglia,M. *et al.* Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes. *PLoS. One.* **4**, e7554 (2009).
22. Holz,T. *et al.* Markers of inflammation and weight change in middle-aged adults: results from the prospective MONICA/KORA S3/F3 study. *Obesity. (Silver. Spring)* **18**, 2347-2353 (2010).

23. Stockl,D. *et al.* Age at menarche and its association with the metabolic syndrome and its components: results from the KORA F4 study. *PLoS. One.* **6**, e26076 (2011).
24. Starr,J.M. & Deary,I.J. Sex differences in blood cell counts in the Lothian Birth Cohort 1921 between 79 and 87 years. *Maturitas* **69**, 373-376 (2011).
25. Deary,I.J., Whiteman,M.C., Starr,J.M., Whalley,L.J., & Fox,H.C. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J Pers. Soc. Psychol.* **86**, 130-147 (2004).
26. Deary,I.J. *et al.* The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC. Geriatr.* **7**, 28 (2007).
27. McIlhagger,R. *et al.* Differences in the haematological profile of healthy 70 year old men and women: normal ranges with confirmatory factor analysis. *BMC. Blood Disord.* **10**, 4 (2010).
28. Stolk,R.P. *et al.* Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur. J Epidemiol.* **23**, 67-74 (2008).
29. Chambers,J.C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet* **43**, 1131-1138 (2011).
30. Berglund,G., Elmstahl,S., Janzon,L., & Larsson,S.A. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern. Med.* **233**, 45-51 (1993).
31. Padmanabhan,S. *et al.* Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS. Genet.* **6**, e1001177 (2010).
32. Pattaro,C. *et al.* The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. *BMC. Med. Genet* **8**, 29 (2007).
33. Penninx,B.W. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J Methods Psychiatr. Res* **17**, 121-140 (2008).
34. Rantakallio,P. The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr. Perinat. Epidemiol.* **2**, 59-88 (1988).
35. Willemsen,G. *et al.* The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet* **13**, 231-245 (2010).
36. Diercks,G.F. *et al.* Rationale, design, and baseline characteristics of a trial of prevention of cardiovascular and renal disease with fosinopril and pravastatin in nonhypertensive, nonhypercholesterolemic subjects with microalbuminuria (the Prevention of REnal and Vascular ENdstage Disease Intervention Trial [PREVEND IT]). *Am. J Cardiol.* **86**, 635-638 (2000).
37. Evans,D.M., Frazer,I.H., & Martin,N.G. Genetic and environmental causes of variation in basal levels of blood cells. *Twin. Res.* **2**, 250-257 (1999).
38. Pilia,G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS. Genet* **2**, e132 (2006).
39. Cao,A. *et al.* Thalassaemia and glucose-6-phosphate dehydrogenase screening in 13- to 14-year-old students of the Sardinian population: preliminary findings. *Community Genet* **11**, 121-128 (2008).
40. Volzke,H. *et al.* Cohort profile: the study of health in Pomerania. *Int. J Epidemiol.* **40**, 294-307 (2011).
41. Tonjes,A. *et al.* Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs. *Hum Mol. Genet* **18**, 4662-4668 (2009).
42. Spector,T.D. & MacGregor,A.J. The St. Thomas' UK Adult Twin Registry. *Twin Res* **5**, 440-443 (2002).
43. Dendrou,C.A. *et al.* Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nat. Genet* **41**, 1011-1015 (2009).
44. Winkelmann,B.R. *et al.* Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics.* **2**, S1-73 (2001).
45. Fraumene,C. *et al.* High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates. *Mol. Biol Evol.* **23**, 2101-2111 (2006).

46. Achterberg,S. *et al.* Patients with coronary, cerebrovascular or peripheral arterial obstructive disease differ in risk for new vascular events and mortality: the SMART study. *Eur. J. Cardiovasc. Prev. Rehabil.* **17**, 424-430 (2010).
47. Simons,P.C., Algra,A., van de Laak,M.F., Grobbee,D.E., & van der Graaf,Y. Second manifestations of ARTerial disease (SMART) study: rationale and design. *Eur. J. Epidemiol.* **15**, 773-781 (1999).
48. Oikonen,M. *et al.* Genetic variants and blood pressure in a population-based cohort: the cardiovascular risk in young Finns study. *Hypertension* **58**, 1079-1085 (2011).
49. Kullo,I.J., Ding,K., Jouni,H., Smith,C.Y., & Chute,C.G. A genome-wide association study of red blood cell traits using the electronic medical record. *PLoS. One.* **5**, (2010).
50. Dixon,A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202-1207 (2007).
51. Dubois,P.C. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet* **42**, 295-302 (2010).
52. Leimeister,C., Schumacher,N., Steidl,C., & Gessler,M. Analysis of HeyL expression in wild-type and Notch pathway mutant mouse embryos. *Mech. Dev.* **98**, 175-178 (2000).
53. Lavery,D.N. *et al.* Repression of androgen receptor activity by HEYL, a third member of the Hairy/Enhancer-of-split-related family of Notch effectors. *J Biol Chem.* **286**, 17796-17808 (2011).
54. Kassouf,M.T. *et al.* Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. *Genome Res* **20**, 1064-1083 (2010).
55. Palii,C.G. *et al.* Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J* **30**, 494-509 (2011).
56. Delaunay,J. The molecular basis of hereditary red cell membrane disorders. *Blood Rev.* **21**, 1-20 (2007).
57. Cuesta,R., Martinez-Sanchez,A., & Gebauer,F. miR-181a regulates cap-dependent translation of p27(kip1) mRNA in myeloid cells. *Mol. Cell Biol* **29**, 2841-2851 (2009).
58. Jazdzewski,K. *et al.* Thyroid hormone receptor beta (THRB) is a major target gene for microRNAs deregulated in papillary thyroid carcinoma (PTC). *J Clin. Endocrinol. Metab* **96**, E546-E553 (2011).
59. Shin,K.H. *et al.* miR-181a shows tumor suppressive effect against oral squamous cell carcinoma cells by downregulating K-ras. *Biochem. Biophys. Res Commun.* **404**, 896-902 (2011).
60. Chen,G. *et al.* MicroRNA-181a sensitizes human malignant glioma U87MG cells to radiation by targeting Bcl-2. *Oncol. Rep.* **23**, 997-1003 (2010).
61. Gao,W. *et al.* Deregulated expression of miR-21, miR-143 and miR-181a in non small cell lung cancer is related to clinicopathologic characteristics or patient prognosis. *Biomed. Pharmacother.* **64**, 399-408 (2010).
62. Tempel,B.L. & Shilling,D.J. The plasma membrane calcium ATPase and disease. *Subcell. Biochem.* **45**, 365-383 (2007).
63. Hopkins,P.C., Sainz-Fuertes,R., & Lovestone,S. The impact of a novel apolipoprotein E and amyloid-beta protein precursor-interacting protein on the production of amyloid-beta. *J Alzheimers. Dis.* **26**, 239-253 (2011).
64. Rozengurt,E., Rey,O., & Waldron,R.T. Protein kinase D signaling. *J Biol Chem.* **280**, 13205-13208 (2005).
65. Nuinoon,M. *et al.* A genome-wide association identified the common genetic variants influence disease severity in beta0-thalassemia/hemoglobin E. *Hum Genet* **127**, 303-314 (2010).
66. Voight,B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet* **42**, 579-589 (2010).
67. Jameson,J.L. Mechanisms by which thyroid hormone receptor mutations cause clinical syndromes of resistance to thyroid hormone. *Thyroid* **4**, 485-492 (1994).

68. Jiang,Y. *et al.* The G protein G alpha12 stimulates Bruton's tyrosine kinase and a rasGAP through a conserved PH/BM domain. *Nature* **395**, 808-813 (1998).
69. Liu,Z. & Gilbert,W. The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure. *Cell* **77**, 1083-1092 (1994).
70. Aisen,P. Transferrin receptor 1. *Int. J Biochem. Cell Biol* **36**, 2137-2143 (2004).
71. Kimura,Y. *et al.* c-Kit-mediated functional positioning of stem cells to their niches is essential for maintenance and regeneration of adult hematopoiesis. *PLoS. One.* **6**, e26918 (2011).
72. Blacque,O.E. *et al.* Loss of C. elegans BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. *Genes Dev.* **18**, 1630-1642 (2004).
73. Pan,J. *et al.* Suppression of cell-cycle progression by Jun dimerization protein-2 (JDP2) involves downregulation of cyclin-A2. *Oncogene* **29**, 6245-6256 (2010).
74. Li,J. *et al.* Crystal structure of human guanosine monophosphate reductase 2 (GMPR2) in complex with GMP. *J Mol. Biol* **355**, 980-988 (2006).
75. Hanson,E.H., Imperatore,G., & Burke,W. HFE gene and hereditary hemochromatosis: a HuGE review. Human Genome Epidemiology. *Am. J Epidemiol.* **154**, 193-206 (2001).
76. Jutabha,P. *et al.* Human sodium phosphate transporter 4 (hNPT4/SLC17A3) as a common renal secretory pathway for drugs and urate. *J Biol Chem.* **285**, 35123-35132 (2010).
77. Roberts,J.D., Jr., Chiche,J.D., Kolpa,E.M., Bloch,D.B., & Bloch,K.D. cGMP-dependent protein kinase I interacts with TRIM39R, a novel Rpp21 domain-containing TRIM protein. *Am. J Physiol Lung Cell Mol. Physiol* **293**, L903-L912 (2007).
78. Reinshagen,M. *et al.* HLA class II gene frequencies in Crohn's disease: a population based analysis in Germany. *Gut* **38**, 538-542 (1996).
79. Cowland,J.B., Andersen,V., Halberg,P., & Morling,N. DNA polymorphism of HLA class II genes in systemic lupus erythematosus. *Tissue Antigens* **43**, 34-37 (1994).
80. Erlich,H. *et al.* HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes* **57**, 1084-1092 (2008).
81. Hor,H. *et al.* Genome-wide association study identifies new HLA class II haplotypes strongly protective against narcolepsy. *Nat. Genet* **42**, 786-789 (2010).
82. Barcellos,L.F. *et al.* High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLoS. Genet* **5**, e1000696 (2009).
83. Galligan,C.L., Baig,E., Bykerk,V., Keystone,E.C., & Fish,E.N. Distinctive gene expression signatures in rheumatoid arthritis synovial tissue fibroblast cells: correlates with disease activity. *Genes Immun.* **8**, 480-491 (2007).
84. Kozar,K. *et al.* Mouse development and cell proliferation in the absence of D-cyclins. *Cell* **118**, 477-491 (2004).
85. Ferrara,N. & Gerber,H.P. The role of vascular endothelial growth factor in angiogenesis. *Acta Haematol.* **106**, 148-156 (2001).
86. Watanabe,O. *et al.* Overproduction of vascular endothelial growth factor/vascular permeability factor is causative in Crow-Fukase (POEMS) syndrome. *Muscle Nerve* **21**, 1390-1397 (1998).
87. Caldwell,R.B. *et al.* Vascular endothelial growth factor and diabetic retinopathy: pathophysiological mechanisms and treatment perspectives. *Diabetes Metab Res Rev.* **19**, 442-455 (2003).
88. Menzel,S. *et al.* The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood* **110**, 3624-3626 (2007).
89. Thein,S.L. *et al.* Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc. Natl. Acad. Sci. U. S. A* **104**, 11346-11351 (2007).

90. Lettre,G. *et al.* DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. U. S. A* **105**, 11869-11874 (2008).
91. Kranc,K.R. *et al.* Cited2 is an essential regulator of adult hematopoietic stem cells. *Cell Stem Cell* **5**, 659-665 (2009).
92. Bamforth,S.D. *et al.* Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. *Nat. Genet* **29**, 469-474 (2001).
93. Feng,Y. & Bankston,A. The star family member QKI and cell signaling. *Adv. Exp. Med. Biol* **693**, 25-36 (2010).
94. Georgopoulos,K., Winandy,S., & Avitahl,N. The role of the Ikaros gene in lymphocyte development and homeostasis. *Annu. Rev. Immunol.* **15**, 155-176 (1997).
95. Jager,R. *et al.* Deletions of the transcription factor Ikaros in myeloproliferative neoplasms. *Leukemia* **24**, 1290-1298 (2010).
96. Kuroda,Y., Oma,Y., Nishimori,K., Ohta,T., & Harata,M. Brain-specific expression of the nuclear actin-related protein ArpNalpha and its involvement in mammalian SWI/SNF chromatin remodeling complex. *Biochem. Biophys. Res Commun.* **299**, 328-334 (2002).
97. Forejtnikova,H. *et al.* Transferrin receptor 2 is a component of the erythropoietin receptor complex and is required for efficient erythropoiesis. *Blood* **116**, 5357-5367 (2010).
98. Folmes,K.D. *et al.* Distinct early signaling events resulting from the expression of the PRKAG2 R302Q mutant of AMPK contribute to increased myocardial glycogen. *Circ. Cardiovasc. Genet* **2**, 457-466 (2009).
99. Bennett,V. Immunoreactive forms of human erythrocyte ankyrin are present in diverse cells and tissues. *Nature* **281**, 597-599 (1979).
100. Gallagher,P.G. *et al.* Mutation of a barrier insulator in the human ankyrin-1 gene is associated with hereditary spherocytosis. *J Clin. Invest* **120**, 4453-4465 (2010).
101. Horn,D.M., Mason,S.L., & Karbstein,K. Rcl1 protein, a novel nuclease for 18 S ribosomal RNA production. *J Biol Chem.* **286**, 34082-34087 (2011).
102. Nakao,M. *et al.* ABO blood group alleles and the risk of pancreatic cancer in a Japanese population. *Cancer Sci.* **102**, 1076-1080 (2011).
103. Nakao,M. *et al.* ABO genotype and the risk of gastric cancer, atrophic gastritis, and Helicobacter pylori infection. *Cancer Epidemiol. Biomarkers Prev.* **20**, 1665-1672 (2011).
104. Plagnol,V. *et al.* Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS. Genet* **7**, e1002216 (2011).
105. Goto,E. *et al.* c-MIR, a human E3 ubiquitin ligase, is a functional homolog of herpesvirus proteins MIR1 and MIR2 and has similar activity. *J Biol Chem.* **278**, 14657-14668 (2003).
106. Bartee,E., Mansouri,M., Hovey Nerenberg,B.T., Gouveia,K., & Fruh,K. Downregulation of major histocompatibility complex class I by human ubiquitin ligases related to viral immune evasion proteins. *J Virol.* **78**, 1109-1120 (2004).
107. Murakami,K., Kanno,H., Tancabelic,J., & Fujii,H. Gene expression and biological significance of hexokinase in erythroid cells. *Acta Haematol.* **108**, 204-209 (2002).
108. van,W.R., Rijksen,G., Huizinga,E.G., Nieuwenhuis,H.K., & van Solinge,W.W. HK Utrecht: missense mutation in the active site of human hexokinase associated with hexokinase deficiency and severe nonspherocytic hemolytic anemia. *Blood* **101**, 345-347 (2003).
109. Tarlinton,D., Light,A., Metcalf,D., Harvey,R.P., & Robb,L. Architectural defects in the spleens of Nkx2-3-deficient mice are intrinsic and associated with defects in both B cell maturation and T cell-dependent immune responses. *J Immunol.* **170**, 4002-4010 (2003).
110. Fisher,S.A. *et al.* Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat. Genet* **40**, 710-712 (2008).
111. Gao,N., Hibi,Y., Cueno,M., Asamitsu,K., & Okamoto,T. A-kinase-interacting protein 1 (AKIP1) acts as a molecular determinant of PKA in NF-kappaB signaling. *J Biol Chem.* **285**, 28097-28104 (2010).

112. Majidi,M., Hubbs,A.E., & Lichy,J.H. Activation of extracellular signal-regulated kinase 2 by a novel Abl-binding protein, ST5. *J Biol Chem.* **273**, 16608-16614 (1998).
113. Gohring,I. *et al.* Disruption of ST5 is associated with mental retardation and multiple congenital anomalies. *J Med. Genet* **47**, 91-98 (2010).
114. Robinson,F.L., Niesman,I.R., Beiswenger,K.K., & Dixon,J.E. Loss of the inactive myotubularin-related phosphatase Mtmr13 leads to a Charcot-Marie-Tooth 4B2-like peripheral neuropathy in mice. *Proc. Natl. Acad. Sci. U. S. A* **105**, 4916-4921 (2008).
115. Cai,L., Holoweckyj,N., Schaller,M.D., & Bear,J.E. Phosphorylation of coronin 1B by protein kinase C regulates interaction with Arp2/3 and cell motility. *J Biol Chem.* **280**, 31913-31923 (2005).
116. Fortin,M. *et al.* Apoptosis mediated through CD45 is independent of its phosphatase activity and association with leukocyte phosphatase-associated phosphoprotein. *J Immunol.* **168**, 6084-6089 (2002).
117. Slattery,M.L., Lundgreen,A., Herrick,J.S., & Wolff,R.K. Genetic variation in RPS6KA1, RPS6KA2, RPS6KB1, RPS6KB2, and PDK1 and risk of colon or rectal cancer. *Mutat. Res* **706**, 13-20 (2011).
118. Rumenapp,U., Freichel-Blomquist,A., Wittinghofer,B., Jakobs,K.H., & Wieland,T. A mammalian Rho-specific guanine-nucleotide exchange factor (p164-RhoGEF) without a pleckstrin homology domain. *Biochem. J* **366**, 721-728 (2002).
119. Bar,I. *et al.* Knockout mice reveal a role for P2Y6 receptor in macrophages, endothelial cells, and vascular smooth muscle cells. *Mol. Pharmacol.* **74**, 777-784 (2008).
120. Riegel,A.K. *et al.* Selective induction of endothelial P2Y6 nucleotide receptor promotes vascular inflammation. *Blood* **117**, 2548-2555 (2011).
121. Perets,T., Blumenstein,Y., Shistik,E., Lotan,I., & Dascal,N. A potential site of functional modulation by protein kinase A in the cardiac Ca2+ channel alpha 1C subunit. *FEBS Lett.* **384**, 189-192 (1996).
122. Decker,T. *et al.* Cell cycle progression of chronic lymphocytic leukemia cells is controlled by cyclin D2, cyclin D3, cyclin-dependent kinase (cdk) 4 and the cdk inhibitor p27. *Leukemia* **16**, 327-334 (2002).
123. Geissler,E.N. *et al.* Stem cell factor (SCF), a novel hematopoietic growth factor and ligand for c-kit tyrosine kinase receptor, maps on human chromosome 12 between 12q14.3 and 12qter. *Somat. Cell Mol. Genet* **17**, 207-214 (1991).
124. Figueroa,K.P. *et al.* Genetic variance in the spinocerebellar ataxia type 2 (ATXN2) gene in children with severe early onset obesity. *PLoS. One.* **4**, e8280 (2009).
125. Daoud,H. *et al.* Association of long ATXN2 CAG repeat sizes with increased risk of amyotrophic lateral sclerosis. *Arch. Neurol.* **68**, 739-742 (2011).
126. Devalliere,J. & Charreau,B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochem. Pharmacol.* **82**, 1391-1402 (2011).
127. Schallus,T. *et al.* Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. *Mol. Biol Cell* **19**, 3404-3414 (2008).
128. Manne,V. *et al.* Identification and preliminary characterization of protein-cysteine farnesyltransferase. *Proc. Natl. Acad. Sci. U. S. A* **87**, 7541-7545 (1990).
129. Luscher,B. Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene* **277**, 1-14 (2001).
130. Vannahme,C. *et al.* Characterization of SMOC-1, a novel modular calcium-binding protein in basement membranes. *J Biol Chem.* **277**, 37977-37986 (2002).
131. Okada,I. *et al.* SMOC1 is essential for ocular and limb development in humans and mice. *Am. J Hum Genet* **88**, 30-41 (2011).
132. Asano,K. *et al.* Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *EMBO J* **20**, 2326-2337 (2001).

