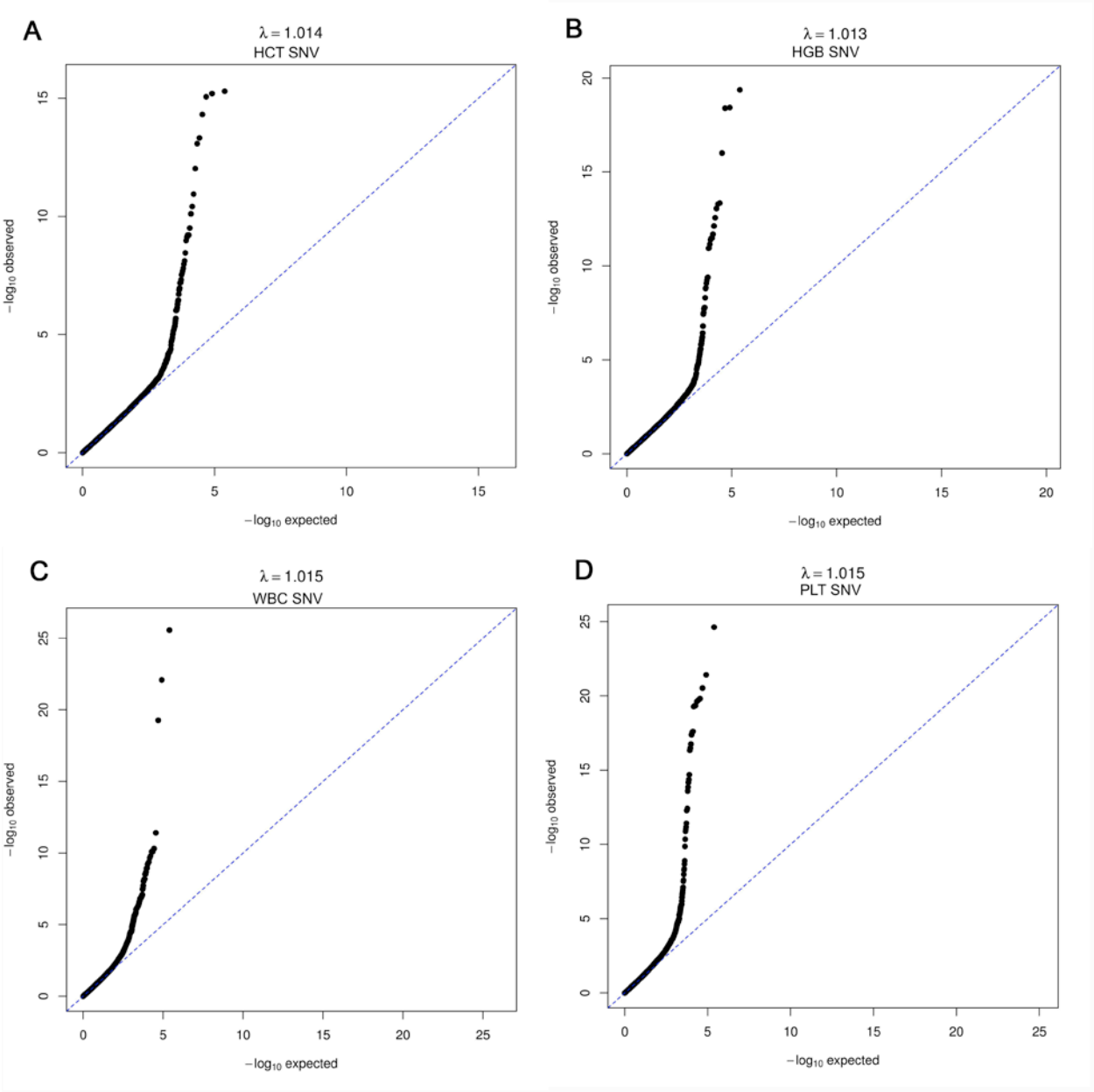
SUPPLEMENTARY INFORMATION

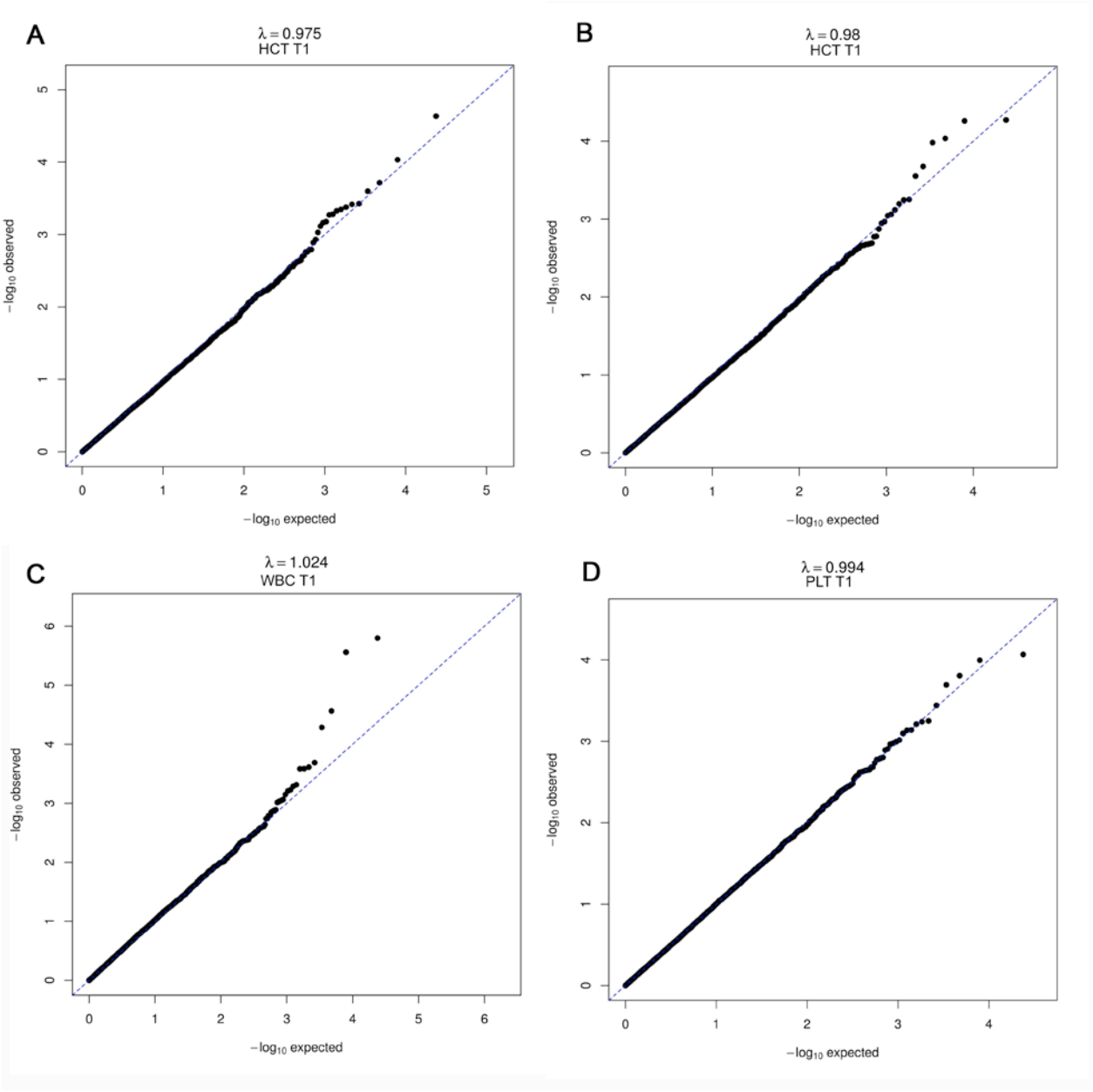
Rare and low-frequency coding variants in *CXCR2* and other genes are associated with hematological traits

Paul L. Auer, Alexander Teumer, Ursula Schick, Andrew O'Shaughnessy, Ken Sin Lo, Nathalie Chami, Chris Carlson, Simon de Denus, Marie-Pierre Dubé, Jeff Haessler, Rebecca D. Jackson, Charles Kooperberg, Louis-Philippe Lemieux Perreault, Matthias Nauck, Ulrike Peters, John D. Rioux, Frank Schmidt, Valérie Turcot, Uwe Völker, Henry Völzke, Andreas Greinacher, Li Hsu, Jean-Claude Tardif, George A. Diaz, Alexander P. Reiner, Guillaume Lettre

