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

Tracking no: ADV-2021-005118R2

Brooke Sadler (Washington University School of Medicine, United States) Charles Minard (Baylor College of Medicine, United States) Gabe Haller (Washington University School of Medicine, United States) Christina Gurnett (Washington University School of Medicine, United States) Sarah O'Brien (Nationwide Children's Hospital, United States) Allison Wheeler (Vanderbilt University Medical Center, United States) Shilpa Jain (Oishei Children's Hospital, United States) Mukta Sharma (University of Missouri Kansas City School of Medicine, United States) Ayesha Zia (The University of Texas Southwestern Medical Center, United States) Roshi Kulkarni (Michigan State University, United States) Eric Mullins (Cincinnati Children's Hospital Medical Center, United States) Margaret Ragni (University of Pittsburgh, United States) Robert Sidonio (Emory University School of Medicine, United States) Jennifer Dietrich (Baylor College of Medicine, United States) Peter Kouides (Mary M. Gooley Hemophilia Center, United States) Jorge Di Paola (Washington University in St. Louis School of Medicine, United States) Lakshmi Srivaths (Baylor College of Medicine, United States)

Abstract:
Adolescents with low von Willebrand factor (VWF) levels and heavy menstrual bleeding (HMB) experience significant morbidity. There is a need to better genetically characterize these patients and improve our understanding of the pathophysiology of bleeding. We performed whole-exome sequencing on 86 post-menarchal patients diagnosed with low-VWF levels (30-50 IU/dL) and HMB and compared them to 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 two most significant genes in the gene-burden analysis, CFB and DNASE2, are associated with atypical hemolytic uremia (aHUS) and severe anemia, respectively. VWF also surpassed exome-wide significance in the gene-burden analysis (p=7.31x10^-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 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 disorders of bleeding and anemia. Variants of variable penetrance in these genes may contribute to the spectrum of phenotypes observed in HMB patients, and could partially explain the bleeding phenotype. By identifying the HMB patients who possess these variants, we may be able to improve risk stratification and patient outcomes.

Conflict of interest: No COI declared

COI notes: The authors declare no conflicts of interest.

Preprint server: No;

Author contributions and disclosures: B.S., L.S., J.D. & P.K. conceived of the manuscript. B.S. & G.H. performed all genetic analyses. C.M. analyzed the clinical phenotype data. G.H. & C.A.G. provided control samples. L.S., P.K., S.H.O, A.W., S.J., M.S., A.A., R.K., M.V.R., R.S. & J.E.D. collected samples. B.S. & J.D. wrote the manuscript. All authors reviewed the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Please email Dr. Jorge Di Paola, dipaolaj@wustl.edu

Clinical trial registration information (if any):
WHOLE EXOME ANALYSIS OF ADOLESCENTS WITH LOW VWF AND HEAVY MENSTRUAL BLEEDING IDENTIFIES NOVEL GENETIC ASSOCIATIONS

Brooke Sadler, Charles G. Minard, Gabe Haller, Christina A. Gurnett, Sarah H. O’Brien, Allison Wheeler, Shilpa Jain, Mutka Sharma, Ayesha Zia, Roshni Kulkarni, Eric Mullins, Margaret V. Ragni, Robert Sidonio, Jennifer E. Dietrich, Peter A. Kouides, Jorge Di Paola, Lakshmi Srivaths

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

Running Title: Exome data identifies novel associations with HMB

Word Count: 3,572
Abstract Word Count:
Number of Figures: 2
Number of Tables: 3
Number of References: 58

Corresponding authors:
Jorge Di Paola, MD
Professor of Pediatrics & Molecular Genetics and Genomics
Elizabeth Finney McDonnell Endowed Chair in Pediatric Hematology Oncology
660 S Euclid Ave Campus Box 8208
Washington University in St Louis
St Louis, MO 63110
Phone (314) 273-3940
dipaolaj@wustl.edu

Lakshmi Srivaths, MD
Professor of Pediatrics
Gulf States Hemophilia & Thrombophilia Center
University of Texas Health Science Center at Houston and McGovern School of Medicine
6655 Travis Street, Suite 400 HMC
Houston, Texas 77030
Phone (713) 500 8322
Lakshmi.V.Srivaths@uth.tmc.edu
KEY POINTS

- Heavy menstrual bleeding is associated with both rare and common variants in genes related to anemias and bleeding disorders.
- These are the first exome-sequencing results of heavy menstrual bleeding patients as well as their comparison to control exomes.

