npg

REVIEW

Common genetic factors for hematological traits in Humans

Yukinori Okada[1,2] and Yoichiro Kamatani[3]

Hematological traits are essential biomedical indicators that are widely used in clinical practice. The elucidation of the etiology that determines an individual's hematological traits would have a substantial impact. Hematological traits are known to be heritable, and it has been suggested that genetic factors contribute significantly to the inter-individual variance of these traits. Here, we review our current knowledge regarding the genetic architecture of hematological traits in humans, most of which has been obtained through recent developments in genome-wide association studies (GWAS). In addition to current knowledge, which is based on the hematological traits of the three major blood-cell lineages (white blood cells; WBC, red blood cells; RBC, and platelets; PLT), we propose future approaches that would be useful as a next step in the post-GWAS era.

INTRODUCTION

Hematological traits are essential medical indicators describing blood cells circulating in the human body.[1,2] Blood cells have an integral role in a variety of physiological processes involved in the maintenance of our vital activities. Blood cells are classified into three general categories: (i) white blood cells (WBC, also known as leucocytes), which mediate the immune system; (ii) red blood cells (RBC, also known as erythrocytes), which transport oxygen and respiratory gases and (iii) platelets (PLT, also known as thrombocytes), which form blood clots.[1] Homeostasis of hematological traits, including the counts and volume of peripheral blood cells and their biological activities, is tightly regulated within narrow physiological ranges, and abnormalities are closely linked to the development of disease. Therefore, laboratory data of hematological traits are widely used as important diagnostic markers in clinical practice. For example, a rapid increase in the WBC counts along with the development of a fever suggests that the patient has an infectious disease, and an acute decrease in RBC volume implies bleeding. Hematological traits can be useful prognostic parameters for several diseases and the overall health of individuals. PLT volume has been proposed as an independent predictor of morbidity and mortality for cardiovascular diseases (CAD).[3] Therefore, the elucidation of the genetic components of an individual's hematological traits would have a substantial impact.

Hematological traits are highly heritable, and it has been suggested that genetic factors as well as environmental factors, such as age, gender, obesity, smoking and drinking behaviors, contribute significantly to inter-individual variance.[4–8] Studies comparing monozygotic

and dizygotic twins have demonstrated that ∼50–90% of inter-individual variance in WBC, RBC and PLT counts could be attributed to genetic components.[4,5] These heritabilities were observed only within healthy individuals, but genetic factors would also give some variance to disease abnormal data, which may get clinicians confused in interpretation of values. The processes by which blood cells differentiate from hematopoietic stem cells and are maintained in the peripheral blood are partly or fundamentally shared among several lineages of blood cells.[9] For this reason, the values of each of the hematological traits are not distributed independently but are significantly correlated.[10] Additionally, significant differences in hematological traits exist between ethnic groups. African Americans are known to have lower WBC counts than Europeans,[11] and Japanese people have, on average, lower levels of RBC-related traits.[12] These observations suggest that there exist both shared and divergent genetic backgrounds for hematological traits and that they might be characterized by ethnicity.[2]

Recently, genome-wide association studies (GWAS), a new approach in genetic epidemiology that comprehensively evaluates hundreds of thousands or millions of polymorphisms in genome, have successfully identified causal genetic loci for many diseases.[13] As the initial GWAS was conducted at the RIKEN institute in Japan in 2002,[14–16] >1000 GWAS have been published until mid 2011 for >200 diseases or traits.[13] At first, GWAS were primarily conducted for binary phenotypes; that is, case–control association studies.[17] Then GWAS for quantitative phenotypes were launched, in which hematological, biological and physical traits were assessed.[18,19] To date, GWAS for a variety of hematological traits have been conducted in several

Table 1 Genetic association studies for hematological traits

Authors	Phenotypes	Total number of subjects	Novel loci identified in the study[a]	Reference
Gieger et al.[43]	PLT, MPV	66 867 Europeans and 24 005 Asians	53 loci	43
Maugeri et al.[42]	Mono	5347 Europeans	ITGA4	42
Okada et al.[41]	Neutro, lympho, mono, eosino, baso	14 792 Japanese	SLC45A3-NUCKS1, ITGA4, GATA2, the MHC region, CDK6, MLZE, NAA-LAD2, STXBP6, ERG	41
Reiner et al.[40]	WBC, neutro, lympho, mono, eosino, baso	16 388 African-Americans	CXCL2, HYDIN[b]	40
Nalls et al.[39]	WBC, neutro, lympho, mono, eosino, baso	31 332 Europeans	ITGA4, GATA2, the MHC region, MLZE, EDG2, EPS15L1	39
Lo et al.[38]	WBC, neutro, lympho, mono, eosino, baso, RBC, Hb, Ht, MCV, MCH, PLT	23 439 Europeans and 7112 African Americans	TPM4, G6PD	38
Kullo et al.[37]	RBC, Hb, MCV, MCH, MCHC	3012 Europeans	SLC17A1	37
Beall et al.[36]	Hb	196 Tibetans	EPAS1	36
Okada et al.[35]	Neutro	7665 Japanese	PSMD3–CSF3, PLCB4	35
Kamatani et al.[34]	WBC, RBC, Hb, Ht, MCV, MCH, MCHC, PLT	14 700 Japanese	43 loci	34
Ferreira et al.[33]	CD4+ T cell, CD8+ T cell, CD56+ T cell (NK), CD4:CD8 ratio	3526 Europeans	the MHC region, 17q12	33
Ferreira et al.[32]	WBC, neutro, lympho, mono, eosino, baso, RBC, Hb, Ht, MCV, MCH, MCHC, PLT	7558 Europeans	CCND3, 6q24, LPAR1	32
Benyamin et al.[31]	Hb, MCV, serum iron/transferrin/ferritin, transferrin saturation	4818 Australians and 3470 Dutch	TMPRSS6	31
Chambers et al.[30]	Hb	6316 Europeans and 9685 Indian Asians	TMPRSS6	30
Ganesh et al.[29]	RBC, Hb, MCV, MCH, MCHC	33 623 Europeans	18 loci	29
Soranzo et al.[28]	WBC, RBC, Hb, MCV, MCH, MCHC, PLT, MPV	13 943 Europeans	13 loci	28
Soranzo et al.[27]	PLT, MPV	8586 Europeans	PIK3CG	27
Gudbjartsson et al.[26]	Eosino	21 510 Europeans and 5212 East Asians	IL1RL1, IKZF2, GATA2, IL5, SH2B3	26
Reich et al.[25]	WBC	6005 African Americans	DARC	25
Meisinger et al.[24]	MPV	10 048 Europeans	ARHGEF3, WDR66, TAOK1	24
Lettre et al.[23]	HbF	1275 African Americans and 350 Brazilians	BCL11A, HBS1L-MYB	23
Uda et al.[22]	RBC, Hb, MCV, MCH, MCHC, HbA$_{1c}$, HbA$_2$, HbF	4826 Sardinians	BCL11A, HBS1L-MYB, HBB, HBA, CSNK2A1	22
Nalls et al.[21]	WBC	3709 African Americans and 1339 Europeans	DARC	21
Thein et al.[20]	HbF	2041 Europeans	HBS1L-MYB	20

Abbreviations: Baso, basophil; CD, cluster of differentiation; Eosino, eosinophil; Hb, hemoglobin; HbA$_{1c}$, hemoglobin A$_{1c}$; HbA$_2$, hemoglobin A$_2$; HbF, fetal hemoglobin; Ht, hematocrit; Lympho, lymphocyte; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; MHC, major histocompatibility complex; Mono, monocyte; MPV, mean platelet volume; Neutro, neutrophil; NK, natural killer cell; PLT, platelet; RBC, red blood cell; WBC, white blood cell.
[a]More than 10 of the newly identified loci are not indicated.
[b]Associations with the HYDIN locus have been suggested to be the result of a probe cross-hybridization artifact because of sequence similarity with the DARC locus.[40]

ethnic populations, and >100 genetic loci that affect individuals' hematological traits have been identified (Table 1).[20–43]

In this review, we describe the current approaches being used to elucidate the genetics of hematological traits in humans, which are primarily based on the three major lineages: WBC, RBC and PLT. We also propose additional approaches that are promising as next steps in the post-GWAS era.