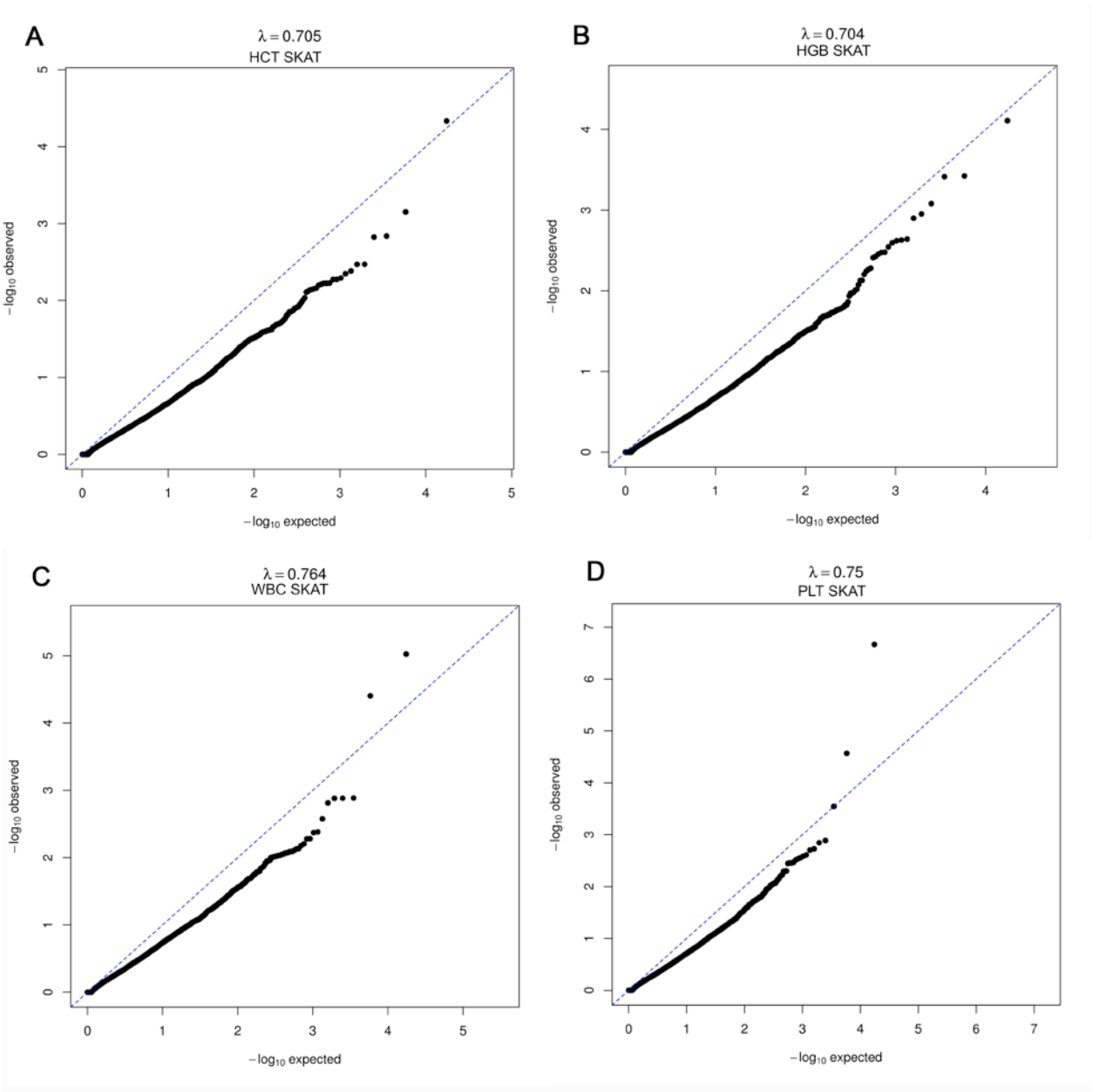
Supplementary Figures 1-9
Supplementary Tables 1-7
Supplementary Note



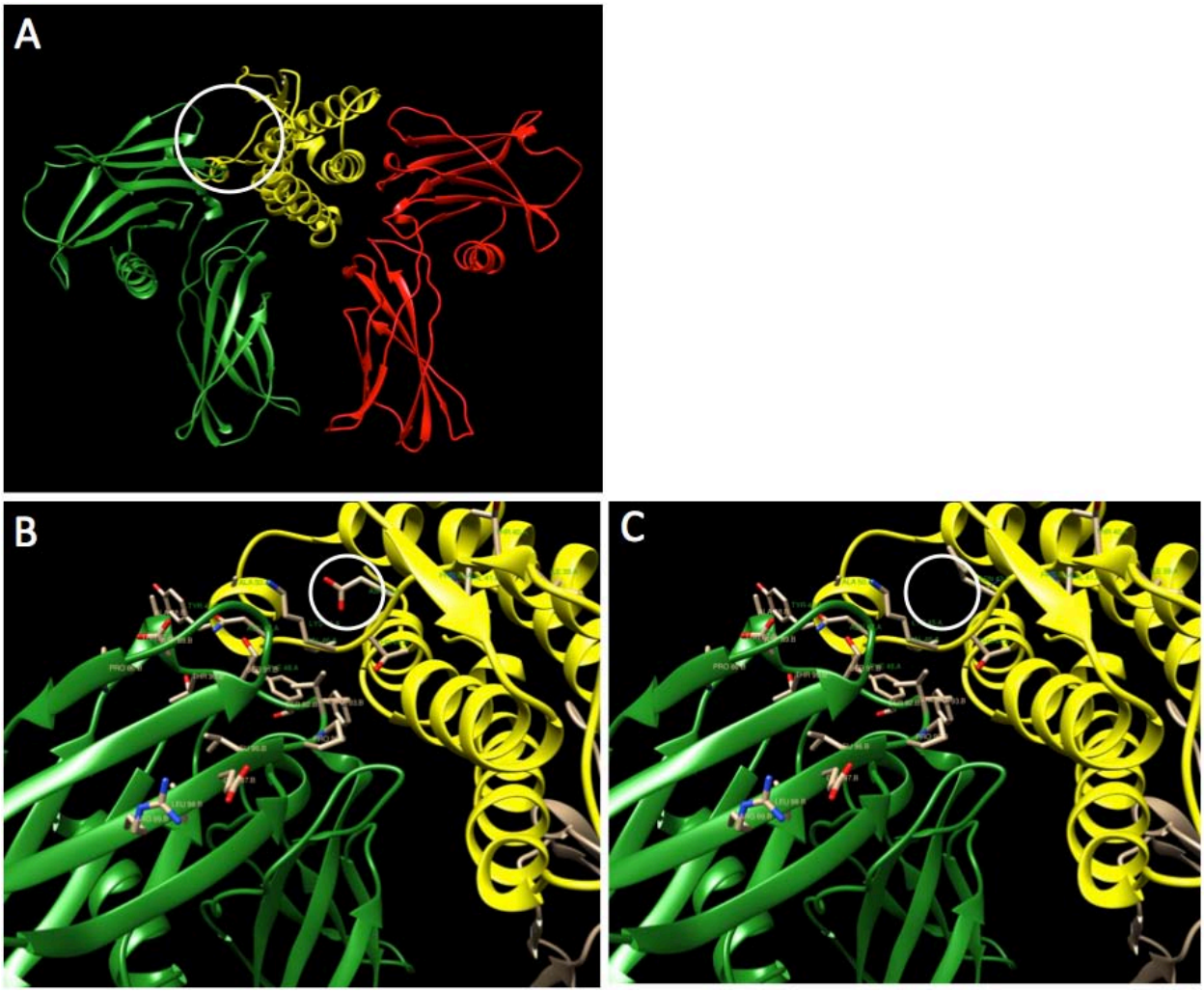
Supplementary Figure 1. Quantile-quantile plots of single variant association results. (A) Hematocrit. (B) Hemoglobin. (C) White blood cell count. (D) platelet count.



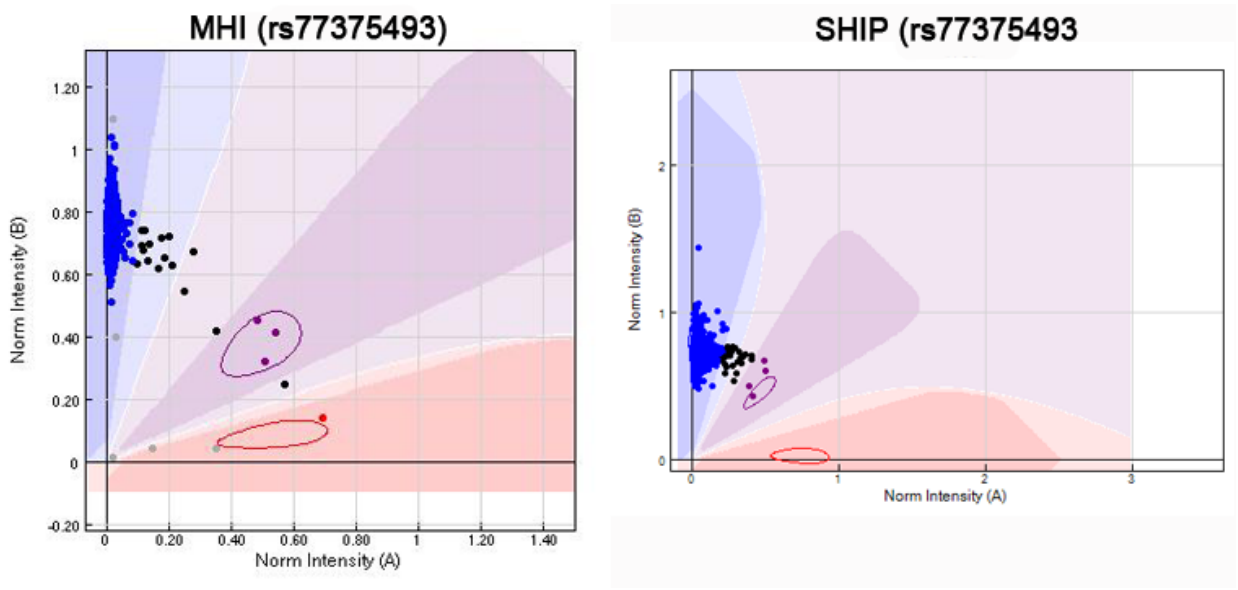
Supplementary Figure 2. Quantile-quantile plots of gene-based burden T1 results. (A) Hematocrit. (B) Hemoglobin. (C) White blood cell count. (D) Platelet count. For each gene, we only considered missense, nonsense and splice site variants with a minor allele frequency <1%.