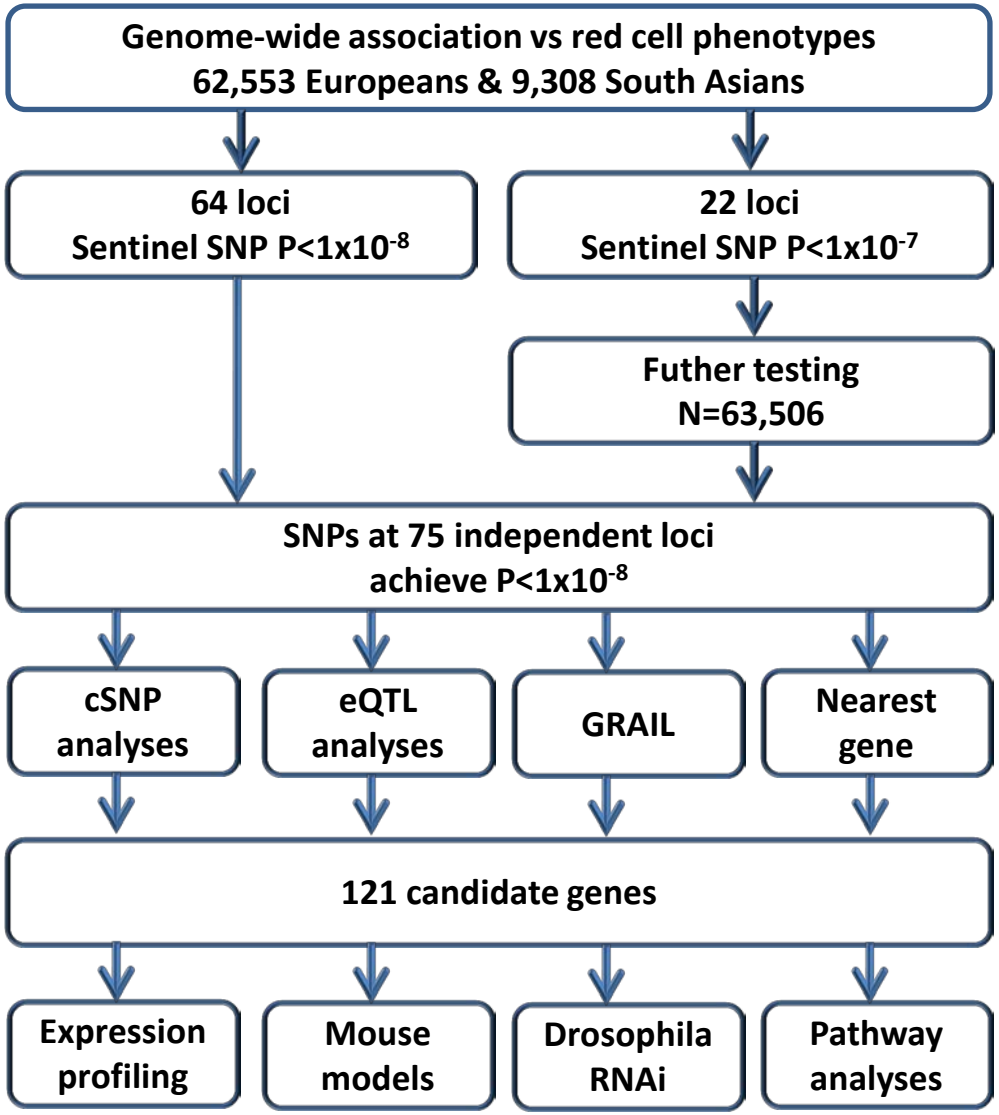
133. Jansen,H., Verhoeven,A.J., & Sijbrands,E.J. Hepatic lipase: a pro- or anti-atherogenic protein? *J Lipid Res* **43**, 1352-1362 (2002).
134. Yu,Y. *et al.* Association of variants in the LIPC and ABCA1 genes with intermediate and large drusen and advanced age-related macular degeneration. *Invest Ophthalmol. Vis. Sci.* **52**, 4663-4670 (2011).
135. Semova,N. *et al.* Molecular cloning, structural analysis, and expression of a human IRLB, MYC promoter-binding protein: new DENN domain-containing protein family emerges small star, filled. *Genomics* **82**, 343-354 (2003).
136. Courilleau,D. *et al.* B-ind1, a novel mediator of Rac1 signaling cloned from sodium butyrate-treated fibroblasts. *J Biol Chem.* **275**, 17344-17348 (2000).
137. Daugherty,M. *et al.* Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *J Biol Chem.* **277**, 21431-21439 (2002).
138. Han,C. *et al.* Human SCAMP5, a novel secretory carrier membrane protein, facilitates calcium-triggered cytokine secretion by interaction with SNARE machinery. *J Immunol.* **182**, 2986-2996 (2009).
139. Noh,J.Y. *et al.* SCAMP5 links endoplasmic reticulum stress to the accumulation of expanded polyglutamine protein aggregates via endocytosis inhibition. *J Biol Chem.* **284**, 11318-11325 (2009).
140. Harari,D. *et al.* Neuregulin-4: a novel growth factor that acts through the ErbB-4 receptor tyrosine kinase. *Oncogene* **18**, 2681-2689 (1999).
141. Robichon,C. *et al.* DnaJA4 is a SREBP-regulated chaperone involved in the cholesterol biosynthesis pathway. *Biochim. Biophys. Acta* **1761**, 1107-1113 (2006).
142. Zhu,B. *et al.* The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev.* **19**, 1668-1673 (2005).
143. Neklesa,T.K. & Davis,R.W. A genome-wide screen for regulators of TORC1 in response to amino acid starvation reveals a conserved Npr2/3 complex. *PLoS. Genet* **5**, e1000515 (2009).
144. Reseland,J.E. *et al.* A novel human chymotrypsin-like digestive enzyme. *J Biol Chem.* **272**, 8099-8104 (1997).
145. Simon,E., Camier,S., & Seraphin,B. New insights into the control of mRNA decapping. *Trends Biochem. Sci.* **31**, 241-243 (2006).
146. Paschal,B.M. & Gerace,L. Identification of NTF2, a cytosolic factor for nuclear import that interacts with nuclear pore complex protein p62. *J Cell Biol* **129**, 925-937 (1995).
147. McHugh,B.J. *et al.* Integrin activation by Fam38A uses a novel mechanism of R-Ras targeting to the endoplasmic reticulum. *J Cell Sci.* **123**, 51-61 (2010).
148. D'Agostino,L. & Giordano,A. Possible functional role of NSPs in cancer. *Cell Cycle* **7**, 1810-1827 (2008).
149. Morerio,C. *et al.* HCMOGT-1 is a novel fusion partner to PDGFRB in juvenile myelomonocytic leukemia with t(5;17)(q33;p11.2). *Cancer Res* **64**, 2649-2651 (2004).
150. Dennerlein,S., Rozanska,A., Wydro,M., Chrzanowska-Lightowlers,Z.M., & Lightowlers,R.N. Human ERAL1 is a mitochondrial RNA chaperone involved in the assembly of the 28S small mitochondrial ribosomal subunit. *Biochem. J* **430**, 551-558 (2010).
151. Holland,P.M. *et al.* Purification, cloning, and characterization of Nek8, a novel NIMA-related kinase, and its candidate substrate Bicd2. *J Biol Chem.* **277**, 16229-16240 (2002).
152. Otto,E.A. *et al.* NEK8 mutations affect ciliary and centrosomal localization and may cause nephronophthisis. *J Am. Soc. Nephrol.* **19**, 587-592 (2008).
153. Forrest,A.R., Taylor,D., & Grimmond,S. Exploration of the cell-cycle genes found within the RIKEN FANTOM2 data set. *Genome Res* **13**, 1366-1375 (2003).
154. Zapata,J.M. *et al.* A diverse family of proteins containing tumor necrosis factor receptor-associated factor domains. *J Biol Chem.* **276**, 24242-24252 (2001).

155. Chen,H.H., Wang,Y.C., & Fann,M.J. Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol. Cell Biol* **26**, 2736-2745 (2006).
156. Ko,T.K., Kelly,E., & Pines,J. CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J Cell Sci.* **114**, 2591-2603 (2001).
157. Bartkowiak,B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303-2316 (2010).
158. McCormick,M.B. *et al.* NeuroD2 and neuroD3: distinct expression patterns and transcriptional activation potentials within the neuroD gene family. *Mol. Cell Biol* **16**, 5792-5800 (1996).
159. Messmer,K., Shen,W.B., Remington,M., & Fishman,P.S. Induction of neural differentiation by the transcription factor NeuroD2. *Int. J Dev. Neurosci.*(2011).
160. Shmukler,B.E., Wilhelm,S., & Alper,S.L. Short sequence repeat polymorphism in the mouse slc4al gene encoding the AE1 Cl-/HCO3-exchanger. *DNA Seq.* **11**, 447-450 (2000).
161. Bruce,L.J. *et al.* Altered structure and anion transport properties of band 3 (AE1, SLC4A1) in human red cells lacking glycophorin A. *J Biol Chem.* **279**, 2414-2420 (2004).
162. Delaunay,J. Molecular basis of red cell membrane disorders. *Acta Haematol.* **108**, 210-218 (2002).
163. Voit,R., Kuhn,A., Sander,E.E., & Grummt,I. Activation of mammalian ribosomal gene transcription requires phosphorylation of the nucleolar transcription factor UBF. *Nucleic Acids Res* **23**, 2593-2599 (1995).
164. Sakakibara,T., Nemoto,Y., Nukiwa,T., & Takeshima,H. Identification and characterization of a novel Rho GTPase activating protein implicated in receptor-mediated endocytosis. *FEBS Lett.* **566**, 294-300 (2004).
165. Kahn,R.A., Kern,F.G., Clark,J., Gelmann,E.P., & Rulka,C. Human ADP-ribosylation factors. A functionally conserved family of GTP-binding proteins. *J Biol Chem.* **266**, 2606-2614 (1991).
166. Aguilera,G., Nikodemova,M., Wynn,P.C., & Catt,K.J. Corticotropin releasing hormone receptors: two decades later. *Peptides* **25**, 319-329 (2004).
167. Grigorenko,A.P., Moliaka,Y.K., Korovaitseva,G.I., & Rogaev,E.I. Novel class of polytopic proteins with domains associated with putative protease activity. *Biochemistry (Mosc.)* **67**, 826-835 (2002).
168. Cai,Y. *et al.* Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem.* **285**, 4268-4272 (2010).
169. Southall,S.M., Wong,P.S., Odho,Z., Roe,S.M., & Wilson,J.R. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol. Cell* **33**, 181-191 (2009).
170. Seeger,M.A. & Rice,S.E. Microtubule-associated protein-like binding of the kinesin-1 tail to microtubules. *J Biol Chem.* **285**, 8155-8162 (2010).
171. Aranda-Abreu,G.E., Hernandez,M.E., Soto,A., & Manzo,J. Possible Cis-acting signal that could be involved in the localization of different mRNAs in neuronal axons. *Theor. Biol Med. Model.* **2**, 33 (2005).
172. Iqbal,K., Liu,F., Gong,C.X., & Grundke-Iqbal,I. Tau in Alzheimer disease and related tauopathies. *Curr. Alzheimer Res* **7**, 656-664 (2010).
173. Hoglinger,G.U. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet* **43**, 699-705 (2011).
174. Gao,L., Tse,S.W., Conrad,C., & Andreadis,A. Saitohin, which is nested in the tau locus and confers allele-specific susceptibility to several neurodegenerative diseases, interacts with peroxiredoxin 6. *J Biol Chem.* **280**, 39268-39272 (2005).

175. Chang,S.C., Heacock,P.N., Clancey,C.J., & Dowhan,W. The PEL1 gene (renamed PGS1) encodes the phosphatidylglycero-phosphate synthase of Saccharomyces cerevisiae. *J Biol Chem.* **273**, 9829-9836 (1998).
176. Le,B.R., Alconada,A., Bauer,U., & Hoflack,B. The mammalian AP-3 adaptor-like complex mediates the intracellular transport of lysosomal membrane glycoproteins. *J Biol Chem.* **273**, 29451-29461 (1998).
177. Lua,B.L. & Low,B.C. Activation of EGF receptor endocytosis and ERK1/2 signaling by BPGAP1 requires direct interaction with EEN/endophilin II and a functional RhoGAP domain. *J Cell Sci.* **118**, 2707-2721 (2005).
178. So,C.W. *et al.* EEN encodes for a member of a new family of proteins containing an Src homology 3 domain and is the third gene located on chromosome 19p13 that fuses to MLL in human leukemia. *Proc. Natl. Acad. Sci. U. S. A* **94**, 2563-2568 (1997).
179. Nagahama,M. *et al.* UBXD1 is a VCP-interacting protein that is involved in ER-associated degradation. *Biochem. Biophys. Res Commun.* **382**, 303-308 (2009).
180. Qiu,Y. & Michalak,M. Transcriptional control of the calreticulin gene in health and disease. *Int. J Biochem. Cell Biol* **41**, 531-538 (2009).
181. Yang,D.C. Mammalian aminoacyl-tRNA synthetases. *Curr. Top. Cell Regul.* **34**, 101-136 (1996).
182. Sen,S. *et al.* Expression of a gene encoding a tRNA synthetase-like protein is enhanced in tumorigenic human myeloid leukemia cells and is cell cycle stage- and differentiation-dependent. *Proc. Natl. Acad. Sci. U. S. A* **94**, 6164-6169 (1997).
183. Costa,Y. *et al.* Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis. *J Cell Sci.* **118**, 2755-2762 (2005).
184. Bolcun-Filas,E. *et al.* SYCE2 is required for synaptonemal complex assembly, double strand break repair, and homologous recombination. *J Cell Biol* **176**, 741-747 (2007).
185. Ofman,R., Speijer,D., Leen,R., & Wanders,R.J. Proteomic analysis of mouse kidney peroxisomes: identification of RP2p as a peroxisomal nudix hydrolase with acyl-CoA diphosphatase activity. *Biochem. J* **393**, 537-543 (2006).
186. Shu,L., Yan,W., & Chen,X. RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript. *Genes Dev.* **20**, 2961-2972 (2006).
187. Ronn,T. *et al.* Genetic variation in ATP5O is associated with skeletal muscle ATP50 mRNA expression and glucose uptake in young twins. *PLoS. One.* **4**, e4793 (2009).
188. Moynihan,T.P. *et al.* Characterization of a human ubiquitin-conjugating enzyme gene UBE2L3. *Mamm. Genome* **7**, 520-525 (1996).
189. Cenciarelli,C. *et al.* Identification of a family of human F-box proteins. *Curr. Biol* **9**, 1177-1179 (1999).
190. Lomonosov,M. *et al.* Expression of Fbxo7 in haematopoietic progenitor cells cooperates with p53 loss to promote lymphomagenesis. *PLoS. One.* **6**, e21165 (2011).
191. Du,X. *et al.* The serine protease TMPRSS6 is required to sense iron deficiency. *Science* **320**, 1088-1092 (2008).
192. Finberg,K.E. *et al.* Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). *Nat. Genet* **40**, 569-571 (2008).
193. Usuki,K. *et al.* Platelet-derived endothelial cell growth factor has thymidine phosphorylase activity. *Biochem. Biophys. Res Commun.* **184**, 1311-1316 (1992).
194. Ono,T. *et al.* Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115**, 109-121 (2003).
195. Gosling,K.M. *et al.* A mutation in a chromosome condensin II subunit, kleisin beta, specifically disrupts T cell development. *Proc. Natl. Acad. Sci. U. S. A* **104**, 12445-12450 (2007).

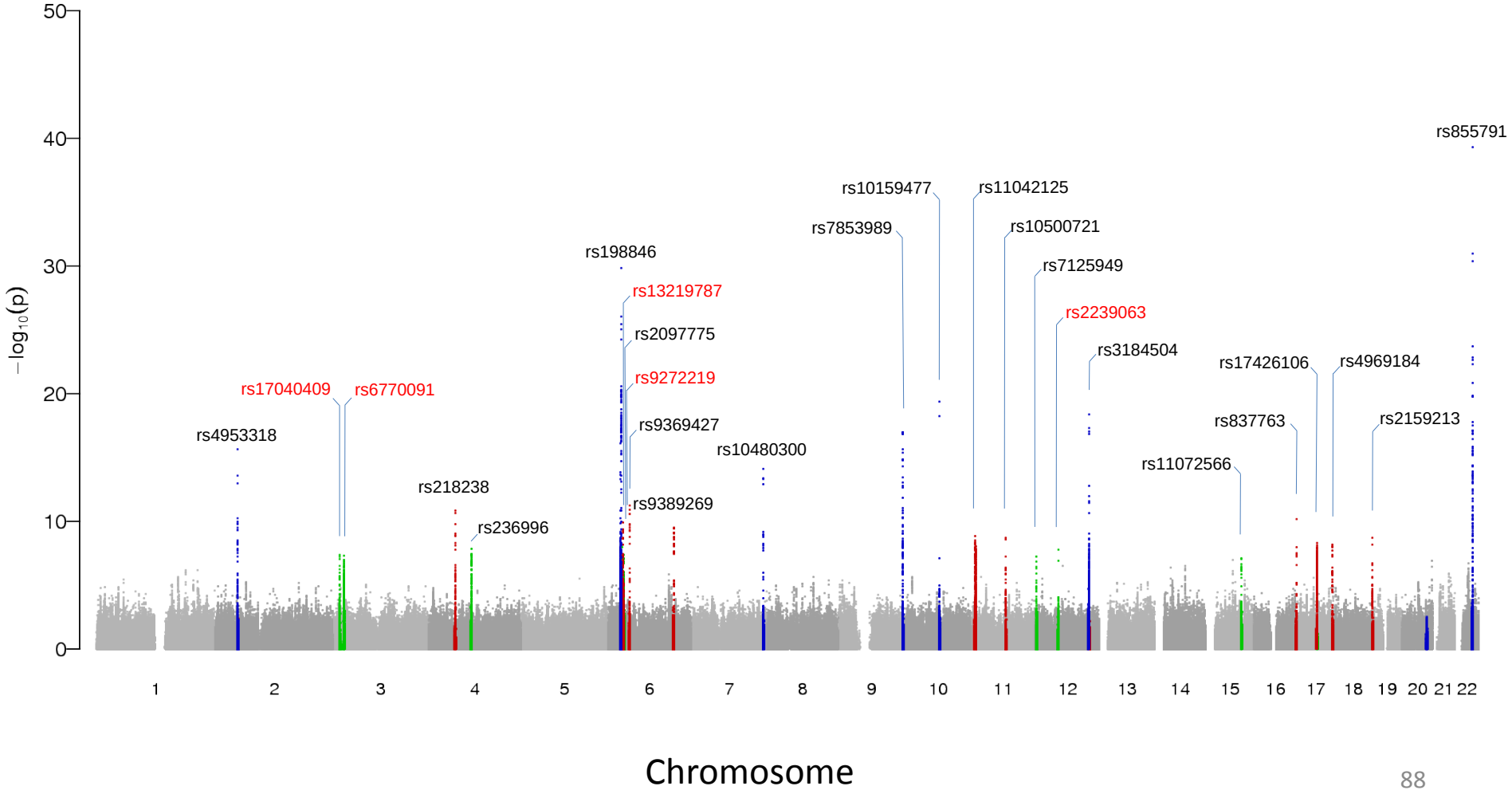
196. Glerum,D.M., Shtanko,A., & Tzagoloff,A. SCO1 and SCO2 act as high copy suppressors of a mitochondrial copper recruitment defect in Saccharomyces cerevisiae. *J Biol Chem.* **271**, 20531-20535 (1996).
197. Verdijk,R.M. *et al.* Phenotypic consequences of a novel SCO2 gene mutation. *Am. J Med. Genet A* **146A**, 2822-2827 (2008).

Supplementary Figure 1. Study design.