GENETIC FACTORS FOR WBC

WBC form a substantial role in the immune system and defend the body against invading foreign microorganisms.[1] WBC include a variety of cells with distinct biological roles and are morphologically classified into five major subtypes: neutrophils, lymphocytes, monocytes, basophils and eosinophils. Although further refined classifications exist, especially for lymphocytes,[44] the five subtypes mentioned above are widely used in routine clinical screenings.

Neutrophils comprise 50–70% of the total WBC population and have an essential part in the innate immune system as an early defenses against microorganisms.[45,46] Inadequate numbers of circulating neutrophils, a hematological disorder called neutropenia, make affected individuals susceptible to bacterial infections. Lymphocytes comprise 20–30% of the total WBC population and have a pivotal role in the adaptive immune system.[1] Lymphocytes are further subdivided into several subsets based on their surface molecules, such as T-cells, B-cells and natural killer cells. Monocytes comprise 5–10% of the total WBC population and mediate the phagocytosis of pathogens and antigen presentation.[47] Eosinophils and basophils comprise 1–5% and 0.5–1% of the total WBC population, respectively. Both are known to be involved in the regulation of allergic reactions.[48,49] Basophils are the rarest of the major WBC subtypes and their role in host immunity has been less clear, although recent studies have indicated that they essentially provide unique functions including acting as antigen-

presenting cells that promote T helper cell 2-associated allergic responses.[48,50]

Associations at the *DARC* locus

Genetic association studies on WBC-related hematological traits firstly focused on the ethnic differences in total WBC counts between European and African populations[11] that could not be explained by known non-genetic factors (Table 1).[7] Admixture and fine mapping scans determined that a common African-derived null variant of rs2814778 at the Duffy antigen receptor for chemokines (*DARC*) gene at 1q23 was responsible for lower WBC and neutrophil counts in African Americans.[21,25] This null variant explained as much as around 20% of the inter-individual variance of total WBC counts in African Americans[21] and have been known to confer selective advantage against malaria.[51] A recent GWAS in African Americans identified an additional significant association between WBC counts and the *HYDIN* locus at 16q22, but this association was considered to be a probe cross hybridization artifact because of sequence similarity with the *DARC* locus that was induced by segmental duplication.[40]

Associations at the *PSMD3–CSF3* locus

Since 2009, GWAS for total WBC and WBC subtypes have identified a number of genetic loci (Table 1).[26,28,32–35,38–42] As expected, both common and unique associations for each WBC subtype were observed (Figure 1a), which reflects their shared and non-shared processes in hematopoiesis. Because neutrophils comprise the majority of WBC, it is not surprising that the associations identified in GWAS for these two phenotypes were mostly shared. This was most evident for the *PSMD3–CSF3* locus at 17q21.[28,34,35] The association at the *PSMD3–CSF3* locus would be interesting from a clinical point of view, because *CSF3*, which is also known as granulocyte colony-stimulating factor (*G-CSF*), encodes a cytokine that regulates granulocyte production[52,53] and its recombinant protein is used in clinical therapy to treat neutropenia.[1] Further analysis assessing relationships between *PSMD3–CSF3* variants and clinical outcomes, such as a response to *G-CSF* therapy and duration of agranulocytosis after chemotherapy, would be desirable.

Associations at the *ITGA4* locus

GWAS for monocytes in European and Japanese populations identified the *ITGA4* locus at 2q31.[39,41,42] *ITGA4* encodes the α4 chain of integrin, which mediates the migration of WBC into the tissues.[54] The variants of the *ITGA4* locus were shown to regulate the expression of *ITGA4* messenger RNA in monocytes.[42,55] Recent studies have demonstrated the clinical benefits of inhibition of α4 integrin in treatments for autoimmune diseases.[56] Although functional relevance between messenger RNA levels of *ITGA4* and their biological effects on immune-related cells needs to be elucidated, the identified variants of the *ITGA4* locus could be promising pharmacogenetic targets for anti-α4 integrin therapy.

Shared associations between eosinophils and basophils

Reflecting the similar biological roles of eosinophils and basophils, several overlapping genetic loci were identified for these two subtypes (Figure 1a). In a Japanese population, the *GATA2* and *ERG* loci demonstrated prominent association with both eosinophils and basophils.[41] In particular, the index single nucleotide protein (SNP) in the *GATA2* locus, rs4328821, explained 2.7% of the inherent correlation between these two phenotypes (Figure 1b).[10,41] Compared with the other WBC subtypes, GWAS for lymphocytes have been less successful.[32,38–41] This may be because the lymphocytes can be further subdivided, and these sub-phenotypes have not been sufficiently examined enough. A GWAS focusing on lymphocyte subsets by Ferreia *et al.*[33] successfully identified associated genetic loci, and additional accumulations of the studies on lymphocytes subsets would be desirable.

Development of trans-ethnic analysis

Notably, genetic association studies for WBC-related hematological traits have been cooperatively conducted to allow for trans-ethnic comparisons of regional associations among different populations. Recently, three large-scale GWAS for WBC and WBC subtypes were concurrently reported for European (by the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) consortium),[39,57] African American (by the COGENT (Continental Origins
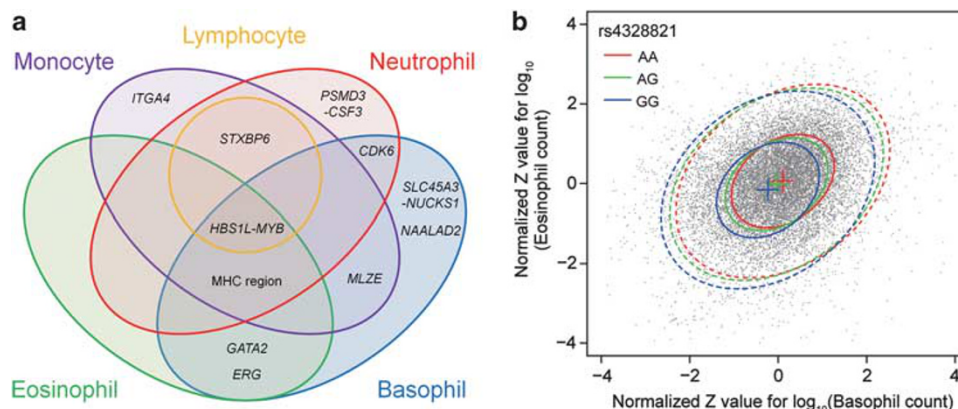

