The genetics of human hematopoiesis and its disruption in disease

Erik L Bao[1,2,3,4,†] (ID), Aaron N Cheng[1,2,3,†] & Vijay G Sankaran[1,2,3,5,*] (ID)

Abstract

Hematopoiesis, or the process of blood cell production, is a paradigm of multi-lineage cellular differentiation that has been extensively studied, yet in many aspects remains incompletely understood. Nearly all clinically measured hematopoietic traits exhibit extensive variation and are highly heritable, underscoring the importance of genetic variation in these processes. This review explores how human genetics have illuminated our understanding of hematopoiesis in health and disease. The study of rare mutations in blood and immune disorders has elucidated novel roles for regulators of hematopoiesis and uncovered numerous important molecular pathways, as seen through examples such as Diamond-Blackfan anemia and the *GATA2* deficiency syndromes. Additionally, population studies of common genetic variation have revealed mechanisms by which human hematopoiesis can be modulated. We discuss advances in functionally characterizing common variants associated with blood cell traits and discuss therapeutic insights, such as the discovery of BCL11A as a modulator of fetal hemoglobin expression. Finally, as genetic techniques continue to evolve, we discuss the prospects, challenges, and unanswered questions that lie ahead in this burgeoning field.

See the Glossary for abbreviations used in this article.

Introduction

Every second, each one of us produces millions of diverse circulating blood cells—including erythrocytes, platelets, and leukocytes—through the coordinated process of hematopoiesis (Fig 1A–C). This dynamic cascade, by which self-renewing stem cells that originate in the embryo go on to generate committed progenitors for the erythroid, megakaryocytic, granulocytic, monocytic, basophilic, eosinophilic, or lymphoid lineages over the course of a lifetime, is one of the best characterized paradigms of cellular differentiation (Orkin & Zon, 2008). However, our understanding of the regulation of hematopoiesis, mediated by transcription factors (TFs), cytokines, and other molecules, remains incomplete and continues to evolve (Jacobsen & Nerlov, 2019).

Early advances in the genetics of hematopoiesis were largely facilitated through the study of model organisms, including mice and zebrafish. These animal models allowed us to characterize gene function through reverse genetic approaches, with knockout experiments having a particularly prominent role over many decades. For example, much of our understanding of the Gata1 TF and its critical role in erythropoiesis stem from initial studies in mice over two decades ago, in which mice deficient in *Gata1* were found to have defective erythropoiesis (Pevny *et al*, 1991; Fujiwara *et al*, 1996). In a similar fashion, mice lacking critical hematopoietic cytokine receptors, such as the erythropoietin receptor (EpoR; Wu *et al*, 1995) or the granulocyte colony-stimulating factor receptor (Csf3r; Liu *et al*, 1996), were shown to have defective production of erythrocytes and neutrophils, respectively. This pattern of model organism-based reverse genetic discovery has been observed for countless other molecules critical for blood cell production (Orkin & Zon, 2008).

Although model organisms enable crucial insights into the functions of specific genes, there are important limitations to translating these findings to human biology and clinical impact. First, in contrast to the binary outcome of knockout models, most human diseases involve a diverse array of allelic variants that tune gene function or expression across a continuous spectrum, thus enabling insights into hypomorphic and other variant alleles. Second, model organisms are usually bred under a homogeneous genetic background. While this is useful for isolating the impact of specific genetic alterations, such isogenic backgrounds can mask the impact of genome-wide genetic variation upon phenotypes of interest. In light of these considerations, a powerful way to gain additional insights is to examine the spectrum of human genetic variation in health and disease. Indeed, studies of human genetic variation have enabled a multitude of important discoveries in hematopoiesis and have been applied to better understand and treat a range of blood diseases.

1   Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
2   Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA
3   Broad Institute of MIT and Harvard, Cambridge, MA, USA
4   Harvard-MIT Health Sciences and Technology, Harvard Medical School, Boston, MA, USA
5   Harvard Stem Cell Institute, Cambridge, MA, USA
     *Corresponding author. Tel: +1 617 919 2558; E-mail: sankaran@broadinstitute.org
     †These authors contributed equally to this work

Glossary

Cis-regulatory element
Genomic regions of transcription factor binding sites and other non-coding DNA that can influence transcription of a nearby gene. Examples include promoters, enhancers, and silencers.

Common variant association studies (CVAS)
Genetic studies which aim to identify common variants (usually defined as minor allele frequency > 1%) associated with a phenotype of interest.

Congenital dyserythropoietic anemia type II (CDA II)
The most common subtype of a group of rare hereditary disorders characterized by congenital anemia, ineffective erythropoiesis, the development of secondary hemochromatosis, and uniquely among CDA II, an abnormal glycosylation of erythrocyte membrane proteins.

Diamond-Blackfan anemia (DBA)
A rare inherited bone marrow failure syndrome characterized by normochromic macrocytic anemia, limited cytopenias of other lineages, low reticulocytes, and decreased erythroid precursor cells in the bone marrow.

Eosinophil
A type of white blood cell that plays important roles in fighting certain parasitic infections, and is also implicated in conditions such as allergies and asthma.

Epistasis
Interactions between genetic loci in their effect on a trait, such that the impact of a particular genotype depends on the genotype at other loci in a non-independent manner.

Expression-quantitative trait locus (eQTL)
Associations of DNA sequence variation with changes in gene expression.

Familial platelet disorder with predisposition to myeloid leukemia (FPDMM)
A rare inherited blood disorder caused by mutations of the *RUNX1* gene, clinically characterized by low platelet count, abnormal platelet function, and an increased risk of developing other blood disorders or cancers such as myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML).

Fanconi anemia (FA)
A heterogeneous genetic syndrome associated with risk of congenital malformations, bone marrow failure, and cancer.

Genome-wide association study (GWAS)
A genetic analysis that tests for genome-wide associations between genetic variants and a phenotype of interest.

Haploinsufficiency
The phenomenon in which a single functional copy of a gene is insufficient to maintain normal function.

Heritability
The proportion of variation in a particular trait that is attributable to genetic factors.

Imputation
The use of linkage patterns in a more densely sequenced reference panel to predict unobserved genotypes in a study dataset.

Linkage
The nonrandom association of alleles at different loci.

Myelodysplastic syndrome (MDS)
A heterogeneous group of malignant hematopoietic stem cell disorders characterized by dysplastic and ineffective blood cell production and a risk of transformation to acute leukemia.

Neutropenia
A decrease in circulating neutrophils.

Pleiotropy
A phenomenon in genetics whereby a DNA mutation or variant has an effect on multiple traits.

Polycythemia
An increased hemoglobin concentration and/or hematocrit in peripheral blood.

Polygenic risk score
A weighted sum of the number of risk alleles for a phenotype carried by an individual, where the risk alleles and their weights are usually defined by association loci and their effect sizes detected from genome-wide association studies.

Population stratification
Sample structure due to differences in genetic ancestry among samples.

Rare variant association studies (RVAS)
Genetic studies which aim to identify rare variants (usually defined as minor allele frequency < 1%) and their effects on a phenotype of interest.

Sickle cell disease (SCD)
A monogenic blood disease caused by a glutamic acid to valine substitution in the β-globin chain of normal adult hemoglobin, which causes polymerization of mutated sickle hemoglobin and deformation of red blood cells under conditions of deoxygenation.

Thrombocytopenia-absent radius (TAR) syndrome
A rare congenital syndrome primarily characterized by limb anomalies and low platelet counts.

Thrombocytopenia
A low number of platelets in the blood.

β-Thalassemia
A group of autosomal recessive hereditary anemias characterized by reduced or absent beta-globin chain synthesis, leading to alpha- and beta-chain imbalances that cause clinical manifestations of hemolytic anemia and impaired iron handling.

In this review, we discuss advances in the genetics of human hematopoiesis in three main sections. We first discuss genetic studies of inherited rare blood disorders, which have provided complementary insights as model organisms into this process through major perturbations. We next review more recent studies of common genetic variation impacting hematopoiesis which have further refined our understanding of this process. Finally, we discuss emerging efforts to combine rare and common genetic studies to achieve a more holistic understanding of hematopoiesis. We describe several clinically relevant vignettes to illustrate how human genetic studies have revealed new knowledge about human hematopoiesis. However, we note that we cannot comprehensively cover every example and instead highlight representative examples. Finally, we look ahead and discuss outstanding questions that will guide the next decade of research in this field.