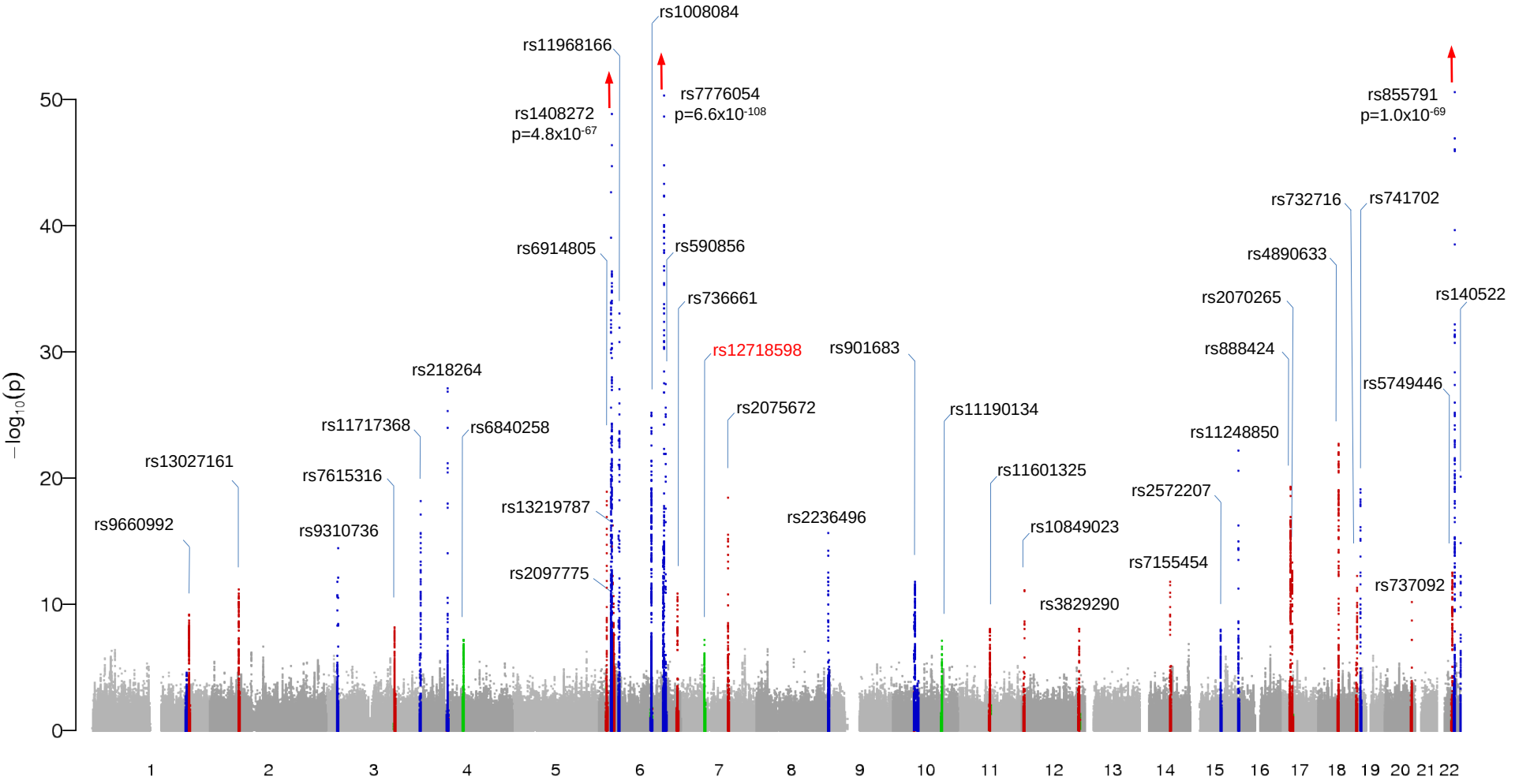


Supplementary Figure 2. SF2.1 to SF2.6: Manhattan plots showing results for genome-wide association with red blood cell traits amongst Europeans. SNPs reaching genome-wide significance ($P<1x10^{-8}$) are coloured red (novel loci) or blue (previously reported loci). SNPs coloured green are at loci which reached $P>1x10^{-8}$ but $P<1x10^{-7}$, and that were carried forward for further testing.

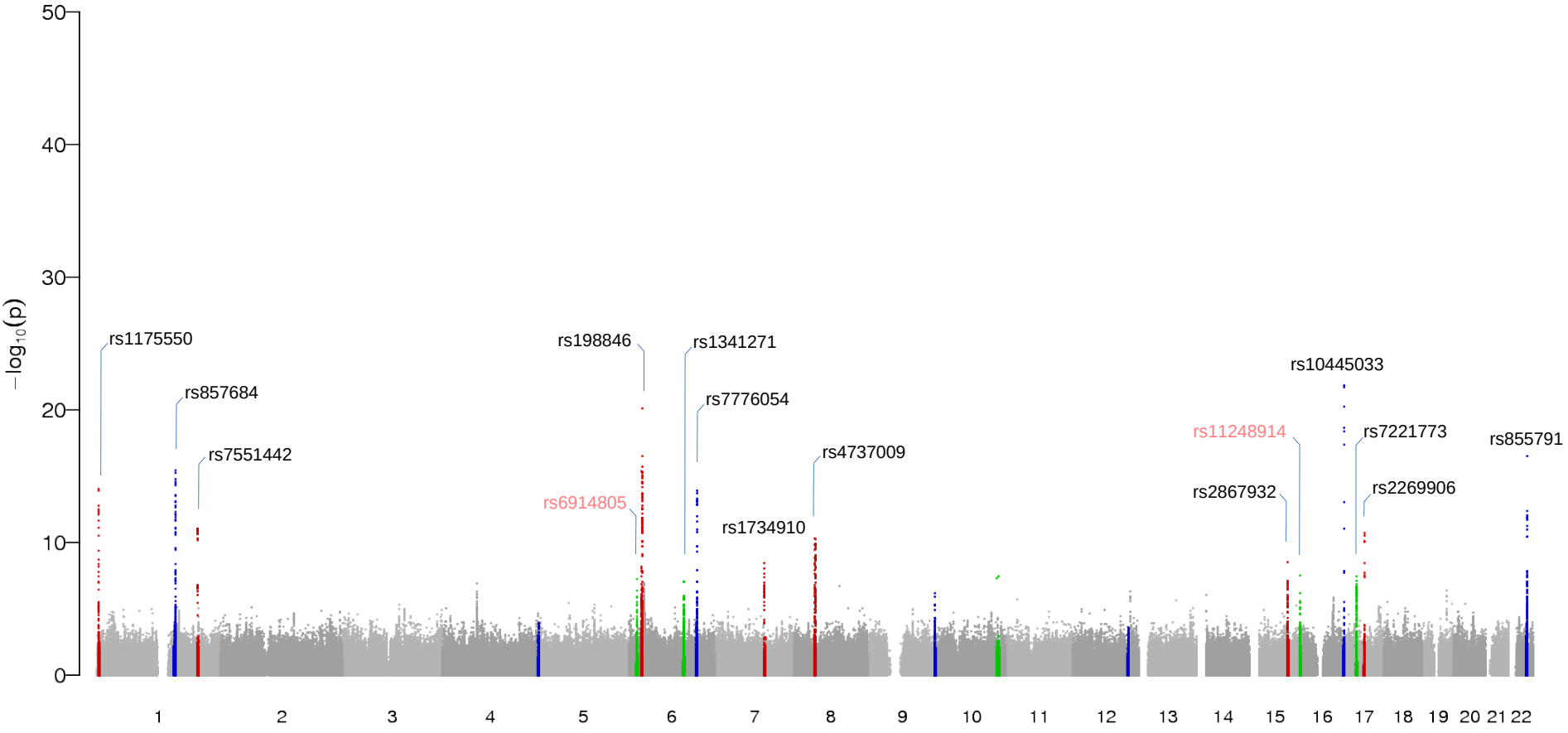
SF2.1: Haemoglobin



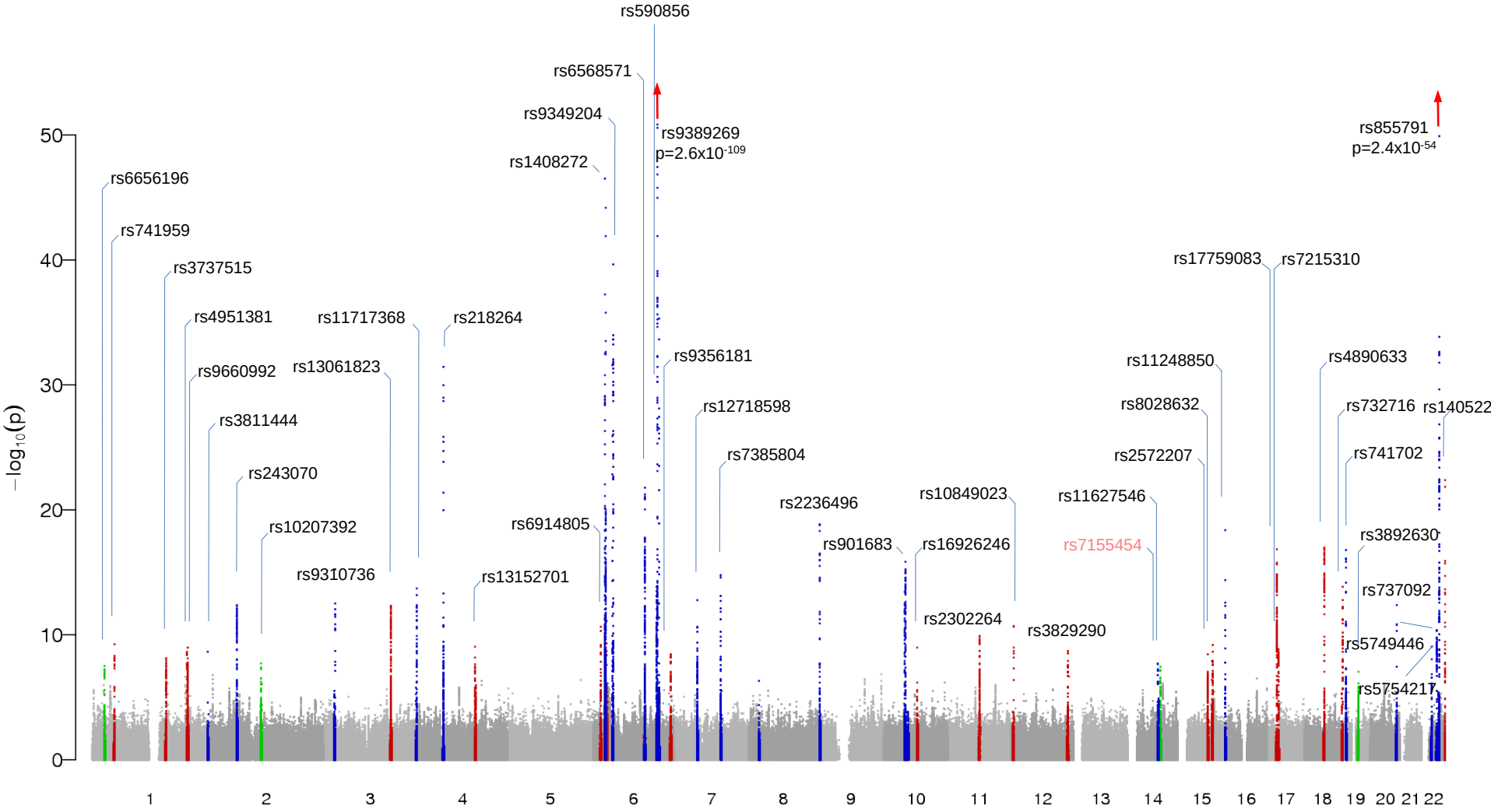
SF2.2: MCH


rs1008084
rs11968166
rs7776054
p=6.6x10-108
rs1408272
p=4.8x10-67
rs855791
p=1.0x10-69
rs732716
rs741702
rs4890633
rs6914805
rs590856
rs2070265
rs140522
rs736661
rs888424
rs12718598
rs901683
rs5749446
rs218264
rs2075672
rs11190134
rs11717368
rs6840258
rs11248850
rs13027161
rs11601325
rs7615316
rs2572207
rs13219787
rs2236496
rs9660992
rs10849023
rs9310736
rs7155454
rs2097775
rs737092
rs3829290
−log10(p)
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

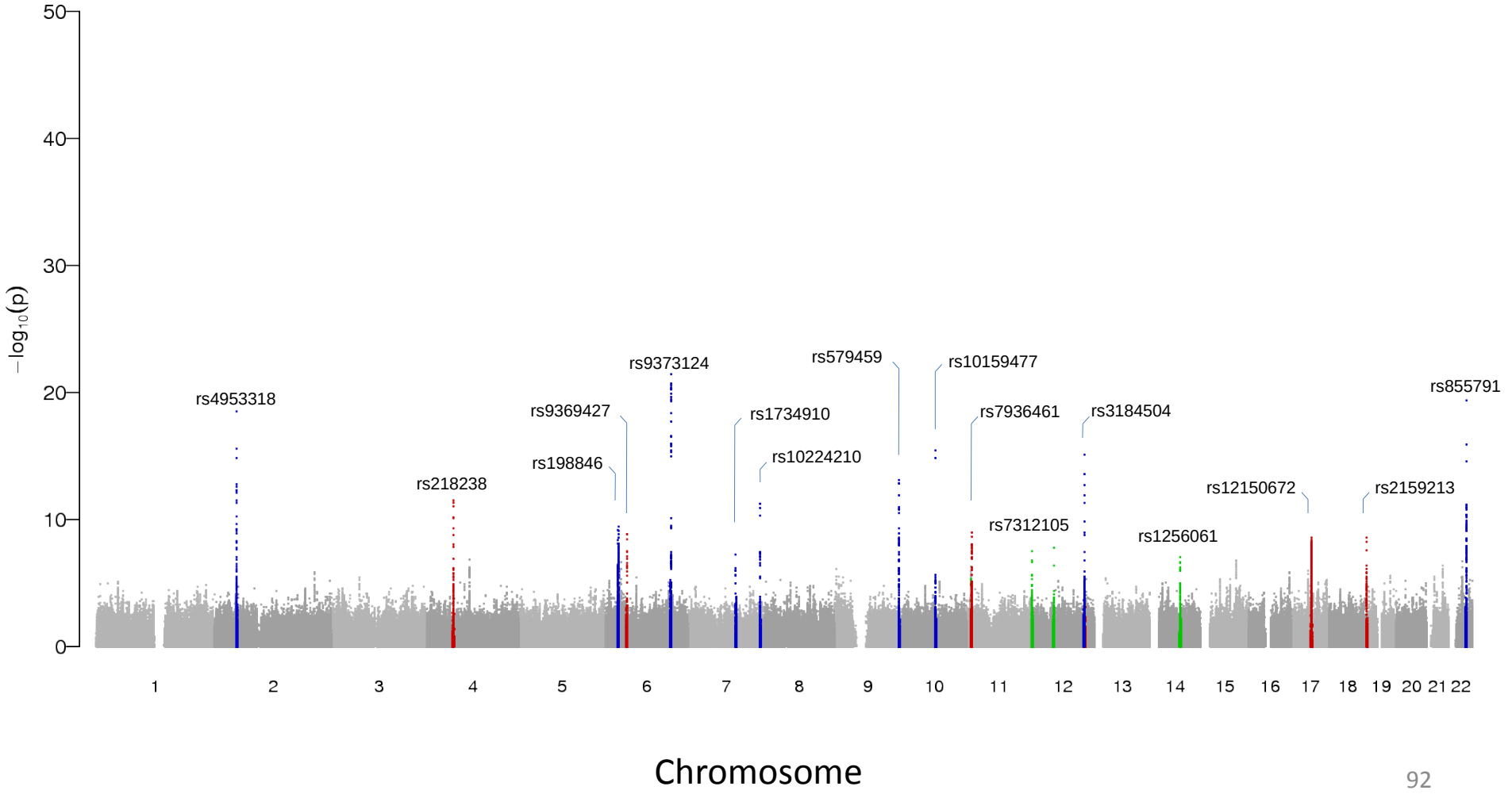
SF2.3: MCHC


50
40
30
20
10
0
$-\log_{10}(p)$
rs1175550
rs857684
rs7551442
rs198846
rs1341271
rs7776054
rs6914805
rs1734910
rs4737009
rs10445033
rs11248914
rs7221773
rs855791
rs2867932
rs2269906
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

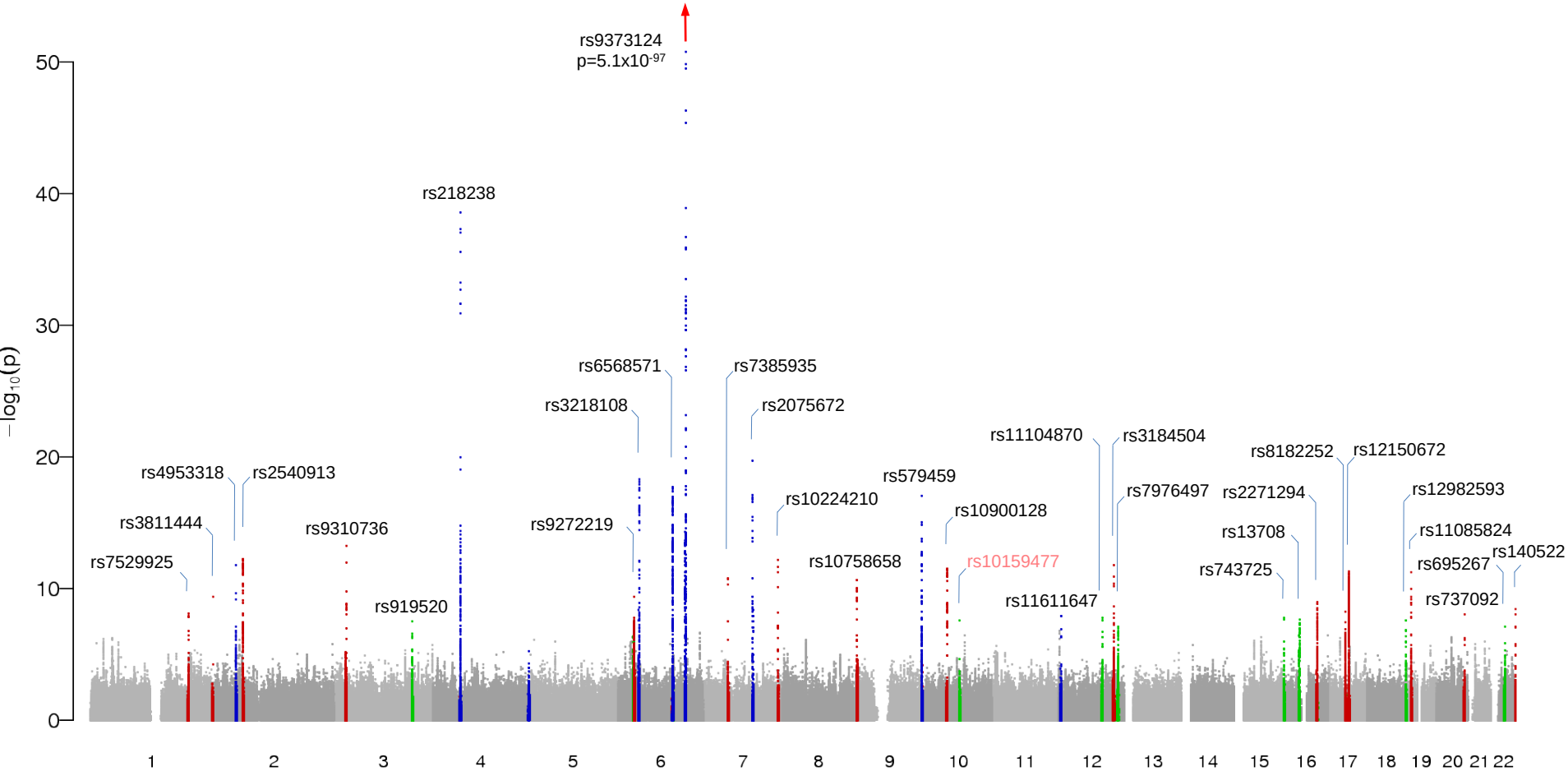
SF2.4: MCV


rs590856
rs6568571
rs9349204
rs9389269
p=2.6x10-109
rs1408272
rs855791
p=2.4x10-54
rs6656196
rs741959
rs3737515
rs4951381
rs11717368
rs218264
rs9660992
rs13061823
rs9356181
rs3811444
rs12718598
rs7385804
rs243070
rs10207392
rs2236496
rs10849023
rs11627546
rs6914805
rs901683
rs16926246
rs7155454
rs9310736
rs13152701
rs2302264
rs3829290
rs17759083
rs7215310
rs11248850
rs4890633
rs8028632
rs732716
rs140522
rs2572207
rs741702
rs3892630
rs737092
rs5749446
rs5754217
−log10(p)
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