Figure 1 Shared associations between white blood cells' (WBC) subtypes. Shared associations of genetic loci with WBC subtypes in a Japanese population.[41] (**a**) A Venn diagram of the shared associations among the WBC subtypes-associated genetic loci. The colors in the Venn diagram (red, orange, purple, green and aqua) correspond to each of the WBC subtypes (neutrophils, lymphocytes, monocytes, eosinophils and basophils, respectively). (**b**) A scatter plot of the normalized *Z* values of the eosinophil and basophil counts in Japanese subjects (*n*=14 654). The center, 50% probability ellipse and 95% probability ellipse of the subjects with the AA (red)/AG (green)/GG (blue) genotypes of rs4328821 at the *GATA2* locus are indicated as crosses, solid ellipses and dashed ellipses, respectively. The subjects who were homozygous for the A allele had 1.19-fold and 1.28-fold higher eosinophil and basophil counts, respectively, than the subjects who were homozygous for the G allele. These figures are taken from Okada *et al.*[41] under a Creative Commons Attribution 2.5 Generic License.

and Genetic Epidemiology Network) consortium)[40] and Japanese populations (by the BioBank Japan Project).[41,58] Comparative analysis of these populations revealed that WBC have both ethnically shared and divergent genetic backgrounds. The association with the *DARC* locus was only observed in African Americans,[40] whereas the associations with the *PSMD3–CSF3* and *CXCL2* loci were shared among all three populations.[39–41] The index SNP (rs4328821) in the *GATA2* locus was associated with eosinophils and basophils in Japanese and with monocytes and basophils in Europeans.[40,41] This observation suggests that rs4328821 would be a candidate of a causal variant in this locus that may have a pivotal role in the regulation of these WBC subtypes. Allelic differences in the binding affinity of nuclear extracts from human basophilic leukemia cells have been demonstrated for this SNP using an electrophoretic mobility shift assay(personal communication with Drs. T. Hirota and M. Tamari at CGM (Center for Genomic Medicine), RIKEN).

Overlap with disease-associated loci
GWAS have also identified substantial overlap between the loci associated with WBC-related hematological traits and those implicated in susceptibility to several diseases. The *PSMD3–CSF3* locus and the major histocompatibility complex class II region have been implicated for rheumatoid arthritis, an autoimmune disease characterized by inflammation of the joints.[59–62] The *PSMD3–CSF3* locus is also a well-known risk locus for asthma.[63] The *ERG* locus, one of the loci associated with eosinophils and basophils,[41] is located in the Down syndrome critical region on chromosome 21.[64] These observations are compatible with our knowledge that WBC are essentially involved in the pathology of the multiple inflammatory diseases[46,65] and should encourage further comparative analysis of WBC-related loci with disease susceptibility.

GENETIC FACTORS FOR RBC
RBC are the most abundant blood cells and comprise approximately half of the total blood volume.[1] RBC are an essential components of the respiration system. They transport oxygen and carbon dioxide into tissues using the circulation of blood flow. The cytoplasm of RBC is rich in hemoglobin (Hb), an iron-containing molecule that is responsible for the red color of blood. Decreases and increases of RBC are called anemia and polycythemia, respectively, and are associated with multiple comorbid conditions, such as chronic bleeding, inflammation, smoking and renal impairment.[12,66] RBC counts, Hb levels and the proportion of RBC in the total blood volume (hematocrit; Ht) are the most common indicators of RBC used in clinical practice. Additionally, the following three indices are calculated using the RBC counts, Hb and Ht:

Mean corpuscular volume; MCV (fL)=Ht (%)/RBC ($\times 10^6 \mu l^{-1}$) $\times$ 10;
Mean corpuscular hemoglobin; MCH (pg)=Hb (g dl^{-1})/RBC ($\times 10^6 \mu l^{-1}$)$\times$10;
Mean corpuscular hemoglobin concentration; MCHC (%)=Hb (g dl^{-1})/Ht (%)$\times$100.

Although RBC, Hb and Ht are useful indicators of the basal status of anemia and polycythemia, MCV, MCH and MCHC provide additional information regarding the underlying pathologies of these disorders.

Associations at the *HBS1L–MYB* locus
Patients affected with blood disorders of sickle cell disease and β-thalassemia have an abnormal proportions in the compositions of Hb subunits, which is characterized by increase of fetal Hb.[67] Studies

investigating this phenomenon have found that the *BCL11A* and *HBS1L–MYB* loci contribute to the inter-individual variance of fetal Hb levels.[20,22,23] The *HBS1L–MYB* locus is now known as a fundamental regulator of blood cells, and it exhibits significant associations with all hematological traits, including those related to the three major lineages: WBC (WBC and WBC subtypes), RBC (RBC, Hb, Ht, MCV, MCH and MCHC) and PLT.[41] Associations in the *HBS1L–MYB* locus were observed in the intergenic region between *HBS1L* and *MYB*. The eQTL (expression quantitative trait locus) analyses demonstrated *HBS1L–MYB* as both cis-eQTL on *HBS1L* within 6q23 and trans-eQTL on *HBG1* and *HBG2* at 11p15 (eQTL Browser; http://eqtl. uchicago.edu/Home.html), whereas Ganesh et al.[29] reported relatively higher expression levels of *MYB* in erythroblasts than other genes in the locus.

Associations at the *TMPRSS6* locus
Large-scale GWAS for RBC-related hematological traits were first conducted in European populations,[28–32] and then in a Japanese population.[34] These studies identified dozens of novel loci, including the *TMPRSS6* locus.[30,31] *TMPRSS6* encodes a type II plasma membrane serine protease that has a pivotal role in iron homeostasis by regulating hepatic hepcidin production.[68] Concurrently, GWAS have been conducted on indicators of serum iron concentrations, such as serum transferrin, iron and ferritin levels, which are important clinical parameters for anemia.[69–72] These studies also identified the associations with the *TMPRSS6* locus.[70,71]

Associations at the *ABO* locus
A GWAS in a Japanese population found that a variant located in the *ABO* locus was associated with certain RBC-related hematological traits (RBC, Hb, Ht and MCHC).[34,73] *ABO* encodes glycosyltransferases that transfer specific sugar residues to the surface substance of RBC to produce blood-type-specific antigens. The combination of the sequence polymorphisms at *ABO* determines ABO blood type of the individual, the most important classification of blood types used in human blood transfusion.[74] The variants identified through the GWAS in the Japanese population were in complete linkage disequilibrium with rs8176743 (nt703),[34] one of the major deterministic variants of B antigens.[74] When the ABO blood types of 14 097 Japanese subjects were classified based on possession of B alleles, as defined by nt703, the subjects with the B allele (B or AB blood types) had ~ 0.15 mg dl^{-1} higher Hb levels than the subjects without the B allele (A or O blood types; $P=1.4 \times 10^{-8}$; Table 2).[34,73] This association was independently validated in 8421 Japanese subjects ($P=3.0 \times 10^{-5}$).[73] Nevertheless, ABO blood type of an individual can be more precisely defined using other polymorphisms in the locus,[75] and further validations is required.