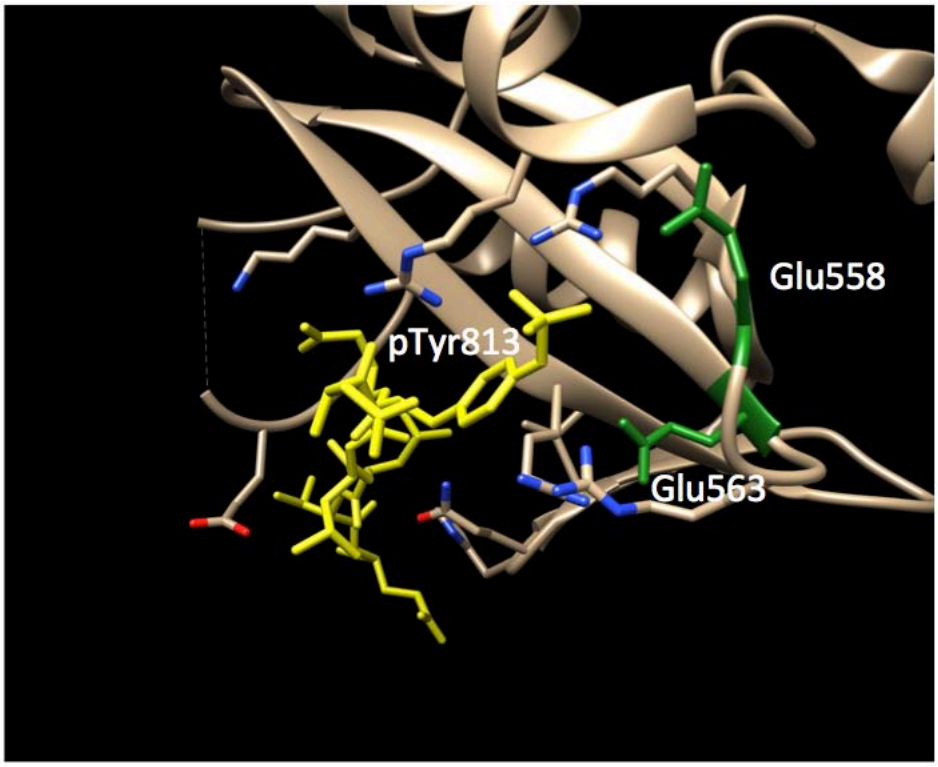
Supplementary Figure 3. Quantile-quantile plots of gene-based SKAT results. (A) Hematocrit. (B) Hemoglobin. (C) White blood cell count. (D) Platelet count. For each gene, we only considered missense, nonsense and splice site variants with a minor allele frequency <5%.



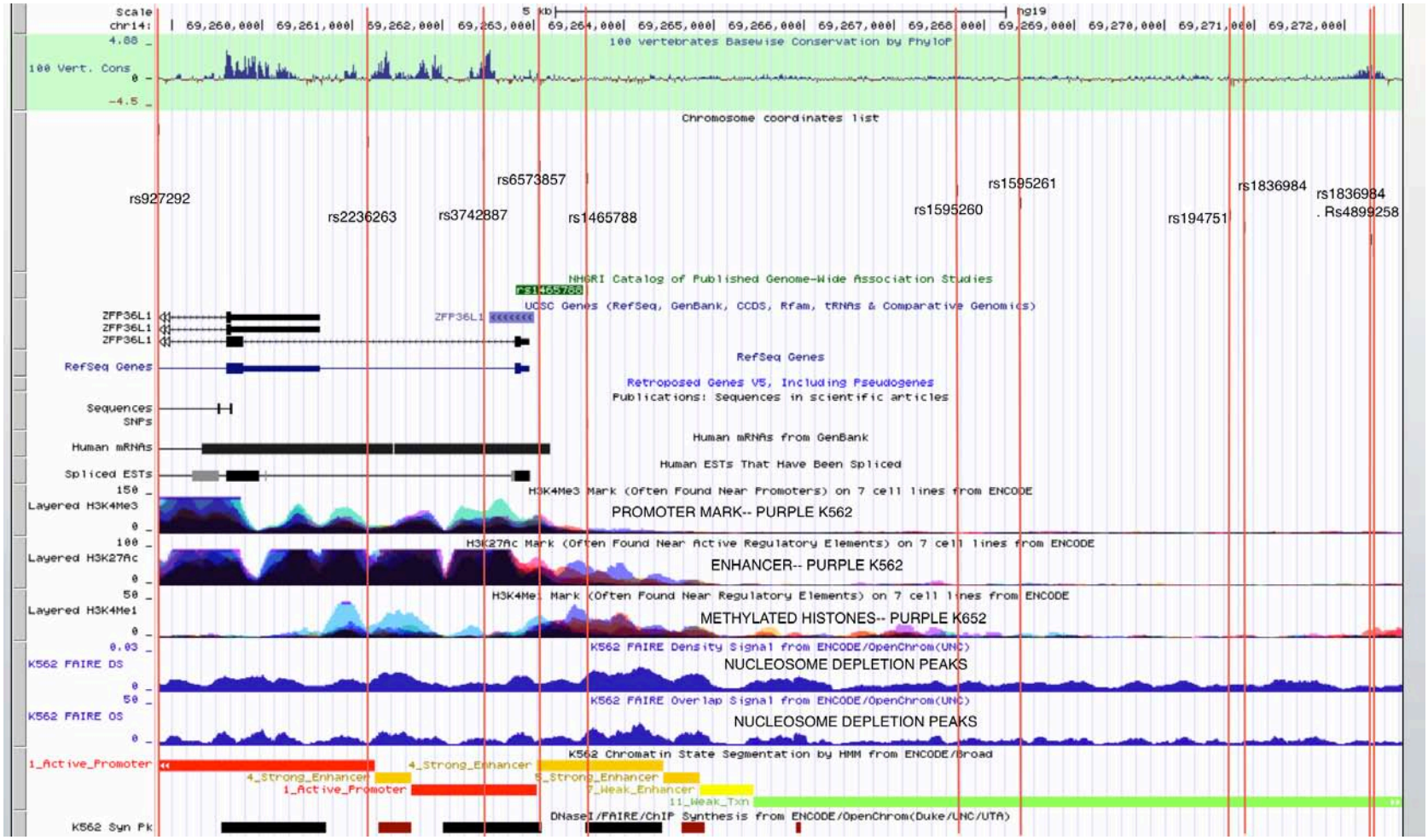
Supplementary Figure 4. Impact of mutation p.Asp70Asn in *EPO*. (A) Crystal structure of erythropoietin (EPO, yellow) with two erythropoietin receptor subunits (EPOR, green and red)[1]. The white circle indicates the interface between EPO and EPOR magnified in (B) and (C). In (B) and (C), the white circles highlight residue 70 (43 after cleavage of the propeptide). Replacing Asp70 (B) by Asn70 (C) may de-stabilize the EPO loop that interacts with EPOR because it disrupts a hydrogen bound between Asp70 and Thr71.



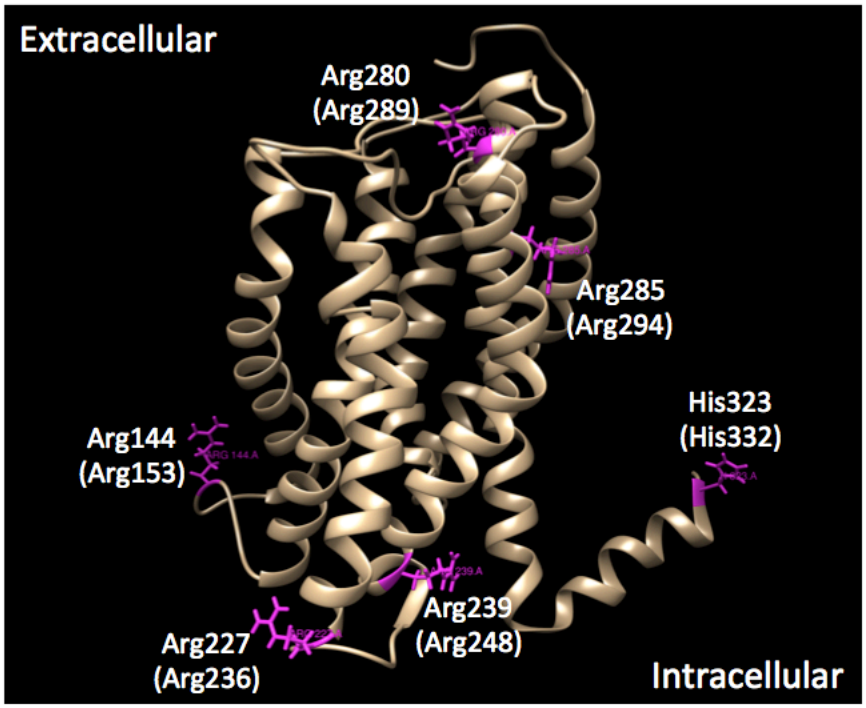
Supplementary Figure 5. Intensity plots for *JAK2* p.Val617Phe in the Montreal Heart Institute (MHI; left) and the Study of Health in Pomerania (SHIP; right) genotyped on the Illumina HumanExome Beadchip. Density plots for the Women's Health Initiative were not available.