SF2.5: PCV

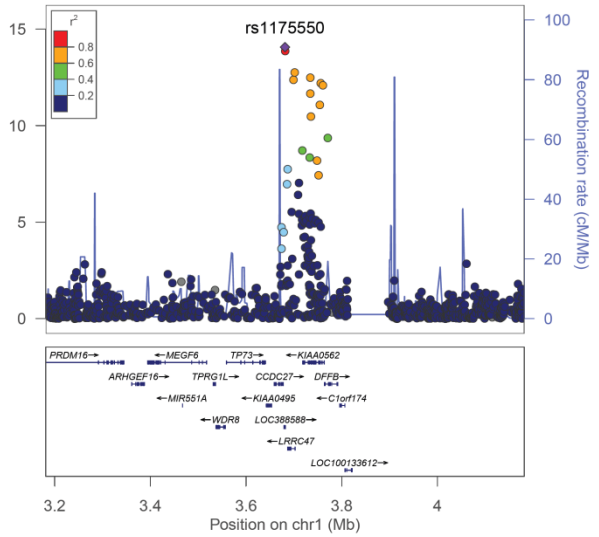

50
40
30
20
10
0
−log10(p)
rs4953318
rs218238
rs9369427
rs198846
rs9373124
rs1734910
rs10224210
rs579459
rs10159477
rs7936461
rs3184504
rs7312105
rs1256061
rs12150672
rs2159213
rs855791
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

SF2.6: RBC

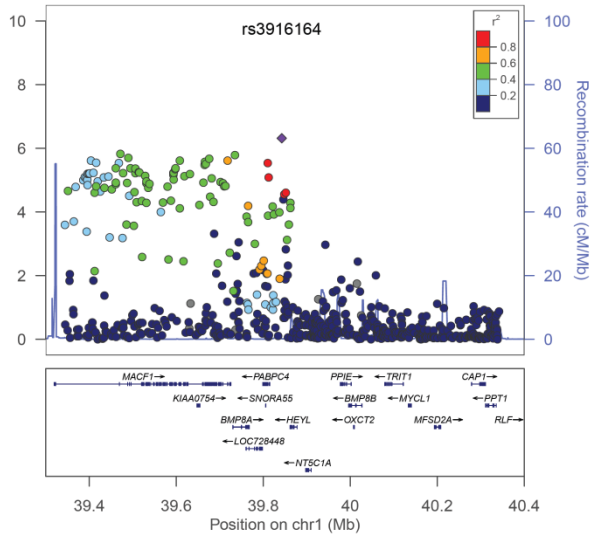

rs9373124
p=5.1x10-97
rs218238
rs6568571
rs7385935
rs3218108
rs2075672
rs11104870
rs3184504
rs8182252
rs12150672
rs4953318
rs2540913
rs579459
rs7976497
rs2271294
rs12982593
rs3811444
rs9310736
rs9272219
rs10224210
rs10900128
rs13708
rs11085824
rs7529925
rs10758658
rs10159477
rs743725
rs695267
rs140522
rs919520
rs11611647
rs737092
−log10(p)
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Supplementary Figure 3. SF3.1 to 3.75: Regional plots for the red blood cell phenotype sentinel SNPs. At each region pairwise LD with the sentinel SNP is indicated.