ABSTRACT

Adolescents with low von Willebrand factor (VWF) levels and heavy menstrual bleeding (HMB) experience significant morbidity. There is a need to better genetically characterize these patients and improve our understanding of the pathophysiology of bleeding. We performed whole-exome sequencing on 86 post-menarchal patients diagnosed with low-VWF levels (30-50 IU/dL) and HMB and compared them to 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 two most significant genes in the gene-burden analysis, CFB and DNASE2, are associated with atypical hemolytic uremia (aHUS) and severe anemia, respectively. VWF also surpassed exome-wide significance in the gene-burden analysis (p=7.31x10^{-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 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 disorders of bleeding and anemia. Variants of variable penetrance in these genes may contribute to the spectrum of phenotypes observed in HMB patients, and could partially explain the bleeding phenotype. By identifying the HMB patients who possess these variants, we may be able to improve risk stratification and patient outcomes.
INTRODUCTION

Heavy menstrual bleeding (HMB), is a pathologic state defined as excessive or prolonged menstrual blood loss in excess of 80 mL per menstrual cycle \(^1-^3\), occurring in approximately one third of adolescent females \(^4,^5\).

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 \(6,^7\). The co-occurrence of bleeding disorders in this population has been shown to be higher than in the general population \(^4,^8\). von Willebrand Disease is the most common inherited bleeding disorder, with a prevalence of 0.1% (defined by low-VWF levels and bleeding) of the general population \(^9\) but a prevalence of 13% in patients with HMB \(^10,^11\). More recently, low von Willebrand factor (VWF) levels independent of a diagnosis of a bleeding disorder have been shown to be a significant risk factor for HMB as well \(^12\).

Although there have been several genome-wide association studies (GWAS) of VWF levels \(^13-^17\), as well as GWAS of hematological parameters such as number and volume of various cell types in blood \(^18-^20\), these do not detect associations with rare variants. One study did validate one 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 \(^21\). To date, there have been no next-generation sequencing studies specifically looking at genetic risk factors for HMB. While 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 low VWF patients, by taking an agnostic approach, other genetic patterns may emerge. Here we describe the results from the first whole-exome sequencing (WES) of 86 adolescent HMB patients with low VWF as well as comparison to 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 2017 to June 2019. Tertiary care centers in North America, with expertise in hemostasis, which 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 have been recently...
published\textsuperscript{12}. Of the 113 patients enrolled, 86 had sufficient blood samples collected for WES, who formed the study population for this analysis.

**Eligibility Criteria**

The study eligibility criteria included post-menarchal females, \(< 21\) years of age diagnosed with HMB, defined as Pictoral Blood Assessment Chart (PBAC) score \(> 100\), and low VWF, defined as having \(\geq 2\) values of VWF activity (VWF: Act) \(> 30\) and \(< 50\) IU/dL (as measured by VWF ristocetin cofactor assay [VWF: RCo] and/or VWF glycoprotein IbM assay [VWF: GpIbM]). 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 HMB patients, 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. De-identified 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 VWF:Ag can be found in Table S1.

**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 (MGI) 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 3-40 years. 57\% of controls (376/660) were female. Analysis of exome sequencing data was performed in-house using our previously described methods\textsuperscript{22,23} with the addition of the Sentieon software package. Briefly, FASTQ formatted sequences were aligned to the hg19 human reference sequence (NCBI GRCh37) using BWA\textsuperscript{24}. 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 76x (range 17-202) and was 71x (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 (BAM) formatted alignments were then processed using the Genome Analysis Toolkit (GATK) Haplotype Caller\textsuperscript{25,26} and genotypes jointly called together with all in-house control exome sequenced individuals. Genotypes were filtered for read-depth (\(> 10\)x), genotype quality (GQ\(> 20\)), and allele balance (AB) \(> 0.3\) and \(< 0.7\). Variants were filtered for GATK-calculated variant quality score recalibration (VQSR) and genotype call-rate \(> 90\%\). Allele frequencies were annotated using the gnomAD database\textsuperscript{27}. Variants were functionally annotated using ANNOVAR.\textsuperscript{28} To reduce the risk of population stratification, only cases and controls with principal components confirmed European ancestry were included in all analysis. Principal components were calculated...
using EIGENSTRAT\textsuperscript{29} from whole-exome SNP data using all common (MAF >5%) SNPs. We did not remove any cases as PCs confirmed self-reported race. For the purposes of all single SNP analyses, the standard cutoff for genome-wide significance ($5 \times 10^{-8}$) and exome-wide significance ($2.5 \times 10^{-6}$) were used. We have reasonably excluded HMB and low-VWF from the control cohort as the estimate of the prevalence of low-VWF associated HMB in the general population is likely around 1% given the prevalence of low-VWF generally is around 2.5% and the prevalence of HMB is around 35% \textsuperscript{4,30,31}.

