

## Whole-exome analysis of adolescents with low VWF and heavy menstrual bleeding identifies novel genetic associations

Brooke Sadler,<sup>1</sup> Charles G. Minard,<sup>2</sup> Gabe Haller,<sup>3,4</sup> Christina A. Gurnett,<sup>4,5</sup> Sarah H. O'Brien,<sup>6</sup> Allison Wheeler,<sup>7</sup> Shilpa Jain,<sup>8</sup> Mutka Sharma,<sup>9</sup> Ayesha Zia,<sup>10</sup> Roshni Kulkarni,<sup>11</sup> Eric Mullins,<sup>12</sup> Margaret V. Ragni,<sup>13</sup> Robert Sidonio,<sup>14</sup> Jennifer E. Dietrich,<sup>15</sup> Peter A. Kouides,<sup>16,17</sup> Jorge Di Paola,<sup>1</sup> and Lakshmi Srivaths<sup>18</sup>

<sup>1</sup>Department of Pediatrics, Washington University in St. Louis, St Louis, MO; <sup>2</sup>Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX; <sup>3</sup>Department of Neurosurgery, <sup>4</sup>Department of Neurology, and <sup>5</sup>Department of Orthopedic Surgery, Washington University in St. Louis, St. Louis, MO; <sup>6</sup>Department of Pediatrics, Nationwide Children's Hospital, Columbus, OH; <sup>7</sup>Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN; <sup>8</sup>Oishei Children's Hospital, Amherst, NY; <sup>9</sup>Department of Pediatrics, University of Missouri Kansas City School of Medicine, Kansas City, MO; <sup>10</sup>Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; <sup>11</sup>Department of Pediatrics and Human Development, Michigan State University, Lansing, MI; <sup>12</sup>Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; <sup>13</sup>Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA; <sup>14</sup>Department of Pediatrics, Emory University School of Medicine, Atlanta, GA; <sup>15</sup>Department of Pediatrics, Baylor College of Medicine, Houston, TX; <sup>16</sup>Department of Medicine, University of Rochester School of Medicine, Rochester, NY; <sup>17</sup>Mary M. Gooley Hemophilia Center, Rochester, NY; and <sup>18</sup>Department of Pediatrics, Baylor College of Medicine, Houston, TX

### Key Points

- HMB is associated with rare and common variants in genes related to anemias and bleeding disorders.
- These are the first exome-sequencing results from patients with HMB, as well as their comparison with control exomes.

Adolescents with low von Willebrand factor (VWF) levels and heavy menstrual bleeding (HMB) experience significant morbidity. There is a need to better characterize these patients genetically and improve our understanding of the pathophysiology of bleeding. We performed whole-exome sequencing on 86 postmenarchal patients diagnosed with low VWF levels (30-50 IU/dL) and HMB and compared them with 660 in-house controls. We compared the number of rare stop-gain/stop-loss and rare ClinVar “pathogenic” variants between cases and controls, as well as performed gene burden and gene-set burden analyses. We found an enrichment in cases of rare stop-gain/stop-loss variants in genes involved in bleeding disorders and an enrichment of rare ClinVar “pathogenic” variants in genes involved in anemias. The 2 most significant genes in the gene burden analysis, *CFB* and *DNASE2*, are associated with atypical hemolytic uremia and severe anemia, respectively. *VWF* also surpassed exome-wide significance in the gene burden analysis ( $P = 7.31 \times 10^{-6}$ ). Gene-set burden analysis revealed an enrichment of rare nonsynonymous variants in cases in several hematologically relevant pathways. Further, common variants in *FERMT2*, a gene involved in the regulation of hemostasis and angiogenesis, surpassed genome-wide significance. We demonstrate that adolescents with HMB and low VWF have an excess of rare nonsynonymous and pathogenic variants in genes involved in bleeding disorders and anemia. Variants of variable penetrance in these genes may contribute to the spectrum of phenotypes observed in patients with HMB and could partially explain the bleeding phenotype. By identifying patients with HMB who possess these variants, we may be able to improve risk stratification and patient outcomes.

Submitted 28 April 2021; accepted 1 November 2021; prepublished online on *Blood Advances* First Edition 22 November 2021; final version published online 17 January 2022. DOI 10.1182/bloodadvances.2021005118.

The data reported in this article have been deposited in the database of Genotypes and Phenotypes.

Requests for data sharing may be submitted to Lakshmi Srivaths (lakshmi.v.srivaths@uth.tmc.edu).

The full-text version of this article contains a data supplement.

© 2022 by The American Society of Hematology. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

## Introduction

Heavy menstrual bleeding (HMB) is a pathologic state defined as excessive or prolonged menstrual blood loss  $> 80$  mL per menstrual cycle<sup>1-3</sup>; it occurs in approximately one third of adolescent females.<sup>4,5</sup> Quality of life can be significantly affected by HMB, in the form of significant iron deficiency anemia (IDA), psychological distress, missed school, limitations on activities, and prolonged periods of required rest.<sup>6,7</sup> The co-occurrence of bleeding disorders in this population has been shown to be higher than in the general population.<sup>4,8</sup> von Willebrand disease (VWD) is the most common inherited bleeding disorder, with a prevalence of 0.1% (defined by low levels of von Willebrand factor [VWF] and bleeding) in the general population<sup>9</sup> but with a prevalence of 13% in patients with HMB.<sup>10,11</sup> More recently, low VWF levels, independent of a diagnosis of a bleeding disorder, have been shown to be a significant risk factor for HMB.<sup>12</sup>

Although there have been several genome-wide association studies (GWAS) of VWF levels,<sup>13-17</sup> as well as GWAS of hematological parameters, such as number and volume of various cell types in blood,<sup>18-20</sup> these do not detect associations with rare variants. One study validated 1 of the top hits from these GWAS and found that homozygosity for a SNP in *TMPRSS6* was protective against IDA, especially in women with HMB.<sup>21</sup> To date, no next-generation sequencing study has specifically looked at genetic risk factors for HMB. Although it might be expected that such a study would find an excess of rare and pathogenic variants in the *VWF* gene in this population of patients with low VWF, by taking an agnostic approach, other genetic patterns may emerge. Here, we describe the results from the first whole-exome sequencing of 86 adolescent patients with HMB and low VWF, as well as comparison with 660 unrelated control exomes.