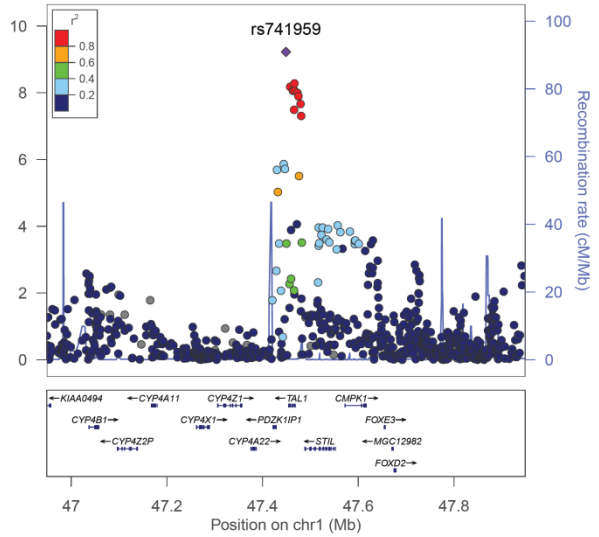
SF3.1: 1p36 - rs1175550 - MCHC



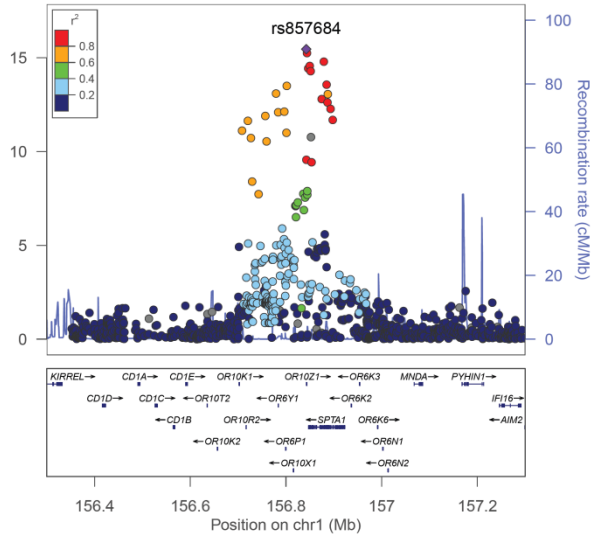
SF3.2: 1p34 - rs3916164 - MCH



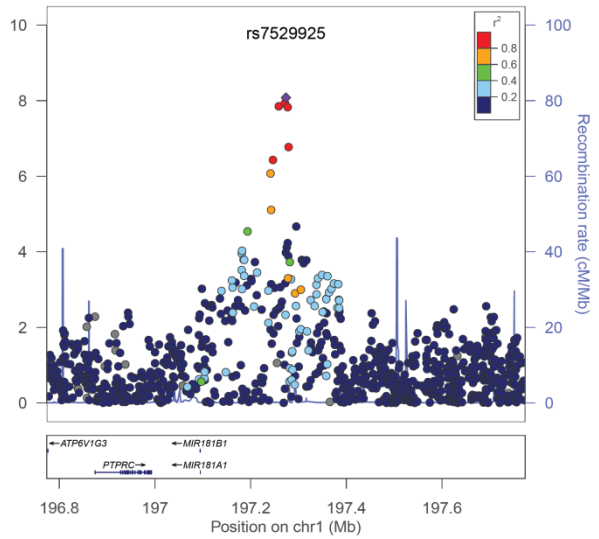
SF3.3: 1p32 - rs741959 - MCV



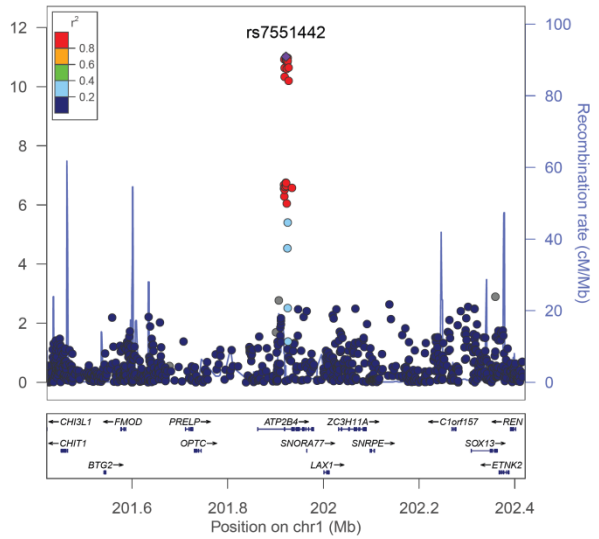
SF3.4: 1q23 - rs857684 - MCHC



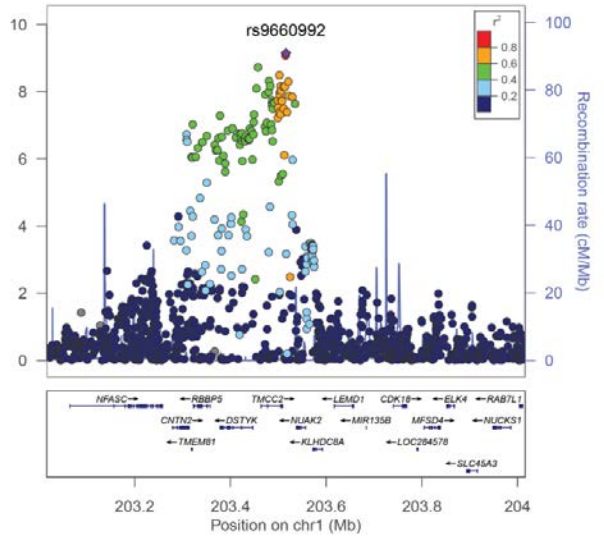
SF3.5: 1q32 - rs7529925 - RBC



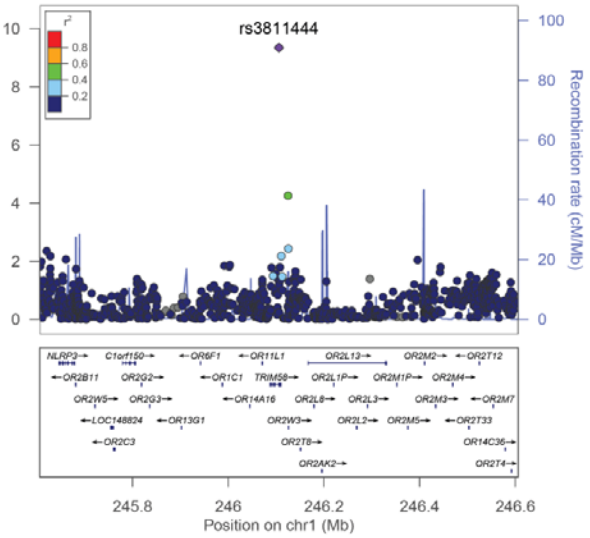
SF3.6: 1q32 - rs7551442 - MCHC



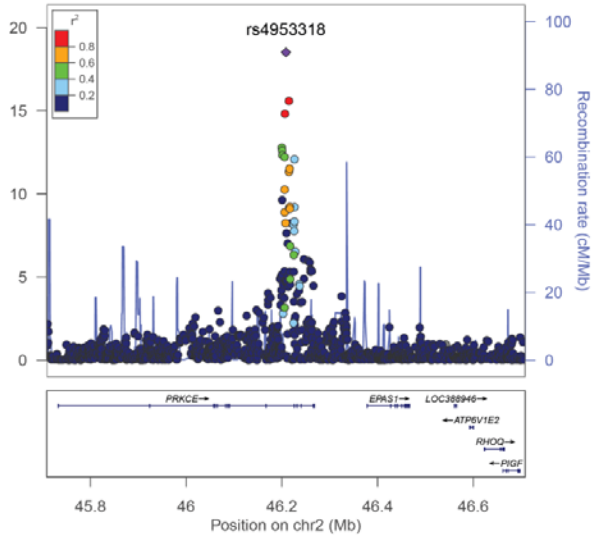
SF3.7: 1q32 - rs9660992 - MCH



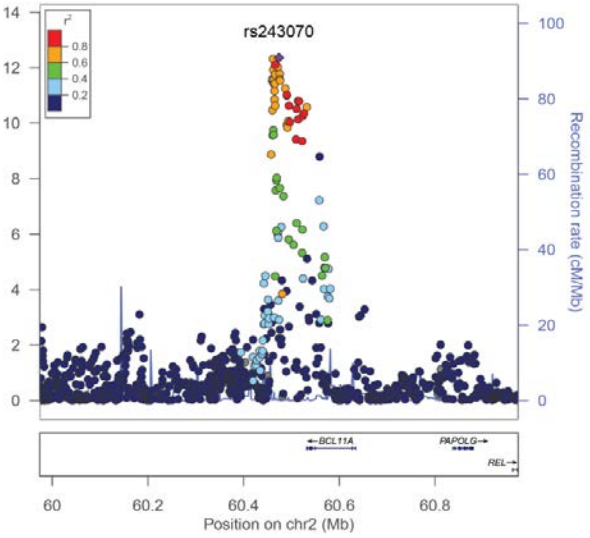
SF3.8: 1q44 - rs3811444 - RBC



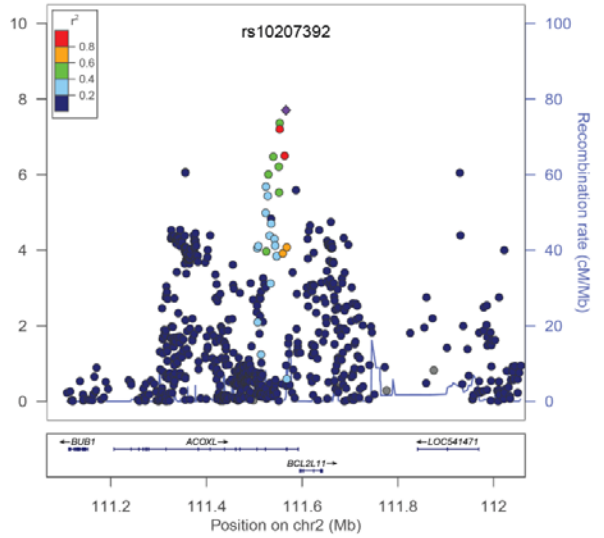
SF3.9: 2p21 - rs4953318 - PCV



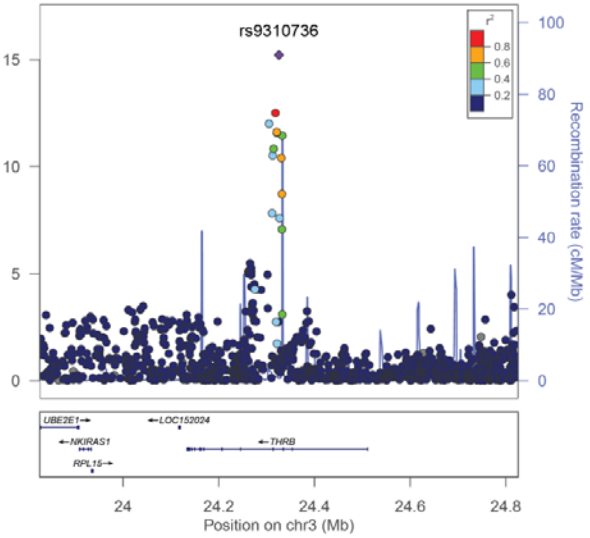
SF3.10: 2p16 - rs243070 - MCV



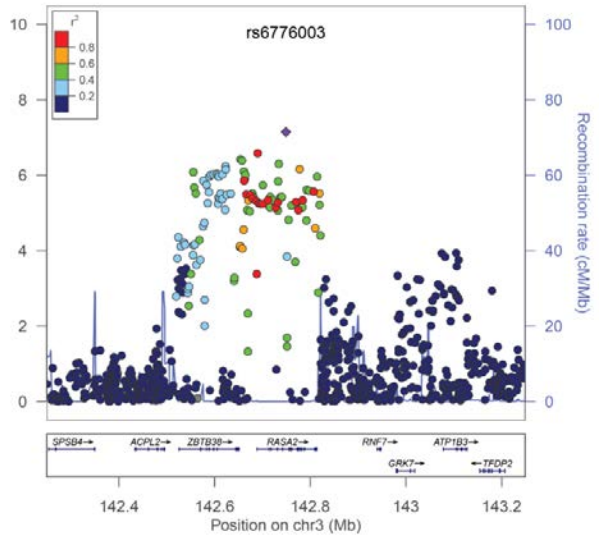
SF3.11: 2q13 - rs10207392 - MCV



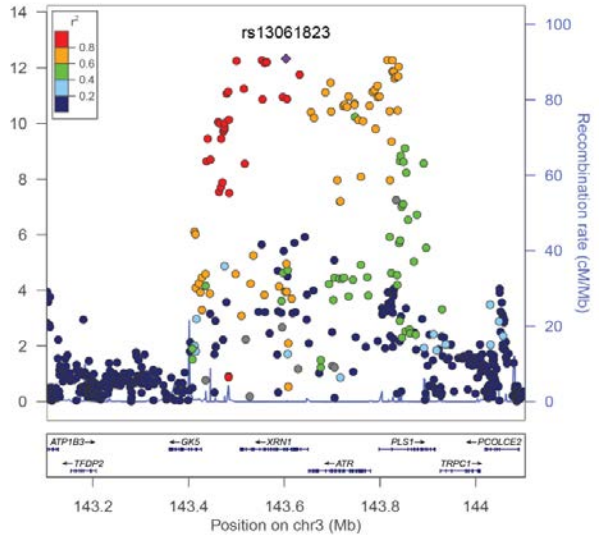
SF3.12: 3p24 - rs9310736 - MCV



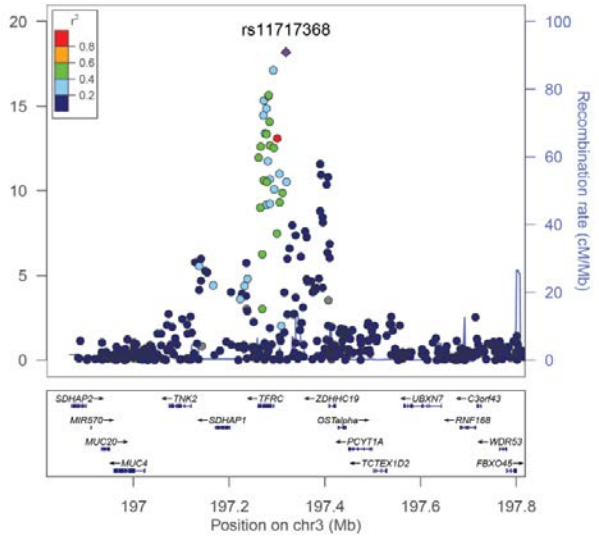
SF3.13: 3q22 - rs6776003 - MCV



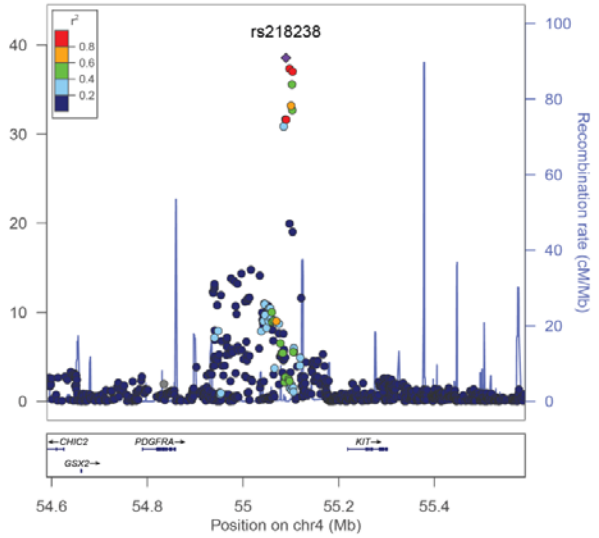
SF3.14: 3q23 - rs13061823 - MCV



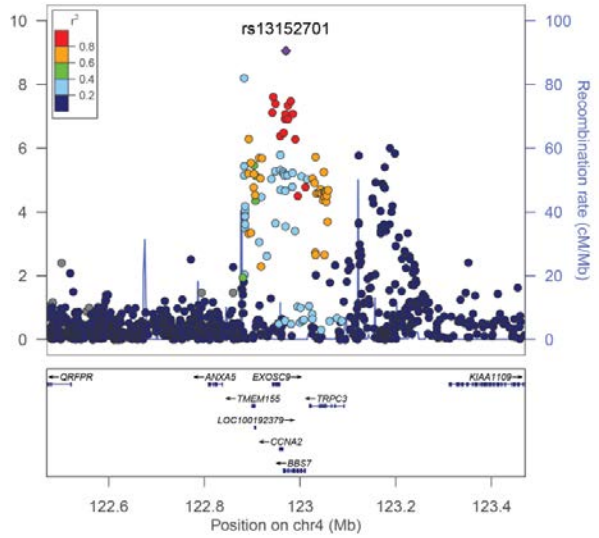
SF3.15: 3q29 - rs11717368 - MCH



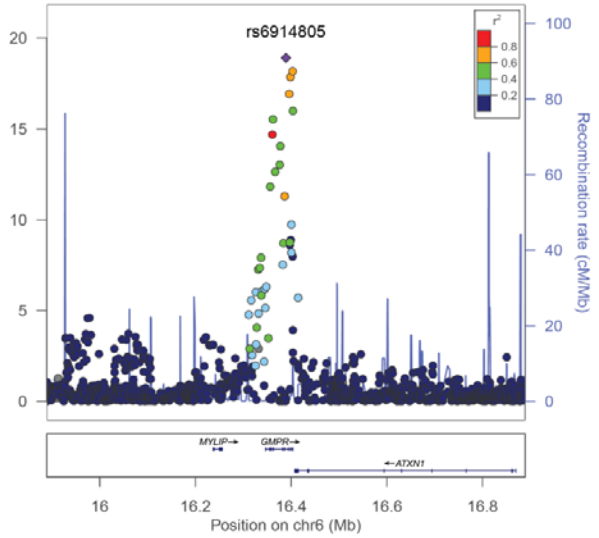
SF3.16: 4q11 - rs218238 - RBC



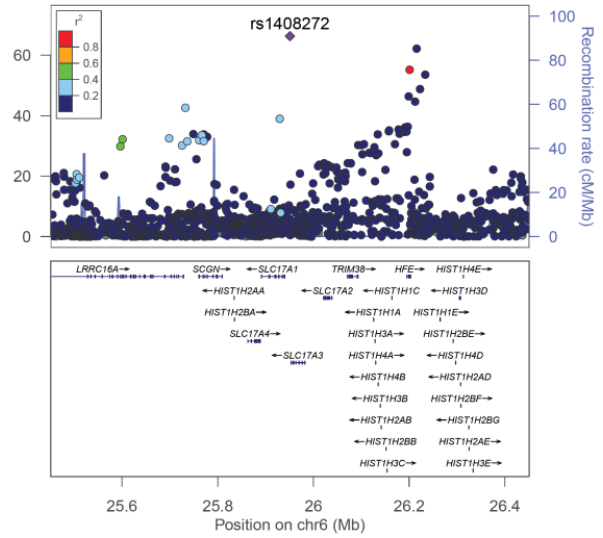
SF3.17: 4q27 - rs13152701 - MCV



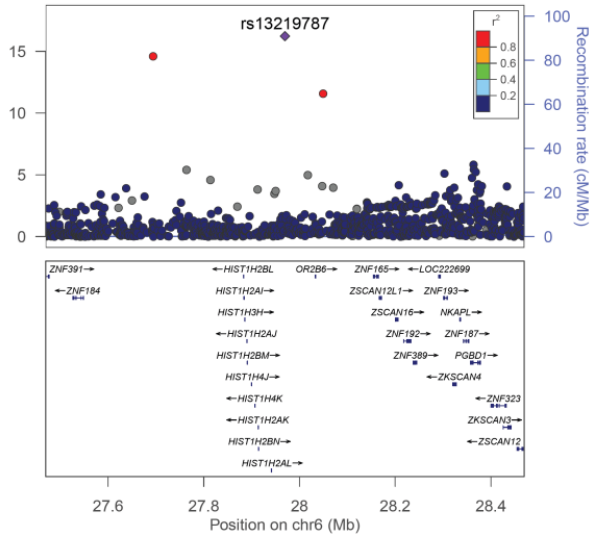
SF3.18: 6p23 - rs6914805 - MCH



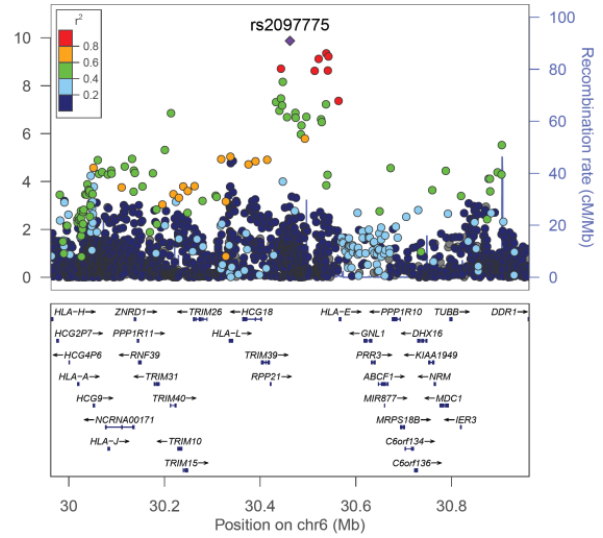
SF3.19: 6p21 - rs1408272 - MCH



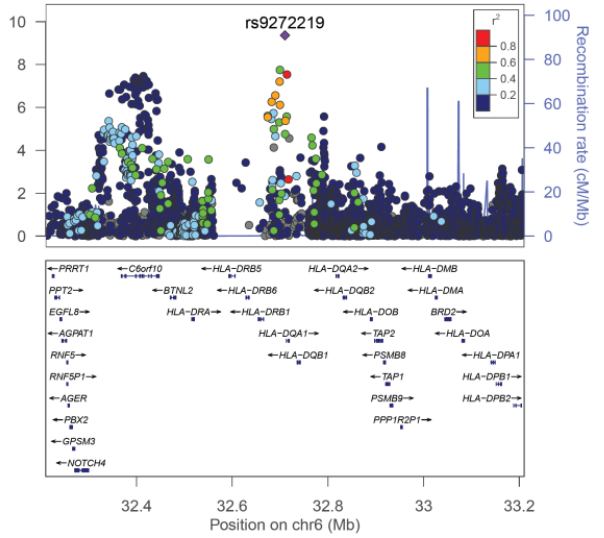
SF3.20: 6p22 - rs13219787 - MCH



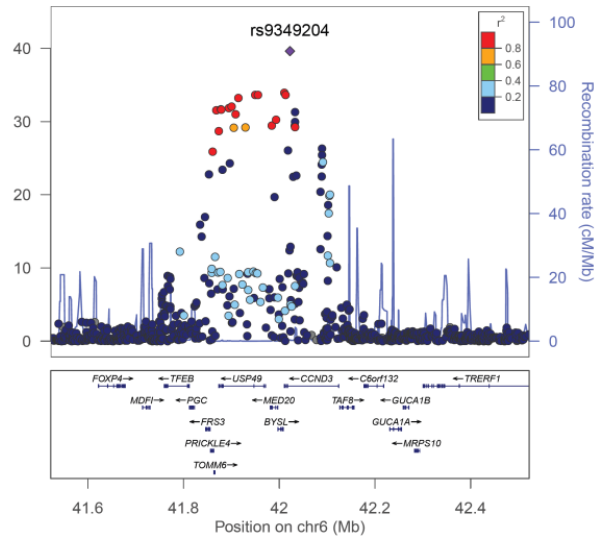
SF3.21: 6p22 - rs2097775 - HB



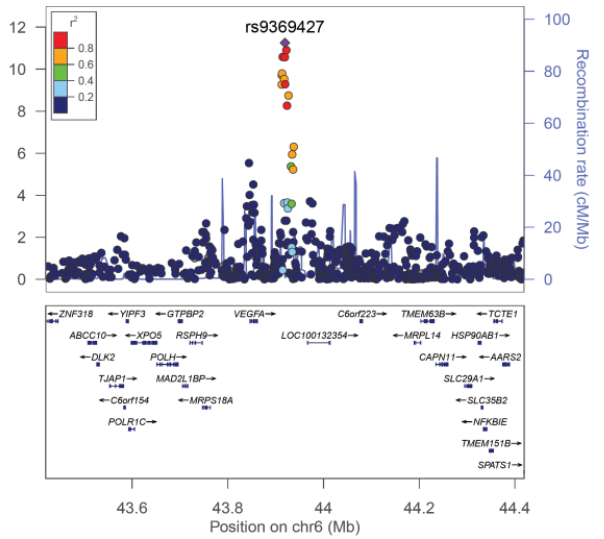
SF3.22: 6p21 - rs9272219 - RBC



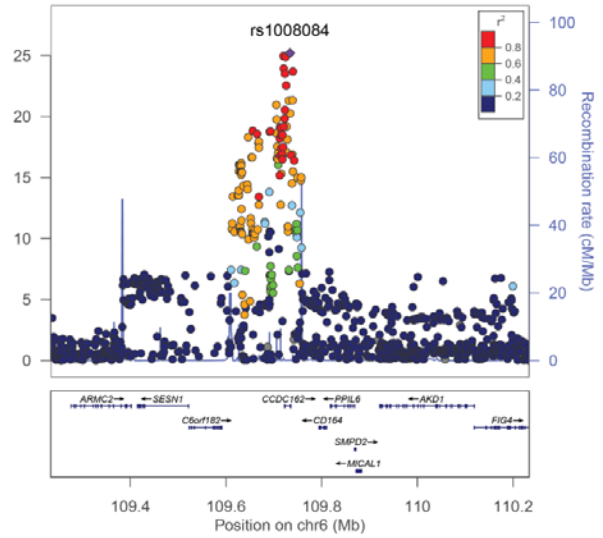
SF3.23: 6p21 - rs9349204 - MCV



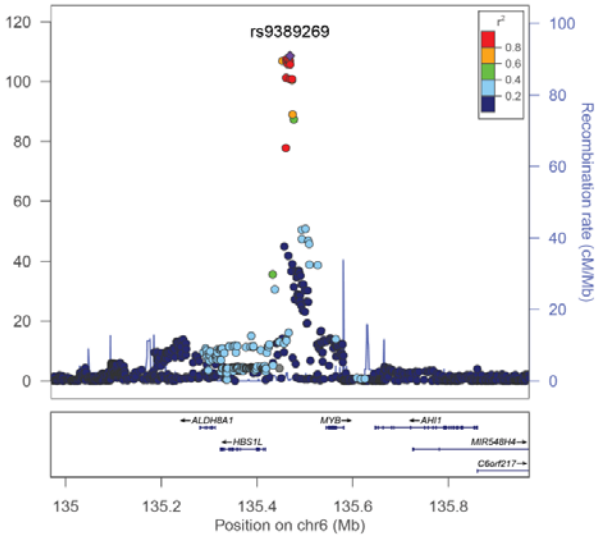
SF3.24: 6p12 - rs9369427 - HB



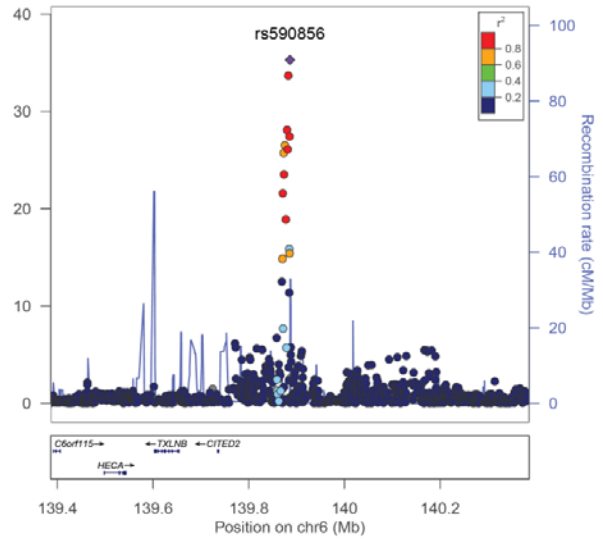
SF3.25: 6q21 - rs1008084 - MCH



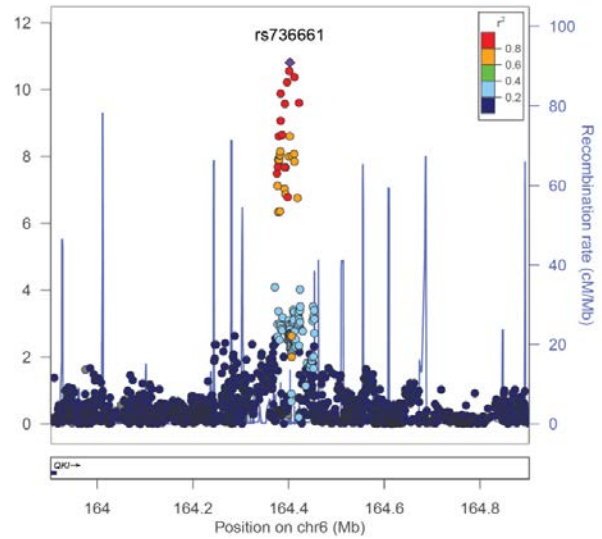
SF3.26: 6q23 - rs9389269 - MCV



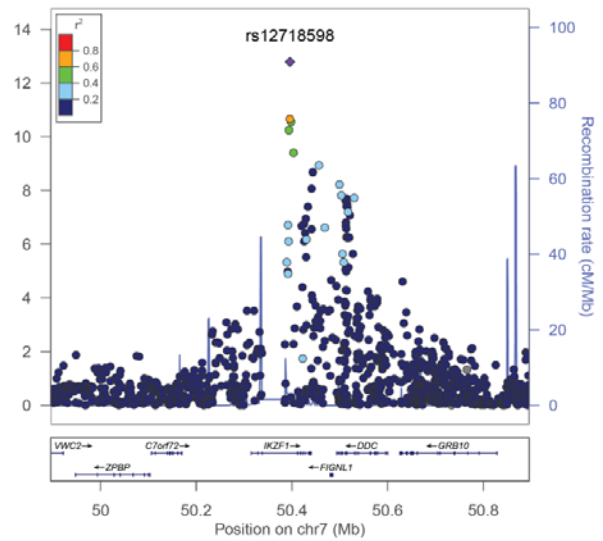
SF3.27: 6q24 - rs590856 - MCV