Framework of human genetic studies

Human genetic studies can be broadly divided into common (allele frequency > 1%) and rare (allele frequency < 1%) variant association studies, each employing different approaches to work up their variants of interest. Common variant association studies (CVAS) usually take the form of genome-wide association studies (GWAS), in which individuals are genotyped using arrays that capture mostly higher-frequency variants. Statistical analyses can then be used to determine whether each variant is associated with a continuous or binary phenotype of interest. CVAS focus on traits with polygenic architectures comprised of many variants with small individual effects and usually include a large proportion of healthy individuals in the study population. Notable current limitations of CVAS include
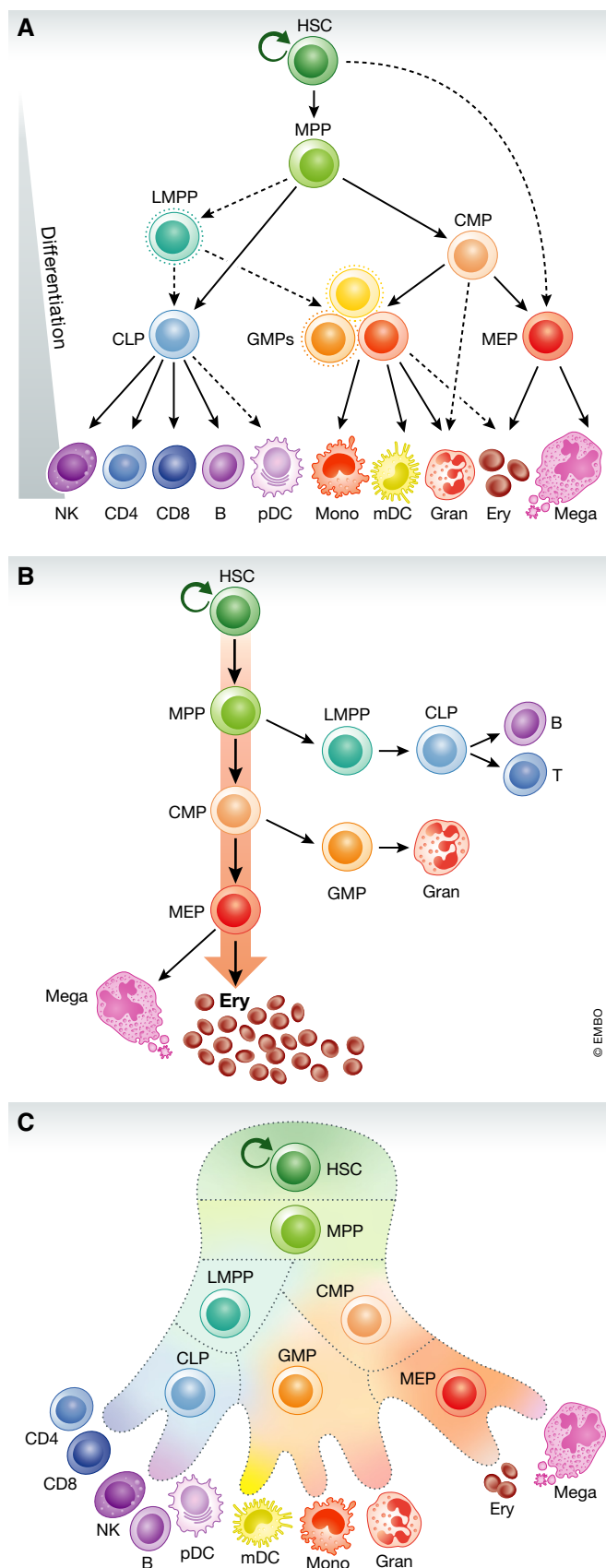
Figure 1. Overview of hematopoiesis.

(A) Schematic of the human hematopoietic hierarchy. Dashed lines indicate recently discovered differentiation paths. mono, monocyte; gran, granulocyte; ery, erythroid; mega, megakaryocyte; CD4, CD4[+] T cell; CD8, CD8[+] T cell; B, B cell; NK, natural killer cell; mDC, myeloid dendritic cell; pDC, plasmacytoid dendritic cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; GMP, granulocyte–macrophage progenitor; MEP, megakaryocyte–erythroid progenitor. Figure adapted from Corces et al (2016). (B) Quantitative depiction of hematopoietic hierarchy, in which erythroid commitment is the predominant and default pathway of differentiation. Figure adapted from Boyer et al (2019). (C) Visualization of hematopoietic hierarchy in which lineage commitment occurs on a continuum rather than in punctuated stages, a perspective motivated by recent single-cell transcriptomic studies. Figure adapted from Grootens et al (2019).

its high multiple testing burden from evaluating millions of variants, its inability to capture a substantial portion of heritability, and the difficulty of functionally characterizing association signals (Tam et al, 2019).

Rare variant association studies (RVAS) often require alternative analytical methods, since single-variant analysis can be underpowered to detect associations if the individual mutation is too rare in the study population. To counteract this, burden tests have been developed, which collapse many variants within a gene or region into a single risk score. This approach thus performs a per-gene or per-region association study as opposed to per-variant association tests in GWAS (Lee et al, 2012; Zuk et al, 2014). Another important difference is that GWAS typically employ single nucleotide polymorphism (SNP) arrays to directly genotype up to a few million common variants. Millions of additional variants can then be inferred via imputation, which is the process of using linkage patterns in a more densely sequenced reference panel to predict unobserved genotypes in the study dataset. However, these methods are ineffective for identifying extremely rare variants, especially when the variants are previously unreported or in low linkage with other variants (preprint: Van Hout et al, 2019). Therefore, RVAS typically use targeted sequencing, whole-exome sequencing (WES), or whole-genome sequencing (WGS), which allow for unbiased variant calling to identify rare or novel variants that would not have been included on genotyping arrays or that are not confidently imputed (preprint: Wainschtein et al, 2019). In addition, RVAS study populations are usually smaller than in CVAS and are more enriched for disease cases. Finally, some limitations of RVAS are that they usually miss non-coding associations due to exclusion (WES) or low sequencing depth (WGS), and they require assumptions about the underlying genetic model when aggregating variants (Lee et al, 2014). Keeping in mind this broad framework of CVAS and RVAS, we now dive into how these approaches have been applied to study hematopoiesis in health and disease.

Genetic studies of rare blood disorders

In the early years of human genetics, prior to the advent of high-throughput sequencing technologies, most efforts revolved around studying rare blood diseases displaying Mendelian or monogenic inheritance patterns, and this continues to be a powerful approach today. What have such studies of rare blood disorders taught us? On

one hand, they have demonstrated how allelic variation in known hematopoietic regulators creates more diverse clinical manifestations compared to the all-or-none knockout studies of model organisms. Secondly, they have revealed how fundamental biological processes can often have distinct and unexpected roles in hematopoiesis. In this section, we describe genetic approaches for studying rare blood diseases and then highlight examples of important biological insights gained from such studies.

Several methods have been employed to map rare blood diseases to causal genetic mutations. In the past, the most common approach was linkage analysis. This approach involves recruiting families with a disease or phenotype of interest, detecting co-segregation of the disease with genetic markers of known chromosomal location, and pinpointing a mutated gene in the linkage window. However, since the development of massively parallel sequencing in the last decade, targeted sequencing, WES, and WGS have emerged as far more scalable and higher-resolution ways to dissect the genetics of rare blood disorders. These approaches have had great success in identifying rare loss (or gain)-of-function coding variants segregating within families with hematologic traits at extremes of the phenotypic distribution (Minelli *et al*, 2004; Shiohara *et al*, 2009; Albers *et al*, 2011; Sankaran *et al*, 2012).

