nature genetics

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

Paul L Auer[1,2], Alexander Teumer[3], Ursula Schick[2], Andrew O'Shaughnessy[4], Ken Sin Lo[5], Nathalie Chami[5], Chris Carlson[2], Simon de Denus[5–7], Marie-Pierre Dubé[5–7], Jeff Haessler[2], Rebecca D Jackson[8], Charles Kooperberg[2], Louis-Philippe Lemieux Perreault[5], Matthias Nauck[9], Ulrike Peters[2,10], John D Rioux[5–7], Frank Schmidt[3], Valérie Turcot[5], Uwe Völker[3], Henry Völzke[11], Andreas Greinacher[12], Li Hsu[2], Jean-Claude Tardif[5–7], George A Diaz[4,13,14], Alexander P Reiner[2,10,14] & Guillaume Lettre[5–7,14]

Hematological traits are important clinical parameters. To test the effects of rare and low-frequency coding variants on hematological traits, we analyzed hemoglobin concentration, hematocrit levels, white blood cell (WBC) counts and platelet counts in 31,340 individuals genotyped on an exome array. We identified several missense variants in *CXCR2* associated with reduced WBC count (gene-based $P = 2.6 \times 10^{-13}$). In a separate family-based resequencing study, we identified a *CXCR2* frameshift mutation in a pedigree with congenital neutropenia that abolished ligand-induced CXCR2 signal transduction and chemotaxis. We also identified missense or splice-site variants in key hematopoiesis regulators (*EPO*, *TFR2*, *HBB*, *TUBB1* and *SH2B3*) associated with blood cell traits. Finally, we were able to detect associations between a rare somatic *JAK2* mutation (encoding p.Val617Phe) and platelet count ($P = 3.9 \times 10^{-22}$) as well as hemoglobin concentration ($P = 0.002$), hematocrit levels ($P = 9.5 \times 10^{-7}$) and WBC count ($P = 3.1 \times 10^{-5}$). In conclusion, exome arrays complement genome-wide association studies in identifying new variants that contribute to complex human traits.

The proliferation and differentiation of hematopoietic progenitor cells into mature blood cells is a tightly regulated process. Erythrocyte, WBC and platelet counts are used in medicine as diagnostic and prognostic biomarkers. Interindividual variation in quantitative blood cell traits is heritable, and genome-wide association studies (GWAS) have implicated hundreds of loci[1–3]. The development of new genotyping arrays that target protein-coding variation offers new opportunities to assess the role of rare (defined here as having minor allele frequency (MAF) of <0.1%) and low-frequency (0.1% ≤ MAF < 1%) coding variants in human complex trait genetics. We genotyped 6,796 participants from the Montreal Heart Institute (MHI) Biobank and 18,018 women from the Women's Health Initiative (WHI) on the Illumina HumanExome BeadChip and analyzed associations between 183,585 polymorphic variants and 4 blood cell phenotypes: hemoglobin concentration, hematocrit levels, WBC count and platelet count. We carried forward variants in 19 genes (gene-based $P < 1 \times 10^{-4}$) and 42 single markers ($P < 1 \times 10^{-5}$; along with $P < 1 \times 10^{-4}$ for coding variants in 3 strong candidate genes: *PKLR*, *BUB1B* and *TUBB1*) for validation of our top association results in 6,526 participants from the Study of Health in Pomerania (SHIP) (**Supplementary Table 1**). We identified six genetic associations for platelet count (*TUBB1*, *SH2B3* and *JAK2*), hematocrit levels and hemoglobin concentration (*EPO* and *HBB*), and WBC count (*CXCR2*) that met our predetermined exome-wide significance threshold ($P < 6.8 \times 10^{-8}$ for single variants and $P < 3.9 \times 10^{-7}$ for gene-based tests) (**Tables 1** and **2**, **Supplementary Figs. 1–3** and **Supplementary Tables 2–4**). These findings highlight the role of rare and low-frequency coding variants in complex traits in large cohorts of individuals not selected on the basis of having a hematological disorder.

We found an association between a low-frequency missense variant (MAF = 0.45%) in the erythropoietin gene *EPO* (rs62483572; p.Asp70Asn), which encodes the main cytokine that controls erythrocyte production, and lower hematocrit levels and hemoglobin concentration (**Fig. 1a** and **Table 1**). Carriers of the *EPO* missense variant had 0.35 g/dl lower hemoglobin concentration and 1% lower hematocrit levels than non-carriers. To our knowledge, the variant encoding p.Asp70Asn is the first naturally occurring *EPO* coding variant associated with any hematological phenotype. In the mature EPO protein, p.Asp70Asn (amino acid 43 after cleavage of the signal peptide) is part

[1]School of Public Health, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin, USA. [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [3]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [5]Montreal Heart Institute, Montreal, Quebec, Canada. [6]Faculty of Medicine, Université de Montréal, Montreal, Québec, Canada. [7]Faculty of Pharmacy, Université de Montréal, Montreal, Québec, Canada. [8]Division of Endocrinology, Diabetes and Metabolism, Ohio State University, Columbus, Ohio, USA. [9]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [10]Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA. [11]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [12]Institute for Immunology and Transfusion Medicine, University Medicine Greifswald, Greifswald, Germany. [13]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [14]These authors jointly directed this work. Correspondence should be addressed to G.A.D. (george.diaz@mssm.edu), A.P.R. (apreiner@u.washington.edu) or G.L. (guillaume.lettre@umontreal.ca).

Received 13 November 2013; accepted 31 March 2014; published online 28 April 2014; doi:10.1038/ng.2962

Table 1 Newly discovered genome-wide significant variants associated with blood cell traits

Variant	Chromosome (position)	A1/A2	Stage	Frequency (A1)	β (s.e.)	P	Heterogeneity P	Gene	Annotation
Hematocrit									
rs62483572	7 (100,319,633)	A/G	Discovery	0.0045	−1.07 (0.21)	3.5×10^{-7}	NA	*EPO*	Missense (p.Asp70Asn)
			Replication	0.0041	−1.11 (0.40)	0.0053	NA		
			Combined	NA	−1.08 (0.19)	6.4×10^{-9}	0.94		
Hemoglobin									
rs62483572	7 (100,319,633)	A/G	Discovery	0.0045	−0.35 (0.07)	9.5×10^{-7}	NA	*EPO*	Missense (p.Asp70Asn)
			Replication	0.0041	−0.35 (0.14)	0.011	NA		
			Combined	NA	−0.35 (0.06)	3.4×10^{-8}	0.96		
rs33971440	11 (5,248,159)	T/C	Discovery	0.0002	−2.41 (0.44)	3.7×10^{-8}	NA	*HBB*	5′ donor splice site (exon 2: c.92+1G>A)
			Replication	Monomorphic			NA		
			Combined	ND					
Platelets									
rs77375493	9 (5,073,770)	T/G	Discovery	0.0005	124.09 (12.83)	3.9×10^{-22}	NA	*JAK2*	Missense (p.Val617Phe)
			Replication	0.0041	16.41 (8.61)	0.057	NA		
			Combined	NA	49.86 (7.15)	3.1×10^{-12}	3.2×10^{-12}		
rs1465788	14 (69,263,599)	T/C	Discovery	0.27	−2.75 (0.58)	2.1×10^{-6}	NA	NA	639 bp upstream of *ZFP36L1*
			Replication	0.27	−3.06 (1.01)	0.0024	NA		
			Combined	NA	−2.82 (0.50)	1.9×10^{-8}	0.78		
rs41303899	20 (57,598,808)	A/G	Discovery	0.0016	−27.70 (6.72)	3.7×10^{-5}	NA	*TUBB1*	Missense (p.Gly109Glu)
			Replication	0.0024	−42.63 (9.07)	2.7×10^{-6}	NA		
			Combined	NA	−32.99 (5.40)	9.9×10^{-10}	0.19		