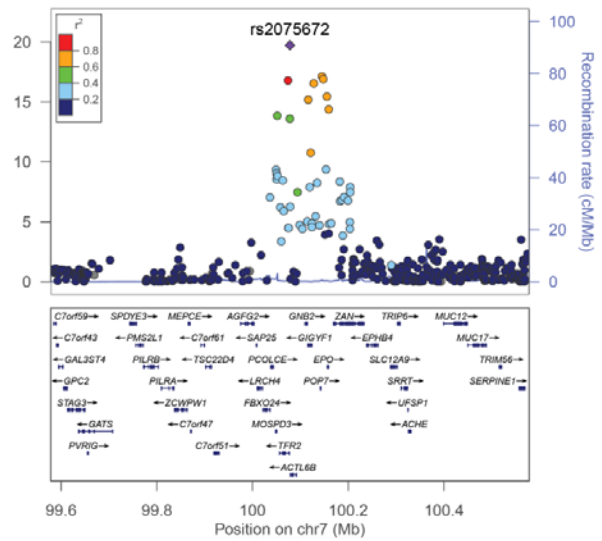
SF3.28: 6q26 - rs736661 - MCH



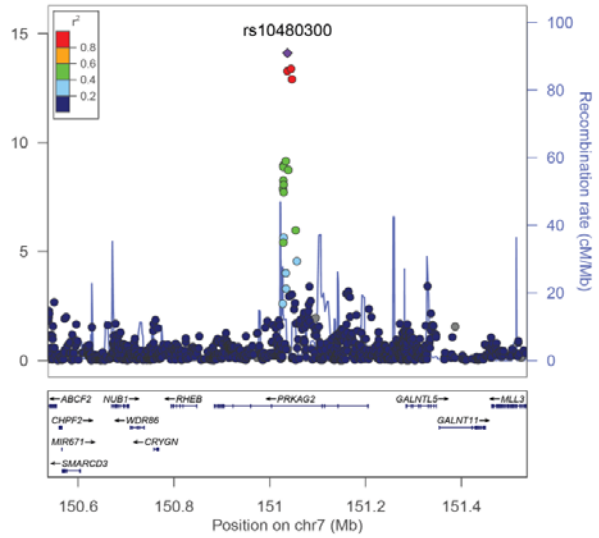
SF3.29: 7p13 - rs12718598 - MCV



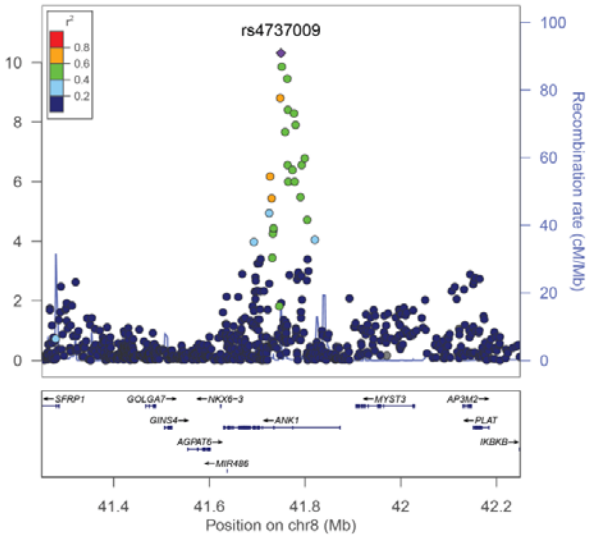
SF3.30: 7q22 - rs2075672 - RBC



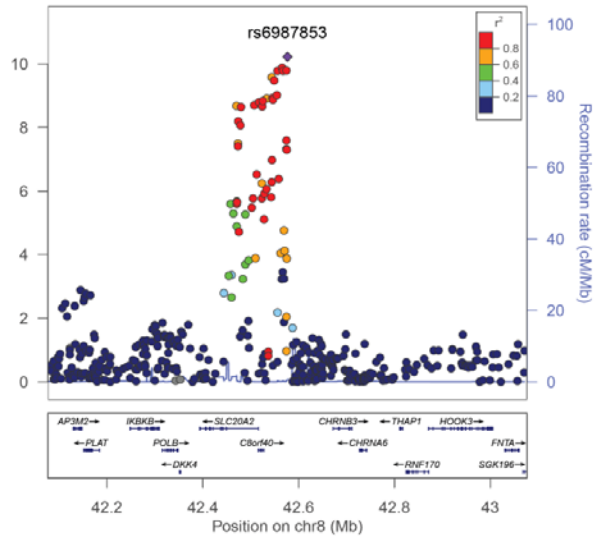
SF3.31: 7q36 - rs10480300 - HB



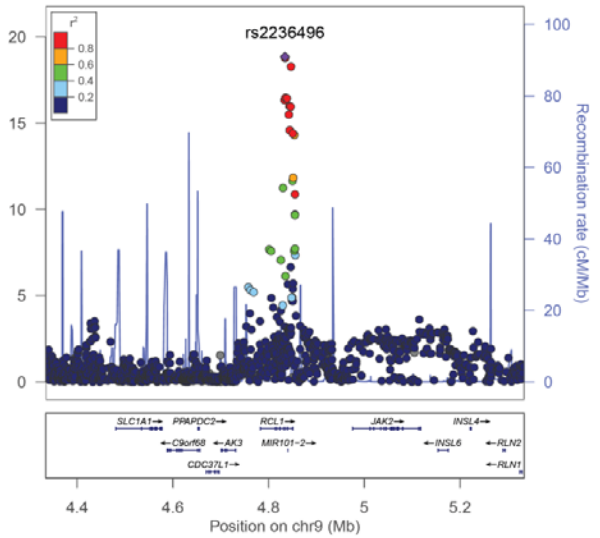
SF3.32: 8p11 - rs4737009 - MCHC



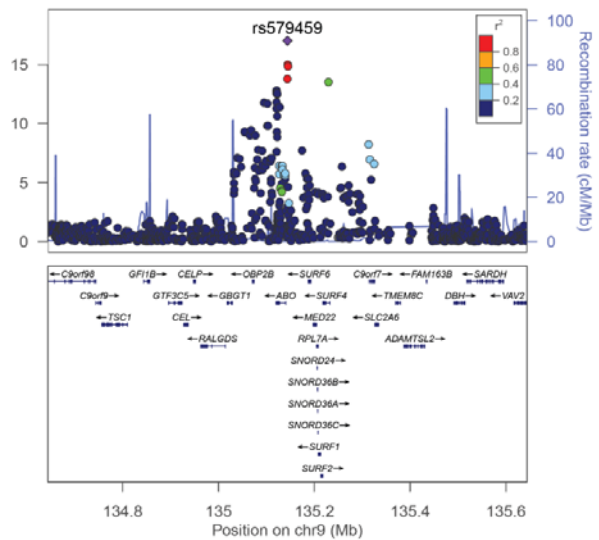
SF3.33: 8p11 - rs6987853 - MCHC



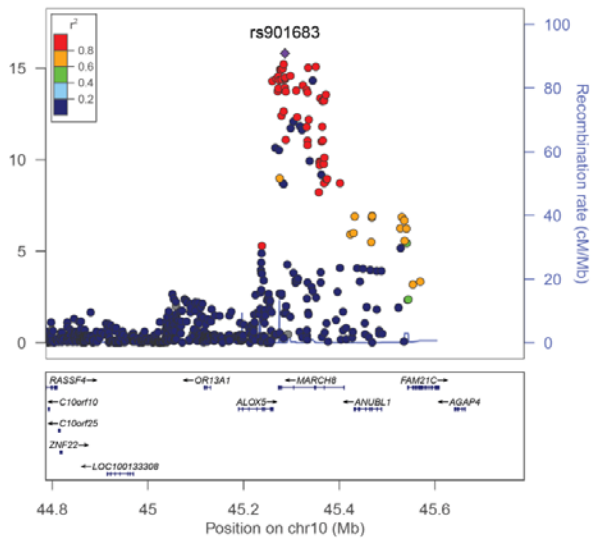
SF3.34: 9p24 - rs2236496 - MCV



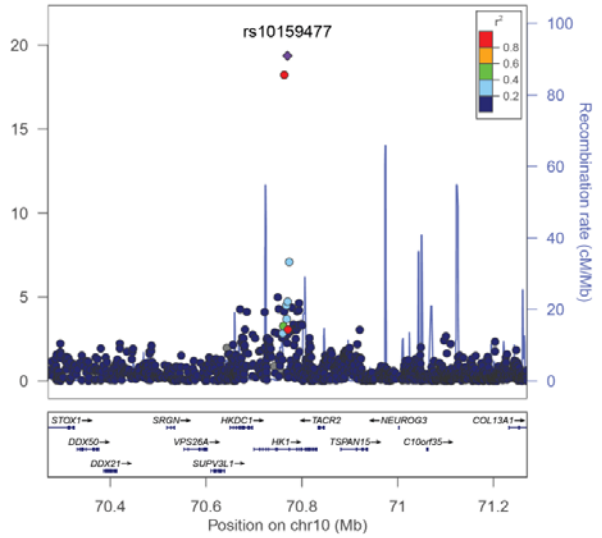
SF3.35: 9q34 - rs579459 - RBC



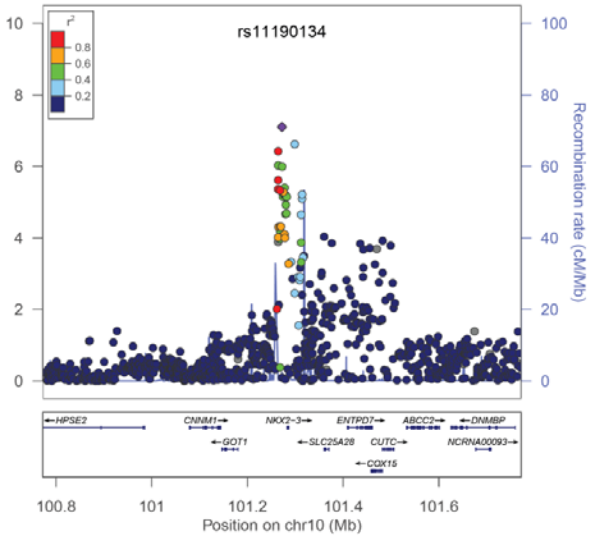
SF3.36: 10q11 - rs901683 - MCV



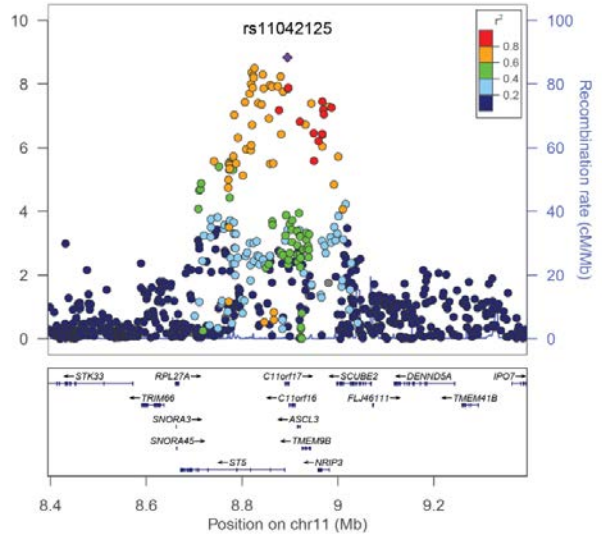
SF3.37: 10q22 - rs10159477 - HB



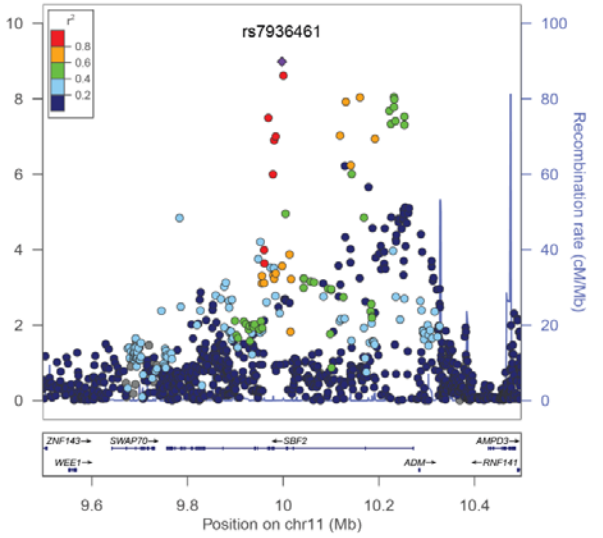
SF3.38: 10q24 - rs11190134 - MCH



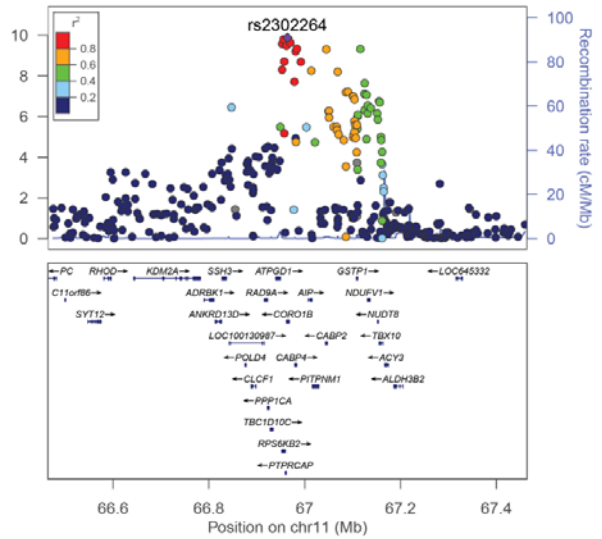
SF3.39: 11p15 - rs11042125 - HB



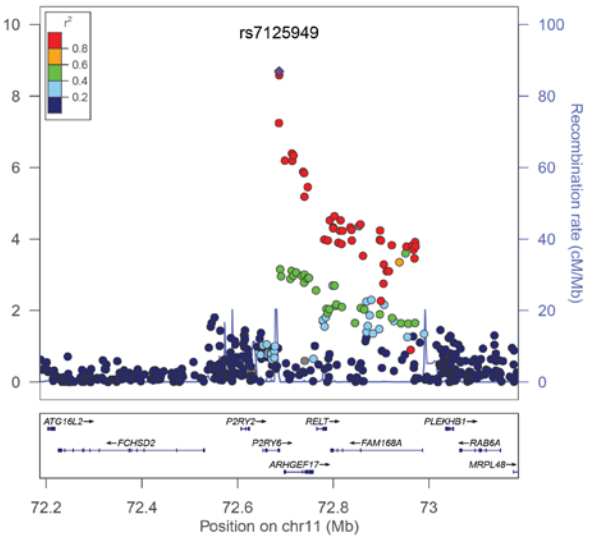
SF3.40: 11p15 - rs7936461 - PCV



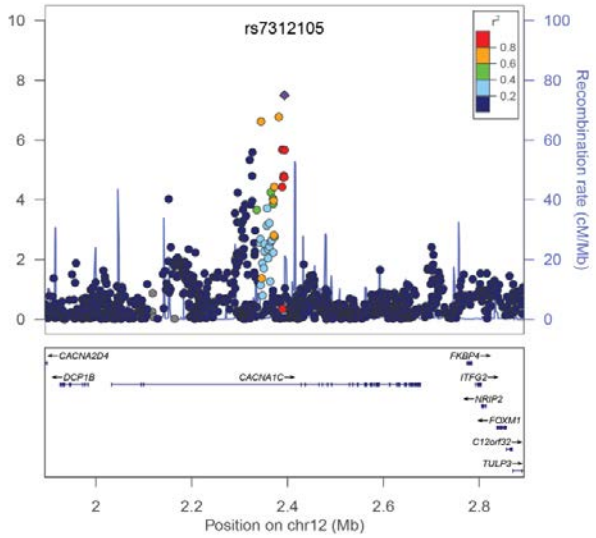
SF3.41: 11q13 - rs2302264 - MCV



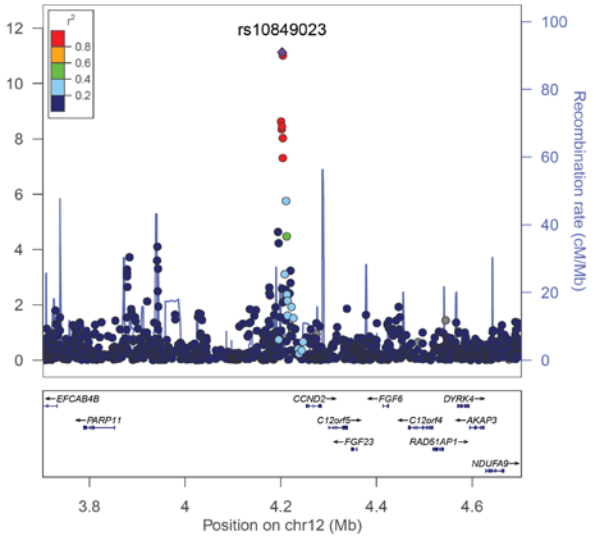
SF3.42: 11q13 - rs7125949 - HB



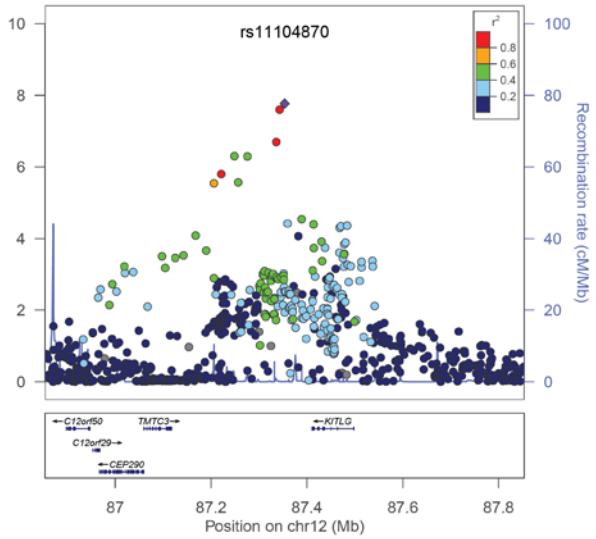
SF3.43: 12p13 - rs7312105 - PCV



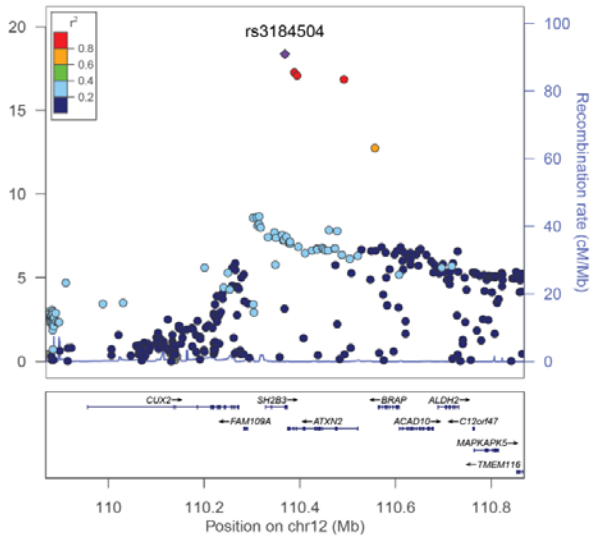
SF3.44: 12p13 - rs10849023 - MCH



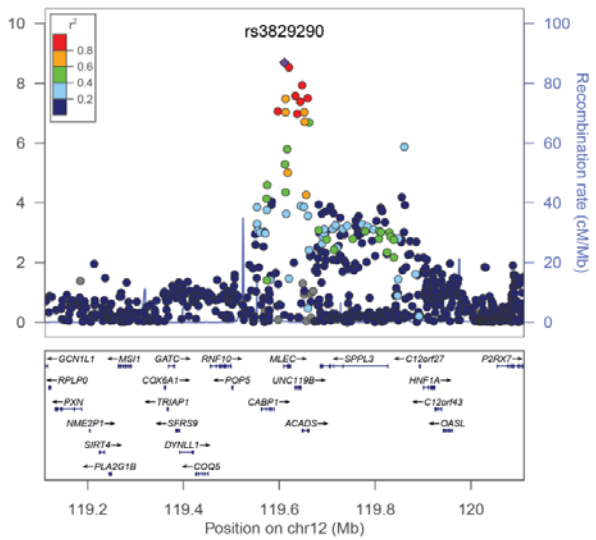
SF3.45: 12q22 - rs11104870 - RBC



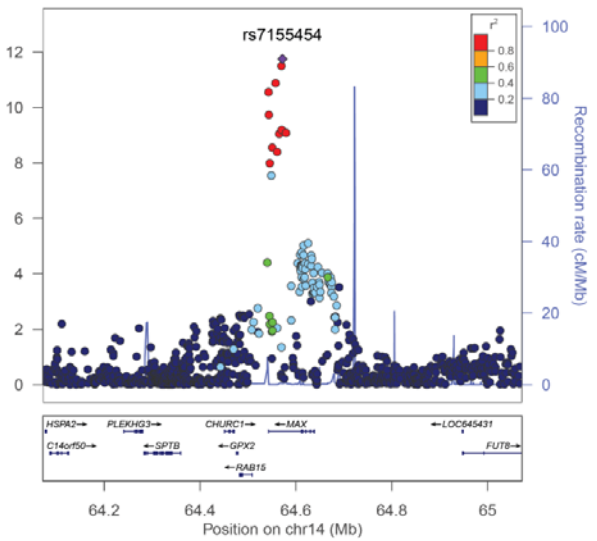
SF3.46: 12q24 - rs3184504 - HB



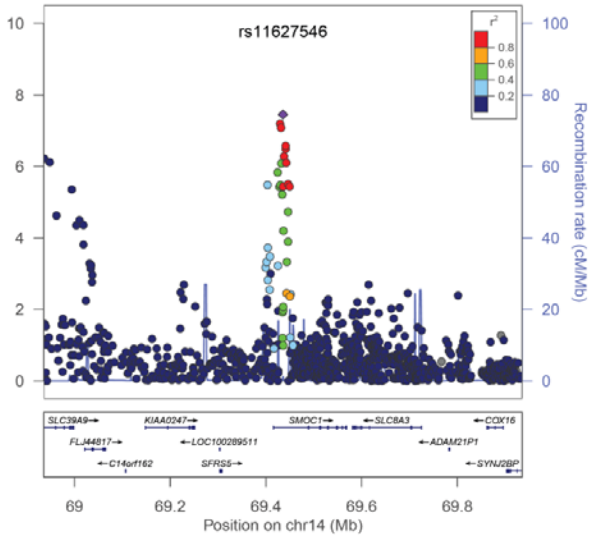
SF3.47: 12q24 - rs3829290 - MCV



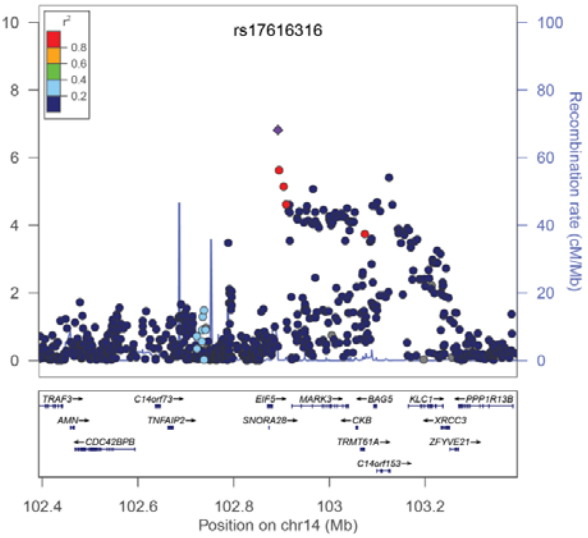
SF3.48: 14q23 - rs7155454 - MCH



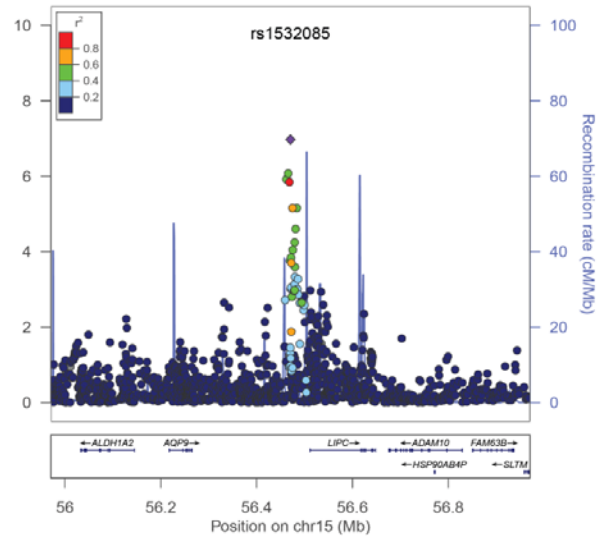
SF3.49: 14q24 - rs11627546 - MCV



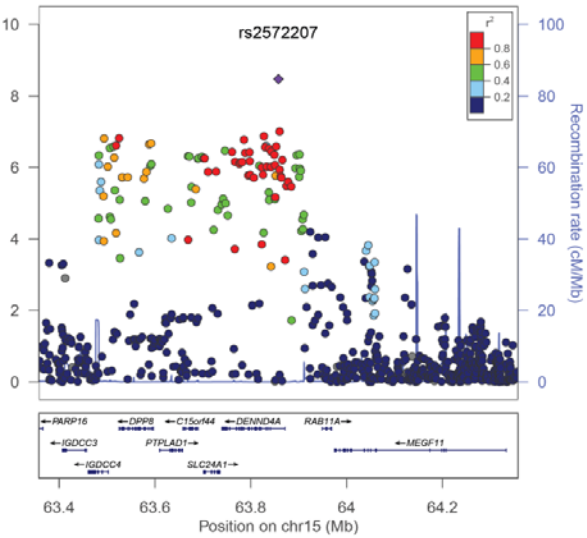
SF3.50: 14q32 - rs17616316 - MCH



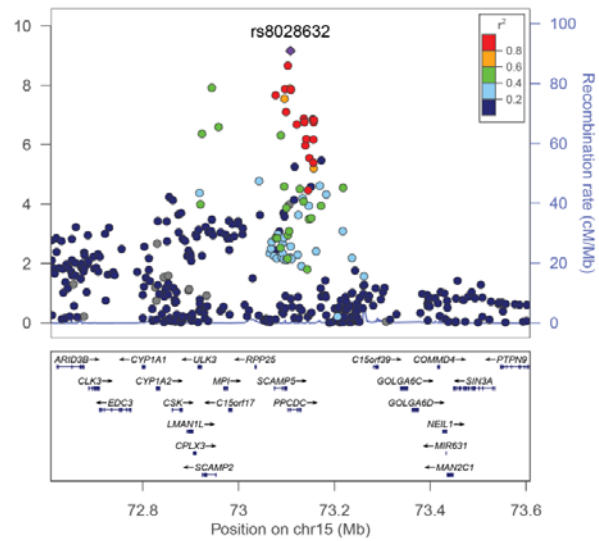
SF3.51: 15q21 - rs1532085 - HB



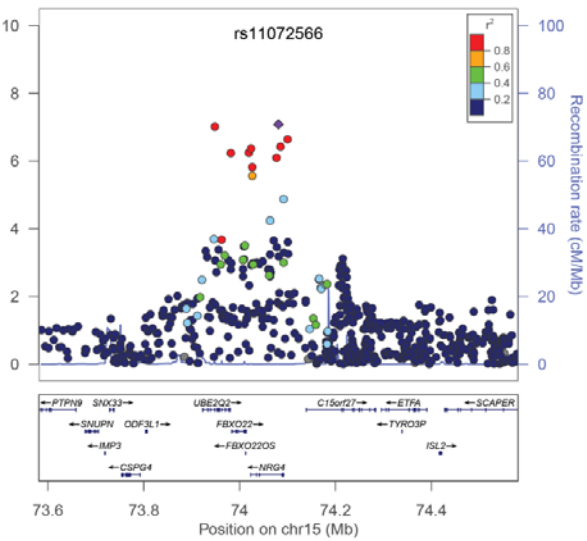
SF3.52: 15q22 - rs2572207 - MCV



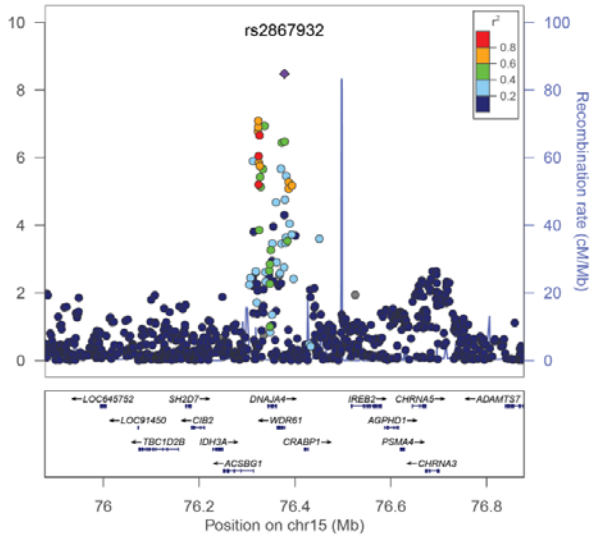
SF3.53: 15q24 - rs8028632 - MCV



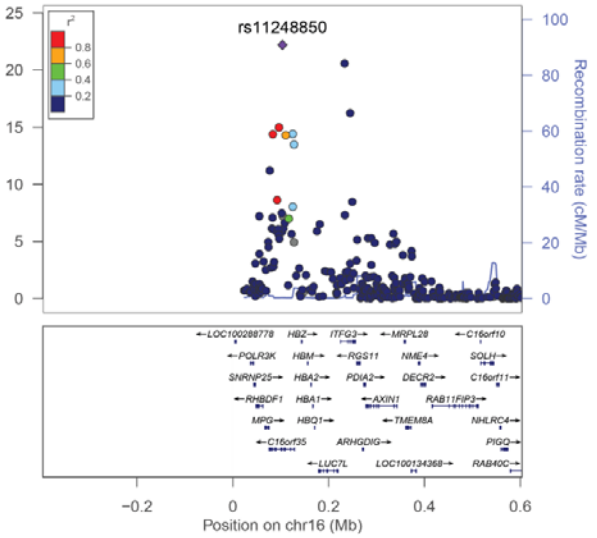
SF3.54: 15q24 - rs11072566 - HB