**Common SNP Analysis**

Fisher's exact tests were performed using PLINK to compare the minor allele frequency of common and rare variants in cases and controls. SNPs were dropped if: (1) genotyping success rate <90% per SNP or per individual; (2) Hardy-Weinberg equilibrium (HWE) (p<1×10^{-6}); (3) Minor allele frequency (MAF) or missingness <0.05.

**Gene Burden and Gene-set Burden Analysis**

To quantify the enrichment of rare, non-synonymous/splice-site variants and ClinVar pathogenic variants in HMB cases compared to controls for each gene, we compared the collapsed minor allele frequency of variants in cases and controls using a Fisher's exact test utilizing only filtered variants. Variants were filtered by rareness and quality: (1) max MAF in all populations <0.01 in gnomAD exomes and genomes, (2) GQ>20, (3) GATK VQSR of ‘PASS’, (4) minimum sequencing depth of 8 reads in each participant, and (5) allele balance (AB) >0.3 and <0.7. For statistical analysis of exome sequencing data, affected individuals were compared to in-house exome controls. For individual gene association analyses, Fisher's exact tests were performed using PLINK to compare the collapsed minor allele frequency of rare variants in cases and controls. T-tests were performed using R to compare the quantitative traits. For gene-set burden analyses, non-synonymous/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. 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 as 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 gene set of genes involved in anemia using OMIM 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’.

**Data Sharing Statement**
For original data, please contact Dr. Lakshmi Srivaths at Lakshmi.v.srivaths@uth.tmc.edu. The dataset can be found in the database of Genotypes and Phenotypes (dbGaP, accession number:XXX).

RESULTS

Exome-Sequencing Reveals Excess of Rare Stop-Gain/Stop-Loss mutations in Bleeding Disorder Genes and Rare Pathogenic Variants in Anemia Genes

As 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 (CD36), hereditary coproporphyria (CPO), von Willebrand Disease (VWF) and bleeding disorder, platelet type 11 (GP6) respectively. In controls, there was one rare stop-gain variant in CPO (0.15%). This difference was statistically significant (Fisher’s Exact p=7.6x10^{-4}, OR=31.8, 95%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 zero, as only one instance of it was observed. This variant was previously identified in one patient with VWD32. This variant is not listed in ClinVar, but is likely to be the cause of the HMB in that individual, and is also possibly contributing to any VWD phenotype they may have as well.

We additionally pooled all rare (<1% in gnomAD) ClinVar ‘Pathogenic’ variants present in cases that were in genes known to cause either 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 p=0.04, OR=2.3, 95%CI=1.1-3.5)(Table 1). Similarly, we performed the same comparison in genes involved in hemostasis and platelet function. We find that 4.6% of cases and 1.4% of controls had rare ClinVar ‘pathogenic’ variants in these genes (Fisher’s Exact p=0.05, 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 to the ISTH VWDdb (Sheffield database). None of the variants we found are listed as ‘segregates’ in the database, but all are listed, giving no specificity with regard to pathogenicity. Additionally, there was only one patient within the cohort that had a VWF:RCo / VWF:Ag ratio of <0.6 from the first set of VWF:Ag measurements, indicating that the cohort was not skewed towards lower VWF:RCo levels due to the presence of this variant. In order to ensure these were not false positive results we also examined five ‘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 (0.08<p<0.99).

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) passed genome-wide significance in the common SNP analysis (p=2.9x10^-9, OR=4.4, 95%CI=3.0-7.8) (Figure 1). Although the minor allele frequency 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. It is known to regulate hemostasis and is required for angiogenesis and blood vessel homeostasis. Additionally, the SNPs we find associated with HMB are eQTLs for FERMT2 according to the Genotype Tissue Expression Project (GTEx) https://www.gtexportal.org/home/ further supporting their potential functional role on the phenotype.

We also discovered SNPs in three additional genes with exome-wide significant p-values (Figure 2). EBAG9 (estrogen receptor binding site associated antigen 9) is mainly produced by macrophages in hematopoietic tissue and has a crucial role in controlling erythropoiesis. 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 (http://www.proteinatlas.org) and TCN1 (transcobalamin 1) is highly expressed in granulocytes, which have been identified as critical determinants of uterine bleeding and tissue remodeling in a mouse menstruation model.