Variants are shown that reached $P < 6.8 \times 10^{-8}$ in the combined analysis and are either outside the GWAS loci or independent of the GWAS sentinel SNPs for hematological traits. Sample size: discovery MHI and WHI cohorts, $n = 24{,}814$; replication SHIP cohort, $n = 6{,}526$; combined, $n = 31{,}340$. The direction of the effect sizes (β) is for the A1 allele. A1 is the alternate allele, and A2 is the reference allele. Effect sizes and standard error (s.e.) values are in the following units: hematocrit, percent; hemoglobin, g/dl; platelet count, $\times 10^9$ cells/l. ND, not determined because the marker is monomorphic in SHIP; NA, not applicable. Genomic positions are on human genome build hg19.

of the high-affinity receptor-binding site (**Supplementary Fig. 4**)[4]. Rare gain-of-function coding variants in the erythropoietin receptor gene (*EPOR*) have been associated with familial erythrocytosis[5]. *EPO* is located near the gene encoding transferrin receptor 2 (*TFR2*), a gene that is important for iron transport and hemoglobin production. Noncoding SNPs in the *TFR2-EPO* region are associated with red blood cell (RBC) phenotypes[3]. Analysis conditioning on the sentinel SNP in *TFR2*, rs7385804, indicated that the rare missense variant in *EPO* is independently associated with hematocrit levels and hemoglobin concentration (**Fig. 1a**). This conditional analysis also identified a low-frequency 5′ donor splice-site variant of *TFR2* (rs139178017) that was independently associated with higher hematocrit levels and hemoglobin concentration (**Fig. 1a** and **Supplementary Table 2**). Other *TFR2* splice-site variants have been reported in individuals with atypical hemochromatosis (negative for *HFE* mutation)[6,7].

A rare 5′ donor splice-site variant in the β-globin gene (*HBB*; rs33971440; MAF = 0.02%) was associated with lower hemoglobin concentration and hematocrit levels (**Table 1** and **Supplementary Table 2**). This rare *HBB* variant was previously identified in Mediterranean individuals with β^0 thalassemia[8], which may explain why it is present in North American cohorts (MHI and WHI), owing to immigration, but absent from the SHIP cohort from northern Germany.

Using criteria from the World Health Organization to define anemia (hemoglobin concentration of <12 g/dl in women and <13 g/dl in men), we confirmed that the *EPO* variant encoding p.Asp70Asn was associated with clinical anemia in the combined analysis of the MHI and WHI cohorts (1,866 cases and 22,397 controls; odds ratio (OR) = 1.7; $P = 0.008$). The rare *HBB* splice-site variant was likewise strongly associated with clinical anemia (OR = 36.1; $P = 1.1 \times 10^{-5}$).

Table 2 Gene-based association results

Phenotype	Test	Gene	MHI + WHI ($n = 24{,}814$) Number of variants	MHI + WHI ($n = 24{,}814$) P	P_{GWAS}[a]	$P_{low\text{-}freq}$[b]	SHIP ($n = 6{,}526$) Number of variants	SHIP ($n = 6{,}526$) P	Combined ($n = 31{,}340$) Number of variants	Combined ($n = 31{,}340$) P
Hematocrit levels	SKAT	*EPO*	5	4.6×10^{-5}	1.0×10^{-4}	0.13	4	0.033	5	1.4×10^{-6}
Hemoglobin concentration	Burden T1	*HBB*	4	9.2×10^{-5}	NA	0.031	2	0.23	5	4.3×10^{-5}
Hemoglobin concentration	SKAT	*EPO*	5	7.8×10^{-5}	9.6×10^{-5}	0.16	4	0.072	5	4.0×10^{-6}
WBC count	Burden T1	*CXCR2*	8	1.6×10^{-6}	NA	8.2×10^{-4}	6	8.3×10^{-9}	8	2.6×10^{-13}
Platelet count	SKAT	*SH2B3*	14	2.2×10^{-7}	3.9×10^{-8}	0.078	11	0.024	16	8.3×10^{-7}
Platelet count	SKAT	*TUBB1*	13	2.2×10^{-5}	3.1×10^{-5}	0.018	8	2.9×10^{-4}	14	6.7×10^{-10}

Listed are genes that are exome-wide significant ($P < 3.9 \times 10^{-7}$) or have both strong statistical evidence of association and biological candidacy. The remaining gene-based results are given in **Supplementary Table 4**. We analyzed association between rare nonsynonymous (missense, nonsense) and splice-site variants and blood cell traits using the Burden T1 and SKAT gene-based tests. When applicable, we also conditioned on GWAS sentinel SNPs or excluded low-frequency variants strongly associated with the traits. NA, not applicable.

[a]For *EPO*, *SH2B3* and *TUBB1*, we conditioned, respectively, on rs7385804, rs3184504 and rs6070697. [b]For *EPO*, we excluded rs62483572; for *HBB*, we excluded rs33971440; for *CXCR2*, we excluded rs61733609; for *SH2B3*, we excluded rs72650673; and, for *TUBB1*, we excluded rs41303899.