Table 2 Mean hemoglobin levels of Japanese individuals based on ABO blood types[a]

Gender	Hemoglobin levels in ABO blood types (mg dl^{-1})[b]			
	**	B*	BB	B*+BB
Female[b]	12.92 (n=3176)	13.04 (n=1503)	13.10 (n=169)	13.05 (n=1672)
Male[b]	14.07 (n=6269)	14.23 (n=2,674)	14.16 (n=320)	14.22 (n=2994)

[a]Based on the results of a genome-wide association studies (GWAS) in a Japanese population.[34,73]
[b]Blood type alleles other than the B allele (i.e., A or O alleles) are denoted as *. The B allele was defined based on the T alleles of rs8176743 (nt703).[74]

Dissection of association patterns among RBC-related traits
Intriguingly, the number of identified loci and the degree of their associations were generally more prominent in the GWAS for MCV, MCH and MCHC than in those for RBC, Hb and Ht.[28,29,34] One probable explanation for this discrepancy is that the associations with medical indicators of physiological aspects of RBC (MCV, MCH and MCHC) are more robust to confounding factors than those with direct morphological measurements of RBC (RBC, HB and Ht). The associations observed for MCV, MCH and MCHC substantially overlapped, which is probably because their trait values were strongly correlated because of their dependent definitions being based on RBC counts, Hb and Ht.[2] Future studies should further explore which phenotype, or phenotypes, are the most responsible for the associations of each of the loci.

GENETIC FACTORS FOR PLT
PLT are the second most abundant blood cells. They have an essential role in hemostasis, the process that stops bleeding.[1] In case of bleeding, PLT adhere to damaged blood vessels and form a plug through aggregation. An abnormal excess of PLT in the blood may induce the formation of clots that obstruct vessels and can result in vascular events, including myocardial or brain infarction or pulmonary embolism. In contrast, a relative decrease in PLT in the blood may cause bruising or hemorrhage. It is known that PLT release multiple growth factors, such as platelet-derived growth factor, vascular endothelial growth factor and transforming growth factor-β, which have broad spectrums of impacts in physiology and medicine, including the repair and regeneration of tissues.[76,77] In clinical practice, two major phenotypes are measured; PLT counts and mean PLT volume (MPV). MPV has been suggested as an independent predictor of CAD.[3]

Associations at the THPO locus
Thrombocythemia is a chronic myeloproliferative syndrome that is caused by the sustained proliferation of megakaryocytes, the bone marrow cells responsible for production of PLT, which results in an increased numbers of circulating PLT. In 1998, a study examining a Dutch family with hereditary thrombocythemia found that THPO was the responsible gene for the disease.[78] THPO encodes thrombopoietin, a glycoprotein hormone that regulates the production of PLT.[53] A splice donor mutation in THPO leads overproduction of thrombopoietin.[78] Associations between the THPO locus with PLT counts were subsequently verified by a linkage analysis on an Asian Indian kindred[79] and a GWAS in a Japanese cohort.[34]

Associations at the WDR66 locus
As with other hematological traits, GWAS have successfully identified the genetic loci associated with PLT-related traits.[24,27,28,32,34,38,43] The initial GWAS for MPV was conducted in European populations and identified WDR66, ARHGEF3 and TAOK1 as candidate genes responsible for the regulation of PLT.[24] Of these, a variant located in the genetic region of WDR66 showed the most significant association, explaining approximately 2.0% of the inter-individual variance in MPV. A recent study reported that another variant of WDR66 was associated with a metabolic trait in human urine (2-hydroxyisobutyrate),[80] although its relevance to the regulation of PLT is unclear. Although we do not discuss the results in this review, GWAS for coagulation and fibrinolysis have also been reported.[81–85]

Associations at 12q24
Given the close biological relationship between PLT and CAD,[3] Soranzo et al.[28] evaluated the risks of CAD in the loci associated with PLT and found an overlap between these two traits at 12q24. The variants identified at 12q24 were encompassed by a long-range haplotype that extended >1.5 Mbps on which significant pressure of natural selection was observed. Interestingly, the association between PLT and CAD and the existence of the long-range haplotype at 12q24 was also observed in East Asian populations.[34,86,87] These results suggest the existence of variant(s) with strong biological effects on multiple traits at 12q24, which may have influenced natural selection in several populations.

Findings from a large-scale GWAS meta-analysis
Recently, Gieger et al.[43] reported a large-scale GWAS meta-analysis by the HaemGen consortium for PLT and MPV. The study incorporated >66 000 individuals of European ancestry and led to the identification of as many as 53 novel associated loci. A detailed comparison of the identified associations with those identified in Asian populations revealed a substantial overlap of the association signals for the two populations. Interestingly, gene silencing of the identified genes using Danio rerio and Drosophila melanogaster demonstrated the functional involvement with hematological phenotypes. This is one of the first successful studies that comprehensively assessed the genetic factors of the traits, including trans-ethnic comparison and functional verification of the identified loci using model organisms. Additional studies of this type with other hematological traits should be warranted.

STRATEGIES FOR THE POST-GWAS ERA
As we have discussed in this review, the recent development of the GWAS approach has provided us with fundamental knowledge regarding the genetic factors that influence hematological traits. However, it has been suggested that the majority of genetic factors could not be explained by the loci identified though GWAS because of the stringent significance threshold used.[88] In fact, despite the large sample size in the GWAS by the HaemGen consortium, the identified loci explained <10 and 20% of the inter-individual variance in PLT and MPV, respectively.[43] Although further accumulations of larger numbers of subjects in the GWAS should be recommended, alternative strategies would also be necessary to uncover these undetected loci in the post-GWAS era. It should also be noted that molecular mechanisms through which the identified variants could influence hematological traits are yet to be uncovered. Here, we would like to introduce several promising approaches (Figure 2).

Meta-analysis of GWAS involving multiple populations
As shown in the previous trans-ethnic studies for WBC subtypes and PLT,[39–41,43] the genetic factors that are associated with traits are partially shared by different populations.[89] Therefore, meta-analysis of GWAS involving multiple populations would yield the identification of novel associated loci that were not detected in studies of single populations and also contribute to fine mapping of the identified loci (Figure 2a). One issue to be resolved is that the different impacts of the variant(s) of the locus among different populations. Even when a significant association of a variant was observed in multiple populations, the effect sizes and allele frequencies would not be identical between the populations.[90] Most of the recent studies have used meta-analysis assuming a fixed-effect model for effect sizes; however, meta-analysis assuming a random-effect model or other sophisticated methods is required for trans-ethnic studies.[91,92]

