genes

Review

Lessons and Implications from Genome-Wide Association Studies (GWAS) Findings of Blood Cell Phenotypes

Nathalie Chami and Guillaume Lettre *

Montreal Heart Institute, Faculté de Médecine, Université de Montréal, 5000 Bélanger Street, Montréal, QC H1T 1C8, Canada; E-Mail: nathalie.chami@mhi-humangenetics.org

* Author to whom correspondence should be addressed; E-Mail: guillaume.lettre@umontreal.ca; Tel.: +1-514-376-3330; Fax: +1-514-593-2539.

Abstract: Genome-wide association studies (GWAS) have identified reproducible genetic associations with hundreds of human diseases and traits. The vast majority of these associated single nucleotide polymorphisms (SNPs) are non-coding, highlighting the challenge in moving from genetic findings to mechanistic and functional insights. Nevertheless, large-scale (epi)genomic studies and bioinformatic analyses strongly suggest that GWAS hits are not randomly distributed in the genome but rather pinpoint specific biological pathways important for disease development or phenotypic variation. In this review, we focus on GWAS discoveries for the three main blood cell types: red blood cells, white blood cells and platelets. We summarize the knowledge gained from GWAS of these phenotypes and discuss their possible clinical implications for common (e.g., anemia) and rare (e.g., myeloproliferative neoplasms) human blood-related diseases. Finally, we argue that blood phenotypes are ideal to study the genetics of complex human traits because they are fully amenable to experimental testing.

1. Genetics of Red Blood Cells, White Blood Cells and Platelets

Blood is mostly composed of plasma and blood cells and plays a major role in a variety of functions involved in general human homeostasis: it transports oxygen, nutrients and hormones to tissues,

removes waste, performs immunological functions and contributes tissue damage repair through coagulation. The main three blood cell types carry out most of these activities: red blood cells (RBC, or erythrocytes) transport oxygen, white blood cells (WBC, or leukocytes) coordinate some of the immune responses, and platelets are the bricks that form blood clots to prevent excessive bleeding. All of these cell types originate through proliferation and differentiation from common precursors (hematopoietic stem cells) [1].

An aberrant number, size or feature of the three main blood cell types characterizes multiple human diseases (Table 1). In many cases, the triggering factor is of environmental origin, often poor nutrition or infections (e.g., malaria, HIV). Germline and somatic mutations can also cause severe blood disorders, such as mutations in glucose-6 phosphate dehydrogenase (*G6PD*) which is responsible for chronic hemolytic anemia or mutations in oncogenes or tumor suppressor genes that result in leukemia. It is also known that blood cell phenotypes vary between healthy individuals, and that some of this inter-individual variation is controlled by genetics. In a large study of healthy Sardinians (N = 6,148), the heritability estimates for RBC, WBC and platelet counts were, respectively, 0.67, 0.38 and 0.53 [2]. Similar heritability estimates were obtained when analyzing phenotype concordance in healthy monozygotic and dizygotic twins from the United Kingdom [3]. These results indicate that a large fraction of the phenotypic variation in these blood traits is controlled by DNA sequence variants segregating in healthy individuals.

Table 1. Main blood cell traits routinely measured in standard complete blood count (CBC).

Trait	Description	Unit
Red blood cell (RBC) count	Count of RBC per microliter	Million cells per microliter ($\times 10^6/\mu L$)
Hemoglobin (HGB)	Hemoglobin concentration	Gram per deciliter (g/dL)
Hematocrit (HCT)	Fraction of blood that contains hemoglobin	Percentage (%)
Mean corpuscular hemoglobin (MCH)	Amount of hemoglobin per RBC	Picogram (pg)
Mean corpuscular volume (MCV)	Average volume of RBC	Femtoliter (fL)
MCH concentration (MCHC)	Hemoglobin divided by hematocrit	Gram per deciliter (g/dL)
RBC distribution width (RDW)	Distribution of RBC volume	Percentage (%)
White blood cell (WBC) count	Number of WBC per liter (include all main subtypes)	Billion cells per liter ($\times 10^9/L$)
Platelet (PLT) count	Number of PLT per liter	Billion cells per liter ($\times 10^9/L$)
Mean platelet volume (MPV)	Average platelet volume	Femtoliter (fL)

The clinical importance of this heritable variation in blood cell phenotypes is unclear. However, it is interesting that epidemiological studies have detected links between WBC or platelet counts and the risk to suffer from cardio- and cerebrovascular diseases [4–6]. As for most epidemiological observations, however, it is difficult to determine if changes in hematological parameters are pathological or reflect consequences of disease manifestation. Using Mendelian randomization methodologies, in which inherited genetic variants associated with hematological traits are used as instruments to test the causal effect of the traits on diseases, may provide an answer to this question [7]. Such an approach was successfully used to determine that LDL-cholesterol and triglyceride levels, but unlikely HDL-cholesterol levels, are causes of coronary artery diseases [8,9]. Understanding how DNA polymorphisms modulate blood cell phenotypes in health (and diseases) could provide new opportunities to study hematopoiesis, improve their use in medicine as biomarkers and maybe even help in the development of new drugs.

To this list, we would also add that hematological traits are ideal phenotypes to further our understanding of the genetics of human complex diseases and traits because experimental systems exist to functionally validate genetic findings.