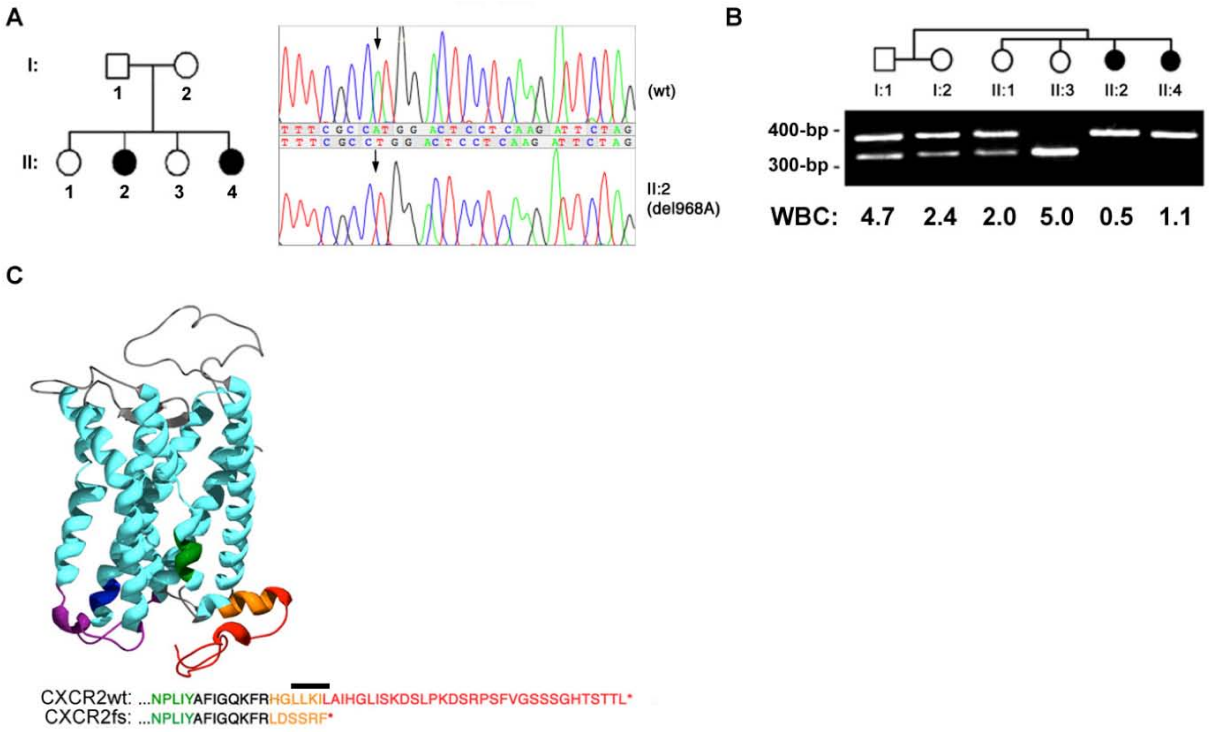
Supplementary Figure 6. Mutations in *SH2B3* may interfere with JAK2 binding. Crystal structure of the SH2-B SH2 domain (beige) with JAK2 phospho-tyrosine peptide (pTyr813, yellow)[2]. JAK2 pTyr813 interacts with arginine residues in the SH2-B BC loop. The glutamate residues identified in our association study (Glu395 and Glu400, green) correspond to Glu558 and Glu563 in the SH2-B sequence. They are directly in the BC loop that interacts with JAK2. Glu563 interacts with Arg578.



Supplementary Figure 7. UCSC ENCODE tracks for the rs1465788 variant near *ZFP36L1* . The rs1465788 variant appears to be in a K562 FAIRE nucleosome depletion peak. The rs1465788 is also in high LD with rs2236263 (r^2=0.91 in CEU), rs3742887 (r^2=0.99 in CEU), and rs6573857 (r^2=1 in CEU) which all lie in an active promoter for *ZFP36L1*.



Supplementary Figure 8. Crystal structure of CXCR1 in a phospholipid bilayer[3]. CXCR2, like CXCR1 and CXCR4, is a seven transmembrane (TM) domains protein. In purple are the CXCR1 residues that correspond to the missense mutations identified in CXCR2 (two CXCR2 residues, Met6 and Ala37, do not have their CXCR1 equivalent on this structure). In parenthesis are the corresponding CXCR2 residues. CXCR2 Arg289 and Arg294 are in TM domains. The remaining four CXCR2 residues (Arg144, Arg236, Arg248, His332) are in intracellular loops.



Supplementary Figure 9. Analysis of CXCR2 in Myelokathexis Pedigree. (A) A myelokathexis pedigree with two affected (filled) and two unaffected (open) daughters is shown at left. *CXCR2* sequence trace shows a single base deletion corresponding to coding sequence position 968 in individual I-2 (unaffected) compared to an unrelated control (arrow). (B) An *Nco*I recognition site (CCATGG) which gives a 325-bp fragment when a wild type amplicon is digested is destroyed by the 968delA mutation so that the 375-bp undigested amplicon persists. Affected siblings II-2 and II-4 are homozygous for the deletion mutation, parents and sibling II-1 are heterozygous for the mutation, and sibling II-3 is homozygous for the wild type allele. White blood cell counts (WBC) for each family member are shown below. (C) A structural model of CXCR2 with the conserved GPCR motifs (intracellular loop 2 - *purple*, DRY - *dark blue*, NPXXY - *green*), the mutant frameshift peptide (orange), and the truncated wild type tail (red) indicated. The frameshift is located in Helix 8, which is predicted to be parallel to the plasma membrane. The amino acid sequences from the wild type and mutant CXCR2 constructs are shown. The CXCR2fs construct contain a six-amino acid frameshift peptide (LDSSRF). Amino acid colors correspond to those used in the structure model. The black bar over the primary sequence indicates the location of a highly conserved LLKIL dileucine motif that is critical for receptor-mediated chemotaxis[4].

Supplementary Table 1. Description of the three cohorts analyzed in this study. Summary statistics of the hematological traits and other covariates measured in each study. Values are presented as mean (standard deviation).

	Montreal Heart Institute (MHI) Biobank	**Women Health Initiative (WHI)**	**Study of Health in Pomerania (SHIP/SHIP-TREND)**
Number of individuals with phenotype and genotype available	6,796	18,072	6,526
Percent women	37	100	51
Age (years)	63 (11)	66 (7)	53 (15)
Hemoglobin (g/dl)	13.8 (1.7)	13.6 (0.9)	14.0 (1.3)
Hematocrit (%)	40.6 (4.9)	40.4 (2.9)	41.6 (3.4)
White blood cell ($x10^9$/l)	7.1 (2.2)	6.1 (1.6)	6.4 (1.8)
Platelet ($x10^9$/l)	228.6 (67.1)	245.9 (55.8)	225.4 (53.5)

Supplementary Table 2. Single variant replication results. The direction of the effect sizes (Beta) is for the A1 allele. Effect sizes and standard errors (SE) are in the following units: for hematocrit (%), for hemoglobin (g/dl), for white blood cell count ($log_{10}(x10^9/l)$) and platelet count ($x10^9/l$). n.d.; not determined because the marker is monomorphic in SHIP.

