Whole-Exome Sequencing Identifies Loci Associated with Blood Cell Traits and Reveals a Role for Alternative *GFI1B* Splice Variants in Human Hematopoiesis

Linda M. Polfus,[1,38] Rajiv K. Khajuria,[2,3,4,38] Ursula M. Schick,[5,38] Nathan Pankratz,[6] Raha Pazoki,[7] Jennifer A. Brody,[8] Ming-Huei Chen,[9] Paul L. Auer,[10] James S. Floyd,[8] Jie Huang,[11] Leslie Lange,[12] Frank J.A. van Rooij,[7] Richard A. Gibbs,[13] Ginger Metcalf,[13] Donna Muzny,[13] Narayanan Veeraraghavan,[13] Klaudia Walter,[11] Lu Chen,[11,14] Lisa Yanek,[15]

(Author list continued on next page)

Circulating blood cell counts and indices are important indicators of hematopoietic function and a number of clinical parameters, such as blood oxygen-carrying capacity, inflammation, and hemostasis. By performing whole-exome sequence association analyses of hematologic quantitative traits in 15,459 community-dwelling individuals, followed by in silico replication in up to 52,024 independent samples, we identified two previously undescribed coding variants associated with lower platelet count: a common missense variant in *CPS1* (rs1047891, MAF = 0.33, discovery + replication $p = 6.38 \times 10^{-10}$) and a rare synonymous variant in *GFI1B* (rs150813342, MAF = 0.009, discovery + replication $p = 1.79 \times 10^{-27}$). By performing CRISPR/Cas9 genome editing in hematopoietic cell lines and follow-up targeted knockdown experiments in primary human hematopoietic stem and progenitor cells, we demonstrate an alternative splicing mechanism by which the *GFI1B* rs150813342 variant suppresses formation of a GFI1B isoform that preferentially promotes megakaryocyte differentiation and platelet production. These results demonstrate how unbiased studies of natural variation in blood cell traits can provide insight into the regulation of human hematopoiesis.

Human genetic studies have provided important insights into hematopoiesis. Genome-wide association studies (GWASs) performed in large, population-based samples have identified associations of genomic regions and common genetic (usually non-coding) variants with inter-individual differences in blood cell traits[1–5], though the causal DNA variants and their functional mechanisms often remain elusive. Whole-exome and targeted sequencing approaches have been used to identify rare, sometimes private, loss (or gain)-of-function coding variants segregating within families with hematologic traits at the extremes of the phenotypic distribution[6–12]. As of yet, whole-exome sequencing has not been applied to large population-based cohorts well-phenotyped for hematologic traits to identify rare, functional variation with moderate-to-large phenotypic effects and to provide new biologic insight.

To this end, we performed exome sequencing in 15,459 unrelated European ancestry (EU) and African American (AA) individuals enrolled in six population-based cohort studies (see Supplemental Note). Replication of significant findings was performed in up to 52,024 additional samples via a combination of whole-exome-based or genome-based sequencing, genotyping, and imputation (Supplemental Note). Our a priori hypothesis was that systematic evaluation of coding variation detected by exome sequence analysis in samples unselected for blood cell traits would identify low-frequency variants influencing hematologic traits and could provide functional insights into hematopoiesis. We analyzed platelet count and 12 other blood cell traits (Table S1). The means of the traits were as expected in a sample of unselected healthy individuals from the population (Table S1). Association results from single-variant and from gene-based burden and sequence kernel association tests (SKATs) meeting our a priori significance thresholds in either EU, AA, or trans-ethnic discovery meta-analyses are summarized for both previously known and novel (which we define as those not reported in the available literature) loci in Table 1 and Tables S2–S5 and described further in the Supplemental Note. Lambda values showed no significant inflation (Table S6).

[1]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [2]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA,; [3]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [4]Berlin-Brandenburg School for Regenerative Therapies, Charité Universitätsmedizin Berlin, Berlin 13353, Germany; [5]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [6]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55454, USA; [7]Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000, the Netherlands; [8]Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, WA 98195, USA; [9]Department of Neurology, School of Medicine, Boston University, Boston, MA 02118, USA; [10]School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53205, USA; [11]Human Genetics, Wellcome Trust Sanger Institute, Hinxton CB10 1HH, UK; [12]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; [13]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [14]Department of Haematology, University of Cambridge, Cambridge CB2 0AH, UK; [15]GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA; [16]Center for Human Genetic Research,

(Affiliations continued on next page)

Lewis C. Becker,[15] Gina M. Peloso,[16] Aoi Wakabayashi,[2,3] Mart Kals,[17] Andres Metspalu,[17] Tõnu Esko,[17] Keolu Fox,[18] Robert Wallace,[19] Nora Franceshini,[20] Nena Matijevic,[21] Kenneth M. Rice,[8] Traci M. Bartz,[8] Leo-Pekka Lyytikäinen,[22] Mika Kähönen,[23] Terho Lehtimäki,[22] Olli T. Raitakari,[24] Ruifang Li-Gao,[25] Dennis O. Mook-Kanamori,[25,26] Guillaume Lettre,[27] Cornelia M. van Duijn,[28] Oscar H. Franco,[7] Stephen S. Rich,[29] Fernando Rivadeneira,[28] Albert Hofman,[28] André G. Uitterlinden,[28] James G. Wilson,[30] Bruce M. Psaty,[8,31] Nicole Soranzo,[11,14] Abbas Dehghan,[7] Eric Boerwinkle,[1] Xiaoling Zhang,[32] Andrew D. Johnson,[33] Christopher J. O'Donnell,[34] Jill M. Johnsen,[35] Alexander P. Reiner,[36,39] Santhi K. Ganesh,[37,39] and Vijay G. Sankaran[2,3,39,*]