## 2. Genome-Wide Association Studies (GWAS) for Blood Cell Phenotypes

Before GWAS, little was known about the role of SNPs and other common DNA sequence variants on normal variation in blood cell phenotypes. Candidate gene DNA sequencing experiments have identified mutations in the globin loci, but also in the erythropoietin receptor (*EPOR*) and hemochromatosis (*HFE*) genes [10,11]. Genome-wide linkage studies also found a few reproducible signals, most notably a linkage peak on chromosome 6q23 that encompasses the MYB transcription factor [12,13]. These findings could not, however, explain the heritability of these blood cell phenotypes in normal individuals.

As for many other complex human traits and diseases, the capacity to test associations with genotypes across the genome by GWAS opened a new world. Prior to the GWAS era, genetic association studies often had sample sizes that were too small and were limited to testing only known genes [14]. With GWAS, it became possible to genotype all genes independently of previous knowledge. Blood cell traits are particularly amenable to the GWAS approach because they are routinely and accurately measured in large cohorts, and initial findings can be tested for replication in other cohorts because it is easy to harmonize these phenotypes (Figure 1) [15]. In general, one of the main challenges for GWAS has been to pinpoint functional genes and variants associated with a given trait. Although this remains a challenge, blood cell traits are particularly well-suited for genetic and functional follow-up. As mentioned earlier, fine-mapping by dense genotyping and DNA re-sequencing is possible because the traits are usually available in most cohorts or biobanks, including participants of different ethnicities (see below). There is also the possibility to test the functions of new genes in cell culture systems or model organisms because the phenotypes are often cell autonomous and the assays already well-developed. Using this approach, investigators showed that SNPs at 6p21.1 modulate erythrocyte traits through a regulatory effect on the cyclin D3 (*CCND3*) gene [16]. Large-scale gene silencing and other functional experiments in fruit flies, zebrafish and mice were also used to validate several new genes involved in platelet and RBC development within loci identified by GWAS [17,18].

All the steps described in Figure 1 now take advantage of powerful bioinformatic tools and other resources freely available on the web. For instance, comparative genomics has identified DNA bases that are conserved through evolution and therefore more likely to be functionally important [19]. There are also software that can predict based on conservation and physicochemical properties whether a DNA polymorphism that changes an amino acid is likely detrimental or not [20,21]. We can also quickly query large gene expression datasets to determine if the genes near an associated SNP are expressed in the relevant tissue(s) for the phenotypes of interest (as an example, see reference [22]). And when genotypes are available, it is possible to test *in silico* if the GWAS SNPs (or SNPs in linkage disequilibrium) control gene expression through regulatory mechanisms; that is, if the variants are expression quantitative trait loci (eQTL) [23]. The ENCODE and Roadmap Epigenomics Projects have used next-generation DNA sequencing applications, including DNAse I hypersensitive sites mapping and chromatin immunoprecipitation with antibodies against several histone tail modifications

(ChIP-seq), to define regulatory sequences in human cell lines and tissues [24–26]. Using a complementary approach (FAIRE-seq), Paul *et al*. identified regions of open chromatin in primary human blood cells and showed that SNPs associated with RBC and platelet phenotypes are enriched in these regions [27]. All this vast genomic information is useful in prioritizing causal genes and variants at GWAS loci, and investigators are developing algorithms to facilitate its integration [28,29].

Figure 1. Ideal study design to identify single nucleotide polymorphisms (SNPs) associated with human complex traits and diseases using genome-wide association studies (GWAS). For blood cell phenotypes, GWAS were particularly successful because sample sizes are large, phenotypes are easy to measure and are accurate, and well-characterized experimental models already exist.
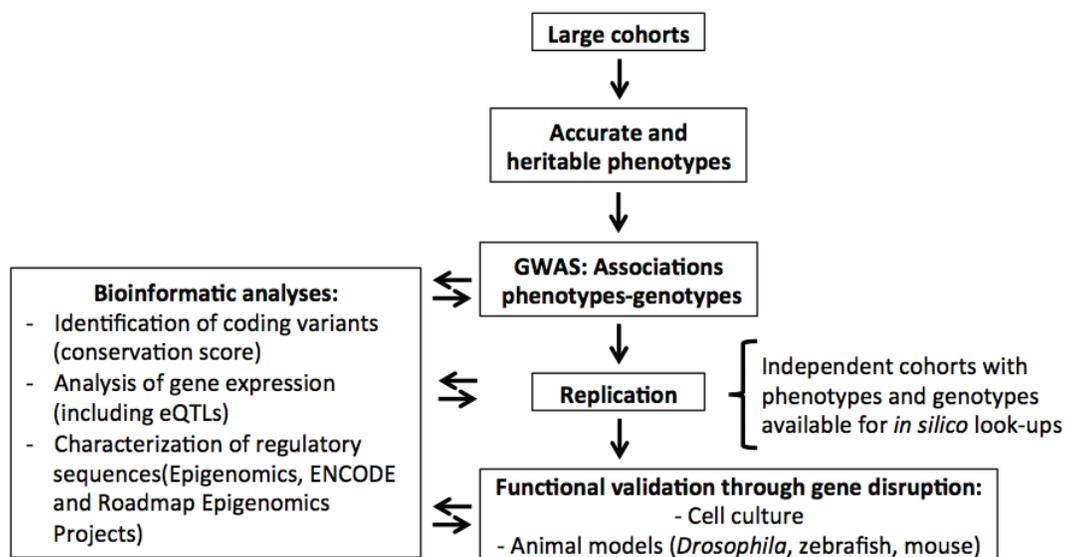
Several GWAS for hematological traits have already been published [17,18,30–46]. The largest studies, carried out in Europeans or individuals of European ancestry, have so far identified at genome-wide significance (*p*-value $< 5 \times 10^{-8}$) 75, 10 and 68 SNPs associated with RBC, WBC and platelet traits respectively [17,18,45]. The lower number of SNPs associated with WBC count could be explained by a lower heritability (see above), but also because the sample size for the WBC GWAS was smaller ($N = 11,823$) in comparison with the GWAS for RBC ($N = 135,367$) and platelet ($N = 66,867$) traits. Despite their large number, these variants only explain a small fraction of the heritable variation in these phenotypes (<10%). They are, however, not random but clustered near genes involved in relevant biological pathways and enriched for regulatory functions by expression quantitative trait loci (eQTL) and epigenomic analyses. Most loci are associated with a single blood cell type but by comparing the different studies, we found seven loci that are associated with at least two different cell types (Table 2). These include *SH2B3*, a gene that encodes the adapter protein LNK that interacts with JAK2 and modulates JAK-STAT signaling in hematopoietic cells, and *MYB*, that encodes a transcription factor essential for definitive hematopoiesis. Both *SH2B3* and *MYB* SNPs are associated with the three main blood cell types. The other loci presented in Table 2 include genes associated with a combination of two phenotypes, maybe suggesting different functions in different hematopoietic lineages.