			Discovery MHI+WHI (N=24,814)			Replication SHIP (N=6,526)			Combined (N=31,340)				
Variant	**Chr (Pos)**	**A1/A2**	**Freq (A1)**	**Beta (SE)**	***P***	**Freq (A1)**	**Beta (SE)**	***P***	**Beta (SE)**	***P***	**Hetero. *P***	**Gene**	**Annotation**
HEMATOCRIT													
rs146600269	1:89,486,344	T/G	0.0005	-3.125 (0.696)	$7.2x10^{-6}$	0.0011	-0.213 (0.854)	0.80	-1.962 (0.54)	$2.8x10^{-4}$	0.0082	*GBP3*	Missense (p.L21M)
rs61755431	1:155,260,382	T/C	0.0025	-1.228 (0.301)	$4.4x10^{-5}$	0.0025	-0.719 (0.506)	0.16	-1.095 (0.259)	$2.3x10^{-5}$	0.39	*PKLR*	Missense (p.R538Q, p.R569Q)
rs200879349	1:169,438,042	T/G	0.9999	14.83 (3.269)	$5.7x10^{-6}$	0.9999	0.989 (2.05)	0.63	4.895 (1.737)	0.0048	$3.3x10^{-4}$	*SLC19A2*	Missense (p.K355Q)
rs17319721	4:77,368,847	A/G	0.43	0.151 (0.028)	$1.1x10^{-7}$	0.45	0.047 (0.051)	0.35	0.126 (0.025)	$3.5x10^{-7}$	0.074	*SHROOM3*	Intronic
rs201657731	7:80,302,104	T/C	0.0002	5.139 (1.164)	$1.0x10^{-5}$	0.0005	-0.062 (1.186)	0.96	2.588 (0.831)	0.0018	0.0018	*CD36*	Nonsense (Q382X)
rs62483572	7:100,319,633	A/G	0.0045	-1.072 (0.211)	$3.5x10^{-7}$	0.0041	-1.105 (0.396)	0.0053	-1.079 (0.186)	$6.4x10^{-9}$	0.94	*EPO*	Missense (p.D70N)
rs77375493	9:5,073770	T/G	0.0004	3.714 (0.758)	$9.5x10^{-7}$	0.0041	-0.116 (0.495)	0.81	1.03 (0.415)	0.013	$2.3x10^{-5}$	*JAK2*	Missense (p.V617F)
rs33971440	11:5,248,159	T/C	0.0002	-5.723 (1.291)	$9.2x10^{-6}$	Monomorphic			n.d.			*HBB*	5' splice site (exon2:c.92+1G>A)
rs2270416	16:89,261,482	A/C	0.035	0.345 (0.077)	$7.0x10^{-6}$	0.032	0.14 (0.143)	0.33	0.299 (0.068)	$9.6x10^{-6}$	0.21	*CDH15*	Nonsense (Y788X)
HEMOGLOBIN													
rs61755431	1:155,260,382	T/C	0.0025	-0.434 (0.102)	$1.9x10^{-5}$	0.0025	-0.417 (0.179)	0.020	-0.43 (0.088)	$1.1x10^{-6}$	0.93	*PKLR*	Missense (p.R538Q, p.R569Q)
rs200879349	1:169,438,042	T/G	0.9999	5.144 (1.134)	$5.7x10^{-6}$	0.9999	0.397 (0.723)	0.58	1.77 (0.61)	0.0037	$4.2x10^{-4}$	*SLC19A2*	Missense (p.K355Q)
rs28664715	4:87,809,025	T/C	0.91	-0.077 (0.016)	$2.6x10^{-6}$	0.92	-0.03 (0.033)	0.36	-0.068 (0.015)	$4.1x10^{-6}$	0.20	*C4orf36*	Missense (p.Y81C)
rs139178017	7:100,225,847	T/C	0.0042	0.31 (0.075)	$3.2x10^{-5}$	0.0025	0.385 (0.181)	0.033	0.321 (0.069)	$3.3x10^{-6}$	0.70	*TRF2*	Synonymous (p.E491E) and 5' donor splice site
rs62483572	7:100,319,633	A/G	0.0045	-0.347 (0.071)	$9.5x10^{-7}$	0.0041	-0.354 (0.14)	0.011	-0.348 (0.063)	$3.4x10^{-8}$	0.96	*EPO*	Missense (p.D70N)
rs33971440	11:5,248,159	T/C	0.0002	-2.41 (0.438)	$3.7x10^{-8}$	Monomorphic			n.d.			*HBB*	5' splice site (exon2:c.92+1G>A)
WHITE BLOOD CELL													
rs1260326	2:27,730,940	T/C	0.43	0.005 (0.001)	$3.8x10^{-6}$	0.40	0.005 (0.002)	0.017	0.005 (0.001)	$8.5x10^{-8}$	0.97	*GCKR*	Missense (p.L446P)
PLATELETS													
rs150282282	1:1,226,312	T/C	0.0001	286.8 (62.538)	$4.5x10^{-6}$	0.0005	-18.06 (19.154)	0.35	8.084 (18.314)	0.66	$3.2x10^{-6}$	*SCNN1D*	Missense (p.P652L)
rs144362185	1:62,579,918	A/G	0.0001	282.9 (62.464)	$5.9x10^{-6}$	0.0002	-31.3 (35.67)	0.38	45.962 (30.975)	0.14	$1.3x10^{-5}$	*INADL*	Missense (p.R1552Q)
		C/G	0.88	-3.878 (0.801)	$1.3x10^{-6}$	0.84	-2.446 (1.216)	0.044	-3.444 (0.669)	$2.6x10^{-7}$	0.33	*PPARG*	Missense (p.P12A)
rs149507188	3:99,513,493	A/G	0.0001	124.5 (24.874)	$5.6x10^{-7}$	Monomorphic			n.d.			*COL8A1*	Missense (p.A250T)

rs149618433	4:3,318,625	T/C	0.0006	101.3 (22.103)	$4.6x10^{-6}$	0.0001	-35.26 (50.472)	0.48	79.325 (20.247)	$8.9x10^{-5}$	0.013	*RGS12*	Missense (p.T243M)
rs202125938	6:151,917,687	A/G	0.9999	-295.1 (62.482)	$2.3x10^{-6}$	0.9999	36.37 (50.423)	0.47	-94.362 (39.239)	0.016	$3.7x10^{-6}$	*C6orf97*	Missense (p.K562R)
rs201367949	8:124,796,805	A/C	0.9999	-278 (62.472)	$8.6x10^{-6}$	Monomorphic			n.d.			*FAM91A1*	Missense (p.N267H)
rs77375493	9:5,073,770	T/G	0.0005	124.092 (12.827)	$3.9x10^{-22}$	0.0041	16.41 (8.61)	0.057	49.857 (7.149)	$3.1x10^{-12}$	$3.2x10^{-12}$	*JAK2*	Missense (p.V617F)
rs146836226	11:8,947,080	C/G	0.9999	-286.8 (62.538)	$4.5x10^{-6}$	0.9999	57.22 (50.459)	0.26	-78.429 (39.27)	0.046	$1.9x10^{-5}$	*C11orf16*	Missense (p.Q378H)
rs3888798	11:88,027,209	T/C	0.96	-5.623 (1.341)	$2.7x10^{-5}$	0.95	-4.122 (1.946)	0.034	-5.14 (1.104)	$3.2x10^{-6}$	0.53	*CTSC*	Missense (p.I453V)
rs72650673	12: 111,885,310	A/G	0.0014	37.261 (6.724)	$3.0x10^{-8}$	0.0007	7.727 (18.254)	0.67	33.732 (6.310)	$9.0x10^{-8}$	0.13	*SH2B3*	Missense (p.E400K)
rs41300592	13:25,459,805	T/C	0.9998	-118 (22.68)	$2.0x10^{-7}$	0.9998	-40.94 (29.139)	0.16	-88.928 (17.897)	$6.7x10^{-7}$	0.037	*CENPJ*	Missense (p.N1102S)
rs1465788	14:69,263,599	T/C	0.27	-2.745 (0.578)	$2.1x10^{-6}$	0.27	-3.063 (1.011)	0.0024	-2.823 (0.502)	$1.9x10^{-8}$	0.78	*ZFP36L1*	Intergenic
rs139734678	14:74,969,470	T/C	0.0001	278 (62.472)	$8.6x10^{-6}$	0.0002	4.954 (29.124)	0.87	53.702 (26.397)	0.042	$7.5x10^{-5}$	*LTBP2*	Missense (p.V1686I)
rs148159407	15:40,477,419	A/G	0.9995	-64.166 (15.371)	$3.0x10^{-5}$	0.9998	15.46 (35.704)	0.67	-51.716 (14.118)	$2.5x10^{-4}$	0.041	*BUB1B*	Missense (p.N269D)
rs150934316	16:1,555,571	A/G	0.0001	282.1 (62.495)	$6.4x10^{-6}$	Monomorphic			n.d.			*TELO2*	Missense (p.R668Q)
rs76931703	19:43,237,080	T/C	0.9999	-120.2 (24.862)	$1.3x10^{-6}$	0.9998	-40.14 (35.68)	0.26	-94.032 (20.399)	$4.0x10^{-6}$	0.066	*PSG3*	Missense (p.M189V)
rs41303899	20:57,598,808	A/G	0.0016	-27.696 (6.717)	$3.7x10^{-5}$	0.0024	-42.63 (9.07)	$2.7x10^{-6}$	-32.986 (5.398)	$9.9x10^{-10}$	0.19	*TUBB1*	Missense (p.G109E)
rs73393075	22:29,537,940	A/G	0.0001	116.199 (25.406)	$4.8x10^{-6}$	0.0002	-23.08 (29.123)	0.43	56.01 (19.145)	0.0034	$3.1x10^{-4}$	*KREMEN1*	Missense (p.R406H, p.R423H)
rs738409	22:44,324,727	C/G	0.77	2.941 (0.612)	$1.6x10^{-6}$	0.78	0.755 (1.065)	0.48	2.398 (0.531)	$6.3x10^{-6}$	0.075	*PNPLA3*	Missense (p.I148M)

Supplementary Table 3. Association results with blood cell phenotypes for SNPs previously identified by genome-wide association studies (GWAS). We only report results from the meta-analysis MHI+WHI (N=24,814) for SNPs present on the Illumina ExomeChip array (or proxies in linkage disequlibrium based on data from the 1000 Genomes Project's European populations). We obtained the list of associated SNPs from the largest meta-analyses published to date: hemoglobin and hematocrit levels[5], white blood cell count[6] and platelet count[7]. Effect sizes and standard errors (SE) are in the following units: for hematocrit (%), for hemoglobin (g/dl), for white blood cell count ($log_{10}(x10^9/l)$) and platelet count ($x10^9/l$).