## Methods

### Subjects and samples

A multicenter observational cohort study to delineate the phenotype and genotype of adolescent females with low VWF-associated HMB was undertaken from February of 2017 to June of 2019. Tertiary care centers in North America, with expertise in hemostasis and that provided care for adolescents with HMB and were members of the Foundation for Women and Girls with Blood Disorders, participated in the study. Institutional Review Board approval was obtained by all participating centers, and parental and/or patient consent/assent was obtained from all patients prior to study participation. Ten centers participated in this multicenter study. One hundred and thirteen adolescent females with a diagnosis of HMB and low VWF, who met the study inclusion and exclusion criteria, were enrolled. One patient was withdrawn because of the subsequent detection of type 1 VWD and factor XI deficiency. The clinical and laboratory characteristics of this cohort were published recently.<sup>12</sup> Of the 113 patients enrolled, 86 had sufficient blood samples collected for whole-exome sequencing and formed the study population for this analysis.

### Eligibility criteria

The study eligibility criteria included postmenarchal females,  $< 21$  years of age who were diagnosed with HMB, defined as a Pictorial

Blood Assessment Chart score  $> 100$  and low VWF, which was defined as having  $\geq 2$  values of VWF activity  $\geq 30$  and  $\leq 50$  IU/dL (as measured by VWF ristocetin cofactor assay [VWF:RCo] and/or VWF glycoprotein IbM assay). Patients who did not meet these criteria or who were diagnosed with other bleeding disorders were ineligible for the study. Adolescent females, seen in hematology clinics in participating centers managing patients with HMB, were screened for study eligibility, eligible patients were approached for study participation, and patients who consented to participate were enrolled in the study. Clinical phenotype data were extracted prospectively from the patients and retrospectively by reviewing the patients' electronic medical records. Deidentified patient data from each participating center were entered into the coordinating center's electronic database; the data were maintained as confidential with access restricted to study investigators by means of password protection. Baseline hemoglobin values, as well as both measures of von Willebrand factor antigen (VWF:Ag), can be found in supplemental Table 1.

### Sequencing analysis and validation

Exome sequence data for 86 unrelated HMB probands and 660 unrelated in-house controls were generated at the McDonnell Genome Institute (Washington University in St. Louis) using IDT xGen Exome Panel V1 capture on Illumina HiSeq 4000 paired-end reads. The average age of the controls was 16, with a range of 3 to 40 years. Fifty-seven percent of controls (376/660) were female. Analysis of exome sequencing data was performed in-house using our previously described methods,<sup>22,23</sup> with the addition of the Sentieon software package. Briefly, FASTQ formatted sequences were aligned to the hg19 human reference sequence (NCBI GRCh37) using Burrows-Wheeler Aligner (BWA).<sup>24</sup> Mapped reads were filtered to remove duplicate reads with the same paired start sites. Median depth per sample at captured bases for HMB cases was  $76\times$  (range, 17-202); it was  $71\times$  (range, 21-121) for in-house controls. All cases and controls had  $> 97\%$  of captured bases covered with  $> 10$  reads. The Binary sequence Alignment/Map formatted alignments were processed using the Genome Analysis Toolkit (GATK) Haplotype Caller<sup>25,26</sup> and genotypes jointly called together with all in-house control exome-sequenced individuals. Genotypes were filtered for read depth ( $> 10\times$ ), genotype quality  $> 20$ , and allele balance  $> 0.3$  and  $< 0.7$ . Variants were filtered for GATK-calculated variant quality score recalibration and genotype call rate  $> 90\%$ . Allele frequencies were annotated using the gnomAD database.<sup>27</sup> Variants were functionally annotated using ANNOVAR.<sup>28</sup> To reduce the risk of population stratification, only cases and controls with principal components-confirmed European ancestry were included in all analyses. Principal components were calculated using EIGENSTRAT<sup>29</sup> from whole-exome single-nucleotide polymorphism (SNP) data using all common (minor allele frequency [MAF]  $> 5\%$ ) SNPs. We did not remove any cases, because principal components confirmed self-reported race. For the purposes of all single SNP analyses, the standard cutoffs for genome-wide significance ( $5 \times 10^{-8}$ ) and exome-wide significance ( $2.5 \times 10^{-6}$ ) were used. We reasonably excluded HMB and low VWF from the control cohort because the estimate of the prevalence of low VWF-associated HMB in the general population is likely  $\sim 1\%$  given that the prevalence of low VWF is generally  $\sim 2.5\%$  and the prevalence of HMB is  $\sim 35\%$ .<sup>4,30,31</sup>

## Common SNP analysis

Fisher's exact tests were performed using PLINK to compare the MAF of common and rare variants in cases and controls. SNPs were dropped if the genotyping success rate was <90% per SNP or per individual, the Hardy-Weinberg equilibrium ( $P < 1 \times 10^{-6}$ ), and MAF or missingness was <0.05.

## Gene burden and gene-set burden analysis

To quantify the enrichment of rare nonsynonymous/splice-site variants and ClinVar "pathogenic" variants in HMB cases compared with controls for each gene, we compared the collapsed MAF of variants in cases and controls using a Fisher's exact test using only filtered variants. Variants were filtered by rareness and quality: (1) max MAF in all populations < 0.01 in gnomAD exomes and genomes, (2) genome quality > 20, (3) GATK variant quality score recalibration of 'PASS,' (4) minimum sequencing depth of 8 reads in each participant, and (5) allele balance > 0.3 and <0.7. For statistical analysis of exome sequencing data, affected individuals were compared with in-house exome controls. For individual gene association analyses, Fisher's exact tests were performed using PLINK to compare the collapsed MAF of rare variants in cases and controls. Student *t* tests were performed using R to compare the quantitative traits. For gene-set burden analyses, nonsynonymous/splice-site variants within a gene were collapsed to obtain the number of rare (<1% in gnomAD) variants per gene. The numbers of variants within groups of genes were then summed based on membership within a given gene set and used as the dependent variable in a linear regression. Gene Ontology (GO) terms were obtained from the UniProt Knowledgebase (<http://www.uniprot.org/help/uniprotkb>). The exome-wide *P*-value cutoff was used in the gene burden analyses, because that is the number of independent tests (0.05/20 000 genes). Additionally, for our disease-associated gene-set burden analyses, we created a disease-defined set of genes involved in anemia using Online Mendelian Inheritance in Man and UniProt disease descriptions by searching for the term "anemia." For the bleeding disorder disease-defined gene set we searched the same databases for the terms "bleeding," "platelet," and "hemostasis."