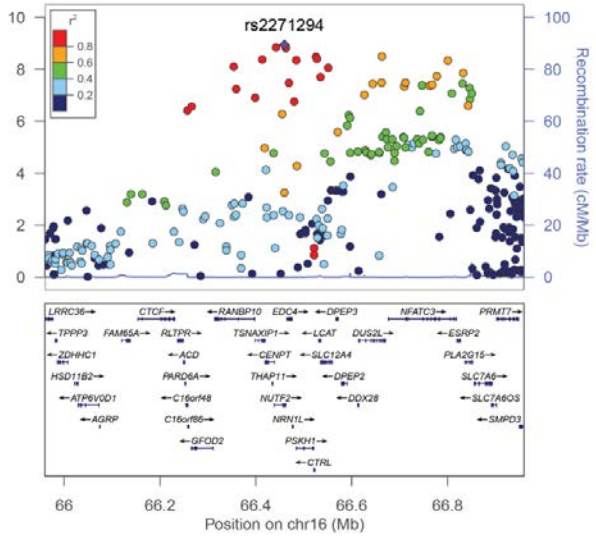
SF3.55: 15q25 - rs2867932 - MCHC



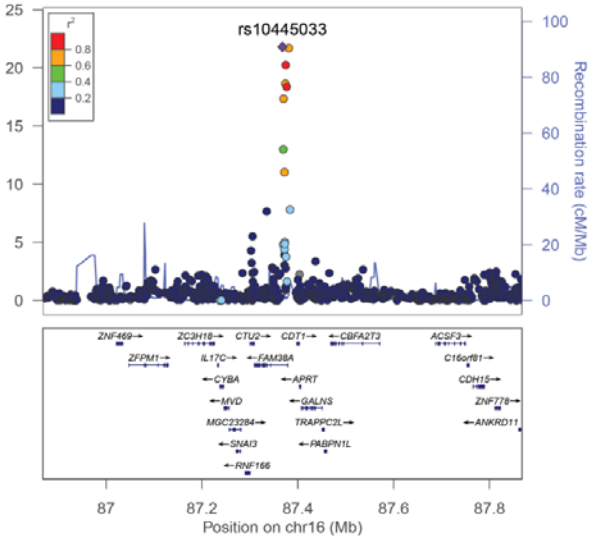
SF3.56: 16p11 - rs11248850 - MCH



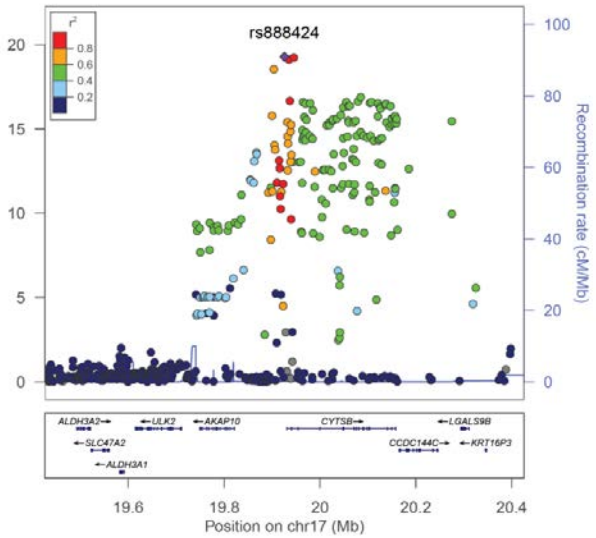
SF3.57: 16q22 - rs2271294 - RBC



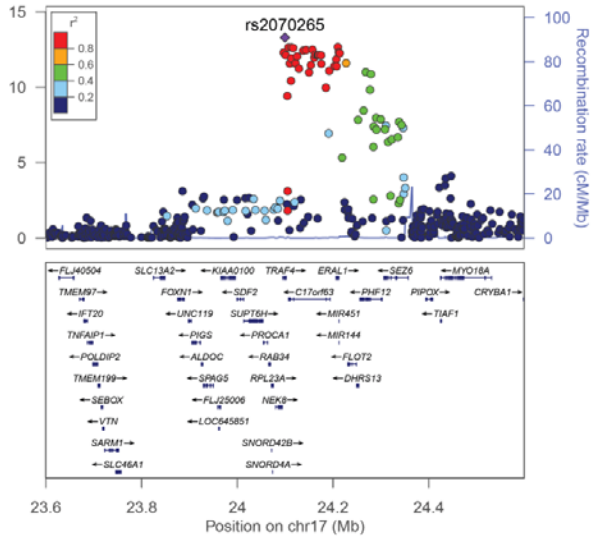
SF3.58: 16q24 - rs10445033 - MCHC



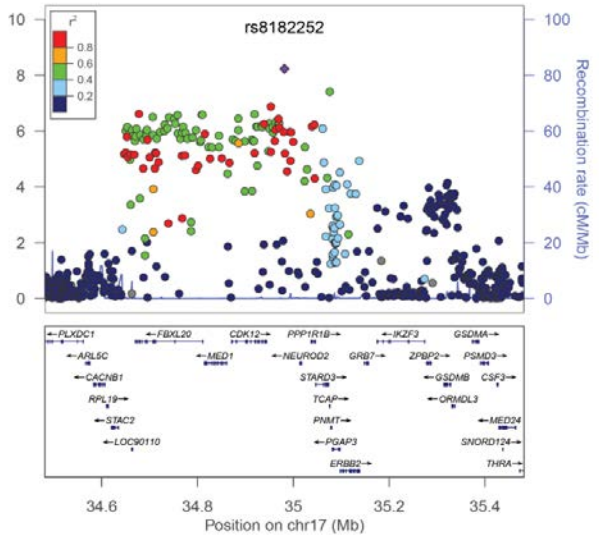
SF3.59: 17p11 - rs888424 - MCH



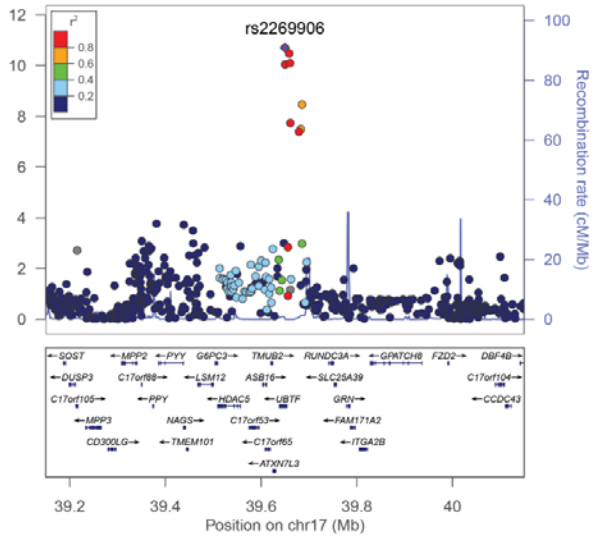
SF3.60: 17q11 - rs2070265 - MCH



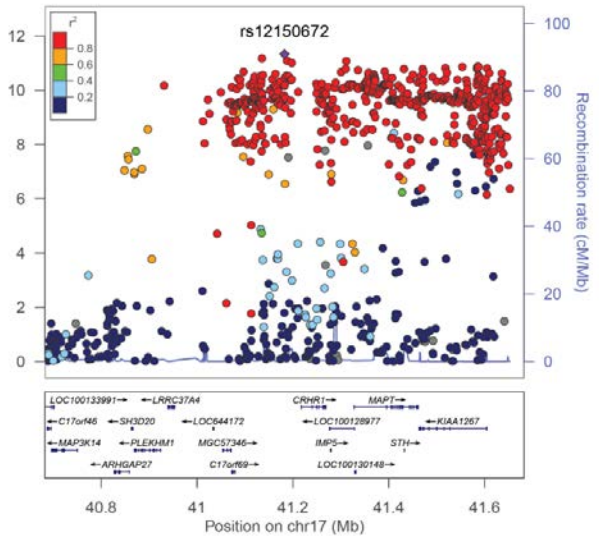
SF3.61: 17q12 - rs8182252 - RBC



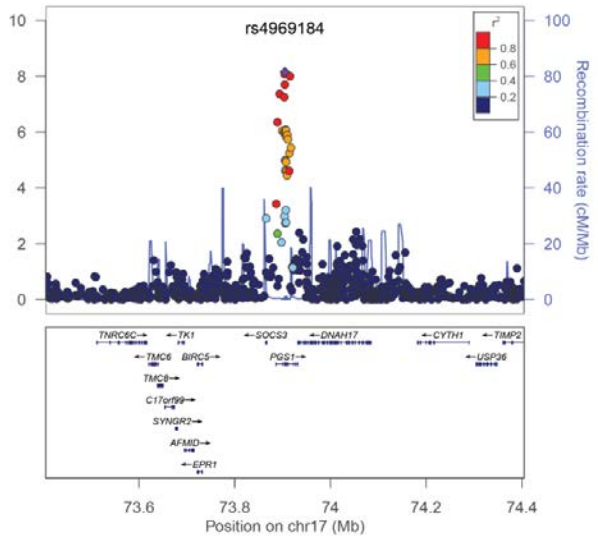
SF3.62: 17q21 - rs2269906 - MCHC



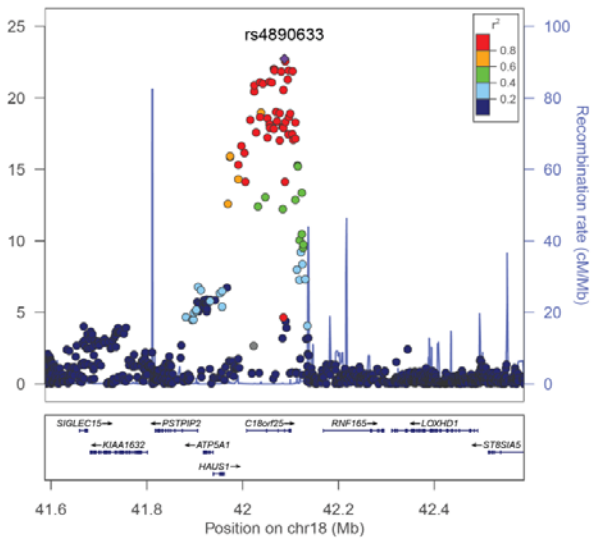
SF3.63: 17q12 - rs12150672 - RBC



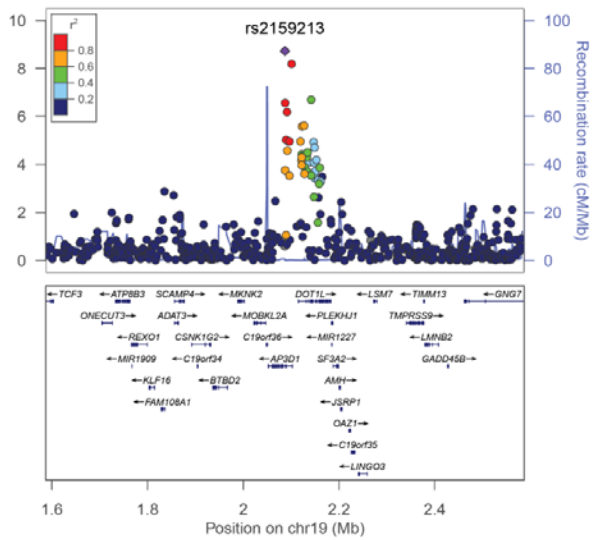
SF3.64: 17q25 - rs4969184 - HB



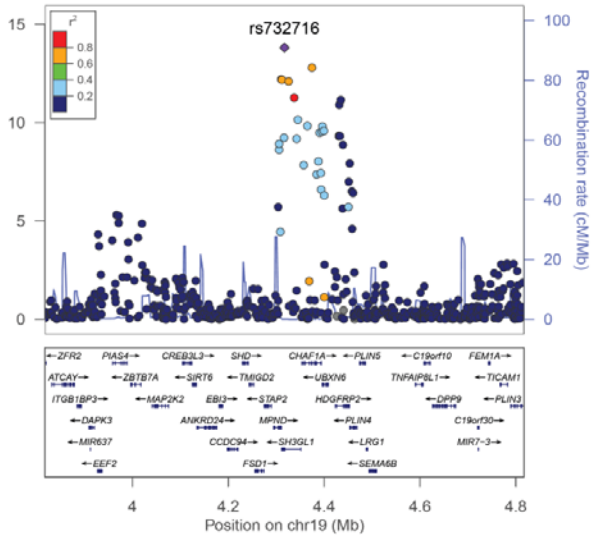
SF3.65: 18q21 - rs4890633 - MCH



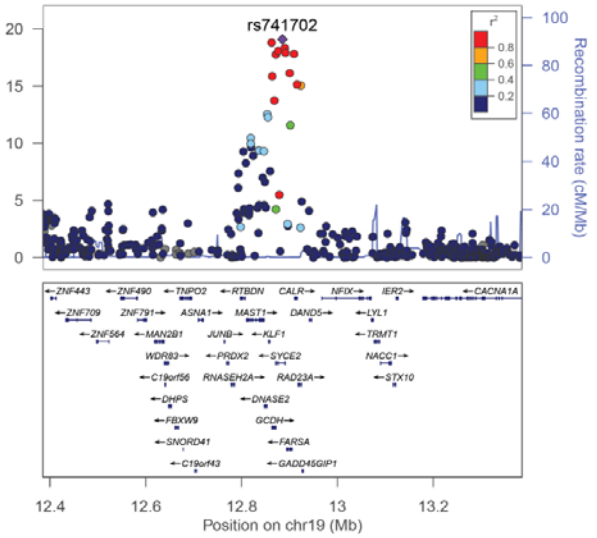
SF3.66: 19p13 - rs2159213 - HB



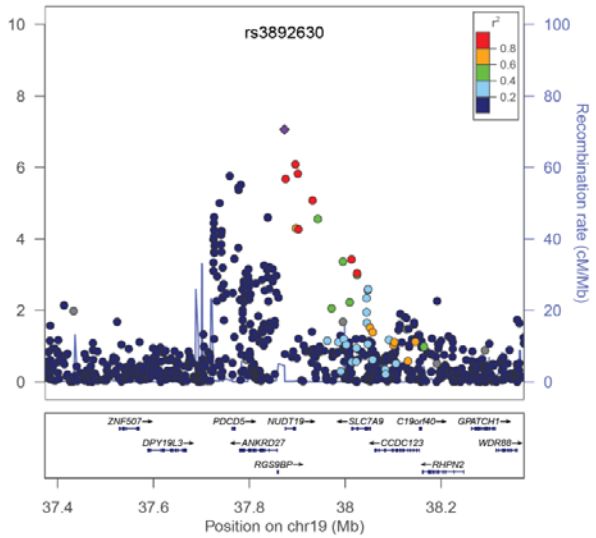
SF3.67: 19p13 - rs732716 - MCV



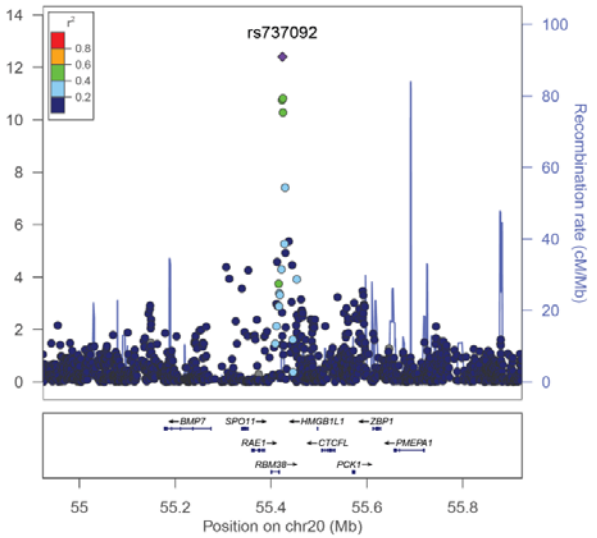
SF3.68: 19p13 - rs741702 - MCH



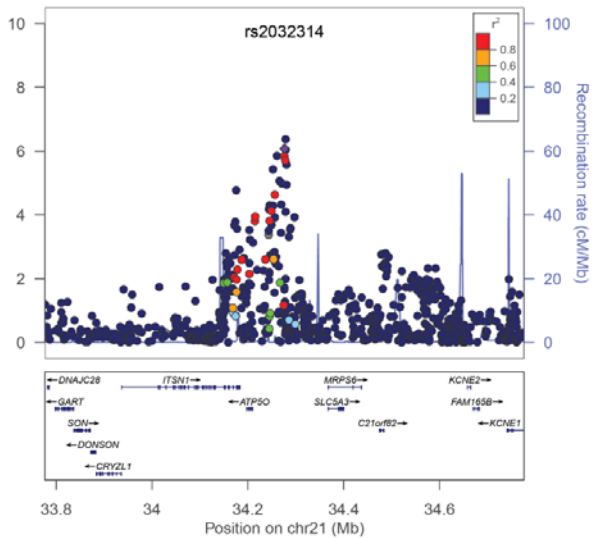
SF3.69: 19q13 - rs3892630 - MCV



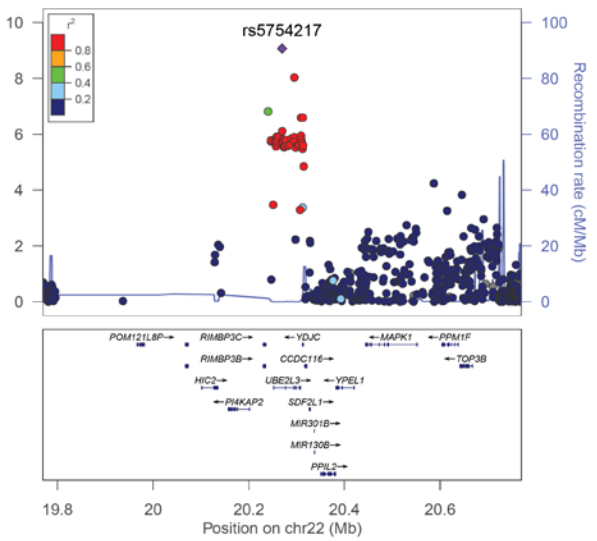
SF3.70: 20q13 - rs737092 - MCV



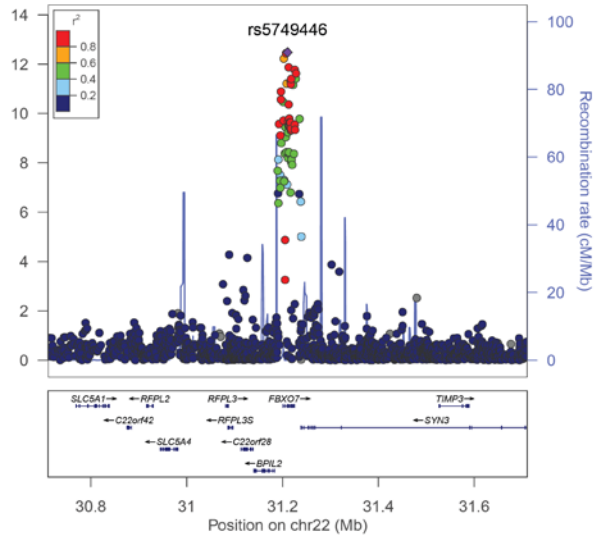
SF3.71: 21q22 - rs2032314 - PCV



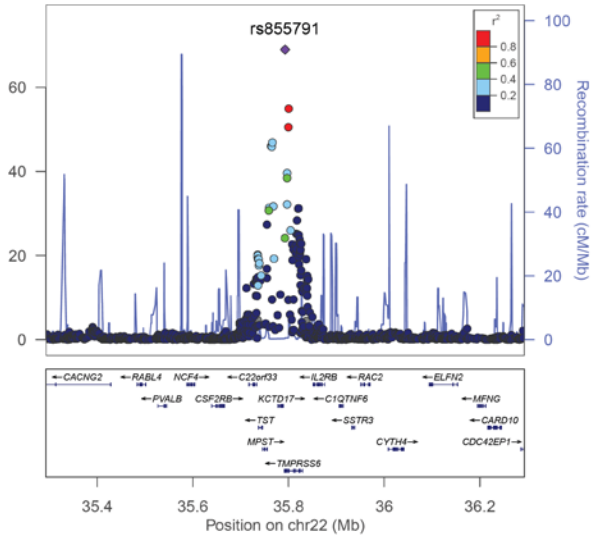
SF3.72: 22q11 - rs5754217 - MCV



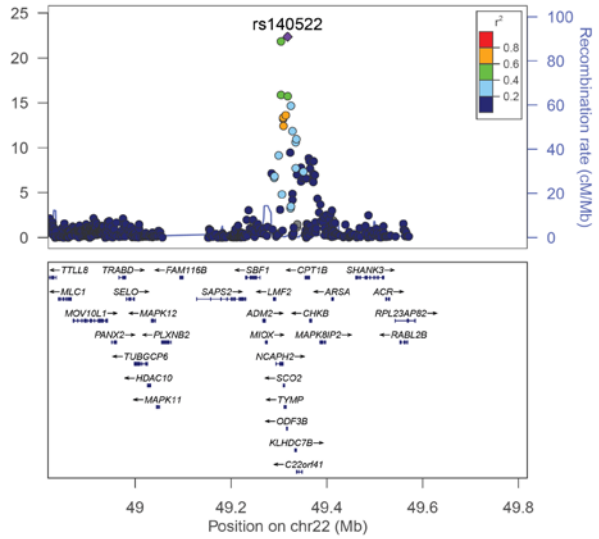
SF3.73: 22q12 - rs5749446 - MCH



SF3.74: 22q12 - rs855791 - MCH

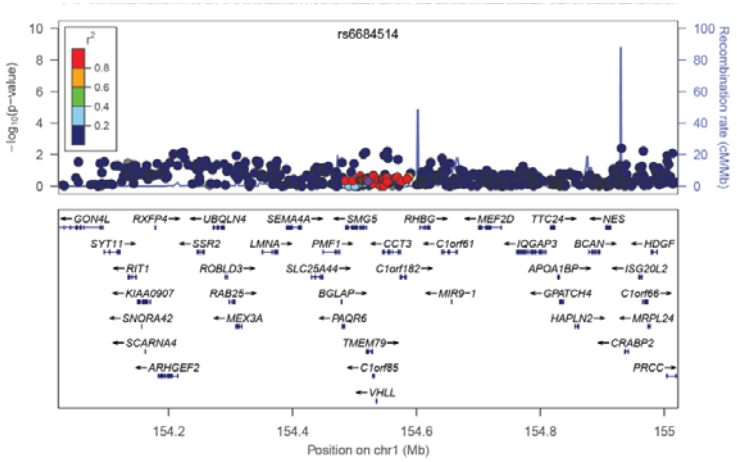


SF3.75: 22q13 - rs140522 - MCV

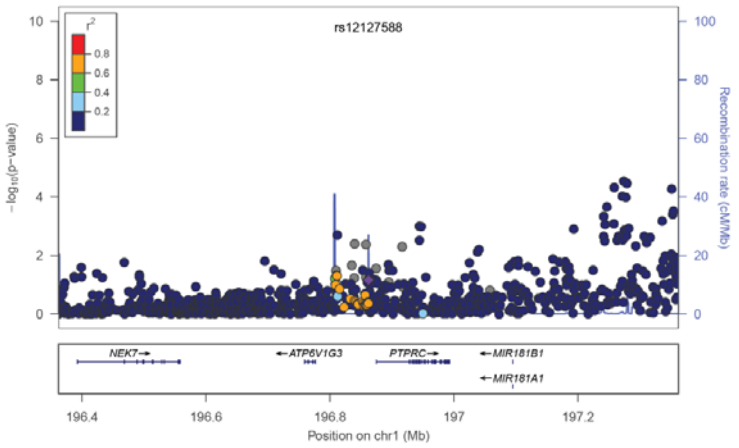


Supplementary Figure 4. SF4.1 to 4.3: Regional plots for the three genetic loci associated with red blood cell phenotypes in an East Asian GWAS. Results are shown for Europeans in the current study. The lead SNP identified in the East Asian GWAS is indicated. Pairwise LD with lead SNP is shown using HapMap2 CEU data.

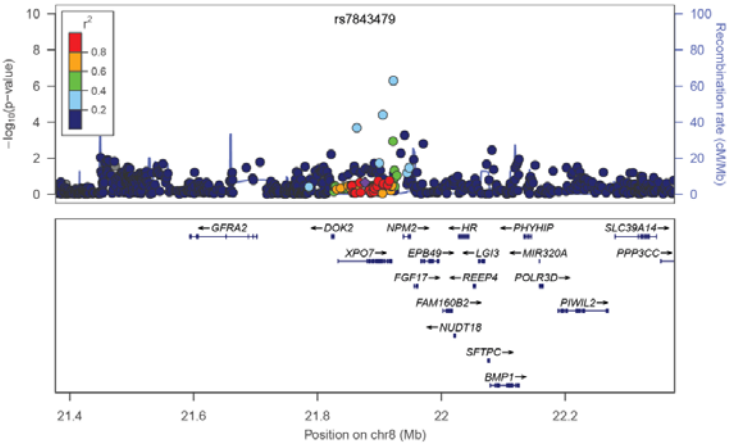
SF4.1: rs6684514 - MCHC



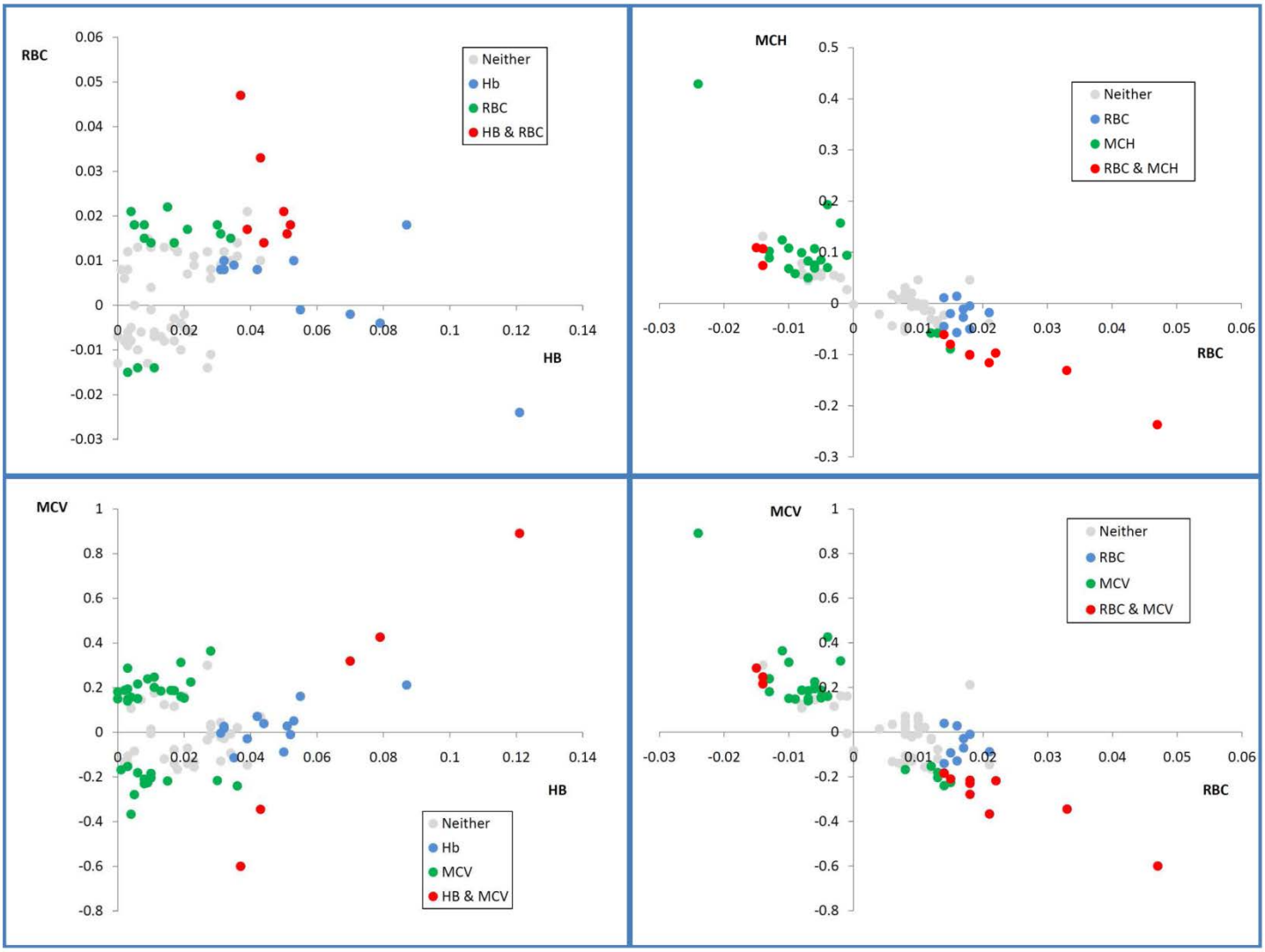