Studies of Diamond-Blackfan anemia (DBA) nicely illustrate how rare variant genetics have illuminated our understanding of human hematopoiesis and how this process can be perturbed in disease in unexpected ways. DBA is an inherited hypoplastic anemia in which erythroid precursors and progenitors are selectively reduced in the bone marrow, while other lineages are ostensibly produced normally (Nathan *et al*, 1978). The first gene mapping studies used linkage analysis of families with DBA to localize a disease-associated region to 1 Mb on chromosome 19 (Gustavsson *et al*, 1997), which was later found to be attributable to loss-of-function mutations in ribosomal protein (RP) gene *RPS19* (Draptchinskaia *et al*, 1999). Subsequent studies identified at least 25 additional RP mutations that explained up to 80% of DBA cases (Landowski *et al*, 2013; Ulirsch *et al*, 2018). However, how heterozygous loss of function of ubiquitously expressed RP genes could cause a selective absence of erythroid cells remained a mystery. New gene discoveries, facilitated by broader methods for genetic interrogation, enabled further insights into this disease. WES of patients who had a clinical diagnosis of DBA, but no known pathogenic mutations, revealed mutations impairing the production of GATA1 in several patients (Sankaran *et al*, 2012; Ludwig *et al*, 2014). Building upon this knowledge, subsequent functional studies solidified the link between RPs, GATA1, and defects in erythropoiesis by showing that RP haploinsufficiency reduces ribosome levels and thus results in reduced *GATA1* mRNA translation (Ludwig *et al*, 2014; Khajuria *et al*, 2018). Therefore, DBA genetics revealed new information about the regulation of GATA1 expression in human hematopoiesis, establishing a novel link between ribosome levels and GATA1 driven by its high translation rate (Khajuria *et al*, 2018).

Another example in which RVAS has enhanced our understanding of a known hematopoietic regulator is exemplified by human variation impacting *RUNX1*. Germline mutations leading to *RUNX1* deficiency cause familial platelet disorder with predisposition to myeloid leukemia (FPDMM). In 1999, linkage analysis of six separate families with FPDMM revealed that all

pedigrees contained heterozygous mutations in *RUNX1* (Song *et al*, 1999). Further analysis of the affected individuals showed a deficiency in megakaryocyte colony formation, implicating *RUNX1* as a regulator of megakaryopoiesis. These cases were particularly intriguing because of the high rate of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) in affected individuals, and demonstrated a key link between *RUNX1* haploinsufficiency and predisposition to malignant hematopoiesis (Owen *et al*, 2008). Given the challenges of studying *Runx1* in mice due to early embryonic lethality (Ichikawa *et al*, 2004), human genetic studies of FPDMM have provided powerful insights into the roles of the RUNX1 TF in normal and malignant hematopoiesis.

Additionally, RVAS of blood disorders have shown how mutations in a single master TF can result in pleiotropic and variable phenotypes, as nicely represented by the study of disorders attributable to deficiency of *GATA2*. In the past decade, researchers have found that GATA2 deficiency can cause a constellation of disparate disorders, including cases of monocytopenia with susceptibility to atypical mycobacterial infection ("MonoMAC"); loss of dendritic cells, monocytes, B, and natural killer (NK) cells (DCML deficiency); and familial MDS and AML (Dickinson *et al*, 2011; Hsu *et al*, 2011; Ostergaard *et al*, 2011). Common across this spectrum of manifestations is the notable evolution of symptoms with age, suggesting that variation in early hematopoietic stem and progenitor function may underlie many of the pleiotropic phenotypes in this disorder (Collin *et al*, 2015). Further studies of these disorders will likely provide more insights into how a master TF of hematopoiesis can lead to such disparate and variable phenotypes.

While rare variant genetics have done much to further characterize factors with known roles in blood cell production, such studies have also connected previously unappreciated molecular pathways with hematopoiesis. For example, extensive work on DBA genetics shed light on the connection between the seemingly distinct pathways of ribosome regulation and erythroid lineage commitment. There are several additional examples of this trend. For instance, investigating cases of thrombocytopenia-absent radius (TAR) syndrome identified biallelic mutations in *RBM8A*, which encodes the Y14 subunit of the exon-junction complex (Albers *et al*, 2012). This revelation linked a general splicing factor to hematopoiesis and suggested that lineage-dependent deficiency of a ubiquitous protein may cause a very specific phenotype. While the exon-junction complex has been shown to play an important role in regulating RNA through alternative splicing and may be involved in fine-tuning gene expression (Michelle *et al*, 2012; Ishigaki *et al*, 2013; Mao *et al*, 2016), the exact basis of the mechanistic connection between the exon-junction complex and platelet production remains unresolved. Studies of Fanconi anemia (FA) have similarly unearthed a previously uncharacterized connection between genetic mutations underlying FA and critical DNA damage repair pathways. In particular, the FA pathway has been found to play critical roles in DNA inter-strand cross-link repair, homologous recombination, and nucleotide excision repair, among other pathways (Ceccaldi *et al*, 2016; Sumpter & Levine, 2017; Niraj *et al*, 2019). Finally, congenital dyserythropoietic anemia type II (CDA II) was found to be caused by mutations in SEC23B, a ubiquitous component of the secretory COPII coat protein complex involved in Golgi trafficking

(Schwarz *et al*, 2009), due to the absence of the paralog SEC23A within the erythroid lineage (Khoriaty *et al*, 2018). All of these examples demonstrate the broad impact of rare variant studies on advancing our understanding of human hematopoiesis and associated fundamental biological processes. There are many more examples of RVAS elucidating new pathways in diverse hematopoietic lineages that we are unable to explore here due to space constraints.

While many of the examples discussed above emerged through traditional family-based linkage or sequencing analyses, as larger cohorts of rare disease patients are being assembled, broader assessments through RVAS and gene burden analyses are occurring. Such approaches have been valuable in the context of DBA (Ulirsch *et al*, 2018), as well as for studies of patients with rare congenital forms of thrombocytopenia and immunodeficiencies (preprint: Downes *et al*, 2018, preprint: Thaventhiran *et al*, 2018). There is no doubt that as larger collaborative efforts are established for rare disease patients, including those with genetic blood disorders, there will be more opportunities to identify additional causal and modifier genes. Moreover, these studies highlight the incomplete penetrance or variable expressivity of many alleles when examined in large cohorts of patients compared with healthy population controls (Ulirsch *et al*, 2018). Such discoveries will pave the way for further insights into human hematopoiesis.

Population-based genetic studies of hematopoiesis and their translation to clinical impact

In addition to the lessons gleaned from studying rare variants, there has been an equally fertile ground on the opposite side of the frequency spectrum. In this section, we will explore a burgeoning array of approaches applied to dissect *common* genetic variation and how they have advanced our understanding of human hematopoiesis.

At the population level, there is a wide spectrum of variation in commonly measured blood traits such as hemoglobin levels and blood cell counts. These traits not only cause disease at extreme ends of the spectrum (e.g., anemia, polycythemia, thrombocytopenia, and neutropenia), but also are independent risk factors for a multitude of non-hematological diseases, including leukocyte count for coronary heart disease (Ensrud & Grimm, 1992; Hoffman *et al*, 2004) and eosinophil count for asthma (Astle *et al*, 2016), highlighting the importance of better understanding how hematopoiesis is regulated. Large family studies have estimated these blood indices to be highly heritable (Pilia *et al*, 2006), meaning that a significant portion of the observed variation in phenotype can be attributed to genetic factors. However, the precise genetic variants responsible for this variation and their mechanisms of action remain poorly understood.

To answer these questions, many groups have leveraged natural variation in blood cell traits in healthy populations to study their genetic underpinnings, most often through GWAS. In part due to the low cost and widespread availability of blood count measurements, many large-scale GWAS have been performed on these traits in various ancestries. Together, these studies have identified thousands of genomic loci linked to blood cell measurements (Table EV1; Fig 2A and B).

Interestingly, multiple GWASs have been performed on the same hematopoietic phenotypes over the past 15 years, with each successive study featuring more and better resolved genetic associations. Why has this been the case? Statistical power to detect true genetic associations is a function of variant allele frequency, the effect size of a variant on the phenotype of interest, and the sample size of the study (Fig 3A and B; Skol *et al*, 2006). Of these, sample size is the most scalable and extrinsic to the variant–phenotype relationship and thus has seen the largest uptick. Other advances contributing to GWAS power and resolution include the generation of larger reference panels and more accurate computational algorithms for imputation, as well as improved statistical models for genetic association testing that correct for population stratification and relatedness. These developments have collectively fueled an explosion in the number of associations identified by GWAS, including those linked to blood cell indices. For example, a study in 2009 identified a single locus associated with mean platelet volume, explaining ~ 1.5% of the total trait variance (Soranzo *et al*, 2009). Seven years later, a GWAS with a > 10-fold increase in sample size discovered 294 significant loci, including many with lower allele frequency and smaller effect sizes, collectively explaining ~ 30% of phenotypic variance in the same trait (Astle *et al*, 2016).