Gene Burden Analysis Uncovers Associations with Anemia and Bleeding Disorder Genes

Only coding variants that caused nonsense, splice-site, missense or insertion/deletion mutations which 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 not only passed exome-wide significance, but was the only gene in this analysis to also pass genome-wide significance (p=3.0x10^-8). This gene encodes for a component of the alternative pathway of complement activation. The collapsed minor allele frequency (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 (aHUS), which is characterized by microangiopathic hemolytic anemia and thrombocytopenia as well as acute renal failure.
The second most significant hit in our results was DNASE2 (deoxyribonuclease 2, lysosomal), which surpassed 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 have been observed to have severe anemia and thrombocytopenia\(^{41}\). As it has been shown that Low VWF patients have an increased rate of rare variants in VWF\(^{42}\), it is perhaps expected that VWF was also significant in the gene-burden analysis. There were 19 cases (22\%) and 47 controls (7\%) with at least one rare nonsynonymous variant in VWF \( (p=7.3 \times 10^{-6}) \).

### Rare Variants in Hematologically Relevant Pathways Collectively Influence HMB Risk

Although some genes surpassed exome-wide significance in the gene burden analyses (Table 2), we utilized our previously developed pathway burden analysis framework\(^{43}\) that, unlike some methods\(^{44}\), preserves power by utilizing 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 Gene Ontology (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, and 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 oxygen transporter activity term.

### DISCUSSION:

It is estimated that up to 30\% of women experience HMB and it accounts for two-thirds of all hysterectomies\(^{45}\). Many systemic disruptions of hemostasis such as liver disease and hypothyroidism are well-described causes of HMB\(^{10}\), yet 50\% of those with hysterectomies due to HMB do not have the underlying cause identified\(^{46}\). Thus, it is clinically important to understand the underlying causes of HMB. While 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\(^2\). The anemia is a direct result of the excessive blood loss during menses and occurs when iron depletion is severe enough to suppress erythropoiesis\(^{47}\). 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 have performed several types of whole-exome analyses on 86 HMB cases compared to our in-house controls. We have mainly focused on rare variant analyses since as a class, rare variants (defined as <5% population frequency) constitute the majority of genetic variation and are four times more likely to be deleterious. 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 controls. Additionally, common variant analysis showed association with one gene known 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 different genes associated with varying types of anemias in cases than controls. A subset of these genes causes dominant forms of anemia, suggesting that it is possible that at least a subset of patients with theses 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 one functional copy of the gene.

Second, the two most significant genes from the gene burden analysis are directly linked to anemia. As previously mentioned, rare missense mutations in \( CFB \) are associated with aHUS. Biallelic mutations in \( DNASE2 \) have been associated with a loss of DNase II endonuclease activity, causing severe neonatal anemia and thrombocytopenia, among other symptoms. 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. This phenotype is also seen in mice as DNase II null mice accumulate undigested DNA in the lysosomes of macrophages, activating the production of type 1 interferon, resulting in lethal perinatal anemia. As the HMB patients 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 association with \( EBAG9 \), which induces apoptosis in normal human erythroid progenitor cells. \( EBAG9 \) has been detected in monocytes and macrophages. When macrophages were stimulated with lipopolysaccharide (LPS), 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 in part negatively regulated by macrophages through production of \( EBAG9 \), contributing to the pathogenesis of anemia in patients with inflammatory disorders.
It has been 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.