SF4.2: rs12127588 - MCH



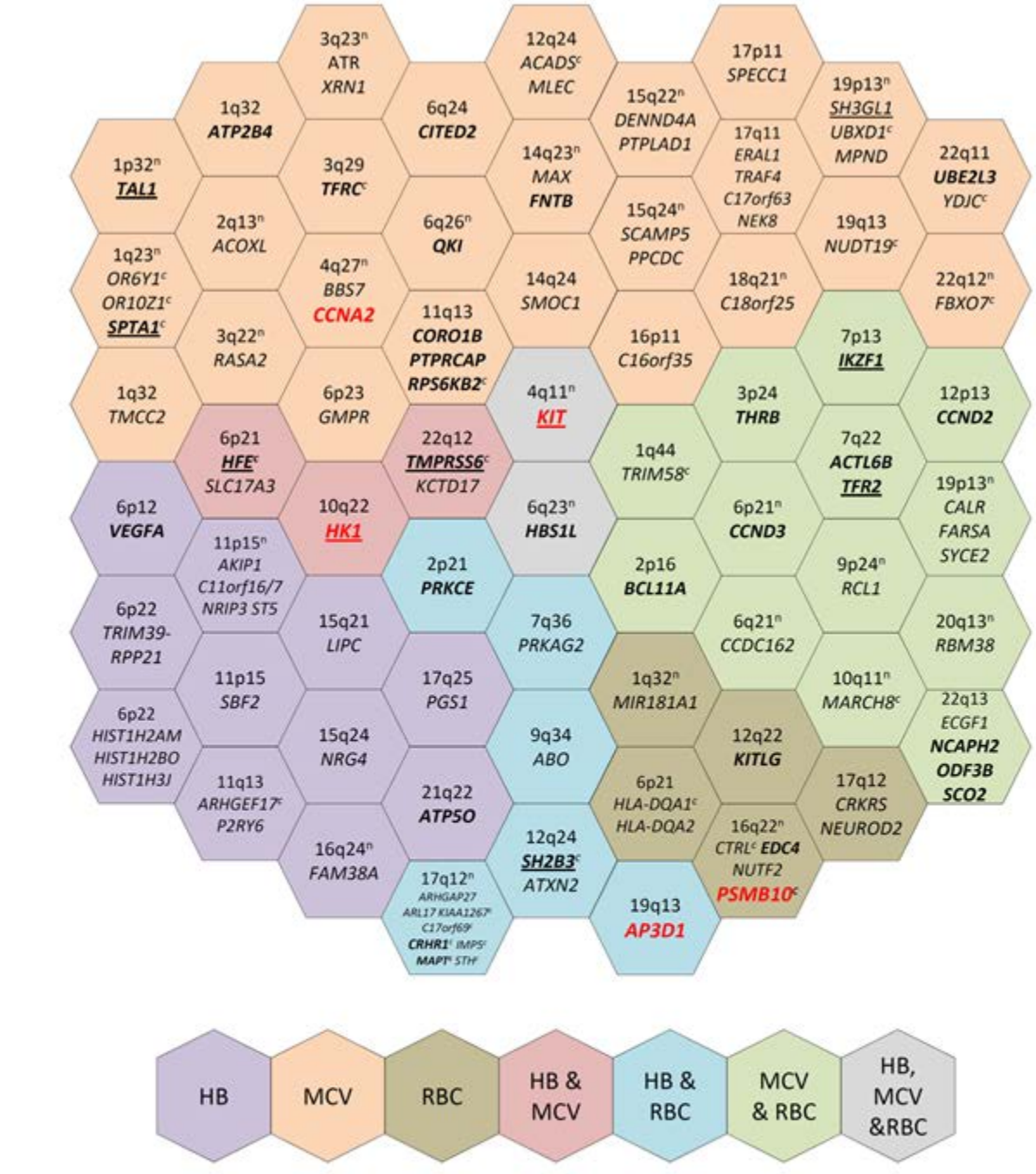
SF4.3: rs7843479 - MCV



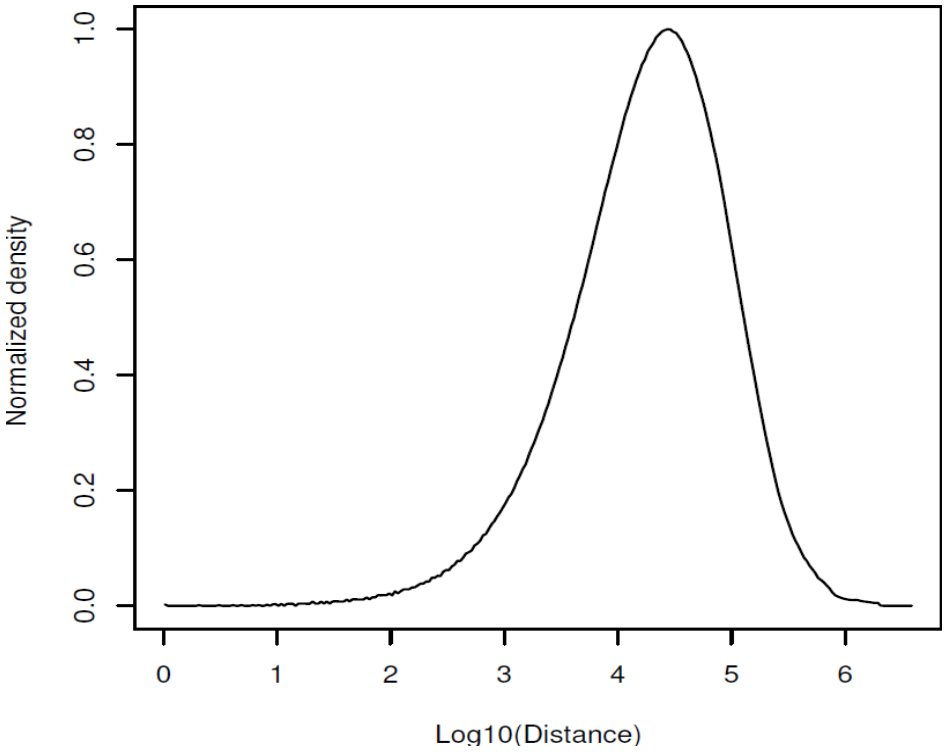
Supplementary Figure 5. Associations of the sentinel SNPs from the 75 genetic loci identified with red blood cell phenotypes. Associations are presented as effect size per copy of variant for pairs of phenotypes: 3a - HB and RBC, 3b - HB and MCV, 3c - RBC and MCH, 3d - RBC and MCV. In each plot, SNP symbols are coloured to indicate whether they are associated with phenotype at $P<1x10^{-8}$.



Supplementary Figure 6. Genomic loci associated with haemoglobin (HB), mean cell volume (MCV), red blood cell concentration (RBC), or combinations of these phenotypes. Candidate genes at each locus are listed. Genes with haematological phenotype in model organisms or humans are highlighted: mouse or drosophila RNAi model (bold); both model organisms (red and bold); humans (OMIM, underlined). [n] Locus contains a nucleosome deplete region that intersects a sentinel SNP or a SNP in LD at r^2>0.8. [c] Gene contains a non-synonymous SNP that is a sentinel SNP or a SNP in LD at r^2>0.8.



Supplementary Figure 7. Distribution for the distance between the HapMap2 SNPs used for the discovery GWAS, and all 1000 Genomes SNPs in high LD (r^2>0.8) in a 4MB window. Results confirm that the great majority (>99.9%) of SNPs in high LD are located within 1Mb of the discovery SNPs, supporting use of a 1Mb distance to define a genetic region.



Supplementary Figure 8. Permutation testing to simulate expectations under the null hypothesis in the Drosophila studies, using a global RNAi screen of blood cell phenotypes. We took a random sample of 121 human genes, identified the Drosophila orthologs and counted the number with a blood cell phenotype. This was repeated 1,000,000 times. The simulation was repeated across the range of calling thresholds; for each threshold the number of red blood cell GWAS candidate genes observed to have haematologic phenotype in the Drosophila global screen is noted.