Prioritization of GWAS results using external biological databases
To effectively explore the genetic loci that were uncaptured by the standard GWAS approach, it would be important to assess the subloci
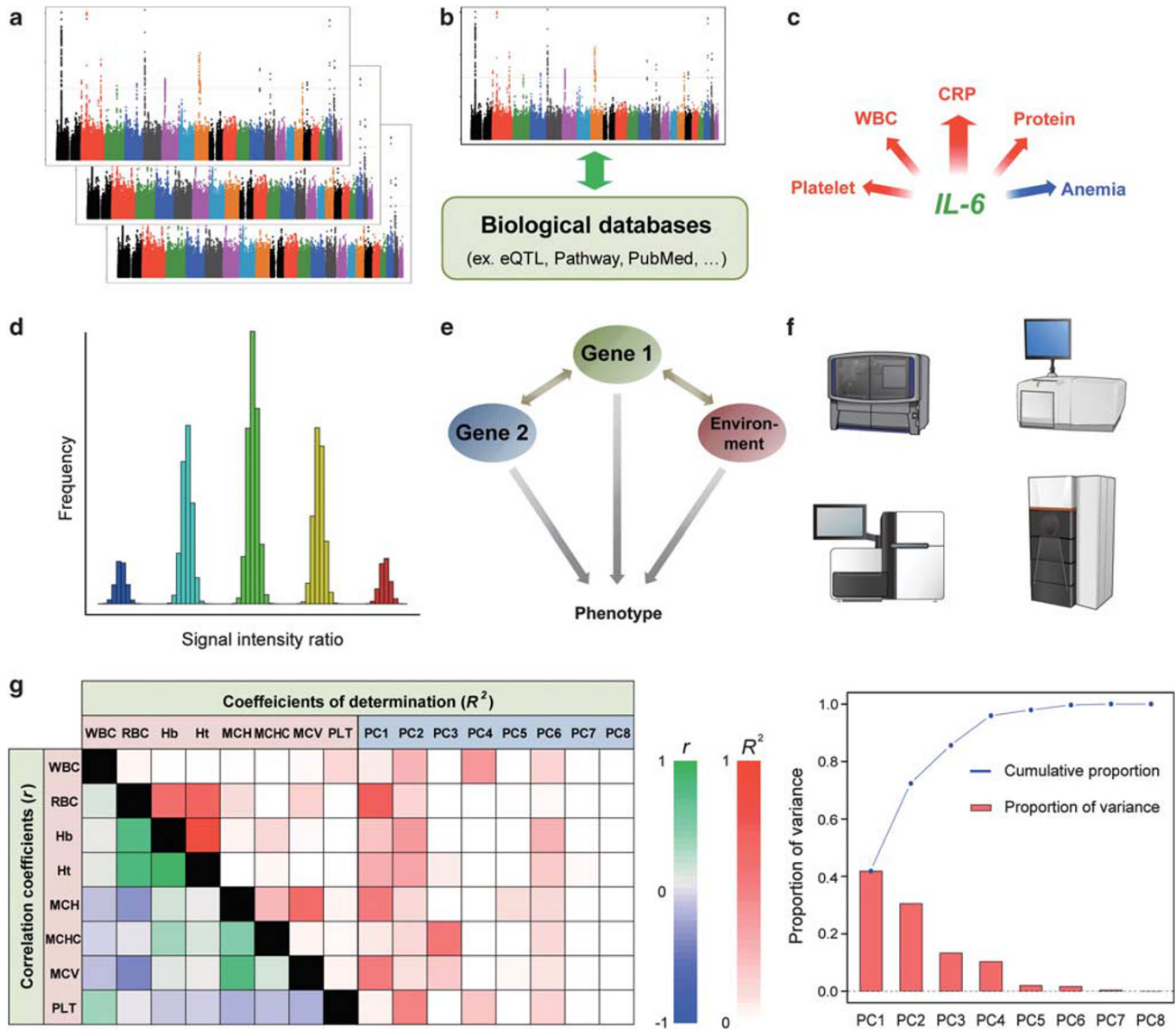
Figure 2 Strategies for detecting genetic factors in the post-genome-wide association studies (GWAS) era. (**a**) Meta-analysis of GWAS involving multiple populations would lead to the identification of novel loci that were not detected by the studies in single populations. (**b**) Prioritization of GWAS results using external biological databases, such as eQTL,[94] biological pathways[95] and the scientific literatures.[96] (**c**) Assessment of pleiotropic associations. A variant in the interleukin (*IL*)*6* locus (rs2097677) increased C-reactive protein (CRP), white blood cells (WBC), platelets (PLT) and serum protein, but decreased anemia-related indices.[98] (**d**) GWAS for structural variants, including copy number polymorphisms (CNPs)[100] and epigenetic variants.[101] (**e**) Evaluation of gene–gene or gene–environmental interactions. (**f**) Detection of rare variants using next generation sequencing (NGS). Pictures of NGS machines were obtained from the Togo picture gallery (http://g86.dbcls.jp/~togoriv/). ©2011 DBCLS Licensed under a Creative Commons Attribution 2.1 Japan License. (**g**) Extraction of shared components from inherently correlated traits. A correlation matrix comparing the covariate adjusted hematological trait values and principal components (PC) 1–8 obtained from a Japanese population (*n*=13 008) is shown in the left panel. The colors in the matrix correspond to the correlation coefficient, *r*, or coefficient of determination, R^2, in the legend. The proportion of variance explained by each PC or its cumulative proportion is shown in the right panel.

of GWAS results by prioritizing or weighting them by incorporating information from external biological databases (Figure 2b).[93] We conducted a highly selective replication study by prioritizing the result of a GWAS for systemic lupus erythematosus, an autoimmune disease of B-cell abnormality.[94] Using eQTL analysis for the lymphoblastoid B-cell lines, we successfully identified a novel susceptibility locus of *AFF1* at 4q21.[94] Because genes involved in specific biological pathways are likely to exhibit enrichments of association signals with related traits, prioritization based on pathway databases would also be useful.[95] Recently, approaches that aimed at understanding the biological

relatedness of the implicated genes and the traits based on published scientific literatures have yielded substantial achievements.[39,60,96]

Assessment for pleiotropic associations
Pleiotropy is a phenomenon that a single genetic locus influences multiple traits.[88,97] Because blood cells mediate a variety of biological processes in a coordinated manner, it would be plausible that some of the genetic loci would have pleiotropic effects on multiple hematological traits or on other related traits. A GWAS for C-reactive protein, a biomarker for the inflammatory

response, identified pleiotropic associations between the interleukin (*IL*)-6 locus at 7p15, and multiple hematological and biochemical traits.[98] The C-reactive protein-increasing allele of the variant in the interleukin-6 locus also increases WBC, PLT and serum protein levels; however, it decreases RBC-related anemia indices (Figure 2c).

GWAS for structural and epigenetic variants
Although SNP-based GWAS have been successful, there exists a concern that not all of the genetic risks are attributable to SNPs. Alternatively, a GWAS investigating structural variants, such as copy number polymorphisms could be a key to dissect this missing pieces (Figure 2d).[99] Kumasaka *et al.*[100] conducted a CNP-based GWAS for hematological traits in Japanese. The study suggested that the association of the major histocompatibility complex region with WBC and PLT previously observed in a SNP-based GWAS[34] may have reflected an association with copy number polymorphisms by linkage disequilibrium. Contributions of epigenetic variants on human diseases are also suggested. Integration of epigenome-wide association study with GWAS would be an effective method to detect origins of functional causality in the identified genetic loci.[101]

Evaluation of gene–gene or gene–environmental interaction
Standard GWAS are performed using the single-locus test that considers the impact of each locus separately. Recently, interest has focused on epistasis that considers interactions between multiple factors, such as gene–gene or gene–environmental interactions (Figure 2e).[102] Although there are few reports on gene–gene or gene–environmental interactions in the field of human quantitative traits, these approaches deserve further evaluation.