However, as a result of the rapid progress in this field, a critical follow-up question has arisen: How can we obtain meaningful biological insight from so many robust genetic signals? In the post-GWAS era, the challenge has shifted to not only identify genetic regions associated with blood cell traits, but also pinpoint the exact variants driving each signal, the genes targeted by these variants, and the cell types in which they act, in order to ultimately better understand mechanisms underlying the regulation of hematopoiesis in health and disease (Gallagher & Chen-Plotkin, 2018). In other words, there is now a pressing need to move from variant to function to achieve the same richness of biological insights that have been derived from studies of rare blood diseases, as discussed above. A variety of computational and experimental approaches have been developed to tackle this multifaceted challenge (Table EV2 and Fig 4).

An estimated 80–90% of GWAS loci tag non-coding regions, where many variants tune gene expression by modulating regulatory elements such as promoters and enhancers (Gusev *et al*, 2014). Many methods have leveraged this connection to map variants to target genes and biological pathways in a tissue-specific manner. For example, expression-quantitative trait locus (eQTL) studies identify variants which influence gene expression (GTEx

Figure 2. Trends in genome-wide association studies (GWAS) of blood traits.
(A) Sample size of GWAS for commonly measured hematopoietic traits, including red cell, platelet, and leukocyte traits, over time. (B) Number of independent genome-wide significant loci discovered for the hemoglobin trait as a function of study sample size. In both panels, the colors of lines and points indicate the ancestry of the study population. The text labels denote the first author of each study.
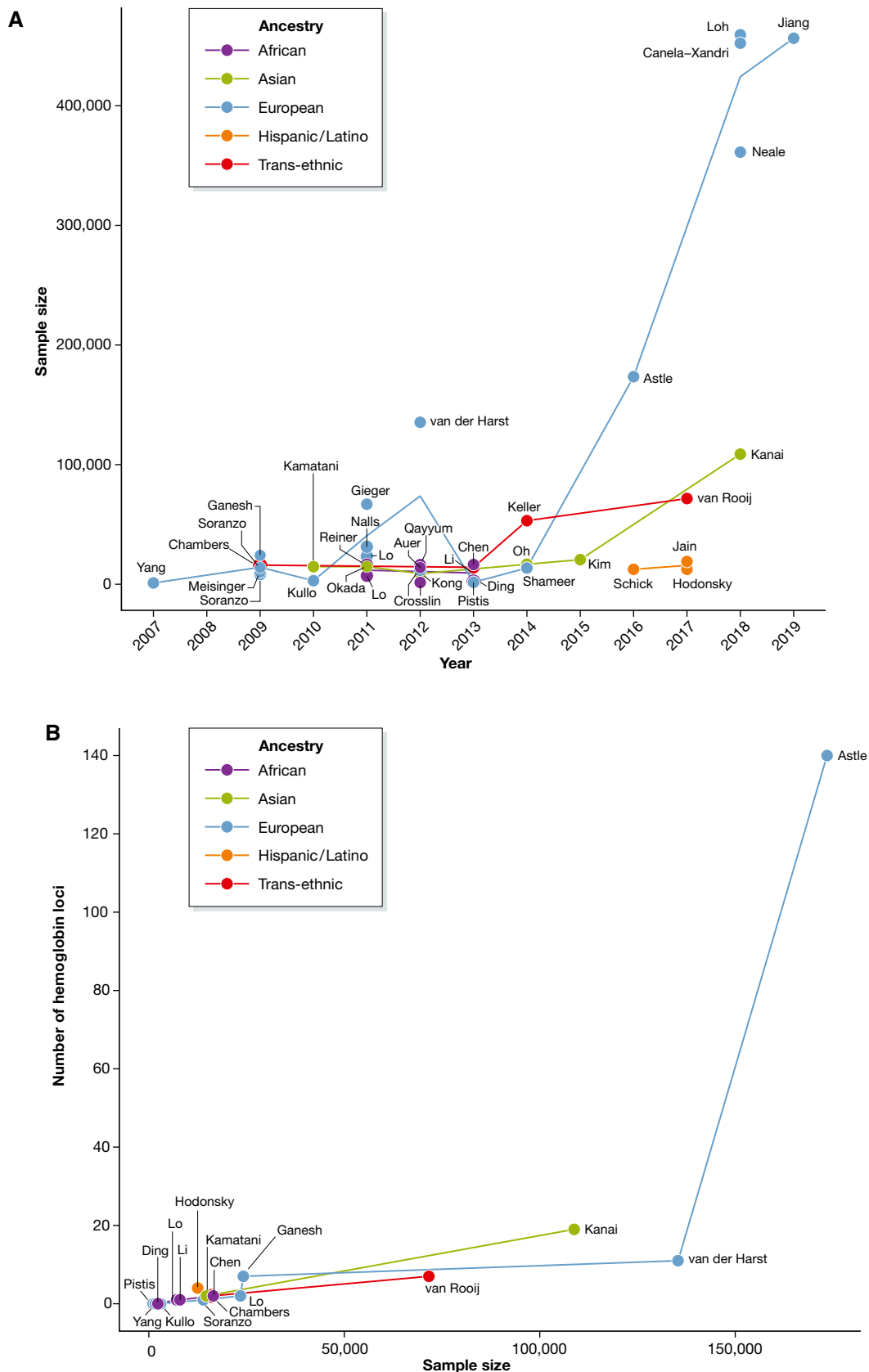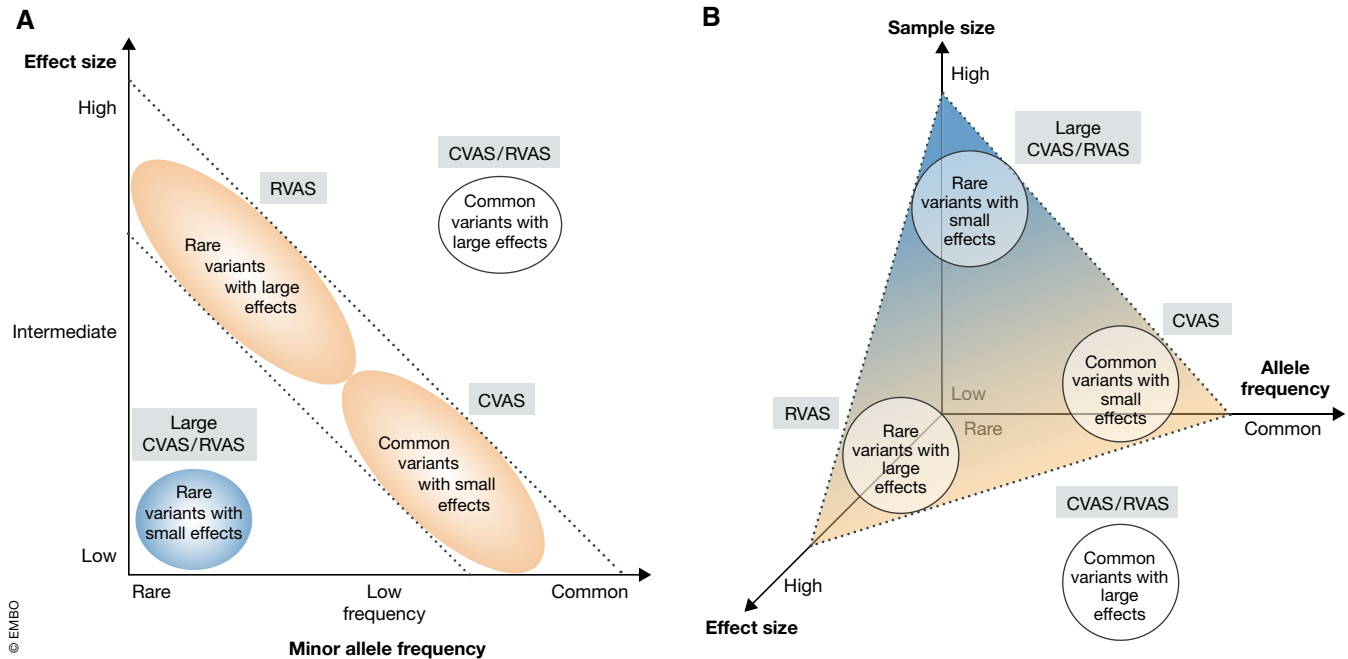
Figure 2.