## Results

### Exome sequencing reveals excess of rare stop-gain/stop-loss mutations in bleeding disorder genes and rare pathogenic variants in anemia genes

Because rare likely pathogenic variants are necessarily enriched for causal variants, we first examined rare stop-gain/stop-loss and rare ClinVar "pathogenic" variants in our sample. Among the 86 patients, we observed 4 (4.6%) rare stop-gain or stop-loss mutations in genes associated with bleeding or hematological diseases: *CD36* (p.I337Vfs\*10, NM\_001371081), *CPO* (p.R184X, NM\_173077), *VWF* (p.E383X, NM\_000552), and *GP6* (p.X340R, NM\_016363). Variants in these genes are known to cause platelet glycoprotein IV deficiency, hereditary coproporphyrinemia, VWD, and bleeding disorder, platelet type 11, respectively. In controls, there was 1 rare stop-gain variant in *CPO* (0.15%). This difference was statistically significant (Fisher's exact test,  $P = 7.6 \times 10^{-4}$ , odds ratio [OR], 31.8; 95% confidence interval [CI], 27.7-35.9). The patient we found with a stop-gain mutation (p.E383X) in *VWF* had a frequency of this allele in gnomAD of nearly 0, because only 1 instance of it was observed.

This variant was previously identified in 1 patient with VWD.<sup>32</sup> This variant is not listed in ClinVar but is likely to be the cause of the HMB in that individual and also may be contributing to any VWD phenotype that they have.

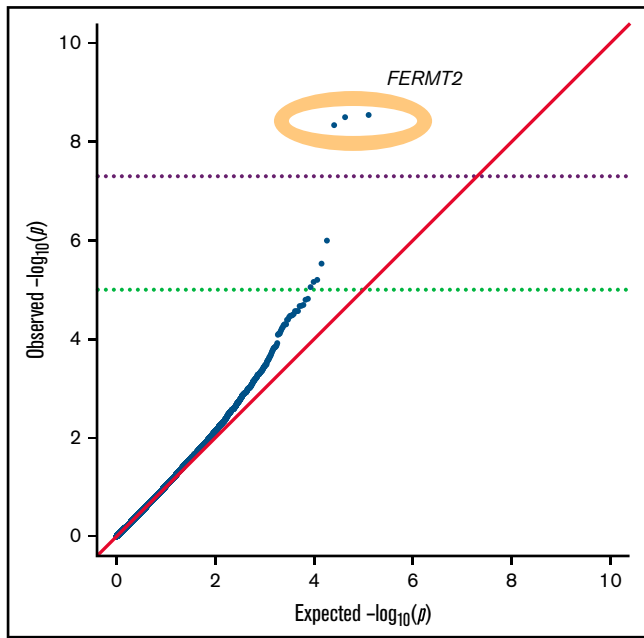
We also pooled all rare (<1% in gnomAD) ClinVar "pathogenic" variants present in cases that were in genes known to cause a dominant or recessive form of anemia or a disease with anemia as a major symptom; 9.3% of cases and 4.2% of unrelated controls harbored such a variant, and this difference was significant (Fisher's exact test,  $P = .04$ ; OR, 2.3; 95% CI, 1.1-3.5) (Table 1). We performed the same comparison in genes involved in hemostasis and platelet function. We found that 4.6% of cases and 1.4% of controls had rare ClinVar "pathogenic" variants in these genes (Fisher's exact test,  $P = .05$ ; OR, 3.5; 95% CI, 2.5-4.5). This association was driven by variants in cases in *VWF* (2), *F2* (1) and *LYST* (1). All significant SNPs and genes remained significant after correcting for the presence of the D1472H variant. We compared our variants with the ISTH VWDdb (Sheffield database). None of the variants that we found are listed as "segregates" in the database, but all are listed, giving no specificity with regard to pathogenicity. Additionally, there was only 1 patient within the cohort that had a VWF:RCo/VWF:Ag ratio < 0.6 from the first set of VWF:Ag measurements, indicating that the cohort was not skewed toward lower VWF:RCo levels because of the presence of this variant. To ensure these were not false positive results, we also examined 5 "control" gene sets in cases and controls including the GO terms "regulation of lipid metabolic processes," "breast/ovarian cancer," "microvascular complications from diabetes," "epilepsy," and "autism." None of these were significant between cases and controls ( $.08 < P < .99$ ).

### Common and rare single SNP analyses

We tested for association with all common SNPs present in the exome and genome data (>5% in gnomAD). Three common SNPs in linkage disequilibrium in or downstream of *FERMT2* (fermitin family member 2) achieved genome-wide significance in the common SNP analysis ( $P = 2.9 \times 10^{-9}$ ; OR, 4.4; 95% CI, 3.0-7.8) (Figure 1). Although the MAF of these SNP is 10% in non-Finnish Europeans in gnomAD, we found these SNPs to be present in 24% of cases and 6% of controls. *FERMT2* encodes a cytoskeletal protein that is a crucial regulator of integrin function.<sup>33</sup> It is known to regulate hemostasis and is required for angiogenesis and blood vessel homeostasis.<sup>34</sup> Additionally, the SNPs that we found to be associated with HMB are expression quantitative trait loci for *FERMT2*, according to the Genotype Tissue Expression Project (<https://www>).