Table 2. Loci identified by GWAS that carry SNPs associated with at least two of the three main blood cell types. For each association, we report the ethnic group in which the genetic associations were found. We also listed only one gene per locus, although for many loci, the causal gene is unknown. RBC: red blood cell; WBC: white blood cell.

Locus	Location	RBC	WBC	Platelet	References
TMCC2	1q32.1	Caucasian		Caucasian	[17,18]
ARHGEF3	3p14.3	African American		Caucasian	[17,30,36,38]
LRRC16A	6p22.2	African American		African American	[31,37]
HBS1L-MYB	6q22-q23.3	African American/ Caucasian/Japanese	Caucasian	African American/ Caucasian	[17,18,31,32,34,35,37]
IL-6	7p21		Japanese	Japanese	[47]
RCL1	9p24.1-p23	Caucasian		Caucasian/Japanese	[17,18,32,34]
SH2B3	12q24	Caucasian	Caucasian	Caucasian/Japanese	[17,32–35,38]

Some Loci Associated with Blood Cell Traits Are Population-Specific

It is difficult to compare association results for hematological traits across different populations because the sample size of the respective GWAS, and thus the statistical power to discover associations, is very different. For instance for RBC phenotypes, the largest studies in Caucasians and African Americans included, respectively, 135,367 and 16,496 participants [18,31]. Despite this caveat, many of the loci found in African Americans or Asians were also present in Caucasians; this general transferability of results across ethnic groups has been observed for other complex human traits [48,49]. For blood cell traits, however, there are notable exceptions. A SNP upstream of the Duffy antigen/receptor for chemokines (*DARC*) gene explains a large fraction of the variation in WBC and neutrophil counts, and is responsible for benign neutropenia [50]. This variant, which is monomorphic in Caucasians, is under positive selection in persons of African ancestry because it provides protection against *Plasmodium vivax* malaria infections. Similarly, genetic variation near the *α-globin*, the *β-globin* and the *G6PD* genes are associated with RBC indices in Africa-derived populations and are relatively common in frequency because they provide a selective advantage against malaria infections. These observations suggest that as we continue to query the human genome for associations with blood cell phenotypes, integrating evidence of natural selection would be a powerful approach.

3. Genetic Modifiers of Disease Severity

Several human diseases, which afflict a large fraction of the human population, are characterized by abnormally low or high counts of the three main blood cell types, or some unusual values for their features or contents. Anemia is a decrease of RBC count and hemoglobin levels (<11 g/dL in women or <13 g/dL in men) and is characterized by a wide spectrum of symptoms from simple fatigue to heart failure [51]. The World Health Organization estimates that anemia affects 1.62 billion people in the World [52]. The main causes of anemia are poor nutrition and iron deficiency, infections (e.g., malaria) and RBC diseases such as the hemoglobinopathies. Although the effect size of an individual SNP associated with RBC count or hemoglobin levels is not sufficient to cause anemia, a combination of

hemoglobin-reducing alleles at many SNPs could have an impact on the risk to develop this disorder. Maybe more importantly, without causing anemia itself, this genetic score could influence clinical severity in at-risk populations (e.g., children with a small number of hemoglobin-increasing alleles that live in a region where malaria is endemic). Since anemia is mostly frequent in Africa and South-East Asia, it is critical to continue to search for genetic associations with hemoglobin levels in these populations [52].

There are many other human diseases that are diagnosed, like anemia, through abnormal counts of the main blood cell types (e.g., cancers). One example is myeloproliferative neoplasms (MPNs), diseases of the bone marrow characterized by excess cell production [53]. By far the main cause of MPNs is a somatic gain-of-function mutation in the kinase gene *JAK2* (Val617Phe), which activates cell proliferation in the myeloid lineage [54,55], and changes platelet formation and reactivity [56]. It has never been tested whether SNPs associated with blood cell counts could modify complication risk in MPN patients with a *JAK2* (Val617Phe) mutation. For instance, MPN patients are at high risk of stroke, but it is unknown if such patients that also carry a large number of platelet-increasing alleles are at even higher stroke risk. Such analyses, on MPNs but also all other diseases characterized by a blood phenotype, are simple and could test the role that SNPs associated with normal variation in hematological traits may have on our risk to develop more severe disorders and related complications [18].