Figure 3. Discovery of association signals across the allelic frequency spectrum.

(A) Traditional depiction of variant discovery power by genetic association studies, as a function of variant effect size and allele frequency. Figure adapted from Manolio *et al* (2009). (B) Revised schematic that illustrates the 3-way relationship between (i) sample size of the study, (ii) effect size of a genetic variant, (iii) and allele frequency of the variant on discovery power. The dashed triangular plane indicates the sample size threshold above which studies are sufficiently powered to detect variants at any given coordinate of allele frequency and effect size. The labeled circles depict categories of variants which are most often studied by the analytical methods listed next to them: common variant association studies (CVAS) and/or rare variant association studies (RVAS).

Consortium *et al*, 2017); chromatin immunoprecipitation (ChIP)-seq can profile DNA methylation, histone modification, and TF binding sites, which can be used to predict the potential for non-coding variants to perturb molecular phenotypes (Chen *et al*, 2016); and chromatin capture methods identify looping interactions between enhancers and gene promoters (Javierre *et al*, 2016). High-throughput gene knockdown or knockout screens have also become a popular experimental approach to identify genes that are important for a phenotype of interest. These screens can be implemented with short hairpin RNA or CRISPR-Cas9 genetic perturbation to systematically test which genes or regulatory elements are essential for a phenotype, such as hematopoietic lineage differentiation (Canver *et al*, 2017; Nandakumar *et al*, 2019).

GWAS of blood cell traits, combined with these functional analyses, have provided tremendous value for understanding the transcriptional regulatory mechanisms underlying hematopoiesis. For example, such studies have identified blood trait-associated variants which predispose or protect from clinical disease (e.g., eosinophil-associated loci increasing risk of rheumatoid arthritis), enabled the discovery of genes and TFs important for the stage-specific regulation of hematopoietic differentiation, and provided insights into how distinct hematopoietic lineages can be co-regulated by pleiotropic variants acting in multipotential progenitor populations (Astle *et al*, 2016; Ulirsch *et al*, 2019).

Having reviewed the major methods used in CVAS, we now present a vignette of how these technologies have been applied to advance our understanding of a key and clinically relevant aspect of human hematopoiesis: fetal hemoglobin regulation and switching. Fetal hemoglobin (HbF) has been shown to be a key modifier of the major β-hemoglobin disorders, sickle cell disease, and β-thalassemia, where it is able to ameliorate symptoms through replacement of the mutated adult β-hemoglobin (Sankaran & Weiss, 2015). Although HbF was found to be highly heritable (Garner *et al*, 2000), little was known about its precise genetic modifiers. In late 2007 and 2008, two GWAS in non-anemic individuals identified three loci associated with variation in fetal hemoglobin levels (Menzel *et al*, 2007; Uda *et al*, 2008). These loci were also shown to be important in ameliorating the severity of symptoms in patients with sickle cell disease and β-thalassemia (Lettre *et al*, 2008; Uda *et al*, 2008). Among these was a locus on chromosome 2 within the *BCL11A* gene, which had been well studied for its role in B lymphopoiesis and neurodevelopment, yet whose role in hemoglobin switching had not been appreciated. As a result, initial functional studies revealed a key role for BCL11A in silencing of HbF (Sankaran *et al*, 2008). In addition, BCL11A was shown to be a critical regulator of fetal hemoglobin switching in humans and mice (Sankaran *et al*, 2009; Xu *et al*, 2011). Recent studies of rare individuals haploinsufficient for BCL11A have provided additional insights into its critical *in vivo* role in silencing HbF in humans (Basak *et al*, 2015; Dias *et al*, 2016). These findings have led to a considerable effort to target BCL11A to achieve HbF induction in patients with the β-hemoglobin disorders. These efforts include gene therapy-based delivery of shRNAs targeting BCL11A and efforts to target an erythroid enhancer of BCL11A using genome editing approaches (Esrick & Bauer, 2018). In addition, groups have begun to perturb
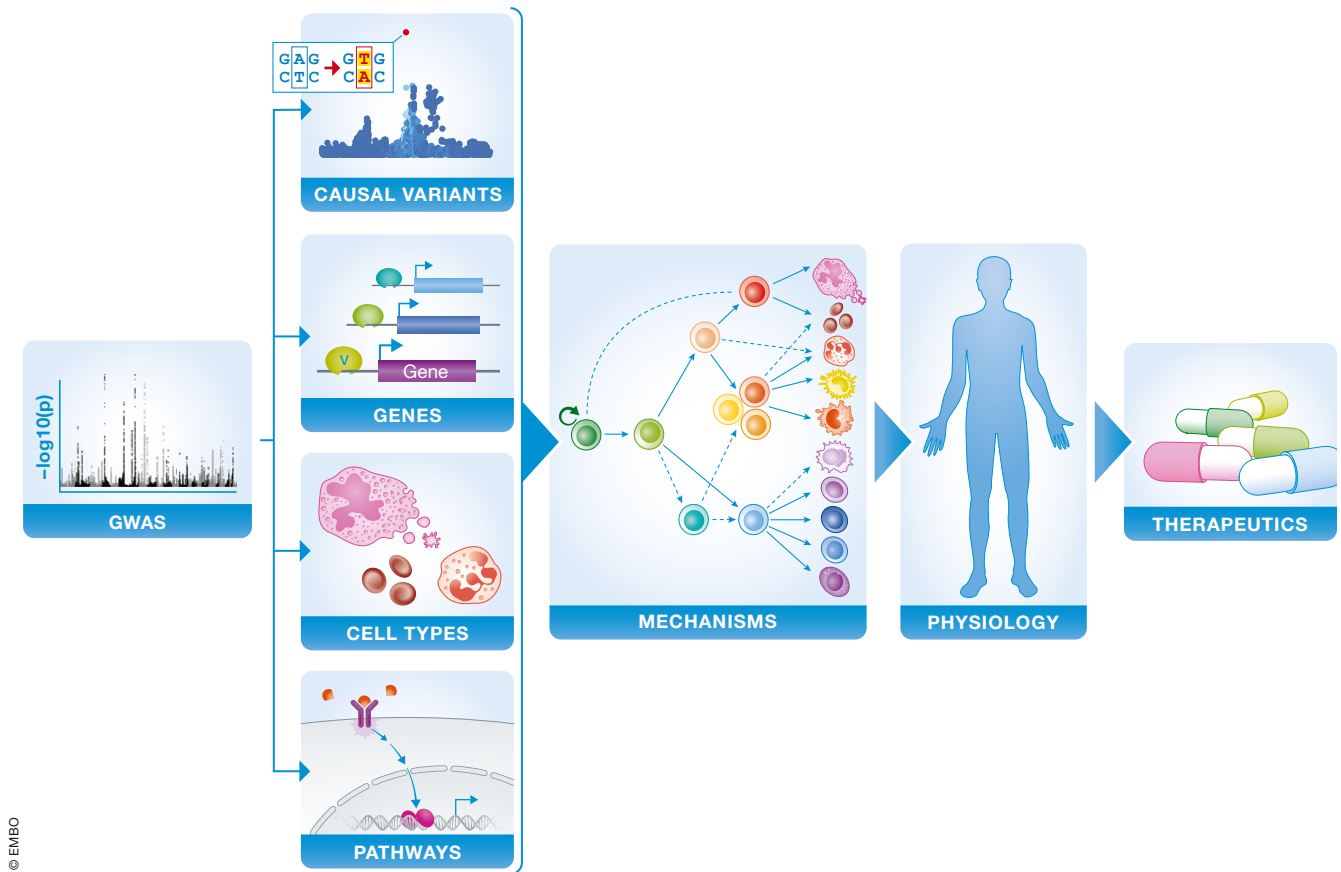
Figure 4. Schematic of moving from variant to function in human genetics research.
Blue boxes indicate key areas involved in characterizing and applying the biological mechanisms of genetic associations.

this pathway and identify more broadly effective therapeutic approaches for HbF induction in these diseases. Collectively, the study of HbF regulation has encompassed the full journey from variation to function and has now moved into the translational realm in an attempt to cure human diseases.