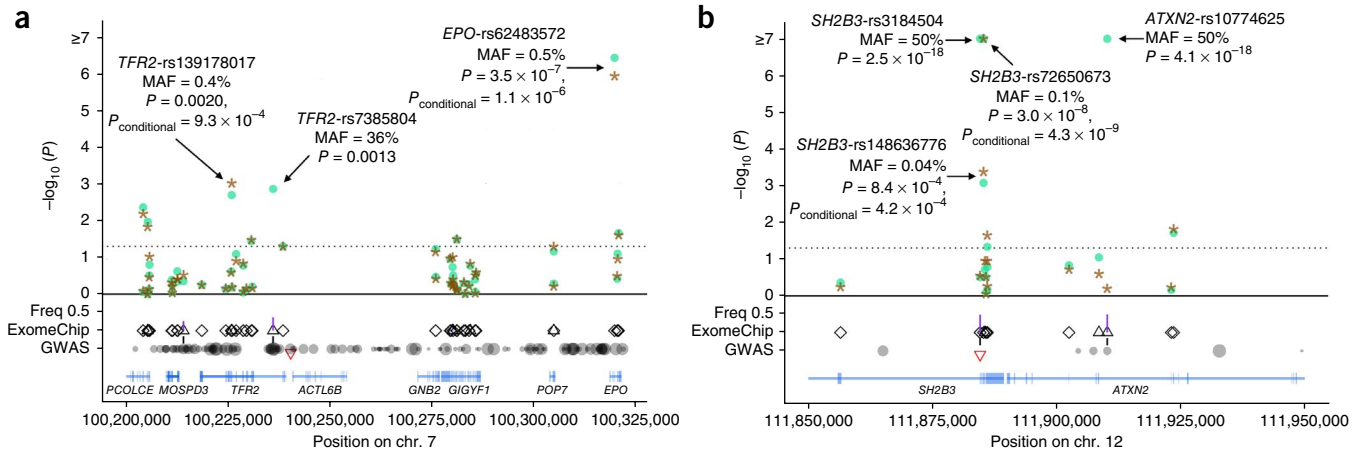

Figure 1 Association results in the meta-analysis of the MHI and WHI cohorts (n = 24,814) for hematocrit levels at the *TFR2-EPO* locus on chromosome 7 and for platelet count at the *SH2B3* locus on chromosome 12. The x axes correspond to genomic coordinates. (**a**) For the *TFR2-EPO* locus and hematocrit levels, we provide association results for each variant in the region before (circles) and after (stars) conditioning on the GWAS sentinel SNP at the locus (rs7385804). (**b**) For *SH2B3* and platelet count, we provide association results without (circles) and with (stars) adjustment for genotypes at the GWAS sentinel SNP rs3184504. In the GWAS row, we represent all markers in LD with the GWAS sentinel SNP (marked by an inverted triangle) on the basis of data from European populations in the 1000 Genomes Project (the size of the circle is proportional to the strength of LD). We added a vertical line connecting the GWAS and ExomeChip rows if GWAS LD proxies were also present on the ExomeChip array. In the ExomeChip row, all markers analyzed in the region are represented as triangles (intronic, intergenic and synonymous variants) or diamonds (missense, nonsense and splice-site variants). Vertical lines connecting the ExomeChip and Freq 0.5 rows represent allele frequencies in the meta-analysis of the MHI and WHI cohorts.

Mutations in *TUBB1* (which encodes a megakaryocyte-specific form of β-tubulin, a microtubule protein involved in pro-platelet production) cause autosomal dominant macrothrombcytopenia in humans (MIM 612901) and Cavalier King Charles Spaniel dogs[9]. We report a new low-frequency *TUBB1* missense variant (rs41303899; p.Gly109Glu; MAF = 0.16%) associated with ~33,000 cells/μl lower mean platelet count (**Table 1**). The association between rs41303899 and platelet count was independent of a common *TUBB1* missense variant (rs6070697; p.Arg307His; MAF = 18%; **Supplementary Table 3**) previously associated with mean platelet volume (for rs41303899, meta-analysis $P = 3.7 \times 10^{-5}$, meta-analysis conditional on rs6070697 $P = 2.4 \times 10^{-4}$).

A somatic *JAK2* mutation encoding p.Val617Phe is a key driver of myeloproliferative neoplasms (MPNs)[10–13]. In our discovery sample, this *JAK2* mutation was strongly associated with platelet count ($P = 3.9 \times 10^{-22}$) and was also associated with hematocrit levels ($P = 9.5 \times 10^{-7}$), hemoglobin concentration ($P = 0.002$) and WBC count ($P = 3.1 \times 10^{-5}$), highlighting the pan-lineage effects of this somatic gain-of-function mutation (**Table 1**). The *JAK2* mutation encoding p.Val617Phe was rare in the MHI and WHI cohorts (MAF = 0.05%), consistent with previous frequency estimates obtained by whole-exome sequencing[14] or allele-specific PCR[15] using peripheral blood samples from unselected individuals. Clinical characteristics, serial blood counts and clinical follow-up information obtained for the 19 mutation carriers in the MHI and WHI studies suggest early-stage MPN in these individuals and also exemplify the phenotypic diversity associated with this somatic *JAK2* mutation (**Supplementary Table 5**). Notably, the frequency of the *JAK2* variant encoding p.Val617Phe appeared to be tenfold higher in SHIP (MAF = 0.4%) (**Table 1**); however, this spuriously high frequency resulted from genotype miscalling, as inspection of the genotyping intensity plots did not identify clearly distinguishable clusters (**Supplementary Fig. 5**). These observations are consistent with both variable sensitivity in the detection of the *JAK2* variant encoding p.Val617Phe across genotyping platforms and heterogeneity in allelic burden across individuals[16,17]. These issues raise a number of complexities to be considered in the decision to return the results of such incidental findings to research participants.

SH2B3 encodes the adaptor protein LNK, which regulates T and B cell development and myelopoiesis[18]. LNK inhibits the JAK2-STAT pathway and other downstream signaling to modulate hematopoietic cytokine receptor signaling[19]. Both somatic and germline mutations in *SH2B3* have been reported in individuals with MPNs[20,21]. In addition to confirming the association of a missense SNP (rs3184504; p.Trp262Arg; MAF = 50%) with all four hematological traits ($P < 4.0 \times 10^{-10}$; **Supplementary Table 3**)[1–3], we identified two independent *SH2B3* missense variants associated with higher platelet count (rs148636776, p.Glu395Lys, MAF = 0.04% and rs72650673, p.Glu400Lys, MAF = 0.1%) (**Fig. 1b**). These two variants affect the functional Src homology 2 domain (SH2) that interacts with JAK2 (**Supplementary Fig. 6**). The variant encoding p.Glu400Lys was previously identified in ~5% of individuals with idiopathic erythrocytosis[22]. However, in our data set, there was no association between the variant encoding p.Glu400Lys or the new missense variant encoding p.Glu395Lys and hematocrit levels or hemoglobin concentration ($P > 0.10$). The *SH2B3* variants encoding p.Glu395Lys and p.Glu400Lys were not associated with RBC count in the MHI Biobank ($P > 0.3$). Unfortunately, data on RBC count are not available in the WHI study.