Four gene-based associations were discovered for red blood cell (RBC) traits (*ACTN4*, *MMACHC*, *MYOM2*, and *MRPL43*). Trans-ethnic discovery meta-analyses are summarized for both previously identified loci, which we verify in this study, and previously unreported loci. A summary of these findings, and driving variants, are provided in the Supplemental Note and Table S3. None of these gene-based SKAT or burden findings could be replicated in independent samples. Nonetheless, a few of the individual rare variants driving the gene-based associations in the discovery sample showed suggestive evidence of association in the replication sample (Supplemental Note and Table S3**)**.

Among the three single-variant associations we identified (Table 1), two coding variants were associated with lower platelet count in our discovery sample: *CPS1* rs1047891, a common missense variant encoding p.Thr1412Asn (EU + AA minor-allele frequency [MAF] = 0.33, EU + AA $p = 5.7 \times 10^{-8}$) and *GFI1B* rs150813342, a rare synonymous variant encoding p.Phe192 and located in alternatively spliced exon 5 (EU MAF = 0.009, EU $p = 4.7 \times 10^{-8}$; EU + AA MAF = 0.008, EU + AA $p = 2.64 \times 10^{-8}$). One single-nucleotide variant (SNV) result (rs9656446; EU + AA MAF = 0.03, EU + AA $p = 1.48 \times 10^{-7}$) associated with basophils in trans-ethnic analyses was in the ATP/GTP binding protein-like 3 (*AGBL3*) gene. However, the allele frequencies in the discovery sample differed by ethnicity (EU MAF = 0.001 and AA MAF = 0.08), and replication in samples of EU ethnicity from the UK10K project was not significant (EU p = 0.71). In our combined replication sample, we replicated the associations of *CPS1* rs1047891 (EU + AA MAF = 0.328, EU + AA $p = 1.02 \times 10^{-4}$) and *GFI1B* rs150813342 (EU + AA $p = 5.71 \times 10^{-21}$) with lower platelet counts. In the combined discovery and replication samples, the p values for *CPS1* rs1047891 and *GFI1B* rs150813342 were 6.38×10^{-10} and 1.79×10^{-27}, respectively. A Manhattan plot for single-variant associations with platelet count and quantile-quantile (Q-Q) plots are shown in Figures S1–S3. Forest plots of the discovery cohorts for the two replicated findings (*GFI1B* rs150813342 and *CPS1* rs1047891) are provided in Figures S4 and S5, as well as regional plots calculating linkage disequilibrium of SNVs in the gene with respect to index SNVs (Figures S6 and S7).

AGBL3 is a metallocarboxypeptidase involved in processing tubulins of the blood cell cytoskeleton. *CPS1* encodes carbamoyl-phosphate synthase 1, a mitochondrial enzyme involved in the urea cycle. The *CPS1* variant (or its LD proxies) has been associated with various cardiometabolic traits, including high-density lipoprotein (HDL) cholesterol, homocysteine, fibrinogen, serum metabolite levels, and kidney function.[13–17] *GFI1B* is a known transcriptional repressor and a key regulator of platelet and red blood cell development. There was no evidence that either *CPS1* rs1047891 or *GFI1B* rs150813342 were significantly associated with other hematologic traits assessed in the discovery sample (Tables S7A and S7B). Moreover, neither *GFI1B* rs150813342 nor *CPS1* rs1047891 was associated with mean platelet volume, platelet aggregation, or expression of platelet surface markers, though these analyses were limited to much smaller numbers of individuals (Supplemental Note, Tables S8 and S10). However, a decrease in

Massachusetts General Hospital, Boston, MA 02114, USA; [17]Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia; [18]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [19]College of Public Health, the University of Iowa, Iowa City, IA 52242, USA; [20]Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA; [21]Department of Surgery, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [22]Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere 33520, Finland; [23]Department of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere 33521, Finland; [24]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital and Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland; [25]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden RC 2300, the Netherlands; [26]Epidemiology Section, Department of Biostatistics, Epidemiology, and Scientific Computing Department, King Faisal Specialist Hospital and Research Centre, Riyadh 11211 Saudi Arabia; [27]Montreal Heart Institute and Université de Montréal, Montreal, QC H1T 1C8, Canada; [28]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3000, the Netherlands; [29]Center for Public Health Genomics, University of Virginia, Charlottesville VA 22908, USA; [30]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson MS 39216, USA; [31]Group Health Research Institute, Group Health Cooperative, Seattle, WA 98101, USA; [32]Departments of Medicine and Biostatistics, Schools of Medicine and Public Health, Boston University, Boston, MA 02118, USA; [33]Cardiovascular Epidemiology and Human Genomics Branch, Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA 01702, USA; [34]Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA 01702, USA; [35]Bloodworks Northwest, Seattle, WA 98102, USA; [36]Women's Health Initiative Clinical Coordinating Center, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [37]Division of Cardiovascular Medicine, Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA
[38]These authors contributed equally to this work
[39]These authors contributed equally to this work
*Correspondence: sankaran@broadinstitute.org (V.G.S.)
http://dx.doi.org/10.1016/j.ajhg.2016.06.016.