**Table 1. List of genes with anemia as a primary disorder or symptom in which a rare ClinVar "pathogenic" variant was observed in cases**

Gene	Disease association	Inheritance mode
ADAMTS13	Thrombotic thrombocytopenic purpura	Recessive
ANK1	Spherocytosis 1	Dominant
FANCA	Fanconi anemia, complementation group A	Recessive
G6PD	Nonspherocytic hemolytic anemia	Dominant
MVK	Mevalonic aciduria	Dominant
PKLR	Pyruvate kinase deficiency of red cells	Dominant
SLC46A1	Hereditary folate malabsorption	Recessive
TF	Atransferrinemia	Recessive



**Figure 1.** QQ plot of the *P* values for common SNPs. Three SNPs in *FERMT2* achieved genome-wide significance (yellow oval).

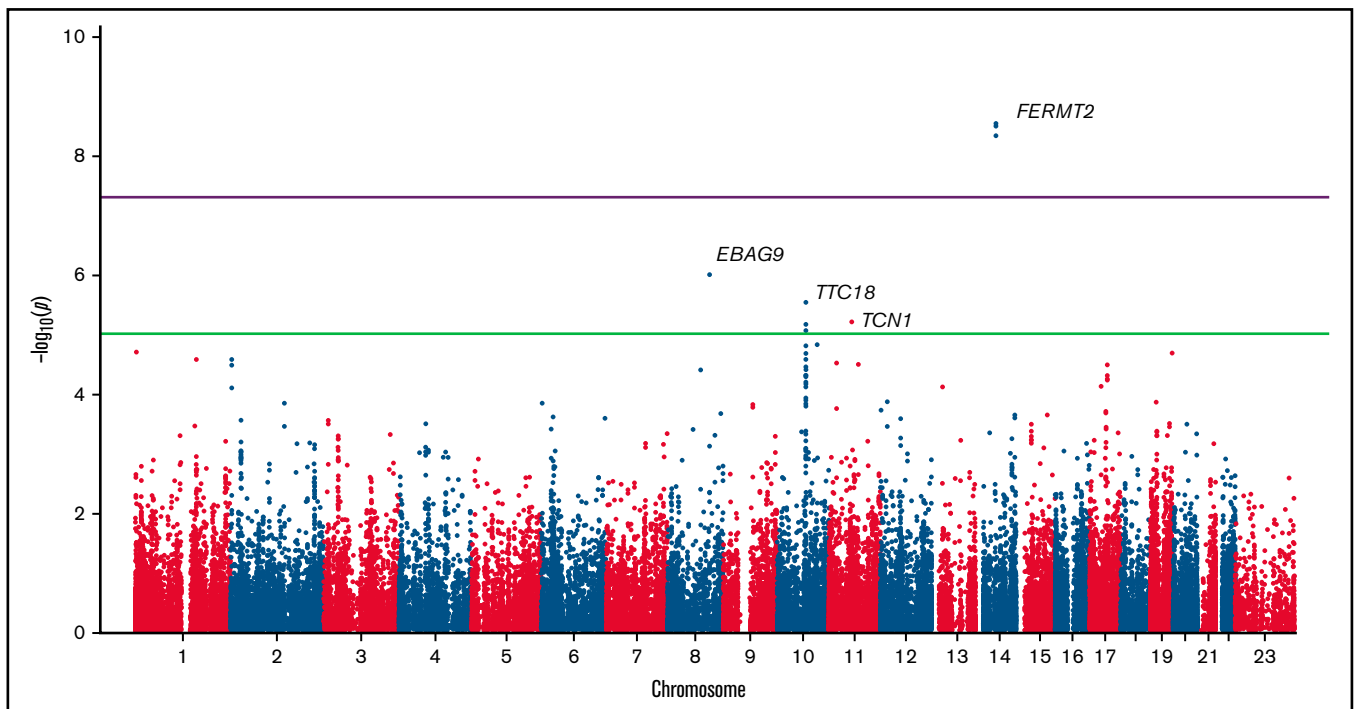
gtexportal.org/home/), further supporting their potential functional role in the phenotype.

We also discovered SNPs in 3 additional genes with *P* values reaching exome-wide significance (Figure 2). *EBAG9* (estrogen receptor

binding site associated antigen 9) is primarily produced by macrophages in hematopoietic tissue and has a crucial role in controlling erythropoiesis.<sup>35</sup> *TTC18* (a.k.a. *CFAP70* [cilia and flagella associated protein 70]) is expressed in the cilia of glandular cells in fallopian tube, endometrium, and respiratory tract<sup>36</sup> (<http://www.proteinatlas.org>), and *TCN1* (transcobalamin 1) is highly expressed in granulocytes,<sup>36</sup> which have been identified as critical determinants of uterine bleeding and tissue remodeling in a mouse menstruation model.<sup>37</sup>

### Gene burden analysis uncovers associations with anemia and bleeding disorder genes

Only coding variants that caused non-sense, splice-site, missense, or insertion/deletion mutations that were rare (defined as <1% in gnomAD) were included in the analysis. This filtering strategy was selected to enrich for mutations that are very rare and most likely to be deleterious. Using a gene burden analysis, we compared the genome-wide frequency of rare coding variants in HMB cases and controls. An increased burden of rare variants that surpassed exome-wide significance was seen in 10 genes (Table 2). The top hit in the results was *CFB* (complement factor B), which achieved exome-wide significance and was the only gene in this analysis to also achieve genome-wide significance ( $P = 3.0 \times 10^{-8}$ ). This gene encodes for a component of the alternative pathway of complement activation. The collapsed MAF (cMAF) for all rare variants in *CFB* was 0.02 in cases ( $n = 86$ ) and 0.0 in controls ( $n = 660$ ). Rare variants in this gene have been associated with atypical hemolytic uremic syndrome,<sup>38,39</sup> which is characterized by microangiopathic hemolytic anemia and thrombocytopenia, as well as acute renal failure.<sup>40</sup>



**Figure 2.** Manhattan plot of the *P* values for common SNPs. Three SNPs in *FERMT2* achieved genome-wide significance (purple line), whereas SNPs in *EBAG9*, *TTC18*, and *TCN1* achieved exome-wide significance (green line).