Bridging rare and common genetics to study hematopoiesis in health and disease

There has traditionally been a division between RVAS (and other rare variant studies) and CVAS, when in reality, there is likely a continuum of genetic effects in human traits (Katsanis, 2016). Any "monogenic" disease likely acts on top of a complex, polygenic structure of variation that confers varying individual severity and risk, and any complex trait can be altered by rare, high-effect mutations. Fortunately, on the heels of major technological and computational advances, we now arrive at an exciting crossroads in which we can begin to appreciate this full allelic spectrum underlying human variation, as the statistician Ronald A. Fisher first provided a theoretical basis for over a hundred years ago (Visscher & Goddard, 2019). In complex traits such as blood traits, schizophrenia, and autism, most of the heritability appears to be due to common alleles (Gaugler *et al*, 2014; Schizophrenia Working Group of the

Psychiatric Genomics Consortium *et al*, 2014), yet recent large-scale sequencing studies have uncovered a concomitant enrichment in rare loss-of-function variants in these and other phenotypes (Chami *et al*, 2016, preprint: Karczewski *et al*, 2019; Weiss *et al*, 2008), suggesting an interplay between rare and common variants. Moreover, in other complex traits including height and body mass index (BMI), WGS studies suggest that variants with low minor allele frequency explain a substantial portion of the total heritability that has not been able to be explained by common variants alone (preprint: Wainschtein *et al*, 2019).

The same convergence can also be observed in other areas of the hematopoietic system. For example, despite being defined as a monogenic disorder by a coding mutation in the β-globin gene, sickle cell disease (SCD) exhibits extensive clinical heterogeneity, ranging from mild phenotypes that remain undetected for decades to severe forms with multiorgan damage and early mortality (Platt *et al*, 1991). Two of the strongest modifiers of SCD severity are HbF levels and white blood cell count (Miller *et al*, 2000; Steinberg & Sebastiani, 2012; Bao *et al*, 2019). Rare and common variants are known to be independently associated with variation in these traits (Galarneau *et al*, 2010), and consequently influence SCD severity. Similarly, in β-thalassemia, a disorder defined by mutations affecting the expression of the *HBB* gene, genetic studies have found that common variants influencing HbF at least partially explain the

significant variation in disease severity (Nuinoon *et al*, 2010). Both of these cases highlight the importance of studying "monogenic" diseases in the context of common genetic variation.

Another example of this convergence can be seen in mutations altering GATA1 activity or function both globally and at specific loci. On one hand, GATA1 bound cis-regulatory elements (CREs) are frequently impacted by common genetic variation associated with hematopoietic traits, but these variants primarily tune the activity of such CREs with effect sizes below what is required to cause frank disease (Ulirsch *et al*, 2016). On the other end of the spectrum, mutations disrupting the GATA1 motif in critical CREs can cause a variety of monogenic blood disorders due to impaired gene expression (Manco *et al*, 2000; Campagna *et al*, 2014; Kaneko *et al*, 2014; Wakabayashi *et al*, 2016). Moreover, mutations in GATA1 itself can result in a range of phenotypes that include a complete absence of erythropoiesis, as is the case in DBA, or more subtle defects in red blood cell and platelet production (Crispino & Horwitz, 2017; Abdulhay *et al*, 2019). Putting these disparate findings together, one can theorize that these disease-causing *GATA1* mutations do not exert their effects in isolation, but rather act on top of a complex, polygenic structure of genetic variation that confers varying predisposition to alteration in blood cell traits (e.g., CREs at specific loci may be more or less sensitive to disruption of key master TFs, like GATA1). Only by considering this complete spectrum of allelic variation can we begin to explain all of the nuances and variation observed in blood disorders.

What mechanisms could mediate this phenomenon in which common variants influence rare disease mutations? Most genetic studies assume an additive effects model, whereby independent risk alleles contribute to a uniform, linear increase in an associated phenotype. From this perspective, individuals who harbor risk alleles across common variants tuning blood production, combined with a rare disease-causing mutation, could manifest with a particularly severe phenotype. However, the biology can also be more complicated. Epistasis refers to interactions between loci in their effect on a trait, such that the impact of a particular genotype depends on the genotype at other loci in a non-independent manner (Wei *et al*, 2014). For example, in a number of rare monogenic disorders, the deleteriousness of the disease-causing mutation can be modulated by neutral or benign alleles in the same haplotype (Jordan *et al*, 2015) or at distinct, but molecularly related, loci (Timberlake *et al*, 2016). Together, these additive and complex genetic interactions likely act in concert to tune the penetrance and expressivity of hematopoietic phenotypes and traits.

An important next step is to assess how well our expanding coverage of human genetic variation can predict complex traits in heterogeneous populations and serve as clinical biomarkers. Polygenic risk scores (PRSs) serve as a prime example of this endeavor. PRSs are a weighted sum of risk alleles carried by an individual, in which the risk alleles and their weights are defined by their effect sizes on a phenotype of interest. In many complex traits and diseases, PRS can already quantify risk more accurately than current clinical models (Sharp *et al*, 2019), at times identifying individuals with an increased risk equivalent to those with rare monogenic mutations (Khera *et al*, 2018). Other studies have found that combining common variant PRSs with other known modifiers, including lifestyle factors and rare mutations, can further improve prediction accuracy for disease risk and severity (Niemi *et al*, 2018).

To our knowledge, there has been no large-scale attempt to implement PRS on hematopoietic traits to date, but given the substantial heritability of these phenotypes (Pilia *et al*, 2006), their wide spectrum of continuous variation, and multifaceted connections to clinical parameters ranging from inflammation to hemostasis, blood traits are promising targets for PRS prediction.

Thus, perhaps we should shift our perspective on allelic variation underlying hematopoiesis from a dichotomous lens involving rare or common variation, to one involving a spectrum of variation that collectively impacts the process of hematopoiesis to varying extents to alter traits or cause disease. From a technical standpoint, we are now poised to fully bridge this gap. In the early years of genetic studies, sample sizes were limited such that for complex traits, only common variants could be detected by GWAS, whereas in presumed monogenic diseases, only highly penetrant and deleterious rare variants could be readily identified. Now, with increasingly large sample sizes, GWAS for common phenotypes can detect a greater proportion of rare variants with real effects, whereas rare disease studies are better powered to assess how common variants tune the penetrance of disease-causing rare variants. Looking ahead, the future is bright for researchers to study the impact of the entire allelic variation spectrum on hematopoiesis.

Concluding remarks and future perspectives

Modern genetic analysis has transformed our understanding of the genetic determinants of human hematopoiesis and the diseases that ensue when these processes go awry. This has enabled us to gain considerable insights beyond the valuable observations made in model organisms. Looking ahead, there is tremendous potential and excitement for the future of genetics and applications to understand human hematopoiesis. One important goal over the next 10 years is to map the full allelic spectrum of how genetic variation regulates hematopoiesis in health and disease. Furthermore, despite the enormous strides the field of genetics has already made, and a critical shortcoming has been that the majority of large studies have been confined to populations of European ancestry (Martin *et al*, 2019). Therefore, an essential part of this challenge is to expand genetic studies to populations of different ancestries. As data for underrepresented subpopulations become more available, it is likely that numerous additional population-specific loci and variants will be uncovered.

As illustrated in this review, we have already started to bear the fruits of increasingly powered rare and common genetic association studies. Merging knowledge of how rare variants contribute to disease with our growing understanding of common variation in human genetics will allow us to more fully characterize and explain the genetic architecture of blood cell production. In addition, with powerful functional tools becoming increasingly available, we can begin to glean rich insights into human hematopoiesis from functional studies that span the allelic spectrum and enable us to move from variant to function.

Expanded View for this article is available online.

Author contributions

ELB: conception of article, article draft, literature research, and preparation of figures; ANC: conception of article, article draft, literature research, and preparation of figures; VGS: conception of article, article draft, and editing.

Conflict of interest

The authors declare that they have no conflict of interest.