Table 1. Single-Variant Association Findings

Trait	Discovery Ethnicity	Gene	SNP Chromosome Position, rs Number, and Function	Discovery p Value	Replication p Value	Discovery MAF	Replication MAF	Discovery Beta Coefficient (SE)	Replication Z Score[a]	Discovery N	Replication N
PLT	EU + AA	*GFI1B*	chr9: 135864513, rs150813342, synonymous	2.64×10^{-8}	5.71×10^{-21}	0.008	0.007	−0.402 (0.07)	−9.40	13,744	48,099[b]
PLT	EU + AA	*CPS1*	chr2: 211540507, rs1047891, missense	5.73×10^{-8}	1.02×10^{-4}	0.328	0.313	−0.07 (0.013)	−3.89	13,744	48,394[b]
BASO	EU + AA	*AGBL3*	chr7: 134717656, rs9656446, synonymous	1.48×10^{-7}	0.71	0.031[c]	0.002	0.271 (0.051)	−0.05 (0.13)	6,877	6,699[d]

AA, African American individuals; BASO, basophil count; EU, European ancestry individuals; MAF, minor-allele frequency; PLT, platelet count.
[a]Z score is reported from N-weighted replication meta-analyses, where more than one replication cohort was available; otherwise, beta coefficient and SE are reported.
[b]UK10K project samples and imputed EU, Cardiovascular Health Study (CHS), and Atherosclerosis Risk in Communities (ARIC) study samples.
[c]EU MAF = 0.001; AA MAF = 0.078; EU + AA MAF = 0.031.
[d]UK10K project samples and imputed EU samples.

the median fluorescence intensity of large, platelet-marker positive ($CD41^+CD61^+$) events[18] was detected by flow cytometry in *GFI1B* variant carriers even after adjustment for circulating platelet count ($p < 0.0001$), which could reflect a decrease in circulating platelet aggregates or a skewing of a platelet subpopulation with regards to platelet-surface-marker expression or size (see Supplemental Note).

We conducted bioinformatic and functional analyses to understand the impact of the *GFI1B* exon 5 synonymous variant and the *CPS1* rs1047891 variant (p.Thr1412Asn) on gene and protein function. The *CPS1* p.Thr1412Asn amino acid substitution is predicted to be benign and tolerated by SIFT and PolyPhen. Moreover, according to the GTEx Portal database, there is no evidence of an expression quantitative trait loci (eQTL) effect for rs1047891. Nonetheless, the CPS1 p.Thr1412Asn missense substitution is located within a region critical for N-acetyl-glutamate binding and has been reported to result in 20%–30% higher enzymatic activity[19] and to influence vascular function.[15]

We initially assessed the association of rs150813342 with *GFI1B* expression by using Affymetrix GeneChip Human Exon 1.0 ST Array data on whole-blood RNA available from 881 Framingham Heart Study participants.[20] There was no evidence for association of the rs150813342 genotype with expression of any *GFI1B* exon, though statistical power is likely limited by the low frequency of the rs150813342 variant allele, which was present in only 7 of the 881 individuals. According to SPANR,[21] rs150813342 had a predicted effect on splicing (difference in the percentage of transcripts with the exon spliced in [dPSI] score of −4.6). rs150813342 was predicted to disrupt a putative exon splicing enhancer (ESE) in exon 5 that contains a consensus SRSF1 binding motif.[22] To functionally evaluate the impact of this variant on GFI1B transcript splicing in a relevant cell type, we used CRISPR/Cas9 genome editing to create multiple independent isogenic K562 hematopoietic cell lines harboring the *GFI1B* synonymous single-nucleotide change (Figure 1A). These cell lines were homozygous for the variant and exhibited inclusion of less than 30% of exon 5 relative to other surrounding exons in the *GFI1B* mRNA (Figure 1B). Semi-quantitative RT-PCR showed that the presence of the synonymous variant resulted in reduced formation of the *GFI1B* isoform containing exon 5 (herein referred to as the long isoform), as well as preferential formation of the isoform lacking exon 5 (herein referred to as the short isoform) (Figures 1C and 1D). No other isoforms or intron inclusion events were detected (Figure 1C, Figure S8).

Although *GFI1B* has been implicated in both RBC and platelet production (erythropoiesis and megakaryopoiesis, respectively),[23–25] only a role for the short isoform in erythroid cells has been suggested previously.[26] We next assessed the effect of the altered splicing of *GFI1B* on lineage-specific hematopoietic differentiation. We chemically induced differentiation of the isogenic K562 cell lines with either hemin (to promote erythroid differentiation) or phorbol 12-myristate 13-acetate (PMA, to promote megakaryocytic differentiation) (Figure 2A). Although erythroid