In addition to the new associations described above, we confirmed many SNPs previously associated with blood cell traits by GWAS (**Supplementary Table 3**). The only new common variant identified, rs1465788 (MAF = 27%), was located in the 5′ flanking region of *ZFP36L1* and was associated with ~3,000 cells/μl lower mean platelet count (**Table 1**). *ZFP36L1* encodes a member of a family of mRNA destabilization proteins involved in regulation of the self-renewal and differentiation of hematopoietic cells[23]. The nearest GWAS index SNP for platelet count is rs11627546, which is located 1.1 Mb away and is not in linkage disequilibrium (LD) with rs1465788 ($r^2 = 0.003$ in Europeans)[2]. According to Encyclopedia of DNA Elements (ENCODE) data from myeloid K562 cells, the rs1465788 variant lies in a nucleosome-depleted region and is in strong LD ($r^2 > 0.90$ in

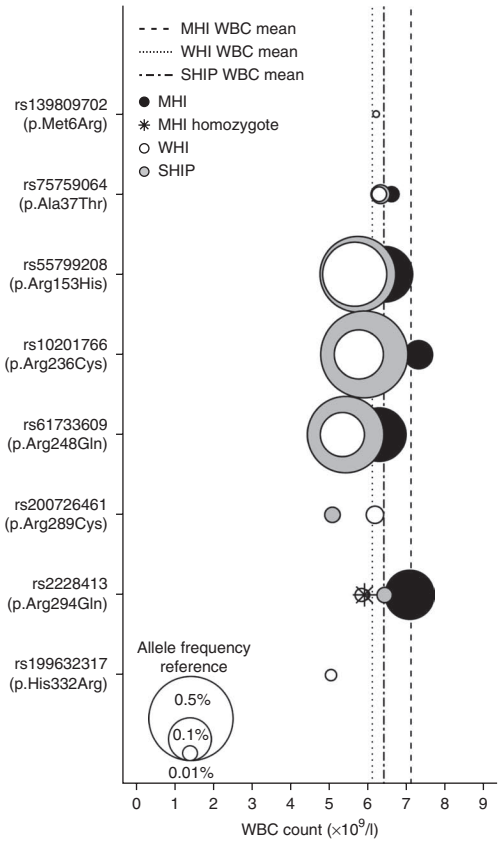

Figure 2 Rare and low-frequency missense variants in *CXCR2* are associated with lower WBC count. A vertical line represents the mean WBC count for each study. The middle of each color-coded circle corresponds to the mean WBC count for individuals who carry the corresponding missense variant, and the size of the circles is correlated with allele frequency. The three circles in the lower left corner are provided as references and represent variants with the indicated MAF values. For rs2228413 (p.Arg294Gln), there is a participant from MHI who is a homozygote for the rare allele (star).

Europeans) with several common variants in the promoter of *ZFP36L1* (**Supplementary Fig. 7**). Unfortunately, there are no megakaryocyte or platelet expression quantitative trait locus (eQTL) data sets publicly available. Using a large eQTL study of monocytes from 1,490 participants[24], we were not able to identify eQTLs for *ZFP36L1*. However, according to analyses performed in lymphoblastoid cell lines[25], rs1465788 is in LD with rs10873217 ($r^2 = 0.38$ in Europeans), an eQTL for *ZFP36L1*. rs1465788 is also located 77 kb downstream of *ACTN1*, a gene mutated in macrothrombocytopenia (MIM 615193), but is neither in LD with markers near *ACTN1* nor an eQTL for *ACTN1* expression. rs1465788 or markers in LD with it have been associated by GWAS with autoimmune disorders[26–28]. Therefore, it is also plausible that the reason for lower platelet count may involve platelet autoantibody formation, which is a common mechanism of thrombocytopenia in adults[29]. Several additional common, low-frequency and rare coding variants did not reach our stringent statistical threshold (**Supplementary Tables 2** and **3**) but are located in strong candidate genes, have been associated with other hematological traits or are in LD with a sentinel SNP and could thus explain GWAS signals. These findings are detailed further in the **Supplementary Note**.

The associations for rare and low-frequency coding variants described here represent examples where a single rare variant of large effect dominates phenotypic association. We report an instance of a newly identified signal for WBC count in a gene previously undetected through GWAS where the association comprises multiple rare and low-frequency missense variants in *CXCR2* (combined gene-based $P = 2.6 \times 10^{-13}$; **Table 2** and **Supplementary Table 4**). The gene-based signal is driven mainly by three independent low-frequency *CXCR2* missense variants (rs55799208, p.Arg153His; rs10201766, p.Arg236Cys; and rs61733609, p.Arg248Gln), each associated with lower WBC count (**Fig. 2** and **Supplementary Table 6**; conditional analysis in **Table 2**). The p.Arg153His, p.Arg236Cys and p.Arg248Gln alterations are located within the second and third intracellular loops of CXCR2, which are important for G protein interactions and receptor activation (**Supplementary Fig. 8**)[30–32]. Moreover, the Ala249 residue adjacent to p.Arg248Gln is important for CXCR2 intracellular signaling and also represents an allosteric binding site for the CXCR2 antagonists currently undergoing clinical trials[33]. Upon CXCR2 binding of the chemokine CXCL2, CXCR2 signaling promotes neutrophil release from the bone marrow, thereby elevating blood neutrophil counts[34]. Accordingly, when we analyzed the association between *CXCR2* missense variants and the counts for different WBC subtypes, the strongest association was with neutrophil count (**Supplementary Table 7**). Common variants of *CXCL2* are associated with WBC, primarily neutrophil, count[35].

In mice, neutrophils lacking CXCR2 are preferentially retained in the bone marrow[36]. A congenital neutropenia in humans resembling this condition, termed myelokathexis, can occur in isolation or as a

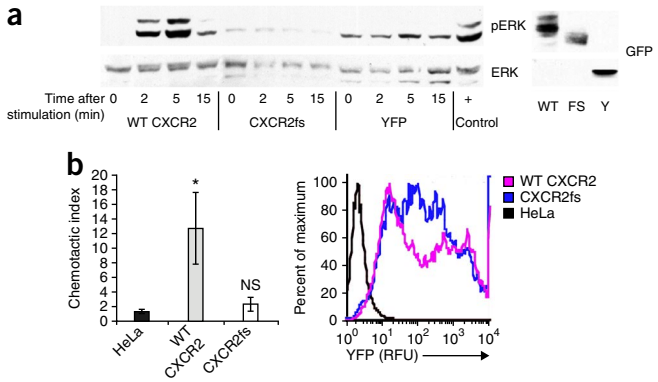

Figure 3 Functional characterization of the CXCR2fs mutant receptor. (**a**) Intracellular ERK1/2 signaling in HeLa cells expressing wild-type or mutant CXCR2 receptor. HeLa cells transfected with constructs encoding wild-type (WT) CXCR2 or CXCR2fs fused to YFP or YFP alone were stimulated with CXCL8 (100 ng/ml), and lysates were collected over the indicated time course to assess the activation of the ERK1/2 pathway (phosphorylated ERK, pERK). Peak phosphorylation after stimulation of wild-type CXCR2 occurred at 5 min, with significant attenuation of activation by 15 min. No activation over baseline was detected in cells expressing CXCR2fs. ERK1/2 activation 5 min after stimulation of endogenous CXCR4 with CXCL12 (100 ng/ml) was used as a positive control (+). The protein abundance of wild-type CXCR2 (WT), CXCR2fs (FS) and the YFP control (Y) detected with antibody to GFP (GFP) in transfected cell lines is shown to the right. (**b**) Chemotaxis assay results for HeLa cells expressing wild-type CXCR2 or CXCR2fs fused to YFP. Transiently transfected HeLa cells were seeded in the top chambers of modified Boyden chambers, and CXCL8 (100 ng/ml) was added to the medium in the bottom chambers. The chemotactic index (migration in the presence of CXCL8/migration in the absence of CXCL8) was calculated from the results of assays performed in triplicate. Robust migration was observed in cells expressing the wild-type CXCR2 protein but not in cells expressing the frameshift mutant or in untransfected cells. Overexpression of each construct was confirmed by detecting YFP fluorescence (expressed in relative fluorescence units, RFU) in aliquots of transfected cells (right). $^*P < 0.02$, Student's *t* test; NS, not significant. Bars representing standard error are shown.