**Table 2. Gene burden analysis results for genes achieving exome-wide significance**

Gene	Gene name	cMAF affected	Cases, n	cMAF unaffected	Controls, n	P
CFB	Complement factor B	0.023	4	0.000	0	$3.03 \times 10^{-8}$
DNASE2	Deoxyribonuclease 2, lysosomal	0.041	7	0.004	5	$3.53 \times 10^{-7}$
FGFRL1	Fibroblast growth factor receptor-like 1	0.041	7	0.004	5	$3.53 \times 10^{-7}$
NEURL3	Neuralized E3 ubiquitin protein ligase 3	0.116	20	0.034	45	$7.05 \times 10^{-7}$
CCL5	C-C motif chemokine ligand 5	0.017	3	0.000	0	$1.62 \times 10^{-6}$
IL5RA	Interleukin 5 receptor subunit $\alpha$	0.017	3	0.000	0	$1.62 \times 10^{-6}$
KCNK18	Potassium 2 pore domain channel subfamily k member 18	0.017	3	0.000	0	$1.62 \times 10^{-6}$
TBCCD1	TBCC domain containing 1	0.017	3	0.000	0	$1.62 \times 10^{-6}$
TBL1X	Transducin $\beta$ like 1X-linked	0.029	5	0.002	3	$6.11 \times 10^{-6}$
VWF	von Willebrand factor	0.111	19	0.036	47	$7.31 \times 10^{-6}$

The second most significant hit in our results was *DNASE2* (deoxyribonuclease 2, lysosomal), which achieved exome-wide significance ( $P = 3.5 \times 10^{-7}$ ). The cMAF for all rare variants in *DNASE2* was 0.04 in cases compared with 0.004 in controls. This gene encodes a protein that mediates the breakdown of DNA during erythropoiesis and apoptosis. Patients with complete loss of function of this gene were observed to have severe anemia and thrombocytopenia.<sup>41</sup> Because it has been shown that patients with low VWF have an increased rate of rare variants in *VWF*,<sup>42</sup> it is perhaps expected that *VWF* was also significant in the gene burden analysis. There were 19 cases (22%) and 47 controls (7%) with  $\geq 1$  rare nonsynonymous variant in *VWF* ( $P = 7.3 \times 10^{-6}$ ).

### Rare variants in hematologically relevant pathways collectively influence HMB risk

Although some genes achieved exome-wide significance in the gene burden analyses (Table 2), we used our previously developed pathway burden analysis framework<sup>43</sup> that, unlike some methods,<sup>44</sup> preserves power by using data from all genes, not only those with significant single-gene associations. With this method, variants are first collapsed at the gene level and then by GO term membership.

Exome-wide pathway burden analysis yielded a strong association between HMB and novel variants in genes within the GO terms “oxygen transporter activity,” “hemoglobin complex,” “platelet degranulation,” “positive regulation of erythrocyte differentiation,” and “platelet  $\alpha$  granule lumen” (Table 3). These terms are not mutually exclusive; in fact, some have a significant amount of overlap. Notably, these and several other top associated GO terms are highly correlated, often consisting of gene lists that are subsets of one another. For example, nearly all of the genes in the term

“hemoglobin complex” are included in the term “oxygen transporter activity.”

## Discussion

It is estimated that up to 30% of women experience HMB, and it accounts for two thirds of all hysterectomies.<sup>45</sup> Many systemic disruptions of hemostasis, such as liver disease and hypothyroidism, are well-described causes of HMB,<sup>10</sup> yet 50% of women with hysterectomies resulting from HMB do not have the underlying cause identified.<sup>46</sup> Thus, it is clinically important to understand the underlying causes of HMB. Although it is known that bleeding disorders, such as VWD, occur more frequently in this subpopulation than in the overall population, IDA is a common, but underappreciated, complication in adolescents with HMB.<sup>2</sup> The anemia is a direct result of the excessive blood loss during menses and occurs when iron depletion is severe enough to suppress erythropoiesis.<sup>47</sup> Genetic causes of anemia do not necessarily imply a mechanism beyond bleeding. In many cases the underlying biological mechanism is unknown and could be a predisposition to bleeding.

### HMB is associated with rare and common variants in anemia genes

We performed several types of whole-exome analyses on 86 HMB cases compared with our in-house controls. We focused primarily on rare variant analyses, because, as a class, rare variants (defined as  $< 5\%$  population frequency) constitute the majority of genetic variation and are 4 times more likely to be deleterious.<sup>48</sup> In single SNP rare variant and rare variant gene burden analyses, HMB cases had significantly more rare variants in genes known to cause different subtypes of anemias than did controls. Additionally, common variant analysis revealed an association with 1 gene that is known

**Table 3. Relevant results of gene-set burden analyses of patients with HMB**

Gene Ontology term/KEGG pathway	Freq Cases	Freq Unaffected	OR	P
Oxygen transporter activity	0.047	0.003	14.13	$2.5 \times 10^{-10}$
Hemoglobin complex	0.047	0.004	12.1	$1.6 \times 10^{-9}$
Platelet degranulation	0.139	0.048	3.205	$7.2 \times 10^{-7}$
Positive regulation of erythrocyte differentiation	0.023	0.001	20.74	$7.4 \times 10^{-7}$
Platelet $\alpha$ granule lumen	0.105	0.031	3.658	$1.2 \times 10^{-6}$

Freq, frequency; KEGG, Kyoto Encyclopedia of Genes and Genomes.

to contribute to anemia. That the HMB patients do not have known diagnoses of these Mendelian disorders suggests that the variants we identified could be causing less severe or incompletely penetrant forms of anemia.

First, we observed significantly more rare ClinVar “pathogenic” variants in 8 genes associated with varying types of anemias in cases than in controls. A subset of these genes causes dominant forms of anemia, suggesting that it is possible that at least a subset of patients with these rare pathogenic variants may actually have undiagnosed disease. In these patients, the anemia may precede the HMB and, thus, would be exacerbated by the blood loss. Patients with variants in the genes causing recessively inherited anemias may also have milder forms of the disease if they have 1 functional copy of the gene.