GWAS SNP	CHR	BP	ExomeChip SNP	CHR	BP	r^2 with GWAS SNP	Allele1	Allele2	Freq (Allele 1)	BETA	SE	P-value	Locus
HEMATOCRIT													
rs857684	1	158575729	rs863362	1	158549492	0.36	t	c	0.4657	0.0735	0.028	0.008771	*OR10X1*
rs4953318	2	46355051	rs10495928	2	46353166	0.8	a	g	0.6602	0.2235	0.0296	4.79E-14	*PRKCE*
rs218238	4	55395024	rs218237	4	55394172	0.53	t	c	0.1296	-0.19	0.0419	5.67E-06	
rs13152701	4	122751061	rs1507995	4	122748308	0.31	a	g	0.2911	0.2149	0.0863	0.01281	*BBS7*
rs1408272	6	25842951	rs1408272	6	25842951	1	t	g	0.9364	-0.2949	0.0582	4.02E-07	*HFE*
rs9349204	6	41914378	rs3218097	6	41905275	0.91	a	g	0.2509	0.0675	0.0325	0.03786	*CCND3*
rs9369427	6	43811430	rs9472138	6	43811762	0.98	t	c	0.2837	-0.101	0.031	0.001141	
rs9389269	6	135427159	rs7776054	6	135418916	0.91	a	g	0.7425	0.2606	0.0321	5.05E-16	*HBS1L-MYB*
rs2075672	7	100240296	rs7385804	7	100235970	0.99	a	c	0.6298	-0.0931	0.0291	0.001349	*TFR2*
rs10480300	7	151406005	rs10224002	7	151415041	0.86	a	g	0.7028	0.178	0.0308	7.66E-09	*PRKAG2*
rs4737009	8	41630405	rs4737009	8	41630405	1	a	g	0.2339	0.0781	0.0382	0.04104	*ANK1*
rs579459	9	136154168	rs635634	9	136155000	0.83	t	c	0.1998	-0.1856	0.0352	1.30E-07	*ABO*
rs2302264	11	67207426	rs1695	11	67352689	0.35	a	g	0.6615	-0.062	0.0296	0.03625	*GSTP1*
rs10849023	12	4332478	rs11611647	12	4333919	1	t	c	0.7902	0.1083	0.0344	0.001656	
rs3184504	12	111884608	rs3184504	12	111884608	1	t	c	0.4964	0.1775	0.0282	3.13E-10	*SH2B3*
rs2271294	16	67902326	rs12449157	16	67708897	0.88	a	g	0.8435	-0.0846	0.0388	0.02923	*GFOD2*
rs4969184	17	76393413	rs2292642	17	76395430	0.64	t	c	0.6096	-0.0632	0.0289	0.02896	*PGS1*
rs741702	19	13024250	rs11085824	19	13001547	0.85	a	g	0.6375	0.068	0.0292	0.01998	*GCDH*
rs855791	22	37462936	rs855791	22	37462936	1	a	g	0.4371	-0.1882	0.0285	3.81E-11	*TMPRSS6*
HEMOGLOBIN													
rs4953318	2	46355051	rs10495928	2	46353166	0.8	a	g	0.6602	0.075	0.01	5.13E-14	*PRKCE*
rs218238	4	55395024	rs218237	4	55394172	0.53	t	c	0.1295	-0.048	0.0141	0.0006523	
rs13152701	4	122751061	rs1507995	4	122748308	0.31	a	g	0.2911	0.0769	0.0299	0.01017	*BBS7*
rs1408272	6	25842951	rs1408272	6	25842951	1	t	g	0.9364	-0.1474	0.0195	4.50E-14	*HFE*
rs9349204	6	41914378	rs3218097	6	41905275	0.91	a	g	0.2509	0.0242	0.0109	0.02666	*CCND3*
rs9369427	6	43811430	rs881858	6	43806609	0.87	a	g	0.7011	0.0378	0.0103	0.0002283	
rs9389269	6	135427159	rs7776054	6	135418916	0.91	a	g	0.7423	0.0675	0.0108	4.23E-10	*HBS1L-MYB*
rs590856	6	139844429	rs643381	6	139839423	0.88	a	c	0.5218	-0.0204	0.0095	0.03157	
rs2075672	7	100240296	rs7385804	7	100235970	0.99	a	c	0.6297	-0.0196	0.0098	0.04489	*TFR2*
rs10480300	7	151406005	rs10224002	7	151415041	0.86	a	g	0.7028	0.0606	0.0104	4.98E-09	*PRKAG2*
rs579459	9	136154168	rs507666	9	136149399	0.83	a	g	0.1996	-0.0732	0.0118	5.93E-10	*ABO*
rs10159477	10	71099888	rs7072268	10	71099913	0.10	t	c	0.50	-0.048	0.009	3.8E-7	*HK1*
rs2302264	11	67207426	rs2276118	11	67288594	0.75	t	c	0.5683	0.0198	0.0095	0.03707	*CABP2*
rs10849023	12	4332478	rs11611647	12	4333919	1	t	c	0.7903	0.036	0.0116	0.001849	
rs3184504	12	111884608	rs3184504	12	111884608	1	t	c	0.4965	0.0693	0.0095	2.74E-13	*SH2B3*

rs2271294	16	67902326	rs12449157	16	67708897	0.88	a	g	0.8437	-0.031	0.013	0.01741	*GFOD2*
rs888424	17	19985427	rs2120282	17	20209474	0.59	t	c	0.5876	-0.0215	0.0096	0.02594	*SPECC1*
rs4969184	17	76393413	rs2292642	17	76395430	0.64	t	c	0.6097	-0.0274	0.0097	0.004916	*PGS1*
rs732716	19	4366219	rs1127888	19	4454083	0.83	t	c	0.2689	-0.0243	0.0107	0.02295	*UBXN6*
rs855791	22	37462936	rs855791	22	37462936	1	a	g	0.4371	-0.0878	0.0096	4.27E-20	*TMPRSS6*
WHITE BLOOD CELLS													
rs546829	4	74956372	rs546829	4	74956372	1	a	t	0.3819	0.0067	0.001	6.94E-12	*CXCL2*
rs2844503	6	31442731	rs2844503	6	31442731	1	a	t	0.4617	0.0063	0.001	5.63E-10	*HLA*
rs445	7	92408370	rs445	7	92408370	1	t	c	0.1112	-0.0094	0.0014	7.30E-11	*CDK6*
rs579459	9	136154168	rs579459	9	136154168	1	t	c	0.7706	0.0061	0.0011	1.42E-08	*ABO*
rs3184504	12	111884608	exm1037423	12	111884608	1	t	c	0.4965	0.0056	0.0009	4.26E-10	*SH2B3*
rs4794822	17	38156712	rs4794822	17	38156712	1	t	c	0.3883	0.011	0.001	5.09E-30	*PSMD3*
PLATELETS													
rs1668871	1	205237137	rs1668873	1	205235990	0.94	a	g	0.357	2.1478	0.5356	6.06E-05	*TMCC2*
rs7550918	1	247675559	rs56043070	1	247719769	0.25	a	g	0.0721	-6.7628	0.9998	1.34E-11	*GCSAML*
rs1260326	2	27730940	rs780094	2	27741237	0.9	t	c	0.415	2.2441	0.5218	1.70E-05	*GCKR*
rs17030845	2	43687879	rs35720761	2	43519977	0.66	t	c	0.1235	-2.3061	0.7815	0.00317	*THADA*
rs7616006	3	12267648	rs1801282	3	12393125	0.07	c	g	0.88	-3.878	0.801	$1.3x10^{-6}$	*PPARG*
rs1354034	3	56849749	rs12485738	3	56865776	0.69	a	g	0.3805	-4.8268	0.527	5.28E-20	*ARHGEF3*
rs17568628	5	76046939	rs34592828	5	75996909	0.94	a	g	0.0383	-6.5492	1.3427	1.08E-06	*IQGAP2*
rs2070729	5	131819921	rs2188962	5	131770805	0.29	t	c	0.421	-1.4387	0.521	0.005754	*C5orf56*
rs441460	6	25548288	rs742132	6	25607571	0.54	a	g	0.7008	1.7683	0.567	0.001815	*LRRC16A*
rs3819299	6	31322367	rs3819299	6	31322367	1	t	g	0.9436	-5.2144	1.1219	3.35E-06	*HLA-B*
rs399604	6	32975014	rs399604	6	32975014	1	t	c	0.5769	-2.2387	0.5226	1.84E-05	*HLA-DOA*
rs210134	6	33540209	rs210139	6	33543409	0.25	a	c	0.6099	-4.4329	0.528	4.66E-17	*BAK1*
rs9399137	6	135419018	rs4895441	6	135426573	0.91	a	g	0.7345	-5.5051	0.5818	2.99E-21	*HBS1L-MYB*
rs342275	7	106359216	rs342293	7	106372219	0.95	c	g	0.5473	3.5766	0.5151	3.83E-12	
rs6993770	8	106581528	rs2343596	8	106593207	0.73	a	c	0.3145	-2.8276	0.5549	3.47E-07	*ZFPM2*
rs6995402	8	145005561	rs11136344	8	145059425	0.22	t	c	0.5673	-2.2594	0.5156	1.18E-05	*PARP10*
rs409801	9	4744743	rs385893	9	4763176	0.46	t	c	0.4913	-5.3356	0.5128	2.36E-25	
rs13300663	9	4814948	rs10758658	9	4856877	0.52	a	g	0.1861	4.5309	0.6601	6.71E-12	*RCL1*
rs10761731	10	65027610	rs12355784	10	65121565	0.8	a	c	0.4893	3.7401	0.5149	3.75E-13	*JMJD1C*
rs505404	11	243268	rs11602954	11	202856	0.95	a	g	0.2078	4.9545	0.6359	6.66E-15	*BET1L*
rs4246215	11	61564299	rs174546	11	61569830	0.9	t	c	0.3338	1.7442	0.5445	0.001359	*FADS1*
rs4938642	11	119099906	rs2239896	11	118983434	0.57	t	c	0.0574	4.0048	1.1123	0.0003177	*C2CD2L*
rs941207	12	57023284	rs2958154	12	57065713	0.62	t	c	0.6533	-1.329	0.5384	0.01357	*PTGES3*
rs3184504	12	111884608	rs3184504	12	111884608	1	t	c	0.4952	4.5008	0.5154	2.49E-18	*SH2B3*
rs17824620	12	113100994	rs233716	12	113039943	0.26	t	c	0.5793	-2.5585	0.5259	1.15E-06	
rs7961894	12	122365583	rs7961894	12	122365583	1	t	c	0.1088	-4.5736	0.8213	2.56E-08	*WDR66*
rs7149242	14	101159416	rs12883126	14	101134008	0.32	t	c	0.769	1.2747	0.6098	0.03658	*LINC00523*
rs2297067	14	103566785	rs2297067	14	103566785	1	t	c	0.2455	1.7078	0.5977	0.004273	*EXOC3L4*
rs6065	17	4836381	rs6065	17	4836381	1	t	c	0.0838	3.3587	0.9335	0.0003205	*GP1BA*
rs397969	17	19804247	rs203462	17	19812541	0.99	t	c	0.6196	-1.2437	0.5271	0.01831	*AKAP10*
rs559972	17	27814496	rs2138852	17	27703349	0.98	t	c	0.521	-3.2703	0.5096	1.38E-10	
rs10512472	17	33884804	rs10512472	17	33884804	1	t	c	0.818	-3.2465	0.6698	1.25E-06	*SLFN14*
rs11082304	18	20720973	rs11082304	18	20720973	1	t	g	0.5018	-3.4896	0.5124	9.78E-12	*CABLES1*
rs4812048	20	55587771	rs6070697	20	57599402	1	a	g	0.1767	2.0876	0.6746	0.001969	*TUBB1*