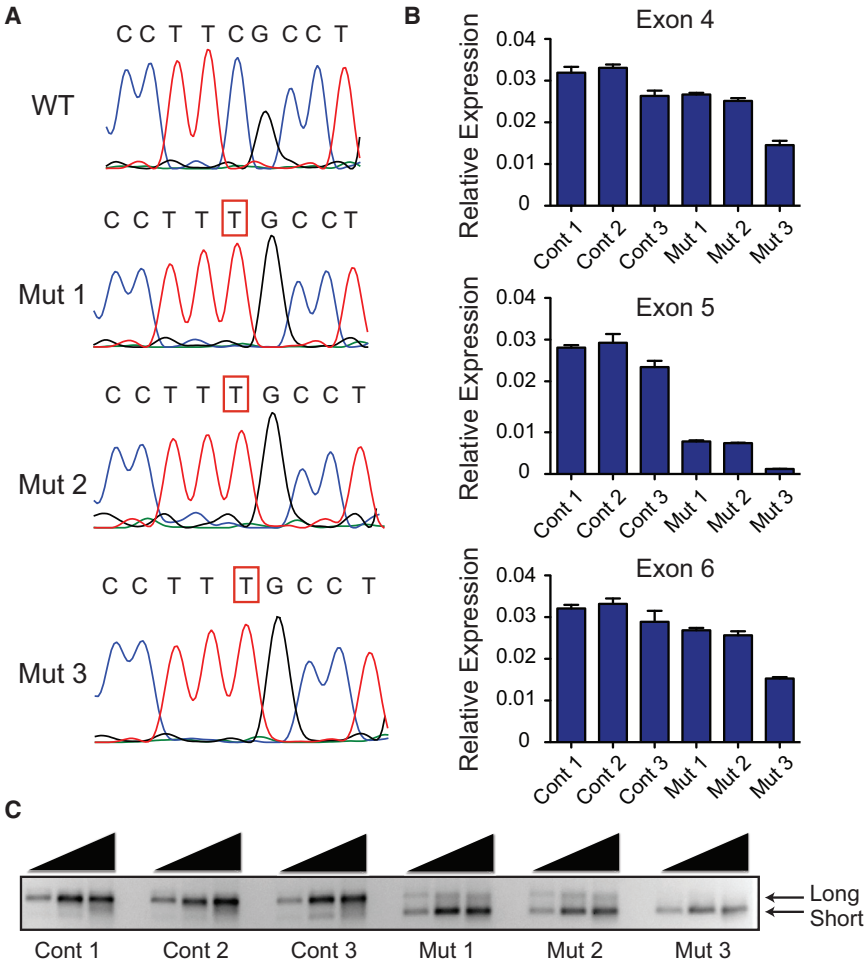

Figure 1. The Variant rs150813342 Results in Reduced Formation of the Long *GFI1B* Isoform and Preferential Formation of the Short Isoform

(A) Chromatograms of the sequence surrounding the altered nucleotide in *GFI1B* exon 5 showing the wild-type (WT) sequence and sequences of isogenic hematopoietic K562 cell mutant clones (Mut 1, Mut 2, and Mut 3) harboring the C>T single-nucleotide variant (SNV) generated via CRISPR/Cas9 mediated homologous repair.

(B) qRT-PCR of *GFI1B* exons 4, 5, and 6 measured from isogenic control (Cont) and mutant K562 cell clones showing inclusion of less than 30% of *GFI1B* exon 5 relative to the surrounding exons in *GFI1B* mRNA from mutant clones (n = 3 per group). Error bars show SD.

(C) Semi-quantitative RT-PCR with *GFI1B* exon 4 forward and exon 6 reverse primers with progressively increasing cycle numbers (26, 28, and 30 cycles) demonstrates reduced formation of the long *GFI1B* isoform and preferential formation of the short isoform, as well as no other intermediate isoforms in the clones harboring the SNV.

(D) rs150813342 is predicted to disrupt a putative exon splicing enhancer (ESE) in exon 5 that contains a consensus SRSF1 binding motif. Disruption of this binding motif results in reduced inclusion of exon 5 and preferential formation of the short isoform. The promotion of alternative splicing by SRSF1 through the spliceosome complex is indicated by an arrow to a light blue circle. Forward (F) and reverse (R) PCR primers of the respective exon are indicated.

differentiation appeared to proceed normally, as assessed morphologically (Figure 2B), and with expression of the surface marker GYPA (CD235a) (Figure 2C) and terminal erythroid marker genes (Figure 2D), megakaryocyte differentiation appeared severely impaired; the cells retained an immature blast-like morphology and failed to upregulate the surface marker of megakaryocyte differentiation, CD41a (encoded by *ITGA2B*), and mRNAs whose expression is characteristic of terminal megakaryopoiesis (Figures 2B–2D, Figure S9). The megakaryocyte genes *PPBP*, *SELP*, and *PF4* were downregulated by an average of 8.6-, 6.7-, and 41.1-fold, respectively, in the isogenic clones ($p = 0.0001$, 0.0013, and 0.0459, respectively) versus in the controls (Figure 2D). These results suggest that the long isoform of GFI1B is necessary for normal megakaryocyte differentiation.