Second, the 2 most significant genes from the gene burden analysis are directly linked to anemia. As previously mentioned, rare missense mutations in *CFB* are associated with atypical hemolytic uremic syndrome.<sup>38,39</sup> Biallelic mutations in *DNASE2* have been associated with a loss of DNase II endonuclease activity, causing severe neonatal anemia and thrombocytopenia, among other symptoms.<sup>41</sup> This loss of DNase II induces interferon signaling, inhibiting macrophages from destroying nuclear DNA expelled from erythroid precursor cells. Increased erythroblasts in peripheral blood are observed, suggestive of ineffective erythropoiesis.<sup>41</sup> This phenotype is also seen in mice, because DNase II–null mice accumulate undigested DNA in the lysosomes of macrophages, activating the production of type 1 interferon and resulting in lethal perinatal anemia.<sup>49,50</sup> Because the patients with HMB in our study did not appear to have biallelic loss of *DNASE2*, it is possible that having haploinsufficiency may cause a milder anemia.

Lastly, common variant analysis revealed an association with *EBAG9*, which induces apoptosis in normal human erythroid progenitor cells.<sup>35</sup> *EBAG9* has been detected in monocytes and macrophages. When macrophages were stimulated with lipopolysaccharide, expression of the protein increased, and cell death of erythroid progenitor cells was induced by this increased expression. It has been suggested that erythropoiesis is negatively regulated, in part, by macrophages through the production of *EBAG9*, contributing to the pathogenesis of anemia in patients with inflammatory disorders.<sup>51</sup>

It was recently shown that elevated hematocrit is associated with increased platelet accumulation at the site of injury in mice and human, establishing the role of red cells in normal hemostasis and further underscoring the association of anemia with increased bleeding.<sup>52</sup>

### Rare and common variants in bleeding genes, including *VWF*, in HMB patients

We also replicate what other investigators have noted in patients with HMB: an enrichment of pathogenic variants in genes known to cause bleeding disorders.<sup>10-12</sup> We found 4 rare stop-gain/stop-loss mutations in our cases. One of these was a rare stop-gain mutation in *VWF* that has not been seen in gnomAD. This mutation may be contributing to the undiagnosed VWD in that patient. Two of these mutations are in genes that are known to cause platelet deficiency (*CD36*) and platelet-type bleeding disorder (*GP6*). The last was a stop-gain mutation in *CPO*, a gene that causes a type of hereditary

porphyria. Porphyrias are a group of inborn errors of heme biosynthesis.<sup>53</sup> We also observed more rare ClinVar “pathogenic” variants in hemostasis and platelet function genes in cases vs controls, including 2 people with the same rare ClinVar “pathogenic” variant in *VWF* (P1266Q, NM\_000552), 1 variant in *F2*, and 1 variant in *LYST*. The variant in *F2* (c.\*97G>A, NM\_000506.5) has been associated with prothrombin-related thrombophilia,<sup>54</sup> and the variant in *LYST* (p.R1563H, NM\_000081.3) has been shown to be associated with Chediak-Higashi syndrome.<sup>55,56</sup> Coagulation defects are 1 of the known symptoms of Chediak-Higashi syndrome.<sup>57</sup> Although the *VWF* variant that we found has been observed in patients with type 2 VWD,<sup>58</sup> the patients with this variant had a normal multimer pattern and normal plasma and platelet *VWF* levels. Of the 2 patients with P1266Q\_NM\_000552, 1 patient had an additional variant in *VWF* (p.P2297L\_NM\_000552). However, there were also 3 controls with the P1266Q variant, suggesting that it is not pathogenic on its own. *VWF* was also significant in the rare variant gene burden analysis, and common variant analysis showed an association with *FERMT2*, which is involved in the regulation of hemostasis, angiogenesis, and blood vessel homeostasis.<sup>34</sup> This gene encodes kindlin-2, a widely distributed cytoskeletal protein that is involved in integrin activation. Its absence is embryonically lethal in mice and causes severe developmental defects in zebrafish.<sup>59</sup> Even partial reduction of kindlin-2 in mice resulted in fewer blood vessels, whereas the vessels that form lack smooth muscle cells and are thinner and shorter than normal.<sup>34,59</sup> Kindlin-2 is present in platelets,<sup>60</sup> and kindlin-2–knockdown mice had prolonged bleeding and vascular occlusion times as a result of the suppression of platelet aggregation caused by the elevated expression of 2 enzymes (*CD39*, *CD73*) on endothelial cells.<sup>34</sup> Therefore, an enrichment of variants in *FERMT2* in cases is consistent with what is known about the pathophysiology of this gene and its encoded protein. Additionally, the gene-set burden analysis revealed an enrichment of rare variants in cases in several pathways relevant to hemostasis.

### Limitations

Although 86 patients are sufficient for detecting an association with variants of modest effect size, we recognize that our study is underpowered to detect variants of smaller effect sizes. We evaluated the clinical phenotypes that we ascertained to associate them with genetic variants, but there was too little variation in the laboratory values to identify any associations. This is due to the fact that all patients have the HMB clinical phenotypes, regardless of any differences in the underlying reason for the HMB. In the future we hope to sequence more patients to further fine-tune the associations, as well as to replicate these associations in other similarly phenotyped data sets.

### Conclusions

We present the first whole-exome sequencing results of patients with HMB, as well as their comparison with control exomes. In addition to observing an excess of rare nonsynonymous variants in genes involved in several bleeding disorders, we observed excess variants in genes causing anemias and disorders with anemia as a symptom. Although most of these patients may not have the full disorder associated with variants in that gene, they may simply have a milder version due to incomplete penetrance of the variants, as well as potential interactions with variants in other genes that are needed to cause the severe form of the disorder. Our findings need

validation in larger cohorts. This work may begin to shed some light on the large proportion of patients with HMB who do not have a cause for their symptoms. Eventually, identifying these patients with HMB and with these risk variants may improve risk stratification and patient outcomes.

## Acknowledgment

This work was supported by an investigator-initiated research grant (IIR[1]USA-BXLT-001980-H16-30985) from Shire US Inc., now part of Takeda (L.S.).