Detection of rare variants using next generation sequencing
Most of the GWAS to date have focused on common variants (minor allele frequency $\geqslant 0.01$–0.05), although recent developments in next generation sequencing technologies have enabled the assessment of rare variants (minor allele frequency < 0.01; Figure 2f).[103,104] Large-scale databases of rare human variants assembled by the 1000 Genomes Project Consortium are now publically available,[105] and association studies assessing rare variants have been launched.[106,107] Family-based linkage analysis is known to capture rare variants that are not detected by case-control GWAS, and the application of next generation sequencing to linkage analysis is also reported.[108]

Extraction of shared components among correlated traits
Finally, we would like to introduce a preliminary approach that would dissect the shared genetic architecture of multiple hematological traits. Because hematological traits are inherently correlated, we performed principal component analysis for the trait values to extract shared components using Japanese subjects obtained from the BioBank Japan Project (*n*=13 008; Figure 2g).[58] Each PC showed unique patterns of correlation with the original hematological traits, and the top four PC (PC 1–4) explained >90% of variance in the covariate-adjusted trait values. We conducted a GWAS in which each PC was assessed as a dependent variable, and using this approach, we identified novel loci that satisfied genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ for a certain PC, but did not satisfy the threshold in any of the GWAS for the original hematological traits (Okada Y, unpublished data). Although the biological meaning of each PC would be difficult to interpret, QTL-GWAS for PC could provide additional information for screening the loci implicated in hematological traits. An approach to evaluate the ratios of the trait pairs as secondary traits, in addition to the primary traits themselves, was also proposed.[109]

SUMMARY
As we have described in this review, the GWAS approach has successfully elucidated the genetics of hematological traits in humans. Additional studies, including assessments of the proposed strategies, are needed to uncover underlying genetic architectures that have yet to be detected.

1 Ronald, H., Edward, J. B., Sanford, J. S., Bruce, F., Leslie, E., Philip, M. *et al.* *Hematology: basic principles and practice* 5th edn (Churchill, Livingstone, 2009).
2 Andrews, N. C. Genes determining blood cell traits. *Nat. Genet.* **41,** 1161–1162 (2009).
3 Martin, J. F., Bath, P. M. & Burr, M. L. Influence of platelet size on outcome after myocardial infarction. *Lancet.* **338,** 1409–1411 (1991).
4 Whitfield, J. B. & Martin, N. G. Genetic and environmental influences on the size and number of cells in the blood. *Genet. Epidemiol.* **2,** 133–144 (1985).
5 Evans, D. M., Frazer, I. H. & Martin, N. G. Genetic and environmental causes of variation in basal levels of blood cells. *Twin. Res.* **2,** 250–257 (1999).
6 Hall, M. A., Ahmadi, K. R., Norman, P., Snieder, H., MacGregor, A. J., Vaughan, R. W. *et al.* Genetic influence on peripheral blood T lymphocyte levels. *Genes. Immun.* **1,** 423–427 (2000).
7 Fisch, I. R. & Freedman, S. H. Smoking, oral contraceptives, and obesity. Effects on white blood cell count. *JAMA.* **234,** 500–506 (1975).
8 Messinezy, M. & Pearson, T. C. Apparent polycythaemia: diagnosis, pathogenesis and management. *Eur. J. Haematol.* **51,** 125–131 (1993).
9 Kawamoto, H., Wada, H. & Katsura, Y. A revised scheme for developmental pathways of hematopoietic cells: the myeloid-based model. *Int. Immunol.* **22,** 65–70 (2010).
10 Elveback, L., Gully, R. J., Halberg, F. & Hamerston, O. Correlation of absolute basophil and eosinophil counts in blood from institutionalized human subjects. *J. Appl. Physiol.* **9,** 205–207 (1956).
11 Bain, B., Seed, M. & Godsland, I. Normal values for peripheral blood white cell counts in women of four different ethnic origins. *J. Clin. Pathol.* **37,** 188–193 (1984).
12 Tsubakihara, Y., Nishi, S., Akiba, T., Hirakata, H., Iseki, K., Kubota, M. *et al.* 2008 Japanese Society for Dialysis Therapy: guidelines for renal anemia in chronic kidney disease. *Ther. Apher. Dial.* **14,** 240–275 (2010).
13 Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 9362–9367 (2009).
14 Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T. *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32,** 650–654 (2002).
15 Thomas, D. C., Haile, R. W. & Duggan, D. Recent developments in genomewide association scans: a workshop summary and review. *Am. J. Hum. Genet.* **77,** 337–345 (2005).
16 Balding, D. J., Bishop, M. & Cannings, C. *Handbook of Statistical Genetics* 3rd edn (Wiley, 2007).
17 The WTCCC. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447,** 661–678 (2007).
18 Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature.* **467,** 832–838 (2010).
19 Okada, Y., Kamatani, Y., Takahashi, A., Matsuda, K., Hosono, N., Ohmiya, H. *et al.* A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum. Mol. Genet.* **19,** 2303–2312 (2010).
20 Thein, S. L., Menzel, S., Peng, X., Best, S., Jiang, J., Close, J. *et al.* Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 11346–11351 (2007).
21 Nalls, M. A., Wilson, J. G., Patterson, N. J., Tandon, A., Zmuda, J. M., Huntsman, S. *et al.* Admixture mapping of white cell count: genetic locus responsible for lower white

blood cell count in the Health ABC and Jackson Heart studies. *Am. J. Hum. Genet.* **82,** 81–87 (2008).

22 Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V. G., Chen, W. *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. U. S. A.* **105,** 1620–1625 (2008).

23 Lettre, G., Sankaran, V. G., Bezerra, M. A., Araujo, A. S., Uda, M., Sanna, S. *et al.* DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11869–11874 (2008).

24 Meisinger, C., Prokisch, H., Gieger, C., Soranzo, N., Mehta, D., Rosskopf, D. *et al.* A genome-wide association study identifies three loci associated with mean platelet volume. *Am. J. Hum. Genet.* **84,** 66–71 (2009).

25 Reich, D., Nalls, M. A., Kao, W. H., Akylbekova, E. L., Tandon, A., Patterson, N. *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.* **5,** e1000360 (2009).

26 Gudbjartsson, D. F., Bjornsdottir, U. S., Halapi, E., Helgadottir, A., Sulem, P., Jonsdottir, G. M. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41,** 342–347 (2009).

27 Soranzo, N., Rendon, A., Gieger, C., Jones, C. I., Watkins, N. A., Menzel, S. *et al.* A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. *Blood.* **113,** 3831–3837 (2009).

28 Soranzo, N., Spector, T. D., Mangino, M., Kuhnel, B., Rendon, A., Teumer, A. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41,** 1182–1190 (2009).

29 Ganesh, S. K., Zakai, N. A., van Rooij, F. J., Soranzo, N., Smith, A. V., Nalls, M. A. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* **41,** 1191–1198 (2009).