To confirm a preferential role for this long GFI1B isoform in megakaryocyte differentiation, we identified two independent short hairpin RNAs (shRNAs) that specifically targeted *GFI1B* exon 5, which would thereby selectively downregulate the long but not the short isoform. We utilized

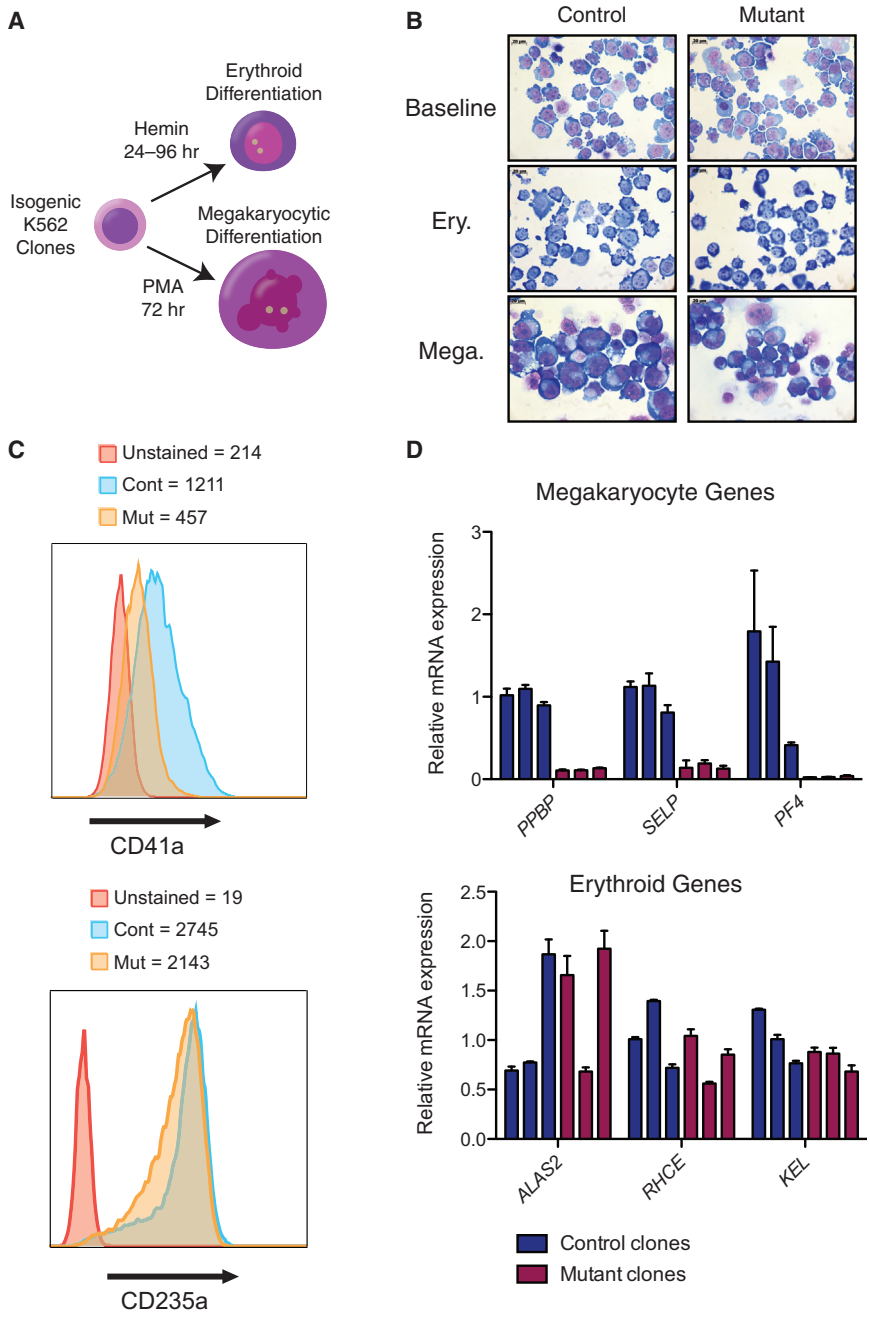

Figure 2. Impaired Megakaryopoiesis and Retained Erythropoiesis in K562 Cells Harboring the rs150813342 SNV in *GFI1B* Exon 5

(A) Scheme of phorbol 12-myristate 13-acetate (PMA)-induced megakaryocytic differentiation and hemin-induced erythroid differentiation of the hematopoietic K562 cell models.

(B) Representative May-Grünwald-Giemsa-stained cytospin images of 72 hr PMA-induced and 96 hr hemin-induced isogenic control and mutant clones showing megakaryocytic differentiation that appears severely impaired, with the cells retaining an immature blast-like morphology in the mutant clones, whereas the erythroid differentiation appears unaffected.

(C) Representative flow cytometry analysis of the megakaryocyte marker CD41a and the erythroid marker CD235a further confirmed the impaired megakaryopoiesis and the retained erythropoiesis as shown by the histogram plots with the mean fluorescence intensity (MFI) for each marker in unstained cells, control, and mutant clones, respectively.

(D) Gene expression analysis by qRT-PCR of the megakaryocyte markers *PPBP*, *SELP*, and *PF4* after 72 hr of PMA-induced differentiation and of the erythroid markers *ALAS2*, *RHCE*, and *KEL* after 24 hr of hemin-induced differentiation (n = 3 per group). Error bars show SD.

lentiviral-mediated shRNA delivery in primary human adult mobilized peripheral-blood hematopoietic stem and progenitor cells (HSPCs), which are capable of differentiation toward the erythroid and megakaryocyte lineages under appropriate culture conditions.[27] We observed a knockdown efficiency of the *GFI1B* long isoform by ~50% for both shRNAs, whereas the short isoform levels increased conversely (Figures 3A and 3B), which resulted in a 1.5- to 1.8-fold reduction in the formation of $CD41a^{+}$ megakaryocytic cells (relative to lineage-marker negative cells) in HSPCs undergoing differentiation (Figure 3C). In contrast, $CD235a^{+}$ erythroid cells appeared to be present in comparable percentages and numbers (Figure 3C). Moreover, whereas numerous morphologically mature erythroblasts could be readily visualized in both groups, fewer mature megakaryocytic cells were seen with knockdown of the long isoform than in the controls (Figure 3D, Figure S10). Overall cell growth appeared comparable between the knockdown and control cells (Figure S10). These findings are in line with our exome-sequence association findings, in which no significant effect was seen on circulating RBC levels.