BCL11A Modifies Clinical Severity in Hemoglobinopathies

In adults, hemoglobin (HbA) is composed of two α- and two β-globin subunits that form a tetramer with the heme moiety to transport oxygen from the lungs to the different organs. Prior to birth, the *β-globin* gene is silent and the β-globin subunits are encoded by the *γ-globin* genes to form fetal hemoglobin (HbF). The switch from HbF to HbA production is a transcriptionally and epigenetically tightly regulated process [57]. For most healthy individuals, the switch itself has no clinical impact. However, for β-thalassemia and sickle cell disease patients with mutations in the *β-globin* gene, understanding and modulating the globin switch is currently the most promising therapeutic strategy. Conceptually, this is easy to appreciate: if the disease-causing mutations are in the *β-globin* gene, then re-activating *γ-globin* gene expression to form "normal" β-globin subunits would bypass the problem. This approach is supported by an extensive literature on the natural history of hemoglobinopathies and epidemiological studies [58]. For instance, it has been shown that sickle cell disease patients that normally produce more HbF have better survival prognostic and less severe disease complications than patients with low HbF levels [59–61].

Although as adults we mostly produce HbA, we continue to make residual levels of HbF. Inter-individual variation in HbF levels is highly heritable ($h^2 \sim 0.6$–0.9) [2,62]. Genetic investigations, including GWAS, have identified common genetic variation at three loci (*BCL11A*, *HBS1L-MYB* and *β-globin*) that have strong phenotypic effects and that together explain almost half of the heritable variation in HbF levels [63–66]. These HbF-associated SNPs are also associated with clinical severity in β-hemoglobinopathy patients: transfusion-dependency in β-thalassemia and painful crises in sickle cell disease [65,67,68]. This again emphasizes the importance of HbF as a strong modifier of severity for these diseases.

BCL11A encodes a transcription factor that had no known function in the globin switch before its discovery in two GWAS for HbF levels [63,65]. Since then, we have learned that BCL11A is a potent transcriptional repressor of *γ-globin* gene expression and that its inactivation in the erythroid lineage can treat a sickle cell disease mouse model through re-activation of HbF production [69,70]. More recently, both genetic and molecular fine-mapping work has determined that HbF-associated SNPs located in a *BCL11A* intron disrupt en erythroid enhancer that controls *BCL11A* expression [71]. This model was confirmed by targeted deletion of the enhancer through genome engineering that blocked *BCL11A* expression and re-activated *γ-globin* gene expression and HbF production [16]. As genome editing methods are rapidly improving, this proof-of-concept experiment suggests a new therapeutic strategy in which the *BCL11A* enhancer would be deleted *ex vivo* in a hemoglobinopathy patient's cells to re-activate HbF production, and the cells would then be transplanted back to the patient [72]. The characterization of *BCL11A* and its role in HbF production serves as a powerful example to illustrate the success of GWAS from new biology to potentially innovative therapy.

4. Orphan Blood Cell Diseases

Although we did not assess the statistical significance of the enrichment, we observed that many of the SNPs associated with blood cell traits are located near genes that are mutated in severe hematological disorders and inherited in a Mendelian fashion. These include SNPs near *HK1* (hemolytic anemia), *TMPRSS6*, *HFE* and *TFR2* (iron deficiency) or *TUBB1* (thrombocytopenia). This observation is similar to the situation of many other complex human phenotypes (e.g., lipids, height, diabetes) where GWAS have identified hypomorphic alleles near human syndrome genes for related phenotypes. As such, the long list of loci found by GWAS provides a framework to investigate human syndromes characterized by aberrant blood features, mapped to a chromosome arm by linkage studies, but where the gene culprit has not been identified yet.

To investigate this hypothesis, we queried the Online Mendelian Inheritance in Man (OMIM) database [73]. In a non-exhaustive search, we identified four such orphan diseases where the genomic locations overlap with SNPs identified by GWAS (Table 3). For three of the diseases, GWAS findings suggest a strong candidate gene (*IL5*, *LIPC*, *NUDT19*) for re-sequencing in affected individuals. As we continue to map these rare blood disorders, cross-referencing with GWAS hits may provide a strong filter to prioritize genes for genetic testing.

Table 3. Orphan human syndromes mapped to a chromosomal band and characterized by a blood cell phenotype. Only such syndromes that overlap with a locus identified by GWAS for the corresponding blood cell trait are included in this table. We generated this list by querying the Online Mendelian Inheritance in Man (OMIM) database with the following keywords: anemia, blood, hemoglobin, leukopenia, neutropenia, platelet, thrombocytopenia.