Supplementary Table 4. Gene-based association results. We analyzed association between rare non-synonymous (missense, nonsense) and splice site variants and blood cell traits using the Burden T1 and SKAT gene-based tests. Markers with minor allele frequency <1% and <5% were considered for the Burden T1 and SKAT tests, respectively. We selected from the MHI+WHI analysis 19 genes with $P<1x10^{-4}$ and attempted to replicate them in SHIP.

				MHI+WHI		SHIP		MHI+WHI+SHIP	
Phenotype	**Test**	**Gene**	**Complete list of variants**	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***
HCT	Burden	*C7orf72*	rs182523159,chr7:50143943,rs149326344,rs61754912	4	2.32E-05	3	0.91	4	1.1E-4
HCT	SKAT	*EPO*	rs62483572, rs137953994, rs11976235, rs149431976, rs73409075	5	$4.6x10^{-5}$	4	0.033	5	$1.4x10^{-6}$
HCT	SKAT	*SLC35B3*	rs184068804,rs143437815,rs190725185,rs143539087,rs141256356,rs138204292,rs34956976,rs142783777	8	8.41E-05	5	0.84	8	0.0033
HCT	SKAT	*TUBGCP4*	chr15:43668300,chr15:43668384,rs188822226,rs191224065,chr15:43678503,rs183013671,chr15:43693969,chr15:43693987	8	8.53E-05	4	0.61	8	0.0029
HCT	SKAT	*RUFY3*	chr4:71634272,rs182772740,chr4:71639274,rs140962303,chr4:71659504,rs149107016,chr4:71659621,chr4:71665919,chr4:71672317,rs114009356	9	9.91E-05	8	1	10	0.024
HGB	Burden	*HBB*	rs33946267, rs33947415, rs1135071, rs33971440, rs33974228	4	$9.2x10^{-5}$	2	0.23	5	$4.3x10^{-5}$
HGB	Burden	*SAMD9L*	chr7:92760773,rs139478067,chr7:92761203,rs141306896,chr7:92762757,chr7:92762855,rs147903234,chr7:92763351,rs73710963,rs142436298,rs143593856,chr7:92763736,rs140536419,rs150070697,rs151204118,rs141294380,chr7:92764714,rs144605831,rs149275726,rs148147724,rs60838691	17	5.33E-05	13	0.92	21	0.0036
HGB	SKAT	*EPO*	rs62483572, rs137953994, rs11976235, rs149431976, rs73409075	5	$7.8x10^{-5}$	4	0.072	5	$4.0x10^{-6}$
HGB	SKAT	*RUFY3*	chr4:71634272,rs182772740,chr4:71639274,rs140962303,chr4:71659504,rs149107016,chr4:71659621,chr4:71665919,chr4:71672317,rs114009356	9	3.86E-05	8	0.96	10	0.022
HGB	SKAT	*ADAP1*	rs145174713,chr7:944803,rs147172089,rs139497517,rs145788459,rs141244581	5	6.92E-05	4	0.22	6	5.72E-05
HGB	SKAT	*TUBGCP4*	chr15:43668300,chr15:43668384,rs188822226,rs191224065,chr15:43678503,rs183013671,chr15:43693969,chr15:43693987	8	9.28E-05	4	0.45	8	0.012
WBC	Burden	*CXCR2*	rs139809702, rs75759064, rs55799208, rs10201766, rs61733609, rs200726461, rs2228413, rs199632317	8	$1.6x10^{-6}$	6	$8.3x10^{-9}$	8	$2.6x10^{-13}$
WBC	Burden	*S1PR4*	rs150230356,rs147876709,rs3826936,rs3746072,chr19:3179931	5	2.76E-06	5	0.20	5	0.013
WBC	Burden	*RANGAP1*	chr22:41647049,rs139571477,rs114673061,rs143560227,rs143866878,chr22:41660673,rs116510625	7	2.72E-05	5	0.95	7	0.0010
WBC	Burden	*IGFL1*	rs147938784,chr19:46733532,rs187146043,rs190483911	4	5.17E-05	2	0.30	4	0.0018
WBC	SKAT	*KRT35*	chr17:39635117,chr17:39635169,rs143386151,rs192328036,rs185963879,chr17:39636962,chr17:39636983,chr17:39637145,rs139838007,chr17:39637271,chr17:39637328	11	2.33E-05	9	0.67	11	4.3E-04
WBC	SKAT	*S1PR4*	rs150230356,rs147876709,rs3826936,rs3746072,chr19:3179931	5	3.94E-05	5	0.14	5	0.016
WBC	SKAT	*MFAP5*	rs148770543,rs139330366	2	8.80E-05	2	0.57	2	8.89E-05
PLT	Burden	*RBM46*	chr4:155719287,rs139756307,rs145321511,chr4:155720617,rs141427057,chr4:155749134,rs147971341	6	2.75E-05	5	0.57	7	4.8E-04
PLT	Burden	*FZD3*	rs140115204,rs139570173,chr8:28413272,rs146950723,rs147574227	5	8.55E-05	2	0.84	5	3.2E-04
PLT	SKAT	*SH2B3*	rs183913232, rs147341899, rs149554298, rs140042617, rs147318193, rs148636776, rs181879167, rs72650673, rs200907236, rs200567433, rs74163669, rs200089302, rs140649197, rs72650662, rs148791142, rs199803113	14	$2.2x10^{-7}$	8	$2.9x10^{-4}$	14	$6.7x10^{-10}$
PLT	SKAT	*TUBB1*	rs145280665, rs144337011, rs415064, rs150551805, rs41303899, rs200931731, rs202177647, rs146846923, rs62639975, rs192115302, rs62639974, rs115253190, chr20:57599645, rs138642232	13	$2.2x10^{-5}$	8	$2.9x10^{-4}$	14	$6.7x10^{-10}$
PLT	SKAT	*CTSC*	rs3888798,rs146182103,chr11:88027277,chr11:88027365,rs142378484	5	2.70E-05	1	0.61	5	3.55E-06

Supplementary Table 5. Clinical characteristics of 19 participants from the MHI Biobank and WHI study that carry *JAK2* p.Val617Phe alleles based on whole blood DNA genotyped on the exome array.