GFI1B private, loss-of-function mutations (nonsense, frameshift) in the DNA-binding fifth and sixth zinc (Zn)-finger domains have recently been identified in families with an autosomal-dominant form of Gray Platelet syndrome (GPS) or related forms of thrombocytopenia, which are characterized by dysmegakaryopoiesis, thrombocytopenia, large platelets, and platelet α-granule deficiency (MIM: 187900)[28,29]. The truncating *GFI1B* mutations reported in GPS appear to have a dominant-negative effect and inhibit transcriptional activity of the *GFI1B* wild-type form. Our population study extends the allelic spectrum of naturally occurring *GFI1B* coding sequence variants associated with a lower circulating platelet count to include a more frequent, synonymous change that alters

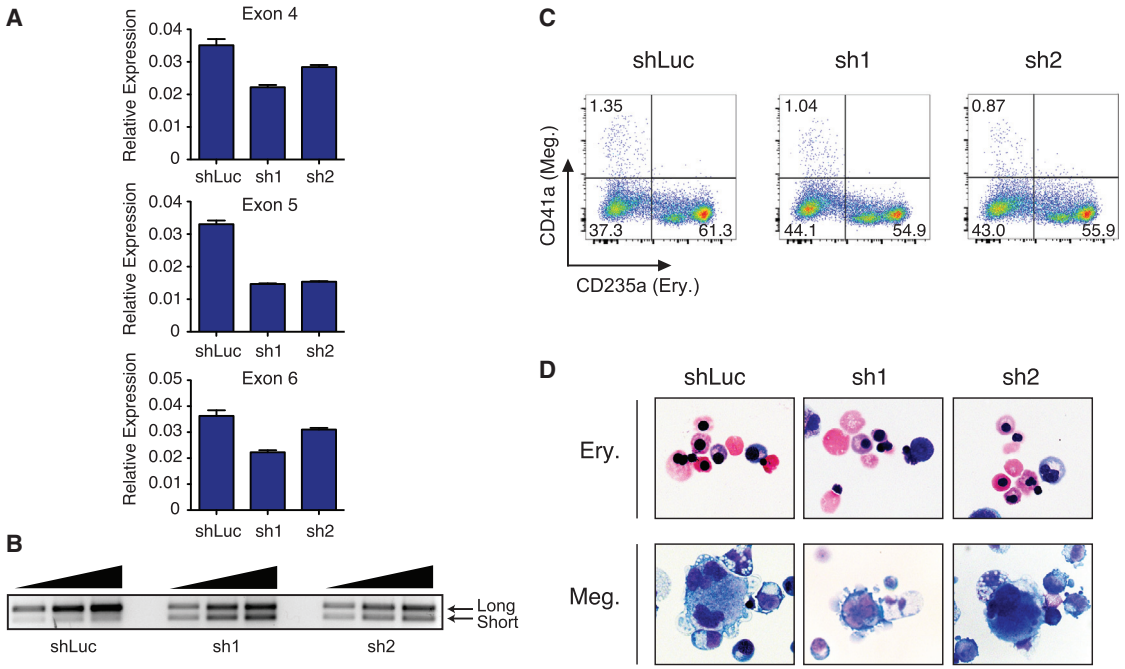

Figure 3. The Long *GFI1B* Isoform is Critical for Megakaryopoiesis in a Human Primary Cell Model
(A) qRT-PCR of *GFI1B* exons 4, 5, and 6 on day 4 after infection showing the identification of two short hairpin RNAs (shRNAs) that specifically target *GFI1B* exon 5 and thereby selectively downregulate the long isoform by ~50%, but not the short isoform (n = 3 per group). Error bars show SD.
(B) Semi-quantitative RT-PCR with *GFI1B* exon 4 forward and exon 6 reverse primers with progressively increasing cycle numbers (26, 28, and 30 cycles) demonstrates reduced formation of the long *GFI1B* isoform and increased formation of the short isoform, as well as no other intermediate isoforms in cells with targeted knockdown of GFI1B exon 5.
(C) Representative flow cytometry analysis of thrombopoietin (TPO)- and erythropoietin (EPO)-stimulated primary human hematopoietic stem and progenitor cells on day 11 of differentiation with assessment of $CD41a^{+}$ megakaryocytic (Meg) cells and $CD235a^{+}$ erythroid (Ery) cells.
(D) Representative May-Grünwald-Giemsa-stained cytospin images of megakaryocytic cells (from day 7 of differentiation) and erythroid cells (from day 13 of differentiation) showing immature megakaryocyte morphology in cells with knockdown of the long *GFI1B* isoform, in comparison with the control. In contrast, maturation of erythroblasts appears unaffected.

an exonic splicing enhancer, resulting in the skipping of exon 5, containing the first and second Zn-finger domains. Heterozygous carriers of the synonymous exon 5 variant in *GFI1B* have an average platelet count that is reduced by 25,000 to 30,000 platelets per microliter, which would be a clinically detectable effect. We also provide additional support for distinct roles of GFI1B long- and short-isoforms, which are differentially expressed at various stages of differentiation during normal hematopoiesis.[23,30] The long GFI1B isoform is expressed in HSPCs and lineage-committed myeloid, erythroid, and megakaryocytic progenitors. The abnormalities in megakaryocyte maturation with reduced formation of the *GFI1B* long isoform in the isogenic K562 cell clones containing the rs150813342 variant and in primary HSPCs with targeted suppression of the long isoform are consistent with an essential role for the GFI1B long isoform in megakaryopoiesis and platelet production. This finding is also congruent with prior work showing that the GFI1B short isoform is required for erythropoiesis[26] and provides insight into how these different splice variants function in distinct aspects of human hematopoiesis.