## Authorship

Contribution: B.S., L.S., J.E.D., and P.A.K. conceived the study; B.S. and G.H. performed all genetic analyses; C.G.M. analyzed the clinical phenotype data; G.H. and C.A.G. provided control samples; L.S., P.A.K., S.H.O., A.W., S.J., M.S., A.Z., R.K., M.V.R., R.S., and

J.E.D. collected samples; and B.S. and J.E.D. wrote the manuscript; and all authors reviewed the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: B.S., 0000-0002-3268-3809; G.H., 0000-0002-3884-5776; S.H.O., 0000-0001-8855-9746; A.W., 0000-0003-3967-4873; S.J., 0000-0001-5523-1939; A.Z., 0000-0003-3283-0415; R.K., 0000-0001-9372-3184; E.M., 0000-0002-1544-9068; P.A.K., 0000-0002-3857-8313.

Correspondence: Jorge Di Paola, 660 South Euclid Ave, Campus Box 8208, Washington University in St. Louis, St. Louis, MO 63110; e-mail: dipaolaj@wustl.edu; and Lakshmi Srivaths, Gulf States Hemophilia and Thrombophilia Center, University of Texas Health Science Center at Houston and McGovern School of Medicine, 6655 Travis St, Suite 400 HMC, Houston, TX 77030; e-mail: lakshmi.v.srivaths@uth.tmc.edu.

## References

1. American Academy of Pediatrics Committee on Adolescence; American College of Obstetricians and Gynecologists Committee on Adolescent Health Care; Diaz A, Laufer MR, Breech LL. Menstruation in girls and adolescents: using the menstrual cycle as a vital sign. *Pediatrics*. 2006; 118(5):2245-2250.
2. Cooke AG, McCavit TL, Buchanan GR, Powers JM. Iron deficiency anemia in adolescents who present with heavy menstrual bleeding. *J Pediatr Adolesc Gynecol*. 2017;30(2):247-250.
3. Dilley A, Drews C, Miller C, et al. von Willebrand disease and other inherited bleeding disorders in women with diagnosed menorrhagia. *Obstet Gynecol*. 2001;97(4):630-636.
4. Borzutzky C, Jaffray J. Diagnosis and management of heavy menstrual bleeding and bleeding disorders in adolescents. *JAMA Pediatr*. 2020; 174(2):186-194.
5. Friberg B, Ornö AK, Lindgren A, Lethagen S. Bleeding disorders among young women: a population-based prevalence study. *Acta Obstet Gynecol Scand*. 2006;85(2):200-206.
6. Azurah AG, Zainuddin AA, Jayasinghe Y. Diagnostic pitfalls in the evaluation and management of amenorrhea in adolescents. *J Reprod Med*. 2013; 58(7-8):324-336.
7. Ragni MV, Machin N, Malec LM, et al. Von Willebrand factor for menorrhagia: a survey and literature review. *Haemophilia*. 2016;22(3):397-402.
8. Venkateswaran L, Dietrich JE. Gynecologic concerns in pubertal females with blood disorders. *J Pediatr Adolesc Gynecol*. 2013;26(2):80-85.
9. Rodeghiero F, Castaman G, Dini E. Epidemiological investigation of the prevalence of von Willebrand's disease. *Blood*. 1987;69(2):454-459.
10. Lukes AS, Kadir RA, Peyvandi F, Kouides PA. Disorders of hemostasis and excessive menstrual bleeding: prevalence and clinical impact. *Fertil Steril*. 2005;84(5):1338-1344.
11. Shankar M, Lee CA, Sabin CA, Economides DL, Kadir RA. von Willebrand disease in women with menorrhagia: a systematic review. *BJOG*. 2004; 111(7):734-740.
12. Srivaths L, Minard CG, O'Brien SH, et al. The spectrum and severity of bleeding in adolescents with low von Willebrand factor-associated heavy menstrual bleeding [published correction appears in *Blood Adv*. 2021;5(7):1952]. *Blood Adv*. 2020;4(13):3209-3216.
13. Antoni G, Morange PE, Luo Y, et al. A multi-stage multi-design strategy provides strong evidence that the BA13 locus is associated with early-onset venous thromboembolism. *J Thromb Haemost*. 2010;8(12):2671-2679.
14. Sabater-Lleal M, Huffman JE, de Vries PS, et al; INVENT Consortium; MEGASTROKE Consortium of the International Stroke Genetics Consortium (ISGC). Genome-wide association transethnic meta-analyses identifies novel associations regulating coagulation factor VIII and von Willebrand factor plasma levels. *Circulation*. 2019;139(5):620-635.
15. Smith NL, Chen MH, Dehghan A, et al; Wellcome Trust Case Control Consortium. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation*. 2010;121(12):1382-1392.
16. Song J, Xue C, Preisser JS, et al. Association of single nucleotide polymorphisms in the ST3GAL4 gene with VWF antigen and factor VIII activity. *PLoS One*. 2016;11(9):e0160757.
17. van Loon J, Dehghan A, Weihong T, et al. Genome-wide association studies identify genetic loci for low von Willebrand factor levels [published correction appears in *Eur J Hum Genet*. 2016;24(7):1096]. *Eur J Hum Genet*. 2016;24(7):1035-1040.