Mendelian genetics: orphan syndromes				Genome-wide association studies				
Locus	Disease	OMIM#	Description	SNP	Position	Phenotype	Candidate-gene(s)	Ref.
5q31	Familial eosinophilia	131400	Characterized by peripheral hypereosinophilia with or without other organ involvement	rs4143832	chr5: 131,862,977	Eosinophil count	*IL5*	[33]

Table 3. *Cont.*

Mendelian genetics: orphan syndromes				Genome-wide association studies				
Locus	Disease	OMIM#	Description	SNP	Position	Phenotype	Candidate-gene(s)	Ref.
6p21	Macroblobulinemia, susceptibility to Waldenstrom	153600	Malignant B-cell neoplasm characterized by lymphoplasmacytic infiltration of the bone marrow and hypersecretion of monoclonal immunoglobulin M (IgM) protein	rs2517524	chr6: 31,025,713	White blood cell	*HLA* region	[45]
15q21	Dyserythropoietic anemia, congenital type III	105600	Characterized by nonprogressive mild to moderate hemolytic anemia, macrocytosis in the peripheral blood, and giant multinucleated erythroblasts in the bone marrow	rs1532085	chr15: 58,683,366	Hemoglobin	*LIPC*	[18]
19q13	Transient erythroblastopenia of childhood	227050	Red blood cell aplasia	rs3892630	chr19: 33,181484	Mean corpuscular volume	*NUDT19*	[18]

5. Conclusions

GWAS have identified hundreds of loci that carry common genetic variants associated with RBC, WBC and platelet phenotypes. Many of these genetic associations still need to be linked to causal genes and genetic variants, yet because tractable cellular and animal models are available, this might be simpler for blood cell traits than it is for most complex human phenotypes. By design, GWAS interrogate common DNA variants, leaving untested low-frequency and rare sequence variation. The development of next-generation DNA sequencing platforms and exome genotyping arrays now provides the tools to test the role of this rarer genetic variation on blood cell phenotypes. Much criticism has been raised against GWAS because identified SNPs have poor predictive value; this is also true for SNPs associated with blood cell traits. However, this observation needs to be counter-balanced by the potential gain in improving our understanding of human biology in health and disease. GWAS blood cell trait loci provide new opportunities to study hematopoiesis, natural selection and the various ways common segregating DNA sequence variants can modify disease severity, paving the way for the development of more specific therapies.

Acknowledgments

Author Contributions

Wrote the paper: NC, GL.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Orkin, S.H.; Zon, L.I. Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* **2008**, *132*, 631–644.
2. Pilia, G.; Chen, W.M.; Scuteri, A.; Orru, M.; Albai, G.; Dei, M.; Lai, S.; Usala, G.; Lai, M.; Loi, P.; *et al*. Heritability of cardiovascular and personality traits in 6,148 sardinians. *PLoS Genet.* **2006**, *2*, e132.
3. Garner, C.; Tatu, T.; Reittie, J.E.; Littlewood, T.; Darley, J.; Cervino, S.; Farrall, M.; Kelly, P.; Spector, T.D.; Thein, S.L. Genetic influences on F cells and other hematologic variables: A twin heritability study. *Blood* **2000**, *95*, 342–346.
4. Hoffman, M.; Blum, A.; Baruch, R.; Kaplan, E.; Benjamin, M. Leukocytes and coronary heart disease. *Atherosclerosis* **2004**, *172*, 1–6.
5. Boos, C.J.; Lip, G.Y. Assessment of mean platelet volume in coronary artery disease—What does it mean? *Thromb. Res.* **2007**, *120*, 11–13.
6. Nieswandt, B.; Kleinschnitz, C.; Stoll, G. Ischaemic stroke: A thrombo-inflammatory disease? *J. Physiol.* **2011**, *589*, 4115–4123.
7. Ebrahim, S.; Davey Smith, G. Mendelian randomization: Can genetic epidemiology help redress the failures of observational epidemiology? *Hum. Genet.* **2008**, *123*, 15–33.
8. Voight, B.F.; Peloso, G.M.; Orho-Melander, M.; Frikke-Schmidt, R.; Barbalic, M.; Jensen, M.K.; Hindy, G.; Holm, H.; Ding, E.L.; Johnson, T.; *et al*. Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study. *Lancet* **2012**, *380*, 572–580.
9. Do, R.; Willer, C.J.; Schmidt, E.M.; Sengupta, S.; Gao, C.; Peloso, G.M.; Gustafsson, S.; Kanoni, S.; Ganna, A.; Chen, J.; *et al*. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **2013**, *45*, 1345–1352.
10. Zeng, S.M.; Yankowitz, J.; Widness, J.A.; Strauss, R.G. Etiology of differences in hematocrit between males and females: Sequence-based polymorphisms in erythropoietin and its receptor. *J. Gend. Specif. Med.: JGSM:* **2001**, *4*, 35–40.
11. McLaren, C.E.; Barton, J.C.; Gordeuk, V.R.; Wu, L.; Adams, P.C.; Reboussin, D.M.; Speechley, M.; Chang, H.; Acton, R.T.; Harris, E.L.; *et al*. Determinants and characteristics of mean corpuscular volume and hemoglobin concentration in white HFE C282Y homozygotes in the hemochromatosis and iron overload screening study. *Am. J. Hematol.* **2007**, *82*, 898–905.
12. Lin, J.P.; O'Donnell, C.J.; Jin, L.; Fox, C.; Yang, Q.; Cupples, L.A. Evidence for linkage of red blood cell size and count: Genome-wide scans in the framingham heart study. *Am. J. Hematol.* **2007**, *82*, 605–610.

13. Menzel, S.; Jiang, J.; Silver, N.; Gallagher, J.; Cunningham, J.; Surdulescu, G.; Lathrop, M.; Farrall, M.; Spector, T.D.; Thein, S.L. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood* **2007**, *110*, 3624–3626.