feature of the rare, autosomal dominant WHIM (warts, hypogammaglobulinemia, immunodeficiency, myelokathexis) syndrome[37]. Most individuals with WHIM syndrome have a gain-of-function mutation in the *CXCR4* gene[36,37]. To further support a role for *CXCR2* mutations in neutropenia and neutrophil trafficking, we studied a pedigree in which two siblings were affected with myelokathexis (**Supplementary Fig. 9**) in the absence of other WHIM syndrome features and in which the involvement of *CXCR4* was excluded[38]. Sequence analysis of *CXCR2* in affected individuals II-2 and II-4 identified a homozygous deletion within the coding sequence (c.968delA) (**Supplementary Fig. 9**). The encoded mutant protein contained a translational frameshift and a premature termination codon after the addition of six novel amino acids (p.His323fs6*, hereinafter referred to as CXCR2fs). We tested the ability of the mutant CXCR2fs receptor to respond to ligand at the level of signal transduction and chemotaxis. Transfected HeLa cells expressing either wild-type CXCR2 or CXCR2fs were stimulated with the ligand CXCL8 over a 15-min time course. The accumulation of serine-phosphorylated extracellular signal–regulated kinases 1 and 2 (ERK1/2) was used to assess the activation of the mitogen-activated protein kinase (MAPK) signaling cascade. Accumulation of phosphorylated ERK1/2 was maximal at the 5-min time point after stimulation of the wild-type CXCR2 receptor, whereas cells expressing CXCR2fs had no detectable response (**Fig. 3a**). Consistent with these results, cells expressing wild-type CXCR2 had a robust chemotactic response to CXCL8, whereas cells expressing CXCR2fs did not, supporting the notion that the frameshift mutation caused a loss of receptor function (**Fig. 3b**). Together with our population-based results, these family-based and functional data support the concept that CXCR2 regulates neutrophil mobilization from the bone marrow and peripheral neutrophil counts.

Our results clearly demonstrate that rare and low-frequency coding variants contribute to phenotypic variation in human populations. We discovered new missense variants in key regulators of hematopoiesis (*EPO*, *SH2B3* and *TUBB1*) that have potential implications for diagnostic screening and drug development in a variety of hematological and inflammatory disorders (cytopenias, MPNs, lung disease and stroke). For example, rare *SH2B3* mutations may account for additional cases of MPN negative for *JAK2* mutation. We also provide further support for the biological role of several molecules in the regulation of hematopoiesis (*CXCR2* and *ZFP36L1*). Collectively, our findings validate exome-wide genotyping in very large samples as an effective tool and complementary approach to GWAS and deep DNA resequencing studies for defining the allelic architecture of complex traits. The association of a collection of rare and low-frequency *CXCR2* coding variants with WBC count emphasizes the relevance of gene-based tests as we continue to query even rarer variants for their role in human phenotypic variation.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all participants and staff of the MHI Biobank and acknowledge the technical support of the Beaulieu-Saucier MHI Pharmacogenomic Center. This work was supported by the Centre of Excellence in Personalized Medicine (CEPMed), Fonds de Recherche du Québec–Santé (FRQS), the Canada Research Chair program and the MHI Foundation. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). Exome chip data and analysis were supported through the Exome Sequencing Project (NHLBI RC2 HL-102924, RC2 HL-102925 and RC2 HL-102926), the Genetics and Epidemiology of Colorectal Cancer Consortium (NCI CA137088), the Genomics and Randomized Trials Network (NHGRI U01-HG005152) and a National Cancer Institute training grant (R25CA094880). The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible. SHIP is part of the Community Medicine Research network of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants 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). Generation of ExomeChip data has been supported by the Federal Ministry of Education and Research (grant 03Z1CN22) and the Federal State of Mecklenburg–West Pomerania. The University of Greifswald is a member of the Center of Knowledge Interchange program of Siemens AG and the Caché Campus program of InterSystems. G.A.D. acknowledges support from US National Institutes of Health award P01AI061093.

AUTHOR CONTRIBUTIONS

P.L.A., G.A.D., A.P.R. and G.L. conceived and designed the experiments. P.L.A., A.T., U.S., A.O., K.S.L., G.A.D., A.P.R. and G.L. performed the experiments. P.L.A., A.T., U.S., A.O., K.S.L., G.A.D., A.P.R. and G.L. analyzed the data. All authors contributed reagents and materials. P.L.A., G.A.D., A.P.R. and G.L. wrote the manuscript with contributions from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Nalls, M.A. *et al.* Multiple loci are associated with white blood cell phenotypes. *PLoS Genet.* **7**, e1002113 (2011).
2. Gieger, C. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**, 201–208 (2011).
3. van der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–375 (2012).
4. Zhou, Z. *et al.* USF and NF-E2 cooperate to regulate the recruitment and activity of RNA polymerase II in the β-globin gene locus. *J. Biol. Chem.* **285**, 15894–15905 (2010).
5. de la Chapelle, A., Sistonen, P., Lehvaslaiho, H., Ikkala, E. & Juvonen, E. Familial erythrocytosis genetically linked to erythropoietin receptor gene. *Lancet* **341**, 82–84 (1993).
6. Biasiotto, G. *et al.* New *TFR2* mutations in young Italian patients with hemochromatosis. *Haematologica* **93**, 309–310 (2008).
7. Pelucchi, S. *et al.* Expression of hepcidin and other iron-related genes in type 3 hemochromatosis due to a novel mutation in transferrin receptor-2. *Haematologica* **94**, 276–279 (2009).
8. Qayyum, R. *et al.* A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African Americans. *PLoS Genet.* **8**, e1002491 (2012).
9. Forand, A. *et al.* EKLF-driven PIT1 expression is critical for mouse erythroid maturation *in vivo* and *in vitro*. *Blood* **121**, 666–678 (2013).
10. Baxter, E.J. *et al.* Acquired mutation of the tyrosine kinase *JAK2* in human myeloproliferative disorders. *Lancet* **365**, 1054–1061 (2005).
11. James, C. *et al.* A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434**, 1144–1148 (2005).
12. Kralovics, R. *et al.* A gain-of-function mutation of *JAK2* in myeloproliferative disorders. *N. Engl. J. Med.* **352**, 1779–1790 (2005).
13. Levine, R.L. *et al.* Activating mutation in the tyrosine kinase *JAK2* in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* **7**, 387–397 (2005).
14. Bohinjec, J. Myelokathexis: chronic neutropenia with hyperplastic bone marrow and hypersegmented neutrophils in two siblings. *Blut* **42**, 191–196 (1981).
15. Xu, X. *et al.* $JAK2^{V617F}$: prevalence in a large Chinese hospital population. *Blood* **109**, 339–342 (2007).
16. Sidon, P., El Housni, H., Dessars, B. & Heimann, P. The JAK2V617F mutation is detectable at very low level in peripheral blood of healthy donors. *Leukemia* **20**, 1622 (2006).
17. Vannucchi, A.M., Pieri, L. & Guglielmelli, P. *JAK2* allele burden in the myeloproliferative neoplasms: effects on phenotype, prognosis and change with treatment. *Ther. Adv. Hematol.* **2**, 21–32 (2011).
18. Oh, S.T. When the brakes are lost: LNK dysfunction in mice, men, and myeloproliferative neoplasms. *Ther. Adv. Hematol.* **2**, 11–19 (2011).