In summary, whole-exome sequence association analysis performed in over 15,000 samples discovered SNVs associated with a lower platelet count in community-dwelling individuals, including a common variant in *CPS1* and a rare, synonymous variant in *GFI1B*. Follow-up genome editing and targeted knockdown experiments identified a mechanism by which alternative splicing associated with the *GFI1B* rs150813342 variant allele suppresses formation of a specific GFI1B long isoform that is required for lineage-specific megakaryocyte differentiation, while being dispensable for erythropoiesis. Functional studies coupled with an association finding demonstrated a previously unappreciated splicing-based mechanism for lineage-specific blood cell production, providing important insights into human hematopoiesis. Genes regulated by the long GFI1B isoform could provide additional understanding of downstream transcriptional events and molecular pathways required for megakaryocyte specification and platelet production. These findings hold promise for the development of therapeutics for altering platelet count without adverse effects on other blood lineages. Further characterization of the role of GFI1B isoforms could have clinical or therapeutic implications for disorders of platelet and other blood cell

production or function, as well as for the prospect of improving the manufacture of ex vivo cell therapies.[31–33]

Supplemental Data

Supplemental Data include a Supplemental Note, ten figures, nine tables, and Supplemental Acknowledgments and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.06.016.

Acknowledgments

The work described in this manuscript was supported in part by NIH grants HL122684 to S.K.G. and DK103794 and HL120791 to V.G.S. N.S. is supported by the Wellcome Trust (grant codes WT098051 and WT091310), the EU 7th Framework Programme (EPIGENESYS grant code 257082 and BLUEPRINT grant code HEALTH-F5-2011-282510) and the NIH Research (NIHR) Blood and Transplant Research Unit (BTRU) in Donor Health and Genomics at the University of Cambridge in partnership with NHS Blood and Transplant (NHSBT).

Received: March 14, 2016
Accepted: June 20, 2016
Published: August 4, 2016

Web Resources:

OMIM, http://www.omim.org/

References

1. Okada, Y., and Kamatani, Y. (2012). Common genetic factors for hematological traits in humans. J. Hum. Genet. *57*, 161–169.
2. Ganesh, S.K., Zakai, N.A., van Rooij, F.J., Soranzo, N., Smith, A.V., Nalls, M.A., Chen, M.H., Kottgen, A., Glazer, N.L., Dehghan, A., et al. (2009). Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat. Genet. *41*, 1191–1198.
3. Soranzo, N., Spector, T.D., Mangino, M., Kühnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet. *41*, 1182–1190.
4. Auer, P.L., Teumer, A., Schick, U., O'Shaughnessy, A., Lo, K.S., Chami, N., Carlson, C., de Denus, S., Dubé, M.P., Haessler, J., et al. (2014). Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat. Genet. *46*, 629–634.
5. Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T.S., and Sankaran, V.G. (2016). Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell *165*, 1530–1545.
6. 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.
7. Minelli, A., Maserati, E., Rossi, G., Bernardo, M.E., De Stefano, P., Cecchini, M.P., 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 Chromosomes Cancer *40*, 165–171.
8. Sankaran, V.G., and Gallagher, P.G. (2013). Applications of high-throughput DNA sequencing to benign hematology. Blood *122*, 3575–3582.
9. Albers, C.A., Cvejic, A., Favier, R., Bouwmans, E.E., Alessi, M.C., Bertone, P., Jordan, G., Kettleborough, R.N., Kiddle, G., Kostadima, M., et al. (2011). Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. Nat. Genet. *43*, 735–737.
10. Johnsen, J.M., Nickerson, D.A., and Reiner, A.P. (2013). Massively parallel sequencing: the new frontier of hematologic genomics. Blood *122*, 3268–3275.
11. Giani, F.C., Fiorini, C., Wakabayashi, A., Ludwig, L.S., Salem, R.M., Jobaliya, C.D., Regan, S.N., Ulirsch, J.C., Liang, G., Steinberg-Shemer, O., et al. (2016). Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells. Cell Stem Cell *18*, 73–78.
12. Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., and Gazda, H.T. (2012). Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J. Clin. Invest. *122*, 2439–2443.
13. Köttgen, A., Pattaro, C., Böger, C.A., Fuchsberger, C., Olden, M., Glazer, N.L., Parsa, A., Gao, X., Yang, Q., Smith, A.V., et al. (2010). New loci associated with kidney function and chronic kidney disease. Nat. Genet. *42*, 376–384.
14. Paré, G., Chasman, D.I., Parker, A.N., Zee, R.R., Mälarstig, A., Seedorf, U., Collins, R., Watkins, H., Hamsten, A., Miletich, J.P., and Ridker, P.M. (2009). Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. Circ Cardiovasc Genet *2*, 142–150.
15. Summar, M.L., Gainer, J.V., Pretorius, M., Malave, H., Harris, S., Hall, L.D., Weisberg, A., Vaughan, D.E., Christman, B.W., and Brown, N.J. (2004). Relationship between carbamoylphosphate synthetase genotype and systemic vascular function. Hypertension *43*, 186–191.
16. Pearson, D.L., Dawling, S., Walsh, W.F., Haines, J.L., Christman, B.W., Bazyk, A., Scott, N., and Summar, M.L. (2001). Neonatal pulmonary hypertension–urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function. N. Engl. J. Med. *344*, 1832–1838.
17. Sabater-Lleal, M., Huang, J., Chasman, D., Naitza, S., Dehghan, A., Johnson, A.D., Teumer, A., Reiner, A.P., Folkersen, L., Basu, S., et al.; VTE Consortium; STROKE Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2); C4D Consortium; CARDIoGRAM Consortium (2013). Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. Circulation *128*, 1310–1324.
18. Catellier, D.J., Aleksic, N., Folsom, A.R., and Boerwinkle, E. (2008). Atherosclerosis Risk in Communities (ARIC) Carotid MRI flow cytometry study of monocyte and platelet markers: intraindividual variability and reliability. Clin. Chem. *54*, 1363–1371.
19. Summar, M.L., Hall, L., Christman, B., Barr, F., Smith, H., Kallianpur, A., Brown, N., Yadav, M., Willis, A., Eeds, A., et al.