14. Lohmueller, K.E.; Pearce, C.L.; Pike, M.; Lander, E.S.; Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **2003**, *33*, 177–182.

15. Lettre, G. The search for genetic modifiers of disease severity in the beta-hemoglobinopathies. *Cold Spring Harbor Perspect. Med.* **2012**, *2*, doi:10.1101/cshperspect.a015032.

16. Sankaran, V.G.; Ludwig, L.S.; Sicinska, E.; Xu, J.; Bauer, D.E.; Eng, J.C.; Patterson, H.C.; Metcalf, R.A.; Natkunam, Y.; Orkin, S.H.; *et al*. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes Dev.* **2012**, *26*, 2075–2087.

17. Gieger, C.; Radhakrishnan, A.; Cvejic, A.; Tang, W.; Porcu, E.; Pistis, G.; Serbanovic-Canic, J.; Elling, U.; Goodall, A.H.; Labrune, Y.; *et al*. New gene functions in megakaryopoiesis and platelet formation. *Nature* **2011**, *480*, 201–208.

18. Van der Harst, P.; Zhang, W.; Mateo Leach, I.; Rendon, A.; Verweij, N.; Sehmi, J.; Paul, D.S.; Elling, U.; Allayee, H.; Li, X.; *et al*. Seventy-five genetic loci influencing the human red blood cell. *Nature* **2012**, *492*, 369–375.

19. Lindblad-Toh, K.; Garber, M.; Zuk, O.; Lin, M.F.; Parker, B.J.; Washietl, S.; Kheradpour, P.; Ernst, J.; Jordan, G.; Mauceli, E.; *et al*. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **2011**, *478*, 476–482.

20. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249.

21. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. *Nat. Protoc.* **2009**, *4*, 1073–1081.

22. Wu, C.; Orozco, C.; Boyer, J.; Leglise, M.; Goodale, J.; Batalov, S.; Hodge, C.L.; Haase, J.; Janes, J.; Huss, J.W., 3rd; *et al*. BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **2009**, *10*, R130.

23. Cookson, W.; Liang, L.; Abecasis, G.; Moffatt, M.; Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **2009**, *10*, 184–194.

24. Bernstein, B.E.; Birney, E.; Dunham, I.; Green, E.D.; Gunter, C.; Snyder, M. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.

25. Bernstein, B.E.; Stamatoyannopoulos, J.A.; Costello, J.F.; Ren, B.; Milosavljevic, A.; Meissner, A.; Kellis, M.; Marra, M.A.; Beaudet, A.L.; Ecker, J.R.; *et al*. The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* **2010**, *28*, 1045–1048.

26. Maurano, M.T.; Humbert, R.; Rynes, E.; Thurman, R.E.; Haugen, E.; Wang, H.; Reynolds, A.P.; Sandstrom, R.; Qu, H.; Brody, J.; *et al*. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **2012**, *337*, 1190–1195.

27. Paul, D.S.; Albers, C.A.; Rendon, A.; Voss, K.; Stephens, J.; van der Harst, P.; Chambers, J.C.; Soranzo, N.; Ouwehand, W.H.; Deloukas, P. Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. *Genome Res.* **2013**, *23*, 1130–1141.

28. Ward, L.D.; Kellis, M. Haploreg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **2012**, *40*, D930–D934.

29. Schaub, M.A.; Boyle, A.P.; Kundaje, A.; Batzoglou, S.; Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **2012**, *22*, 1748–1759.

30. Auer, P.L.; Johnsen, J.M.; Johnson, A.D.; Logsdon, B.A.; Lange, L.A.; Nalls, M.A.; Zhang, G.; Franceschini, N.; Fox, K.; Lange, E.M.; *et al*. Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI go exome sequencing project. *Am. J. Hum. Genet.* **2012**, *91*, 794–808.

31. Chen, Z.; Tang, H.; Qayyum, R.; Schick, U.M.; Nalls, M.A.; Handsaker, R.; Li, J.; Lu, Y.; Yanek, L.R.; Keating, B.; *et al*. Genome-wide association analysis of red blood cell traits in African Americans: The cogent network. *Hum. Mol. Genet.* **2013**, *22*, 2529–2538.

32. Ganesh, S.K.; Zakai, N.A.; van Rooij, F.J.; Soranzo, N.; Smith, A.V.; Nalls, M.A.; Chen, M.H.; Kottgen, A.; Glazer, N.L.; Dehghan, A.; *et al*. Multiple loci influence erythrocyte phenotypes in the charge consortium. *Nat. Genet.* **2009**, *41*, 1191–1198.

33. Gudbjartsson, D.F.; Bjornsdottir, U.S.; Halapi, E.; Helgadottir, A.; Sulem, P.; Jonsdottir, G.M.; Thorleifsson, G.; Helgadottir, H.; Steinthorsdottir, V.; Stefansson, H.; *et al*. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **2009**, *41*, 342–347.

34. Kamatani, Y.; Matsuda, K.; Okada, Y.; Kubo, M.; Hosono, N.; Daigo, Y.; Nakamura, Y.; Kamatani, N. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **2010**, *42*, 210–215.

35. Lo, K.S.; Wilson, J.G.; Lange, L.A.; Folsom, A.R.; Galarneau, G.; Ganesh, S.K.; Grant, S.F.; Keating, B.J.; McCarroll, S.A.; Mohler, E.R., 3rd; *et al*. Genetic association analysis highlights new loci that modulate hematological trait variation in caucasians and African Americans. *Hum. Genet.* **2011**, *129*, 307–317.