30 Chambers, J. C., Zhang, W., Li, Y., Sehmi, J., Wass, M. N., Zabaneh, D. *et al.* Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat. Genet.* **41,** 1170–1172 (2009).

31 Benyamin, B., Ferreira, M. A., Willemsen, G., Gordon, S., Middelberg, R. P., McEvoy, B. P. *et al.* Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nat. Genet.* **41,** 1173–1175 (2009).

32 Ferreira, M. A., Hottenga, J. J., Warrington, N. M., Medland, S. E., Willemsen, G., Lawrence, R. W. *et al.* Sequence variants in three loci influence monocyte counts and erythrocyte volume. *Am. J. Hum. Genet.* **85,** 745–749 (2009).

33 Ferreira, M. A., Mangino, M., Brumme, C. J., Zhao, Z. Z., Medland, S. E., Wright, M. J. *et al.* Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.* **86,** 88–92 (2010).

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

35 Okada, Y., Kamatani, Y., Takahashi, A., Matsuda, K., Hosono, N., Ohmiya, H. *et al.* Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count. *Hum. Mol. Genet.* **19,** 2079–2085 (2010).

36 Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J. *et al.* Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11459–11464 (2010).

37 Kullo, I. J., Ding, K., Jouni, H., Smith, C. Y. & Chute, C. G. A genome-wide association study of red blood cell traits using the electronic medical record. *PLoS. One.* **5,** e13011 (2010).

38 Lo, K. S., Wilson, J. G., Lange, L. A., Folsom, A. R., Galarneau, G., Ganesh, S. K. *et al.* Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum. Genet.* **129,** 307–317 (2011).

39 Nalls, M. A., Couper, D. J., Tanaka, T., van Rooij, F. J., Chen, M. H., Smith, A. V. *et al.* Multiple Loci are associated with white blood cell phenotypes. *PLoS. Genet.* **7,** e1002113 (2011).

40 Reiner, A. P., Lettre, G., Nalls, M. A., Ganesh, S. K., Mathias, R., Austin, M. A. *et al.* Genome-Wide Association Study of White Blood Cell Count in 16388 African Americans: the Continental Origins and Genetic Epidemiology Network (COGENT). *PLoS. Genet.* **7,** e1002108 (2011).

41 Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N. *et al.* Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS. Genet.* **7,** e1002067 (2011).

42 Maugeri, N., Powell, J. E., t Hoen, P. A., de Geus, E. J., Willemsen, G., Kattenberg, M. *et al.* LPAR1 and ITGA4 regulate peripheral blood monocyte counts. *Hum. Mutat.* **32,** 873–876 (2011).

43 Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature.* **480,** 201–208 (2011).

44 Okamura, T., Fujio, K., Shibuya, M., Sumitomo, S., Shoda, H., Sakaguchi, S. *et al.* CD4+CD25-LAG3+ regulatory T cells controlled by the transcription factor Egr-2. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13974–13979 (2009).

45 Witko-Sarsat, V., Rieu, P., Descamps-Latscha, B., Lesavre, P. & Halbwachs-Mecarelli, L. Neutrophils: molecules, functions and pathophysiological aspects. *Lab. Invest.* **80,** 617–653 (2000).

46 Fahy, J. V. Eosinophilic and neutrophilic inflammation in asthma: insights from clinical studies. *Proc. Am. Thorac. Soc.* **6,** 256–259 (2009).

47 Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D. N. *et al.* Nomenclature of monocytes and dendritic cells in blood. *Blood.* **116,** e74–e80 (2010).

48 Sullivan, B. M. & Locksley, R. M. Basophils: a nonredundant contributor to host immunity. *Immunity.* **30,** 12–20 (2009).

49 Gauvreau, G. M., Ellis, A. K. & Denburg, J. A. Haemopoietic processes in allergic disease: eosinophil/basophil development. *Clin. Exp. Allergy.* **39,** 1297–1306 (2009).

50 Yoshimoto, T., Yasuda, K., Tanaka, H., Nakahira, M., Imai, Y., Fujimori, Y. *et al.* Basophils contribute to T(H)2-IgE responses *in vivo* via IL-4 production and presentation of peptide-MHC class II complexes to CD4+ T cells. *Nat. Immunol.* **10,** 706–712 (2009).

51 Allison, A. C. Genetic control of resistance to human malaria. *Curr. Opin. Immunol.* **21,** 499–505 (2009).

52 Nagata, S., Tsuchiya, M., Asano, S., Kaziro, Y., Yamazaki, T., Yamamoto, O. *et al.* Molecular cloning and expression of cDNA for human granulocyte colony-stimulating factor. *Nature.* **319,** 415–418 (1986).

53 Kaushansky, K. Lineage-specific hematopoietic growth factors. *N. Engl. J. Med.* **354,** 2034–2045 (2006).

54 Rose, D. M., Han, J. & Ginsberg, M. H. Alpha4 integrins and the immune response. *Immunol. Rev.* **186,** 118–124 (2002).

55 Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R. *et al.* Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PLoS. One.* **5,** e10693 (2010).

56 von Andrian, U. H. & Engelhardt, B. Alpha4 integrins as therapeutic targets in autoimmune disease. *N. Engl. J. Med.* **348,** 68–72 (2003).

57 Psaty, B. M., O'Donnell, C. J., Gudnason, V., Lunetta, K. L., Folsom, A. R., Rotter, J. I. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2,** 73–80 (2009).

58 Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5,** 696–697 (2007).

59 Raychaudhuri, S., Remmers, E. F., Lee, A. T., Hackett, R., Guiducci, C., Burtt, N. P. *et al.* Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40,** 1216–1223 (2008).

60 Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42,** 508–514 (2010).

61 Okada, Y., Yamada, R., Suzuki, A., Kochi, Y., Shimane, K., Myouzen, K. *et al.* Contribution of a haplotype in the HLA region to anti-cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. *Arthritis. Rheum.* **60,** 3582–3590 (2009).

62 Okada, Y., Suzuki, A., Yamada, R., Kochi, Y., Shimane, K., Myouzen, K. *et al.* HLA-DRB1*0901 lowers anti-cyclic citrullinated peptide antibody levels in Japanese patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **69,** 1569–1570 (2010).

63 Moffatt, M. F., Kabesch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature.* **448,** 470–473 (2007).

64 Loughran, S. J., Kruse, E. A., Hacking, D. F., de Graaf, C. A., Hyland, C. D., Willson, T. A. *et al.* The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat. Immunol.* **9,** 810–819 (2008).

65 Fahey, J. L., Taylor, J. M., Detels, R., Hofmann, B., Melmed, R., Nishanian, P. *et al.* The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. *N. Engl. J. Med.* **322,** 166–172 (1990).

66 Zakai, N. A., Katz, R., Hirsch, C., Shlipak, M. G., Chaves, P. H., Newman, A. B. *et al.* A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. *Arch. Intern. Med.* **165,** 2214–2220 (2005).

67 Olivieri, N. F. & Weatherall, D. J. The therapeutic reactivation of fetal haemoglobin. *Hum. Mol. Genet.* **7,** 1655–1658 (1998).

68 Du, X., She, E., Gelbart, T., Truksa, J., Lee, P., Xia, Y. *et al.* The serine protease TMPRSS6 is required to sense iron deficiency. *Science.* **320,** 1088–1092 (2008).