For more information

- https://www.ebi.ac.uk/gwas/home
- https://clinicaltrials.gov/ct2/show/NCT03282656

References

Abdulhay NJ, Fiorini C, Verboon JM, Ludwig LS, Ulirsch JC, Zieger B, Lareau CA, Mi X, Roy A, Obeng EA *et al* (2019) Impaired human hematopoiesis due to a cryptic intronic GATA1 splicing mutation. *J Exp Med* 216: 1050–1060

Albers CA, Cvejic A, Favier R, Bouwmans EE, Alessi M-C, Bertone P, Jordan G, Kettleborough RNW, Kiddle G, Kostadima M *et al* (2011) Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nat Genet* 43: 735

Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA, Jolley JD, Cvejic A, Kostadima M, Bertone P *et al* (2012) Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 44: 435

Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA *et al* (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 167: 1415–1429 e19

Bao EL, Lareau CA, Brugnara C, Fulcher IR, Barau C, Moutereau S, Habibi A, Badaoui B, Berkenou J, Bartolucci P *et al* (2019) Heritability of fetal hemoglobin, white cell count, and other clinical traits from a sickle cell disease family cohort. *Am J Hematol* 94: 522–527

Basak A, Hancarova M, Ulirsch JC, Balci TB, Trkova M, Pelisek M, Vlckova M, Muzikova J, Cermak J, Trka J *et al* (2015) BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. *J Clin Invest* 125: 2363–2368

Boyer SW, Rajendiran S, Beaudin AE, Smith-Berdan S, Muthuswamy PK, Perez-Cunningham J, Martin EW, Cheung C, Tsang H, Landon M *et al* (2019) Clonal and quantitative *in vivo* assessment of hematopoietic stem

cell differentiation reveals strong erythroid potential of multipotent cells. *Stem Cell Rep* 12: 801–815

Campagna DR, de Bie CI, Schmitz-Abe K, Sweeney M, Sendamarai AK, Schmidt PJ, Heeney MM, Yntema HG, Kannengiesser C, Grandchamp B *et al* (2014) X-linked sideroblastic anemia due to ALAS2 intron 1 enhancer element GATA-binding site mutations. *Am J Hematol* 89: 315–319

Canver MC, Bauer DE, Orkin SH (2017) Functional interrogation of non-coding DNA through CRISPR genome editing. *Methods* 121–122: 118–129

Ceccaldi R, Sarangi P, D'Andrea AD (2016) The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol* 17: 337

Chami N, Chen M-H, Slater Andrew J, Eicher John D, Evangelou E, Tajuddin Salman M, Love-Gregory L, Kacprowski T, Schick Ursula M, Nomura A *et al* (2016) Exome genotyping identifies pleiotropic variants associated with red blood cell traits. *Am J Hum Genet* 99: 8–21

Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yan Y, Kundu K, Ecker S *et al* (2016) Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 167: 1398–1414.e24

Collin M, Dickinson R, Bigley V (2015) Haematopoietic and immune defects associated with GATA2 mutation. *Br J Haematol* 169: 173–187

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ *et al* (2016) Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 48: 1193–1203

Crispino JD, Horwitz MS (2017) GATA factor mutations in hematologic disease. *Blood* 129: 2103

Dias C, Estruch Sara B, Graham Sarah A, McRae J, Sawiak Stephen J, Hurst Jane A, Joss Shelagh K, Holder Susan E, Morton Jenny EV, Turner C *et al* (2016) BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. *Am J Hum Genet* 99: 253–274

Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, Lakey JH, Rahman T, Wang X-N, McGovern N *et al* (2011) Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* 118: 2656

Downes K, Megy K, Duarte D, Vries M, Gebhart J, Hofer S, Shamardina O, Deevil SVV, Stephens J, Mapeta R *et al* (2018) Diagnostic high-throughput sequencing of 2,390 patients with bleeding, thrombotic and platelet disorders. *bioRxiv* https://doi.org/10.1101/504142 [PREPRINT]

Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Willig T-N, Dianzani I, Ball S, Tchernia G, Klar J, Matsson H *et al* (1999) The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet* 21: 169

Ensrud K, Grimm RH (1992) The white blood cell count and risk for coronary heart disease. *Am Heart J* 124: 207–213

Esrick EB, Bauer DE (2018) Genetic therapies for sickle cell disease. *Semin Hematol* 55: 76–86

Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci USA* 93: 12355–12358

Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G (2010) Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat Genet* 42: 1049

Gallagher MD, Chen-Plotkin AS (2018) The post-GWAS era: from association to function. *Am J Hum Genet* 102: 717–730

Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL (2000) Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood* 95: 342

Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Mahajan M, Manaa D, Pawitan Y, Reichert J *et al* (2014) Most genetic risk for autism resides with common variation. *Nat Genet* 46: 881

Grootens J, Ungerstedt JS, Ekoff M, Rönnberg E, Klimkowska M, Amini R-M, Arock M, Söderlund S, Mattsson M, Nilsson G *et al* (2019) Single-cell analysis reveals the KIT D816V mutation in haematopoietic stem and progenitor cells in systemic mastocytosis. *EBioMedicine* 43: 150–158

GTEx Consortium, Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park Y, Parsana P *et al* (2017) Genetic effects on gene expression across human tissues. *Nature* 550: 204

Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson Bjarni J, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E *et al* (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* 95: 535–552

Gustavsson P, Willig T-N, Haeringen AV, Tchernia G, Dianzani I, Donnér M, Elinder G, Renter J-I, Nilsson P-G, Gordon L *et al* (1997) Diamond-Blackfan anaemia: genetic homogeneity for a gene on chromosome 19q13 restricted to 1.8 Mb. *Nat Genet* 16: 368–371

Hoffman M, Blum A, Baruch R, Kaplan E, Benjamin M (2004) Leukocytes and coronary heart disease. *Atherosclerosis* 172: 1–6

Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, Frucht DM, Vinh DC, Auth RD, Freeman AF *et al* (2011) Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* 118: 2653

Ichikawa M, Asai T, Saito T, Yamamoto G, Seo S, Yamazaki I, Yamagata T, Mitani K, Chiba S, Hirai H *et al* (2004) AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med* 10: 299–304

Ishigaki Y, Nakamura Y, Tatsuno T, Hashimoto M, Shimasaki T, Iwabuchi K, Tomosugi N (2013) Depletion of RNA-binding protein RBM8A (Y14) causes cell cycle deficiency and apoptosis in human cells. *Exp Biol Med* 238: 889–897

Jacobsen SEW, Nerlov C (2019) Haematopoiesis in the era of advanced single-cell technologies. *Nat Cell Biol* 21: 2–8

Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ *et al* (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 167: 1369–1384 e19

Jordan DM, Frangakis SG, Golzio C, Cassa CA, Kurtzberg J, Task Force for Neonatal G, Davis EE, Sunyaev SR, Katsanis N (2015) Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* 524: 225

Kaneko K, Furuyama K, Fujiwara T, Kobayashi R, Ishida H, Harigae H, Shibahara S (2014) Identification of a novel erythroid-specific enhancer for the ALAS2 gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia. *Haematologica* 99: 252

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP *et al* (2019) Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* https://doi.org/10.1101/531210 [PREPRINT]

Katsanis N (2016) The continuum of causality in human genetic disorders. *Genome Biol* 17: 233

Khajuria RK, Munschauer M, Ulirsch JC, Fiorini C, Ludwig LS, McFarland SK, Abdulhay NJ, Specht H, Keshishian H, Mani DR *et al* (2018) Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. *Cell* 173: 90–103.e19

Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT *et al* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 50: 1219–1224

Khoriaty R, Hesketh GG, Bernard A, Weyand AC, Mellacheruvu D, Zhu G, Hoenerhoff MJ, McGee B, Everett L, Adams EJ *et al* (2018) Functions of the COPII gene paralogs SEC23A and SEC23B are interchangeable *in vivo*. *Proc Natl Acad Sci USA* 115: E7748

Landowski M, O'Donohue M-F, Buros C, Ghazvinian R, Montel-Lehry N, Vlachos A, Sieff CA, Newburger PE, Niewiadomska E, Matysiak M *et al* (2013) Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond-Blackfan anemia. *Hum Genet* 132: 1265–1274

Lee S, Wu MC, Lin X (2012) Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13: 762–775

Lee S, Abecasis Gonçalo R, Boehnke M, Lin X (2014) Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 95: 5–23

Lettre G, Sankaran VG, Bezerra MAC, Araújo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN *et al* (2008) 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* 105: 11869–11874

Liu F, Wu HY, Wesselschmidt R, Kornaga T, Link DC (1996) Impaired production and increased apoptosis of neutrophils in granulocyte colony-stimulating factor receptor-deficient mice. *Immunity* 5: 491–501

Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA *et al* (2014) Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat Med* 20: 748