36. Meisinger, C.; Prokisch, H.; Gieger, C.; Soranzo, N.; Mehta, D.; Rosskopf, D.; Lichtner, P.; Klopp, N.; Stephens, J.; Watkins, N.A.; *et al*. A genome-wide association study identifies three loci associated with mean platelet volume. *Am. J. Hum. Genet.* **2009**, *84*, 66–71.

37. Qayyum, R.; Snively, B.M.; Ziv, E.; Nalls, M.A.; Liu, Y.; Tang, W.; Yanek, L.R.; Lange, L.; Evans, M.K.; Ganesh, S.; *et al*. A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African Americans. *PLoS Genet.* **2012**, *8*, e1002491.

38. Soranzo, N.; Spector, T.D.; Mangino, M.; Kuhnel, B.; Rendon, A.; Teumer, A.; Willenborg, C.; Wright, B.; Chen, L.; Li, M.; *et al*. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the haemgen consortium. *Nat. Genet.* **2009**, *41*, 1182–1190.

39. Soranzo, N.; Rendon, A.; Gieger, C.; Jones, C.I.; Watkins, N.A.; Menzel, S.; Doring, A.; Stephens, J.; Prokisch, H.; Erber, W.; *et al*. A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. *Blood* **2009**, *113*, 3831–3837.

40. Chambers, J.C.; Zhang, W.; Li, Y.; Sehmi, J.; Wass, M.N.; Zabaneh, D.; Hoggart, C.; Bayele, H.; McCarthy, M.I.; Peltonen, L.; *et al*. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat. Genet.* **2009**, *41*, 1170–1172.

41. Reiner, A.P.; Lettre, G.; Nalls, M.A.; Ganesh, S.K.; Mathias, R.; Austin, M.A.; Dean, E.; Arepalli, S.; Britton, A.; Chen, Z.; *et al*. Genome-wide association study of white blood cell count in 16,388 African Americans: The continental origins and genetic epidemiology network (cogent). *PLoS Genet.* **2011**, *7*, e1002108.

42. Crosslin, D.R.; McDavid, A.; Weston, N.; Nelson, S.C.; Zheng, X.; Hart, E.; de Andrade, M.; Kullo, I.J.; McCarty, C.A.; Doheny, K.F.; *et al*. Genetic variants associated with the white blood cell count in 13,923 subjects in the emerge network. *Hum. Genet.* **2012**, *131*, 639–652.

43. Okada, Y.; Hirota, T.; Kamatani, Y.; Takahashi, A.; Ohmiya, H.; Kumasaka, N.; Higasa, K.; Yamaguchi-Kabata, Y.; Hosono, N.; Nalls, M.A.; *et al*. Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **2011**, *7*, e1002067.

44. Li, J.; Glessner, J.T.; Zhang, H.; Hou, C.; Wei, Z.; Bradfield, J.P.; Mentch, F.D.; Guo, Y.; Kim, C.; Xia, Q.; *et al*. GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum. Mole. Genet.* **2013**, *22*, 1457–1464.

45. Nalls, M.A.; Couper, D.J.; Tanaka, T.; van Rooij, F.J.; Chen, M.H.; Smith, A.V.; Toniolo, D.; Zakai, N.A.; Yang, Q.; Greinacher, A.; *et al*. Multiple loci are associated with white blood cell phenotypes. *PLoS Genet.* **2011**, *7*, e1002113.

46. Shameer, K.; Denny, J.C.; Ding, K.; Jouni, H.; Crosslin, D.R.; de Andrade, M.; Chute, C.G.; Peissig, P.; Pacheco, J.A.; Li, R.; *et al*. A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum. Genet.* **2014**, *133*, 95–109.

47. Okada, Y.; Takahashi, A.; Ohmiya, H.; Kumasaka, N.; Kamatani, Y.; Hosono, N.; Tsunoda, T.; Matsuda, K.; Tanaka, T.; Kubo, M.; *et al*. Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum. Mol. Genet.* **2011**, *20*, 1224–1231.

48. Monda, K.L.; Chen, G.K.; Taylor, K.C.; Palmer, C.; Edwards, T.L.; Lange, L.A.; Ng, M.C.; Adeyemo, A.A.; Allison, M.A.; Bielak, L.F.; *et al*. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat. Genet.* **2013**, *45*, 690–696.

49. N'Diaye, A.; Chen, G.K.; Palmer, C.D.; Ge, B.; Tayo, B.; Mathias, R.A.; Ding, J.; Nalls, M.A.; Adeyemo, A.; Adoue, V.; *et al*. Identification, replication, and fine-mapping of loci associated with adult height in individuals of African ancestry. *PLoS Genet.* **2011**, *7*, e1002298.

50. Reich, D.; Nalls, M.A.; Kao, W.H.; Akylbekova, E.L.; Tandon, A.; Patterson, N.; Mullikin, J.; Hsueh, W.C.; Cheng, C.Y.; Coresh, J.; *et al*. Reduced neutrophil count in people of African descent is due to a regulatory variant in the duffy antigen receptor for chemokines gene. *PLoS Genet.* **2009**, *5*, e1000360.