69 Benyamin, B., McRae, A. F., Zhu, G., Gordon, S., Henders, A. K., Palotie, A. *et al.* Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. *Am. J. Hum. Genet.* **84,** 60–65 (2009).

70 Tanaka, T., Roy, C. N., Yao, W., Matteini, A., Semba, R. D., Arking, D. *et al.* A genome-wide association analysis of serum iron concentrations. *Blood.* **115,** 94–96 (2010).

71 Oexle, K., Ried, J. S., Hicks, A. A., Tanaka, T., Hayward, C., Bruegel, M. *et al.* Novel association to the proprotein convertase PCSK7 gene locus revealed by analysing soluble transferrin receptor (sTfR) levels. *Hum. Mol. Genet.* **20,** 1042–1047 (2011).

72 Pichler, I., Minelli, C., Sanna, S., Tanaka, T., Schwienbacher, C., Naitza, S. *et al.* Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. *Hum. Mol. Genet.* **20,** 1232–1240 (2011).

73 Okada, Y., Kamatani, Y., Takahashi, A., Matsuda, K., Honoso, N., Ohmiya, H. *et al.* Identification of genetic loci associated with quantitative traits using GWAS in a Japanese population. *The 55th meeting of the Japan Society of Human Genetics* (2010).

74 Yip, S. P. Sequence variation at the human ABO locus. *Ann. Hum. Genet.* **66,** 1–27 (2002).

75 Shintani-Ishida, K., Zhu, B. L., Maeda, H., Uemura, K. & Yoshida, K. A new method for ABO genotyping to avoid discrepancy between genetic and serological determinations. *Int. J. Legal. Med.* **122,** 7–9 (2008).

76 Andrae, J., Gallini, R. & Betsholtz, C. Role of platelet-derived growth factors in physiology and medicine. *Genes. Dev.* **22,** 1276–1312 (2008).

77 Miyazono, K. Shear activates platelet-derived latent TGF-beta. *Blood.* **112,** 3533–3534 (2008).

78 Wiestner, A., Schlemper, R. J., van der Maas, A. P. & Skoda, R. C. An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat. Genet.* **18,** 49–52 (1998).

79 Garner, C., Best, S., Menzel, S., Rooks, H., Spector, T. D. & Thein, S. L. Two candidate genes for low platelet count identified in an Asian Indian kindred by genome-wide linkage analysis: glycoprotein IX and thrombopoietin. *Eur. J. Hum. Genet.* **14,** 101–108 (2006).

80 Suhre, K., Wallaschofski, H., Raffler, J., Friedrich, N., Haring, R., Michael, K. *et al.* A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* **43,** 565–569 (2011).

81 Smith, N. L., Chen, M. H., Dehghan, A., Strachan, D. P., Basu, S., Soranzo, N. *et al.* Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation.* **121,** 1382–1392 (2010).

82 Tang, W., Basu, S., Kong, X., Pankow, J. S., Aleksic, N., Tan, A. *et al.* Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood.* **116,** 5032–5036 (2010).

83 Danik, J. S., Pare, G., Chasman, D. I., Zee, R. Y., Kwiatkowski, D. J., Parker, A. *et al.* Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: the Women's Genome Health Study. *Circ. Cardiovasc. Genet.* **2,** 134–141 (2009).

84 Dehghan, A., Yang, Q., Peters, A., Basu, S., Bis, J. C., Rudnicka, A. R. *et al.* Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. *Circ. Cardiovasc. Genet.* **2,** 125–133 (2009).

85 Johnson, A. D., Yanek, L. R., Chen, M. H., Faraday, N., Larson, M. G., Tofler, G. *et al.* Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. *Nat. Genet.* **42,** 608–613 (2010).

86 Ozaki, K., Sato, H., Inoue, K., Tsunoda, T., Sakata, Y., Mizuno, H. *et al.* SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nat. Genet.* **41,** 329–333 (2009).

87 Kato, N., Takeuchi, F., Tabara, Y., Kelly, T. N., Go, M. J., Sim, X. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* **43,** 531–538 (2011).

88 Stranger, B. E., Stahl, E. A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics.* **187,** 367–383 (2011).

89 Kurreeman, F., Liao, K., Chibnik, L., Hickey, B., Stahl, E., Gainer, V. *et al.* Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am. J. Hum. Genet.* **88,** 57–69 (2011).

90 Kochi, Y., Okada, Y., Suzuki, A., Ikari, K., Terao, C., Takahashi, A. *et al.* A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat. Genet.* **42,** 515–519 (2010).

91 Okada, Y., Mori, M., Yamada, R., Suzuki, A., Kobayashi, K., Kubo, M. *et al.* SLC22A4 polymorphism and rheumatoid arthritis susceptibility: a replication study in a Japanese population and a metaanalysis. *J. Rheumatol.* **35,** 1723–1728 (2008).

92 Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88,** 586–598 (2011).

93 Cantor, R. M., Lange, K. & Sinsheimer, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* **86,** 6–22 (2010).

94 Okada, Y., Shimane, K., Kochi, Y., Tahira, T., Suzuki, A., Higasa, K. *et al.* A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus in Japanese. *PLoS. Genet* doi:10.1371/journal.pgen.1002455.

95 Wang, K., Li, M. & Hakonarson, H. Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* **11,** 843–854 (2010).

96 Raychaudhuri, S., Plenge, R. M., Rossin, E. J., Ng, A. C., Purcell, S. M., Sklar, P. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS. Genet.* **5,** e1000534 (2009).

97 Okada, Y., Yamazaki, K., Umeno, J., Takahashi, A., Kumasaka, N., Ashikawa, K. *et al.* HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. *Gastroenterology.* **141,** 864–871 (2011).

98 Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N. *et al.* Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum. Mol. Genet.* **20,** 1224–1231 (2011).

99 The WTCCC. Genome-wide association study of CNVs in 16 000 cases of eight common diseases and 3000 shared controls. *Nature* **464,** 713–720 (2010).

100 Kumasaka, N., Fujisawa, H., Hosono, N., Okada, Y., Takahashi, A., Nakamura, Y. *et al.* PlatinumCNV: a Bayesian Gaussian mixture model for genotyping copy number polymorphisms using SNP array signal intensity data. *Genet. Epidemiol.* **35,** 831–844 (2011).

101 Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12,** 529–541 (2011).

102 Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10,** 392–404 (2009).

103 Fujimoto, A., Nakagawa, H., Hosono, N., Nakano, K., Abe, T., Boroevich, K. A. *et al.* Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nat. Genet.* **42,** 931–936 (2010).

104 Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **11,** 685–696 (2010).

105 1000 Genomes Project,. Consortium A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).

106 Holm, H., Gudbjartsson, D. F., Sulem, P., Masson, G., Helgadottir, H. T., Zanon, C. *et al.* A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat. Genet.* **43,** 316–320 (2011).

107 Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* **43,** 1066–1073 (2011).

108 Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nat. Rev. Genet.* **12,** 465–474 (2011).

109 Suhre, K., Shin, S. Y., Petersen, A. K., Mohney, R. P., Meredith, D., Wagele, B. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature.* **477,** 54–60 (2011).