18. Ganesh SK, Zakai NA, van Rooij FJ, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet.* 2009;41(11):1191-1198.
19. Shameer K, Denny JC, Ding K, et al. A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum Genet.* 2014;133(1):95-109.
20. Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet.* 2009;41(11):1182-1190.
21. Pei SN, Ma MC, You HL, et al. TMPRSS6 rs855791 polymorphism influences the susceptibility to iron deficiency anemia in women at reproductive age. *Int J Med Sci.* 2014;11(6):614-619.
22. Buchan JG, Alvarado DM, Haller GE, et al. Rare variants in FBN1 and FBN2 are associated with severe adolescent idiopathic scoliosis. *Hum Mol Genet.* 2014;23(19):5271-5282.
23. Cruchaga C, Karch CM, Jin SC, et al; Alzheimer's Research UK (ARUK) Consortium. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature.* 2014;505(7484):550-554.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
25. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
26. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43(1110):11.10.1-11.10.33.
27. Karczewski KJ, Francioli LC, Tiao G, et al; Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans [published correction appears in *Nature* 2021;590(7846):E53]. *Nature.* 2020;581(7809):434-443.
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-909.
30. O'Donnell JS. Low VWF: insights into pathogenesis, diagnosis, and clinical management. *Blood Adv.* 2020;4(13):3191-3199.
31. Sadler JE. Low von Willebrand factor: sometimes a risk factor and sometimes a disease. *Hematology (Am Soc Hematol Educ Program).* 2009;2009:106-112.
32. Mannucci PM, Kempton C, Millar C, et al; rVWF Ad Hoc Study Group. Pharmacokinetics and safety of a novel recombinant human von Willebrand factor manufactured with a plasma-free method: a prospective clinical trial. *Blood.* 2013;122(5):648-657.
33. Moser M, Legate KR, Zent R, Fässler R. The tail of integrins, talin, and kindlins. *Science.* 2009;324(5929):895-899.
34. Pluskota E, Ma Y, Bledzka KM, et al. Kindlin-2 regulates hemostasis by controlling endothelial cell-surface expression of ADP/AMP catabolic enzymes via a clathrin-dependent mechanism. *Blood.* 2013;122(14):2491-2499.
35. Matsushima T, Nakashima M, Oshima K, et al. Receptor binding cancer antigen expressed on SiSo cells, a novel regulator of apoptosis of erythroid progenitor cells. *Blood.* 2001;98(2):313-321.
36. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
37. Menning A, Walter A, Rudolph M, Gashaw I, Fritzeimer KH, Roese L. Granulocytes and vascularization regulate uterine bleeding and tissue remodeling in a mouse menstruation model. *PLoS One.* 2012;7(8):e41800.
38. Alfakkeh K, Azar M, Alfadhel M, Abdullah AM, Aloudah N, Alsaad KO. Rare genetic variant in the CFB gene presenting as atypical hemolytic uremic syndrome and immune complex diffuse membranoproliferative glomerulonephritis, with crescents, successfully treated with eculizumab. *Pediatr Nephrol.* 2017;32(5):885-891.
39. Tawadrous H, Maga T, Sharma J, Kupferman J, Smith RJ, Schoeneman M. A novel mutation in the complement factor B gene (CFB) and atypical hemolytic uremic syndrome. *Pediatr Nephrol.* 2010;25(5):947-951.
40. Józsi M, Licht C, Strobel S, et al. Factor H autoantibodies in atypical hemolytic uremic syndrome correlate with CFHR1/CFHR3 deficiency. *Blood.* 2008;111(3):1512-1514.
41. Rodero MP, Tesser A, Bartok E, et al. Type I interferon-mediated autoinflammation due to DNase II deficiency. *Nat Commun.* 2017;8(1):2176.
42. Lavin M, Aguila S, Schneppenheim S, et al. Novel insights into the clinical phenotype and pathophysiology underlying low VWF levels. *Blood.* 2017;130(21):2344-2353.
43. Haller G, Alvarado D, McCall K, et al. A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. *Hum Mol Genet.* 2016;25(1):202-209.
44. Sherman BT, Huang W, Tan Q, et al. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics.* 2007;8(1):426.
45. Oehler MK, Rees MC. Menorrhagia: an update. *Acta Obstet Gynecol Scand.* 2003;82(5):405-422.
46. Clarke A, Black N, Rowe P, Mott S, Howle K. Indications for and outcome of total abdominal hysterectomy for benign disease: a prospective cohort study. *Br J Obstet Gynaecol.* 1995;102(8):611-620.
47. Johnson S, Lang A, Sturm M, O'Brien SH. Iron deficiency without anemia: a common yet under-recognized diagnosis in young women with heavy menstrual bleeding. *J Pediatr Adolesc Gynecol.* 2016;29(6):628-631.



48. Tennessen JA, Bigham AW, O'Connor TD, et al; NHLBI Exome Sequencing Project. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337(6090):64-69.
49. Kawane K, Fukuyama H, Kondoh G, et al. Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science*. 2001;292(5521):1546-1549.
50. Yoshida H, Okabe Y, Kawane K, Fukuyama H, Nagata S. Lethal anemia caused by interferon-beta produced in mouse embryos carrying undigested DNA. *Nat Immunol*. 2005;6(1):49-56.
51. Suehiro Y, Muta K, Nakashima M, et al. A novel mechanism in suppression of erythropoiesis during inflammation: a crucial role of RCAS1. *Eur J Haematol*. 2005;74(5):365-373.
52. Walton BL, Lehmann M, Skorzewski T, et al. Elevated hematocrit enhances platelet accumulation following vascular injury. *Blood*. 2017;129(18):2537-2546.
53. Schmitt C, Gouya L, Malonova E, et al. Mutations in human CPO gene predict clinical expression of either hepatic hereditary coproporphyrin or erythropoietic sideroporphyria. *Hum Mol Genet*. 2005;14(20):3089-3098.
54. Kujovich JL. Prothrombin thrombophilia. In: Adam MP, Ardinger HH, Pagon RA, et al, eds. GeneReviews [Internet]. Seattle, WA: University of Washington, Seattle; 1993-2021. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK1148/>. Accessed 27 May 2020
55. Toro C, Nicoli ER, Malicdan MC, Adams DR, Introne WJ. Chediak-Higashi syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al, eds. GeneReviews [Internet]. Seattle, WA: University of Washington, Seattle; 1993-2021. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK5118/>. Accessed 1 June 2020.
56. Karim MA, Suzuki K, Fukai K, et al. Apparent genotype-phenotype correlation in childhood, adolescent, and adult Chediak-Higashi syndrome. *Am J Med Genet*. 2002;108(1):16-22.
57. Kaplan J, De Domenico I, Ward DM. Chediak-Higashi syndrome. *Curr Opin Hematol*. 2008;15(1):22-29.
58. Casonato A, Daidone V, Galletta E, Bertomoro A. Type 2B von Willebrand disease with or without large multimers: a distinction of the two sides of the disorder is long overdue. *PLoS One*. 2017;12(6):e0179566.
59. Pluskota E, Dowling JJ, Gordon N, et al. The integrin coactivator kindlin-2 plays a critical role in angiogenesis in mice and zebrafish. *Blood*. 2011;117(18):4978-4987.
60. Ma YQ, Qin J, Wu C, Plow EF. Kindlin-2 (Mig-2): a co-activator of beta3 integrins. *J Cell Biol*. 2008;181(3):439-446.