19. Bersenev, A., Wu, C., Balcerek, J. & Tong, W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. *J. Clin. Invest.* **118**, 2832–2844 (2008).
20. Beck, L. *et al.* The phosphate transporter PiT1 (Slc20a1) revealed as a new essential gene for mouse liver development. *PLoS ONE* **5**, e9148 (2010).
21. Gerstein, M.B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
22. McMullin, M.F., Wu, C., Percy, M.J. & Tong, W. A nonsynonymous *LNK* polymorphism associated with idiopathic erythrocytosis. *Am. J. Hematol.* **86**, 962–964 (2011).
23. Macari, E.R. & Lowrey, C.H. Induction of human fetal hemoglobin via the NRF2 antioxidant response signaling pathway. *Blood* **117**, 5987–5997 (2011).
24. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
25. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
26. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
27. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
28. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
29. Stasi, R. Immune thrombocytopenia: pathophysiologic and clinical update. *Semin. Thromb. Hemost.* **38**, 454–462 (2012).
30. Johnston, C.M. *et al.* Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genet.* **4**, e9 (2008).
31. Cernelc, P. *et al.* Effects of molgramostim, filgrastim and lenograstim in the treatment of myelokathexis. *Pflugers Arch.* **440**, R81–R82 (2000).
32. Park, S.H. *et al.* Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491**, 779–783 (2012).
33. Salchow, K. *et al.* A common intracellular allosteric binding site for antagonists of the CXCR2 receptor. *Br. J. Pharmacol.* **159**, 1429–1439 (2010).
34. Marioni, J.C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8**, R228 (2007).
35. Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
36. Bird, C.P. *et al.* Fast-evolving noncoding sequences in the human genome. *Genome Biol.* **8**, R118 (2007).
37. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
38. Hernandez, P.A. *et al.* Mutations in the chemokine receptor gene *CXCR4* are associated with WHIM syndrome, a combined immunodeficiency disease. *Nat. Genet.* **34**, 70–74 (2003).

ONLINE METHODS

Study participants and phenotypes. The MHI Biobank is a longitudinal cohort with the aim of recruiting 30,000 patients of the MHI for clinical and genetic research[39]. Participants are recruited from different departments within the MHI and at its affiliated EPIC center (Centre de Médecine Préventive et d'Activité Physique de l'Institut de Cardiologie de Montréal), the largest fitness center in Canada for coronary patients and research in primary and secondary disease prevention. The MHI ethics committee approves the project, and informed consent is obtained from all participants. The MHI cohort collects data by using a 35-page questionnaire administered by a research nurse at baseline, including information of demographics, personal and family medical history, physical activity, diet, tobacco use and medication use, as well as depression and hostility questionnaires. Vital signs (heart rate, blood pressure, weight, height and waist circumference) are obtained by the nurse, and blood, DNA and plasma are collected and stored at the Beaulieu-Saucier Pharmacogenomics Center. The health status of participants is confirmed by the nurse using the hospital's health record for retrospective and prospective follow-up. The database for the cohort is updated frequently with participants' medical information from the hospital's electronic records. A follow-up study questionnaire is administered every 4 years. Blood cell counts and other related phenotypes were automatically generated with the UniCel DxH 800 cellular analysis system from Beckman Coulter.

The WHI is one of the largest (n = 161,808) US studies of women's health. This project was approved by the ethics committee at the Fred Hutchinson Cancer Research Center. The WHI consists of 2 main components: (i) a clinical trial that enrolled 68,132 post-menopausal women aged 50–79 years and randomized them to 1 of 3 placebo-controlled clinical trials of hormone therapy, dietary modification or supplementation with calcium and vitamin D and (ii) an observational study that enrolled 93,676 women of the same age range in a parallel prospective study[40]. Of the women in the WHI cohort who were eligible and consented to genetic research, 18,072 were included in the current study. Samples for blood count were collected at baseline by venipuncture into tubes containing EDTA. Blood counts were performed with automated hematology cell counters and standardized quality assurance procedures. Hemoglobin concentration, hematocrit levels, WBC count and platelet count were the only complete blood count values recorded at data entry during the WHI baseline examination, which was conducted nationwide during 1993–1998.

The Study of Health in Pomerania consists of two independent prospectively collected population-based cohorts in West Pomerania (SHIP and SHIP-TREND), a region in the northeast of Germany, assessing the prevalence and incidence of common population-based diseases and their risk factors. The study followed the recommendations of the Declaration of Helsinki. The study protocol for SHIP was approved by the medical ethics committee of the University of Greifswald. Written informed consent was obtained from each of the study participants. The study design has previously been described in detail[41]. Briefly, a sample from the population aged 20 to 79 years was drawn from population registries. First, the three cities of the region (with 17,076 to 65,977 inhabitants) and the 12 towns (with 1,516 to 3,044 inhabitants) were selected, and 17 of the 97 smaller towns (with fewer than 1,500 inhabitants) were then drawn at random. Second, from each of the selected communities, subjects were drawn at random in proportion to the population size of each community and were stratified by age and sex. Only individuals with German citizenship and main residency in the study area were included.

For SHIP, baseline examinations were carried out from 1997 until 2001, and the sample finally comprised 4,308 participants. Baseline examinations for SHIP-TREND were carried out between 2008 and 2012, with the study finally comprising 4,420 participants. Blood count was measured within 60 min of sample collection. Samples were analyzed at the hospital laboratory in Greifswald with a Coulter Max M analyzer (Coulter Electronics) and with a Coulter T660 analyzer (Coulter Electronics) at the hospital laboratory in Stralsund. Both analyzers were calibrated and maintained according to the manufacturer's instructions. Quality control was performed internally as well as externally by participation in external proficiency testing programs. For this project, data from SHIP-TREND and the first 5-year follow-up of SHIP were included. In all association analyses, the cohort was included as an additional covariate in the model.