51. Greenburg, A.G. Pathophysiology of anemia. *Am. J. Med.* **1996**, *101*, 7S–11S.

52. Worldwide Prevalence of Anaemia 1993–2005 (WHO Global Database on Anaemia). Available online:http://whqlibdoc.Who.Int/publications/2008/9789241596657_eng.pdf (accessed on 19 November 2013).

53. Skoda, R. The genetic basis of myeloproliferative disorders. *Am. Soc. Hematol. Educ. Program* **2007**, *2007*, 1–10.

54. Oh, S.T.; Gotlib, J. JAK2 V617F and beyond: Role of genetics and aberrant signaling in the pathogenesis of myeloproliferative neoplasms. *Expert Rev. Hematol.* **2010**, *3*, 323–337.

55. 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.* **2011**, *2*, 21–32.

56. Hobbs, C.M.; Manning, H.; Bennett, C.; Vasquez, L.; Severin, S.; Brain, L.; Mazharian, A.; Guerrero, J.A.; Li, J.; Soranzo, N.; *et al*. JAK2V617F leads to intrinsic changes in platelet formation and reactivity in a knock-in mouse model of essential thrombocythemia. *Blood* **2013**, *122*, 3787–3797.

57. Sankaran, V.G.; Xu, J.; Orkin, S.H. Advances in the understanding of haemoglobin switching. *Br. J. Haematol.* **2010**, *149*, 181–194.

58. Sankaran, V.G.; Lettre, G.; Orkin, S.H.; Hirschhorn, J.N. Modifier genes in mendelian disorders: The example of hemoglobin disorders. *Ann. N. Y. Acad. Sci.* **2010**, *1214*, 47–56.

59. Platt, O.S.; Brambilla, D.J.; Rosse, W.F.; Milner, P.F.; Castro, O.; Steinberg, M.H.; Klug, P.P. Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N. Engl. J. Med.* **1994**, *330*, 1639–1644.

60. Platt, O.S.; Thorington, B.D.; Brambilla, D.J.; Milner, P.F.; Rosse, W.F.; Vichinsky, E.; Kinney, T.R. Pain in sickle cell disease. Rates and risk factors. *N. Engl. J. Med.* **1991**, *325*, 11–16.

61. Castro, O.; Brambilla, D.J.; Thorington, B.; Reindorf, C.A.; Scott, R.B.; Gillette, P.; Vera, J.C.; Levy, P.S. The acute chest syndrome in sickle cell disease: Incidence and risk factors. The cooperative study of sickle cell disease. *Blood* **1994**, *84*, 643–649.

62. Thein, S.L.; Craig, J.E. Genetics of HB F/F cell variance in adults and heterocellular hereditary persistence of fetal hemoglobin. *Hemoglobin* **1998**, *22*, 401–414.

63. Menzel, S.; Garner, C.; Gut, I.; Matsuda, F.; Yamaguchi, M.; Heath, S.; Foglio, M.; Zelenika, D.; Boland, A.; Rooks, H.; *et al*. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.* **2007**, *39*, 1197–1199.

64. Thein, S.L.; Menzel, S.; Peng, X.; Best, S.; Jiang, J.; Close, J.; Silver, N.; Gerovasilli, A.; Ping, C.; Yamaguchi, M.; *et al*. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11346–11351.

65. Uda, M.; Galanello, R.; Sanna, S.; Lettre, G.; Sankaran, V.G.; Chen, W.; Usala, G.; Busonero, F.; Maschio, A.; Albai, G.; *et al*. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1620–1625.

66. Galarneau, G.; Palmer, C.D.; Sankaran, V.G.; Orkin, S.H.; Hirschhorn, J.N.; Lettre, G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat. Genet.* **2010**, *42*, 1049–1051.

67. Lettre, G.; Sankaran, V.G.; Bezerra, M.A.; Araujo, A.S.; Uda, M.; Sanna, S.; Cao, A.; Schlessinger, D.; Costa, F.F.; Hirschhorn, J.N.; *et al*. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11869–11874.

68. Nuinoon, M.; Makarasara, W.; Mushiroda, T.; Setianingsih, I.; Wahidiyat, P.A.; Sripichai, O.; Kumasaka, N.; Takahashi, A.; Svasti, S.; Munkongdee, T.; *et al*. A genome-wide association identified the common genetic variants influence disease severity in beta(0)-thalassemia/hemoglobin E. *Hum. Genet.* **2010**, *127*, 303–314.

69. Sankaran, V.G.; Menne, T.F.; Xu, J.; Akie, T.E.; Lettre, G.; van Handel, B.; Mikkola, H.K.; Hirschhorn, J.N.; Cantor, A.B.; Orkin, S.H. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **2008**, *322*, 1839–1842.

70. Xu, J.; Peng, C.; Sankaran, V.G.; Shao, Z.; Esrick, E.B.; Chong, B.G.; Ippolito, G.C.; Fujiwara, Y.; Ebert, B.L.; Tucker, P.W.; *et al*. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science* **2011**, *334*, 993–996.

71. Bauer, D.E.; Kamran, S.C.; Lessard, S.; Xu, J.; Fujiwara, Y.; Lin, C.; Shao, Z.; Canver, M.C.; Smith, E.C.; Pinello, L.; *et al*. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **2013**, *342*, 253–257.

72. Hardison, R.C.; Blobel, G.A. Genetics. Gwas to therapy by genome edits? *Science* **2013**, *342*, 206–207.

73. Online Mendelian Inheritance in Man. Available online: http://omim.org/ (accessed on 18 November 2013).