Manco L, Ribeiro ML, Máximo V, Almeida H, Costa A, Freitas O, Barbot J, Abade A, Tamagnini G (2000) A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency. *Br J Haematol* 110: 993–997

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al* (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747

Mao H, McMahon JJ, Tsai Y-H, Wang Z, Silver DL (2016) Haploinsufficiency for core exon junction complex components disrupts embryonic neurogenesis and causes p53-mediated microcephaly. *PLoS Genet* 12: e1006282

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 51: 584–591

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

Michelle L, Cloutier A, Toutant J, Shkreta L, Thibault P, Durand M, Garneau D, Gendron D, Lapointe E, Couture S *et al* (2012) Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol Cell Biol* 32: 954

Miller ST, Sleeper LA, Pegelow CH, Enos LE, Wang WC, Weiner SJ, Wethers DL, Smith J, Kinney TR (2000) Prediction of adverse outcomes in children with sickle cell disease. *N Engl J Med* 342: 83–89

Minelli A, Maserati E, Rossi G, Bernardo ME, De Stefano P, Cecchini MP, Valli R, Albano V, Pierani P, Leszl A *et al* (2004) Familial platelet disorder with propensity to acute myelogenous leukemia: genetic heterogeneity and progression to leukemia via acquisition of clonal chromosome anomalies. *Genes Chromosom Cancer* 40: 165–171

Nandakumar SK, McFarland SK, Mateyka LM, Lareau CA, Ulirsch JC, Ludwig LS, Agarwal G, Engreitz JM, Przychodzen B, McConkey M *et al* (2019) Gene-centric functional dissection of human genetic variation uncovers regulators of hematopoiesis. *Elife* 8: e44080

Nathan DG, Clarke BJ, Hillman DG, Alter BP, Housman DE (1978) Erythroid precursors in congenital hypoplastic (Diamond-Blackfan) anemia. *J Clin Invest* 61: 489 − 498

Niemi MEK, Martin HC, Rice DL, Gallone G, Gordon S, Kelemen M, McAloney K, McRae J, Radford EJ, Yu S *et al* (2018) Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* 562: 268 − 271

Niraj J, Färkkilä A, D'Andrea AD (2019) The Fanconi anemia pathway in cancer. *Annu Rev Cancer Biol* 3: 457 − 478

Nuinoon M, Makarasara W, Mushiroda T, Setianingsih I, Wahidiyat PA, Sripichai O, Kumasaka N, Takahashi A, Svasti S, Munkongdee T *et al* (2010) A genome-wide association identified the common genetic variants influence disease severity in β0-thalassemia/hemoglobin E. *Hum Genet* 127: 303 − 314

Orkin SH, Zon LI (2008) Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132: 631 − 644

Ostergaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, Dafou D, Kilo T, Smithson S, Lunt P *et al* (2011) Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* 43: 929

Owen CJ, Toze CL, Koochin A, Forrest DL, Smith CA, Stevens JM, Jackson SC, Poon M-C, Sinclair GD, Leber B *et al* (2008) Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy. *Blood* 112: 4639

Pevny L, Simon MC, Robertson E, Klein WH, Tsai S-F, D'Agati V, Orkin SH, Costantini F (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* 349: 257 − 260

Pilia G, Chen W-M, Scuteri A, Orrú M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P *et al* (2006) Heritability of cardiovascular and personality traits in 6,148 sardinians. *PLoS Genet* 2: e132

Platt OS, Thorington BD, Brambilla DJ, Milner PF, Rosse WF, Vichinsky E, Kinney TR (1991) Pain in sickle cell disease. *N Engl J Med* 325: 11 − 16

Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HKA, Hirschhorn JN, Cantor AB, Orkin SH (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* 322: 1839

Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, Fujiwara Y, Ito M, Groudine M, Bender MA *et al* (2009) Developmental and species-divergent globin switching are driven by BCL11A. *Nature* 460: 1093 − 1097

Sankaran VG, Ghazvinian R, Do R, Thiru P, Vergilio J-A, Beggs AH, Sieff CA, Orkin SH, Nathan DG, Lander ES *et al* (2012) Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest* 122: 2439 − 2443

Sankaran VG, Weiss MJ (2015) Anemia: progress in molecular mechanisms and therapies. *Nat Med* 21: 221

Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA *et al* (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421

Schwarz K, Iolascon A, Verissimo F, Trede NS, Horsley W, Chen W, Paw BH, Hopfner K-P, Holzmann K, Russo R *et al* (2009) Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat Genet* 41: 936

Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, Schneider DA, Locke JM, Tyrrell J, Weedon MN *et al* (2019) Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes Care* 42: 200

Shiohara M, Shigemura T, Saito S, Tanaka M, Yanagisawa R, Sakashita K, Asada H, Ishii E, Koike K, Chin M *et al* (2009) Ela2 mutations and clinical manifestations in familial congenital neutropenia. *J Pediatr Hematol Oncol* 31: 319 − 324

Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38: 209

Song W-J, Sullivan MG, Legare RD, Hutchings S, Tan X, Kufrin D, Ratajczak J, Resende IC, Haworth C, Hock R *et al* (1999) Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* 23: 166 − 175

Soranzo N, Rendon A, Gieger C, Jones CI, Watkins NA, Menzel S, Döring A, Stephens J, Prokisch H, Erber W *et al* (2009) A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. *Blood* 113: 3831

Steinberg MH, Sebastiani P (2012) Genetic modifiers of sickle cell disease. *Am J Hematol* 87: 795 − 803

Sumpter R, Levine B (2017) Emerging functions of the Fanconi anemia pathway at a glance. *J Cell Sci* 130: 2657

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D (2019) Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 20: 467 − 484

Thaventhiran JED, Lango Allen H, Burren OS, Farmery JHR, Staples E, Zhang Z, Rae W, Greene D, Simeoni I, Maimaris J *et al* (2018) Whole Genome Sequencing of Primary Immunodeficiency reveals a role for common and rare variants in coding and non-coding sequences. *bioRxiv* https://doi.org/10.1101/499988 [PREPRINT]

Timberlake AT, Choi J, Zaidi S, Lu Q, Nelson-Williams C, Brooks ED, Bilguvar K, Tikhonova I, Mane S, Yang JF *et al* (2016) Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *Elife* 5: e20125

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

Ulirsch JC, Nandakumar Satish K, Wang L, Giani Felix C, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen Tarjei S *et al* (2016) Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* 165: 1530 − 1545

Ulirsch JC, Verboon JM, Kazerounian S, Guo MH, Yuan D, Ludwig LS, Handsaker RE, Abdulhay NJ, Fiorini C, Genovese G *et al* (2018) The genetic landscape of Diamond-Blackfan anemia. *Am J Hum Genet* 103: 930 − 947

Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, Satpathy AT, Kartha VK, Salem RM, Hirschhorn JN *et al* (2019) Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet* 51: 683 − 693

Van Hout CV, Tachmazidou I, Backman JD, Hoffman JX, Ye B, Pandey AK, Gonzaga-Jauregui C, Khalid S, Liu D, Banerjee N *et al* (2019) Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv* https://doi.org/10.1101/572347 [PREPRINT]

Visscher PM, Goddard ME (2019) From R.A. Fisher's 1918 paper to GWAS a century later. *Genetics* 211: 1125

Wainschtein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, McKnight B, Shoemaker BM, Mitchell BD, Psaty BM *et al* (2019) Recovery of trait heritability from whole genome sequence data. *bioRxiv* https://doi.org/10.1101/588020 [PREPRINT]

Wakabayashi A, Ulirsch JC, Ludwig LS, Fiorini C, Yasuda M, Choudhuri A, McDonel P, Zon LI, Sankaran VG (2016) Insight into GATA1 transcriptional activity through interrogation of cis elements disrupted in human erythroid disorders. *Proc Natl Acad Sci USA* 113: 4434 – 4439

Wei W-H, Hemani G, Haley CS (2014) Detecting epistasis in human complex traits. *Nat Rev Genet* 15: 722

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MAR, Green T *et al* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358: 667 – 675

Wu H, Liu X, Jaenisch R, Lodish HF (1995) Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* 83: 59 – 67

Xu J, Peng C, Sankaran VG, Shao Z, Esrick EB, Chong BG, Ippolito GC, Fujiwara Y, Ebert BL, Tucker PW *et al* (2011) Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science* 334: 993 – 996

Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES (2014) Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci USA* 111: E455