Genotyping and quality control steps. We attempted to genotype 10,856 participants from the MHI Biobank on the Illumina ExomeChip array (version Infinium HumanExome v1.0 DNA Analysis BeadChip). Genotyping was performed at the MHI Pharmacogenomics Centre. We initially called genotypes with Illumina GenomeStudio software and, after data quality control, repeated calling of missing genotypes with zCall software[42]. We carried out most quality control steps in PLINK[43] and developed additional custom scripts when needed. We excluded markers with genotyping success rates of <95% and Hardy-Weinberg equilibrium $P < 1 \times 10^{-4}$. We excluded samples with genotyping success rates of <95%%, abnormal heterozygosity (inbreeding F values of <−0.2 or >0.1) and extensive low-level identity-by-descent (IBD) sharing with a large number of samples. Genotype concordance calculated from samples genotyped in duplicate or samples also sequenced by the 1000 Genomes Project[44] was >99.99%. To identify population outliers, we used principal-component analysis as implemented in EIGENSOFT[45] and anchored our analysis on continental populations from the 1000 Genomes Project[44]. For this study, we only analyzed individuals of European ancestry. We used GCTA software[46] to detect cryptic relatedness, removing one individual from each pair of samples that shared >18% of their genomes (corresponding to duplicates and first- and second-degree relatives). Our final data set included 9,656 individuals and 136,997 polymorphic markers. Blood cell phenotypes were available for 6,796 MHI Biobank participants (**Supplementary Table 1**).

DNA samples from the WHI clinical coordinating center were sent to the Broad Institute or the Translational Genomics Research Institute (TGEN) for genotyping and were placed on 96-well plates for processing using the Illumina HumanExome v1.0 SNP array. Genotypes were assigned using GenomeStudio v2010.3. A master file was then created containing all genotypes from the Broad Institute and TGEN. Quality control was performed on this master file using the PLINK and R[47] computing platforms. We excluded markers with a genotyping success rate of less than 99%. We excluded samples with a genotyping success rate of less than 98%. With the resulting sample set, we performed a principal-component analysis as well as an analysis of relatedness using PLINK IBS/IBD functionality. Outlier samples on the principal-component plots were excluded, as were samples that were determined to be contaminated via relatedness analysis (that were apparently related to hundreds of other samples). For each related or duplicate pair of samples, we excluded the sample with the lower call rate. Unexpected duplicate samples were also filtered out to prevent potential sample swaps from entering the analysis. For intentionally duplicated samples, we removed samples with low relatedness estimates, as we expected them to be close to 1. We excluded samples with hemoglobin concentrations greater than 20 g/dl or with a hematocrit (%) to hemoglobin (g/dl) ratio of greater than 5. Samples with WBC count ($\times 10^9$ cells/l) greater than 100 were also excluded from the analysis. After quality control, there were 18,072 samples and 166,836 polymorphic markers.

SHIP and SHIP-TREND samples were genotyped using the Illumina ExomeChip array (version Infinium HumanExome v1.0 DNA Analysis BeadChip). Hybridization of genomic DNA was carried out in accordance with the manufacturer's standard recommendations at the Helmholtz Zentrum München. Genotype calling was performed according to ExomeChip quality control standard operating procedure (SOP) version 5. Initial genotypes were determined using GenomeStudio Genotyping Module v1.0 (GenCall algorithm) with the HumanExome-12v1_A manifest file and the standard Illumina cluster file (HumanExome-12v1.egt). Contaminated samples, samples with a call rate of <90% and samples showing extreme heterozygosity, extensive estimated IBD sharing with a large number of samples or mismatch between reported and genotype-determined sex were excluded. Next, calling of missing genotypes was repeated using zCall software version 3.3 with default values. In both cohorts together, 7,366 individuals were successfully genotyped with an average call rate of 99.97%.

Trait modeling and study-level association testing. We used untransformed hemoglobin concentration, hematocrit levels, platelet counts and $\log_{10}$-transformed WBC counts for association testing. Because the analysis of rare variants is particularly sensitive to phenotypic outliers, we winsorized our data such that all individuals with a phenotype below the 0.5 percentile or above the 99.5 percentile of the trait distribution were assigned, respectively, the

phenotype corresponding to the 0.5 and 99.5 percentile values. In our analysis, phenotype winsorization reduced inflation while maintaining phenotype scale. In the MHI cohort, we used sex, age^2 and the first ten principal components as covariates. In the WHI cohort, we used age and the first two principal components as covariates. We also included a term in the linear model to account for the different WHI substudies that contributed to this project.

We used PLINK[43] to test association between phenotypes and single-variant genotypes (including single-variant conditional analyses) under an additive genetic model. For the gene-based analyses, each cohort ran an analysis with RVtests software[48]. This software calculates a score statistic for each variant and a covariance matrix for markers within sliding windows.

Meta-analysis and statistical significance. We combined single-variant results with Metal using the inverse variance method[49]. For gene-based tests, we combined score test results from RVtests with RareMetal software using default parameters[48]. *A priori*, we decided to focus exclusively on missense and nonsense variants as well as on variants within donor or acceptor splice sites for these analyses. For each trait, we ran two gene-based tests: a simple burden test with a MAF cutoff of <1% (burden T1) and a sequence kernel association test (SKAT) with a MAF cutoff of <5%.

We defined statistical significance using Bonferroni corrections. For single-variant analyses, we tested 183,585 DNA sequence variants and 4 phenotypes, yielding $\alpha = 0.05/183{,}585 \text{ variants}/4 \text{ phenotypes} = 6.8 \times 10^{-8}$. For gene-based analyses, we tested 15,930 genes (genes with no or only 1 missense, nonsense or splice-site variant were not tested), 4 phenotypes and 2 different statistical tests, yielding $\alpha = 0.05/15{,}930 \text{ genes}/4 \text{ phenotypes}/2 \text{ tests} = 3.9 \times 10^{-7}$.

Functional characterization of CXCR2. A previously described family in which two sisters were affected with isolated myelokathexis[14] was evaluated for mutations in *CXCR2*. Clinical manifestations in this family included neutropenia without lymphopenia or warts leading to recurrent bacterial infections, including septic thrombophlebitis and subacute bacterial endocarditis. At age 43, one of the siblings was reported to have had >80 infectious episodes[31]. Her younger sister had fewer infectious episodes, which was speculated to result from more robust transient release of neutrophils into the peripheral blood during infection. The family was lost to follow-up and not available for further functional studies on primary cells.