(2004). Environmentally determined genetic expression: clinical correlates with molecular variants of carbamyl phosphate synthetase I. Mol. Genet. Metab. *81* (*Suppl 1*), S12–S19.
20. Zhang, X., Joehanes, R., Chen, B.H., Huan, T., Ying, S., Munson, P.J., Johnson, A.D., Levy, D., and O'Donnell, C.J. (2015). Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat. Genet. *47*, 345–352.
21. Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., et al. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science *347*, 1254806.
22. Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc. Natl. Acad. Sci. USA *108*, 8233–8238.
23. Foudi, A., Kramer, D.J., Qin, J., Ye, D., Behlich, A.S., Mordecai, S., Preffer, F.I., Amzallag, A., Ramaswamy, S., Hochedlinger, K., et al. (2014). Distinct, strict requirements for Gfi-1b in adult bone marrow red cell and platelet generation. J. Exp. Med. *211*, 909–927.
24. Randrianarison-Huetz, V., Laurent, B., Bardet, V., Blobe, G.C., Huetz, F., and Duménil, D. (2010). Gfi-1B controls human erythroid and megakaryocytic differentiation by regulating TGF-beta signaling at the bipotent erythro-megakaryocytic progenitor stage. Blood *115*, 2784–2795.
25. Saleque, S., Cameron, S., and Orkin, S.H. (2002). The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. Genes Dev. *16*, 301–306.
26. Laurent, B., Randrianarison-Huetz, V., Frisan, E., Andrieu-Soler, C., Soler, E., Fontenay, M., Dusanter-Fourt, I., and Duménil, D. (2012). A short Gfi-1B isoform controls erythroid differentiation by recruiting the LSD1-CoREST complex through the dimethylation of its SNAG domain. J. Cell Sci. *125*, 993–1002.
27. Ludwig, L.S., Gazda, H.T., Eng, J.C., Eichhorn, S.W., Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., et al. (2014). Altered translation of GATA1 in Diamond-Blackfan anemia. Nat. Med. *20*, 748–753.
28. Stevenson, W.S., Morel-Kopp, M.C., Chen, Q., Liang, H.P., Bromhead, C.J., Wright, S., Turakulov, R., Ng, A.P., Roberts, A.W., Bahlo, M., and Ward, C.M. (2013). GFI1B mutation causes a bleeding disorder with abnormal platelet function. J. Thromb. Haemost. *11*, 2039–2047.
29. Monteferrario, D., Bolar, N.A., Marneth, A.E., Hebeda, K.M., Bergevoet, S.M., Veenstra, H., Laros-van Gorkom, B.A., MacKenzie, M.A., Khandanpour, C., Botezatu, L., et al. (2014). A dominant-negative GFI1B mutation in the gray platelet syndrome. N. Engl. J. Med. *370*, 245–253.
30. Chen, L., Kostadima, M., Martens, J.H., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., et al.; BRIDGE Consortium (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science *345*, 1251033.
31. Vassen, L., Khandanpour, C., Ebeling, P., van der Reijden, B.A., Jansen, J.H., Mahlmann, S., Dührsen, U., and Möröy, T. (2009). Growth factor independent 1b (Gfi1b) and a new splice variant of Gfi1b are highly expressed in patients with acute and chronic leukemia. Int. J. Hematol. *89*, 422–430.
32. Koldehoff, M., Zakrzewski, J.L., Beelen, D.W., and Elmaagacli, A.H. (2013). Additive antileukemia effects by GFI1B- and BCR-ABL-specific siRNA in advanced phase chronic myeloid leukemic cells. Cancer Gene Ther. *20*, 421–427.
33. Thon, J.N., Medvetz, D.A., Karlsson, S.M., and Italiano, J.E., Jr. (2015). Road blocks in making platelets for transfusion. J. Thromb. Haemost. *13* (*Suppl 1*), S55–S62.