ID	Baseline				Year 3 (follow-up)				Coronary heart disease	Stroke	Leukemia	Cause of death	Number of *JAK2* p.Val617Phe alleles
	Hematocrit (%)	Hemoglobin (g/dl)	Platelet (10^9/l)	White blood cell (10^9/l)	Hematocrit (%)	Hemoglobin (g/dl)	Platelet (10^9/l)	White blood cell (10^9/l)					
1	46	15.9	658	9	50.6	16.8	662	12.7	No	No	No	NA	1
2	40	12.5	502	5.9	NA	NA	NA	NA	No	Yes	No	Other Cancer	1
3	46	13	796	23.4	54	15.4	647	67.5	No	Yes	No	Cerebrovascular	2
4	40.7	12.9	482	10.7	NA	NA	NA	NA	No	Yes	Yes	NA	1
5	41.1	14	437	3.9	NA	NA	NA	NA	Yes	Yes	No	NA	1
6	46.6	16.1	310	8.3	47.2	15.4	259	11.5	No	Yes	No	NA	1
7	48	15.9	357	6.9	NA	NA	NA	NA	No	No	No	NA	1
8	45.4	14.8	411	8.6	NA	NA	NA	NA	Yes	No	No	Other Known Cause	1
9	39.9	13.5	227	5.4	NA	NA	NA	NA	No	Yes	No	NA	1
10	43.8	14.8	365	5.8	NA	NA	NA	NA	Yes	No	No	NA	1
11	42.5	14.7	279	5.8	NA	NA	NA	NA	Yes	No	No	NA	1
12	49.5	16.5	529	7.3	NA	NA	NA	NA	No	No	No	NA	1
13	42.5	14.2	411	7	NA	NA	NA	NA	No	No	No	NA	1
14	49.3	16.7	266	9.1	NA	NA	NA	NA	No	No	No	NA	1
15	49.4	16.4	697	8	NA	NA	NA	NA	Yes	No	No	NA	1
16	45.1	14.2	412	9.5	NA	NA	NA	NA	No	Yes	No	NA	1
17	37.5	13.1	239	11.4	NA	NA	NA	NA	Yes	No	No	NA	1
18	42.9	14.4	201	11.7	NA	NA	NA	NA	Yes	Yes	No	NA	1
19	42.3	14.6	291	14.5	NA	NA	NA	NA	Yes	No	No	NA	1

Supplementary Table 6. Missense variants in *CXCR2* and white blood cell count. For each variant by study, we list allele count and frequency, white blood cell mean (s.d.) and association P-value using the score test implemented in raremetal (see above). The mean (s.d.) white blood cell count in each study is ($x10^9$/l): MHI 7.1 (2.2), WHI 6.1 (1.7), SHIP 6.4 (1.8).

	MHI (N=6,796)			**WHI (N=18,018)**			**SHIP (N=6,526)**			
***CXCR2* variant**	**Allele count (het/homo)**	**Allele frequency**	**WBC ($x10^9$/l)**	**Allele count (het/homo)**	**Allele frequency**	**WBC ($x10^9$/l)**	**Allele count (het/homo)**	**Allele frequency**	**WBC ($x10^9$/l)**	**Association *P*-value**
rs139809702 (p.M6R)	Monomorphic			1 (1/0)	$2.8x10^{-5}$	6.2 (NA)	Monomorphic			0.91
rs75759064 (p.A37T)	2 (2/0)	$1.5x10^{-4}$	6.6 (1.7)	5 (5/0)	$1.4x10^{-4}$	6.3 (1.4)	3 (3/0)	$2.3x10^{-4}$	6.3 (1.2)	0.63
rs55799208 (p.R153H)	28 (28/0)	0.0021	6.4 (1.5)	99 (99/0)	0.0027	5.6 (1.8)	50 (50/0)	0.0038	5.7 (1.7)	$2.5x10^{-6}$
rs10201766 (p.R236C)	7 (7/0)	$5.2x10^{-4}$	7.3 (1.6)	58 (58/0)	0.0016	5.8 (1.5)	66 (66/0)	0.0051	5.9 (1.7)	0.0036
rs61733609 (p.R248Q)	26 (26/0)	0.0019	6.3 (1.3)	47 (47/0)	0.0013	5.3 (1.4)	51 (51/0)	0.0039	5.4 (1.6)	$2.2x10^{-9}$
rs200726461 (p.R289C)	Monomorphic			7 (7/0)	$1.9x10^{-4}$	6.2 (0.8)	2 (2/0)	$1.5x10^{-4}$	5.1 (1.6)	0.95
rs2228413 (p.R294Q)	24 (22/1)	0.0017	7.0 (2.1)	4 (4/0)	$1.1x10^{-4}$	5.8 (1.9)	2 (2/0)	$1.5x10^{-4}$	6.4 (0.7)	0.87
rs199632317(p.H332R)	Monomorphic			3 (3/0)	$8.3x10^{-5}$	5.0 (2.1)	Monomorphic			0.15

Supplementary Table 7. Association of *CXCR2* missense variants with white blood cell subtypes in the MHI Biobank (N=6,796) and a subset of SHIP (N=3,438). WBC subtypes are not available for WHI. We present burden T1 results for the *CXCR2* missense variants listed in **Table 2** of the main text (rs139809702, rs75759064, rs55799208, rs10201766, rs61733609, rs200726461, rs2228413, rs199632317). The effect size is calculated using a meta-analysis of the single variant results and the direction of the effect is for the rare alleles. For basophils and eosinophils, the units are standard deviations of counts. For lymphocytes, monocytes and neutrophils, the units are $\log_{10}(x10^9/l)$.

Phenotype	**MHI Burden T1**		**SHIP Burden T1 *P*-value**	
	Effect size	***P*-value**	**Effect size**	***P*-value**
Basophils	0.004	0.97	-0.268	0.012
Eosinophils	-0.165	0.12	-0.204	0.056
Lymphocytes	0.526	0.44	-1.533	0.061
Monocytes	-1.222	0.087	-3.490	$1.2x10^{-5}$
Neutrophils	-1.697	0.013	-3.493	$6.7x10^{-7}$

SUPPLEMENTARY NOTE

Coding variants associated with hematological traits identified by GWAS or within candidate genes

The single variant meta-analyses of association results for MHI and WHI confirmed many single nucleotide polymorphisms (SNPs) previously associated with blood cell traits by GWAS (**Supplementary Fig. 1** and **Supplementary Table 3**). Although most GWAS SNPs are non-coding, we found three examples where a coding or splice site variant in linkage disequilibrium (LD) with a sentinel SNP could explain a platelet GWAS signal: a 5' donor splice variant in *GCSAML* (rs56043070, minor allele frequency (MAF)=7%, P=1.3x10^{-11}), a missense variant in *TUBB1* (rs6070697, p.Arg307His, MAF=18%, P=0.002), and the *PPARG* p.Pro12Ala missense variant also associated with type 2 diabetes (rs1801282, MAF=12%, P=1.3x10^{-6})(**Supplementary Table 3**). Mutations in *TUBB1* (which encodes a megakarocyte lineage-specific isoform of beta-tubulin) are linked to low platelet counts in mouse models, and cause autosomal dominant macrothrombocytopenia in humans (MIM 612901) and Cavalier King Charles Spaniel dogs[8]. *PPARG* is highly expressed in platelets[9] and thiazolidinediones, PPARγ agonists that are used to treat type 2 diabetes, also improve platelet function and coagulation[10,11].

To identify coding variants associated with blood cell traits that lie outside of known GWAS loci, we selected 42 markers with P<1x10^{-5} (along with P<1x10^{-4} for coding variants in *PKLR*, *BUB1B* and *TUBB1*, three strong candidate genes) in the MHI+WHI discovery sample for validation in SHIP (**Supplementary Table 2**). We identified six new genetic associations with hematocrit, hemoglobin and platelet traits that met our predetermined statistical threshold of P<6.8x10^{-8} (discussed in the main text). There are also several coding variants in strong candidate genes that did not reach our stringent statistical threshold (**Supplementary Table 2**). This includes a rare missense variant in *PKLR* (rs61755431, p.Arg538Gln or p.Arg569Gln, MAF=0.3%) associated with hematocrit (P=2.3x10^{-5}) and hemoglobin (P=1.1x10^{-6}) levels. *PKLR* encodes a pyruvate kinase present in red blood cells and rare familial mutations in this gene cause hemolytic anemia (MIM 609712). We found an association between a common missense variant (rs1260326, p.Leu446Pro, MAF=43%) in the glucokinase regulator (*GCKR*) gene and WBC count (P=8.5x10^{-8}): although the same SNP has been associated with many human traits by GWAS, including metabolic[12], lipid[13] and platelet[7] traits, this is the first report of an association with WBC.

We can increase power to find associations between complex traits and rare variants by combining them into functional units (*e.g.* genes)[14]. We used two tests to detect association between rare/low-frequency variants and hematological traits: a burden (T1) test and the sequence kernel association test (SKAT), and a P<3.9x10^{-7} threshold to declare significance. We only considered nonsense, missense and splice site variants, and used a minor allele frequency inclusion cutoff of 1% and 5% for the T1 test and SKAT, respectively. Using this strategy, we prioritized from the MHI+WHI meta-analysis 23 gene-phenotype pairs for replication in SHIP (**Supplementary Table 4** and **Supplementary Figs. 2-3**). Of the top five genes, four were already highlighted by our analysis of rare/low-

frequency variants (*EPO*, *HBB*, *SH2B3* and *TUBB1*) and when we excluded the top associated rare/low-frequency variant, the gene-based signal disappeared (**Table 2**).

References

1. Syed, R.S. *et al.* Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* **395**, 511-6 (1998).

2. Hu, J. & Hubbard, S.R. Structural basis for phosphotyrosine recognition by the Src homology-2 domains of the adapter proteins SH2-B and APS. *J Mol Biol* **361**, 69-79 (2006).

3. Park, S.H. *et al.* Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491**, 779-83 (2012).

4. Fan, G.H., Yang, W., Wang, X.J., Qian, Q. & Richmond, A. Identification of a motif in the carboxyl terminus of CXCR2 that is involved in adaptin 2 binding and receptor internalization. *Biochemistry* **40**, 791-800 (2001).

5. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-5 (2012).

6. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

7. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol* **1019**, 215-36 (2013).

8. Forand, A. *et al.* EKLF-driven PIT1 expression is critical for mouse erythroid maturation in vivo and in vitro. *Blood* **121**, 666-78 (2013).

9. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* **10**, R130 (2009).

10. van der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369-75 (2012).

11. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).

12. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

13. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).

14. Bansal, V., Libiger, O., Torkamani, A. & Schork, N.J. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet* **11**, 773-85 (2010).