Human CXCR2 is encoded by a single exon with multiple upstream non-coding exons giving rise to a number of differentially spliced isoforms. The coding sequence of the gene was amplified on three overlapping PCR fragments and analyzed by automated sequencing. Exonic boundaries were identified using publicly available genomic database information[50], and primers were designed to amplify coding exons and flanking intronic sequences using Primer3 software[51]. PCR amplicons were sequenced using the ABI3700 automated platform. Confirmatory restriction enzyme analysis was performed using NcoI (New England BioLabs) under standard conditions.

IMAGE clone 5752441 containing the full-length *CXCR2* ORF was obtained from Invitrogen. The insert was confirmed by sequencing the clone using *CXCR2*-specific primers. The c.968delA (CXCR2fs), c.[967C>T]+[969T>G] (p.His323*) and c.[968delA]+[986insC] (CXCR2fs wild type) mutations were introduced by primers for site-directed mutagenesis and amplified by PfuTurbo (Stratagene). PCR amplification on the resulting clones of the *CXCR2* ORF using the IMAGE clone as a template was used to introduce sequence encoding epitopes on the C-terminal tail of the receptor. Primers for amplification contained exogenous restriction sites at their 3′ and 5′ ends and were digested using EcoRI-XhoI (Flag) or EcoRI-SalI (YFP). After digestion with the appropriate enzymes, fragments were ligated in frame upstream of the sequence encoding the 3×Flag epitope in pIRES-3×Flag-hrGFP (Invitrogen) or the YFP full-length coding sequence in pEYFP-N1 (Clontech). The full-length ORFs for Flag-tagged and YFP-fused proteins were verified by sequencing.

HEK293 cells (CRL-1573) and HeLa cells (CRL-1651) were purchased from the American Type Culture Collection and were tested for mycoplasma. HEK293 and HeLa cell cultures were maintained in 5% CO_2 at 37 °C in DMEM (Cellgro) supplemented with 10% FBS (Gibco), 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (Cellgro). Cultures were transfected using FuGENE-6 (Roche) according to the manufacturer's protocol. Transgene expression was evaluated 36–72 h after transfection by indirect immunofluorescence or protein blot. Primary antibodies used in these studies (antibodies to Flag (M2, Sigma), CXCR2 (E2, Santa Cruz Biotechnology), GFP (A-6455, Molecular Probes) and phosphorylated ERK and total ERK (sc-7383 and sc-93, Santa Cruz Biotechnology)) were used at the dilutions suggested by the manufacturers. Horseradish peroxidase (HRP)-conjugated secondary antibodies to rabbit and mouse (Pierce) were used for protein blots, and FITC-labeled secondary antibody to mouse (Santa Cruz Biotechnology), AF594- and AF488-labeled secondary antibodies to rabbit (Invitrogen) and direct-labeled phalloidin-AF594 (Molecular Probes) were used for immunofluorescence microscopy. Fluorescein-labeled monoclonal antibody to CXCR2 (48311, R&D Systems) and labeled isotype control antibodies (5 µl) were used for flow cytometry experiments.

Transfected HeLa cells were rendered quiescent in serum-free, antibiotic-free DMEM overnight before the addition of ligand. Cells (1×10^6) were stimulated with 100 ng/ml CXCL8 or 100 ng/ml CXCL12 (Cell Sciences) or kept in serum-free medium for the times described. Cells were either collected for protein blot analysis or fixed for immunofluorescence staining.

Before collection, cells were rinsed with ice-cold PBS and were lysed on ice in RIPA lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1.0% Triton X-100, 0.5% DOC, 0.1% SDS, 0.025% NaN_3) supplemented with Complete protease inhibitor cocktail (Roche) for a minimum of 30 min. After determination of protein concentration by Bradford assay, lysates were denatured in 1× Laemmli sample buffer with 100 mM DTT, and equivalent amounts were separated by 10% SDS-PAGE. Proteins were transferred to nitrocellulose (Pall/GE) using the Bio-Rad Semi-Dry Trans-Blot system for 1 h, blocked in 5% milk in PBS for 1 h and incubated overnight at 4 °C in primary antibody. Blots were washed three times for 10 min per wash in PBS with Tween-20 (0.05%; PBST). Incubation with species-specific HRP-conjugated antibody for 1 h was followed by three more 10-min washes in PBST. Membranes were developed using SuperSignal West Pico and Chemiluminescent Substrate (Pierce) on HyBlot CL film (Denville). For deglycosylation experiments, 30 µg of total protein was incubated with Endo Hf (New England BioLabs), PNGase F (New England BioLabs) or buffer alone for 1 h at 37 °C as recommended by the manufacturer.

Cells were transfected with *CXCR2* constructs as described, an aliquot of cells resuspended by trypsin treatment was fixed for 15 min using 2.2% paraformaldehyde (Electron Microscopy Sciences) in PBS, and cells were spun down at 500*g* for 5 min. Cells were washed twice in Flow Buffer (PBS with 0.5% BSA and 5 mM EDTA) to inhibit the clumping of cells. Approximately 2.5×10^5 cells in 50-µl aliquots were used for each assay. Flow cytometry data for the quantification of CXCR2 expression were collected using a FACS LSR II instrument (Becton Dickson). All data were analyzed using FlowJo software (Tree Star). At least three independent experiments were performed for each condition tested.

HeLa cells at 70% confluence in 10-cm dishes were transfected with either pEYFP-N1-CXCR2wt or pEYFP-N1-CXCR2wt and cultured for 3 d. After overnight serum starvation, cells were collected on the fourth day and resuspended in DMEM with 0.5% BSA at a density of 5×10^5 cells/100 µl. Aliquots of 100 µl were added to the upper chamber of 24-well Transwell plates (Corning-Costar) with collagen-coated 8.0-µm pore polycarbonate membranes. Chemotaxis was assayed by the addition of CXCL8 (100 ng/ml) to DMEM in the bottom chambers of the plates. After 2 h in a tissue culture incubator at 37 °C and 5% CO_2, inserts were removed, loose cells were scraped off, and transmigrated cells were fixed and stained in crystal violet. Cells that migrated onto the lower surface of the membrane were counted in five low-power (10×) fields using a bright-field inverted microscope to determine the average number of cells that migrated per field. The chemotactic index was calculated by dividing the average number of cells that migrated per field under conditions of chemokine addition by the average number of cells that migrated under conditions of saline addition. Experiments were performed in triplicate at least three times each.

39. Beaudoin, M. *et al.* Pooled DNA resequencing of 68 myocardial infarction candidate genes in French Canadians. *Circ. Cardiovasc. Genet.* **5**, 547–554 (2012).
40. Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control. Clin. Trials* **19**, 61–109 (1998).
41. Völzke, H. *et al.* Cohort profile: the study of health in Pomerania. *Int. J. Epidemiol.* **40**, 294–307 (2011).

doi:10.1038/ng.2962

42. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543–2545 (2012).
43. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
44. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
45. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
46. 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).
47. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2013).
48. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46**, 200–204 (2014).
49. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
50. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
51. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).