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

We also replicate what others have noted in patients with HMB – an enrichment of pathogenic variants in genes known to cause bleeding disorders. We found 4 rare stop-gain/stop-loss in our cases. One of these was a rare, never before seen in gnomAD stop-gain mutation in VWF. This mutation is possibly contributing to undiagnosed VWD in that patient. Two of these mutations are in genes known to cause platelet deficiency (CD36) and platelet type bleeding disorder (GP6). The last was a stop-gain in CPO, a gene that causes a type of hereditary porphyria. Porphyrias are a group of inborn errors of heme biosynthesis. We also observed more rare Clinvar ‘pathogenic’ variants in hemostasis and platelet function genes in cases versus controls, including 2 people with the same rare Clinvar ‘pathogenic’ variant in VWF (P1266Q, NM_000552), one in F2 and one in LYST. The variant in F2 (c.*97G>A, NM_000506.5) has been associated with prothrombin-related thrombophilia and the variant in LYST (p.R1563H, NM_000081.3) has been shown to be associated with Chediak-Higashi Syndrome (CHS). Coagulation defects are one of the known symptoms of CHS. While the VWF variant we find has been observed in VWD Type 2B patients, those patients with this variant had a normal multimer pattern and normal plasma and platelet VWF levels. Of the two patients with P1266Q_NM_000552, one 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 association with FERMT2, involved in regulation of hemostasis, angiogenesis and blood vessel homeostasis. This gene encodes kindlin-2, a widely distributed cytoskeletal protein involved in integrin activation. Its absence is embryonic lethal in mice and causes severe developmental defects in zebrafish. Even partial reduction of kindlin-2 in mice resulted in fewer blood vessels, while the vessels that form lack smooth muscle cells and are thinner and shorter than normal. Kindlin-2 is present in platelets, and kindlin-2 knockdown mice had prolonged bleeding and vascular occlusion times due to suppression of platelet aggregation from elevated expression of 2 enzymes (CD39, CD73) on endothelial cells. 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**
While 86 patients are sufficient for detecting association with variants of modest effect size, we recognize that we will be underpowered to detect variants of smaller effect sizes. Relatedly, we did evaluated the clinical phenotypes we ascertained in order to associate them with genetic variants, but there was too little variation in the laboratory values to determine any associations. This is due to the fact that all the 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 in order to further fine-tune the associations as well as replicate these associations in other similarly phenotyped datasets.

Conclusions

We present the first whole-exome sequencing results of HMB patients as well as their comparison to control exomes. In addition to observing an excess of rare nonsynonymous variants in genes involved in several bleeding disorders, we similarly observed excess variants in genes causing anemias and disorders with anemia as a symptom. While 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 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 HMB patients that do not have a cause for their symptoms. Eventually, identifying these HMB patients with these risk variants may improve risk stratification and patient outcomes.
ACKNOWLEDGEMENTS
This work was supported by investigator-initiated research grant (IIR[1]USA-BXLT-001980-H16-30985) from Shire US Inc., now part of Takeda (LS).

AUTHORSHIP CONTRIBUTIONS
B.S., L.S., J.D. & P.K. conceived of the manuscript. B.S. & G.H. performed all genetic analyses. C.M. analyzed the clinical phenotype data. G.H. & C.A.G. provided control samples. L.S., P.K., S.H.O, A.W., S.J., M.S., B.A., R.K., M.V.R., R.S. & J.E.D. collected samples. B.S. & J.D. wrote the manuscript. All authors reviewed the manuscript.

DISCLOSURE OF CONFLICTS OF INTEREST
The authors declare no conflicts of interest.

Data Sharing Statement
For original data, please contact Dr. Lakshmi Srivaths at Lakshmi.v.srivaths@uth.tmc.edu. The dataset can be found in the database of Genotypes and Phenotypes (dbGaP, accession number: XXX).
REFERENCES

1. American Academy of Pediatrics Committee on A, American College of O, Gynecologists Committee on Adolescent Health C, 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, Orno 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. 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. 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. 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, Preiss 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, Roodenburg L, et al. Genome-wide association studies identify genetic loci for low von Willebrand factor levels. Eur J Hum Genet. 2016;24(7):1096.

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. Shameek D, 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. 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 Auwerda 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;43:11 10 11-11 10 33.

27. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. 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. 2009:106-112.

32. Mannucci PM, Kempton C, Millar C, et al. 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, Fassler 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. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260-1265.

37. Menning A, Walter A, Rudolph M, Gashaw I, Fritzemeier KH, Roese L. Granulocytes and vascularization regulate uterine bleeding and tissue remodeling in a mouse menstruation model. PLoS One. 2012;7(8):e41800.

38. Alfaeeh 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):1260-1269.

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. Jozsi 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 da 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;8:426.
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. Indicators 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. 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, Skorczewski 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 coproporphyria or erythropoietic harderoporphryia. Hum Mol Genet. 2005;14(20):3089-3098.
54. Kujovich JL. Prothrombin-Related Thrombophilia. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews(R). Seattle (WA); 1993.
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(R). Seattle (WA); 1993.
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.
### 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       | Non-spherocytic hemolytic anemia                         | Dominant         |
| MVK        | Mevalonic aciduria                                       | Dominant         |
| PKLR       | Pyruvate kinase deficiency of red cells                  | Dominant         |
| SLC46A1    | Hereditary folate malabsorption                          | Recessive        |
| TF         | Atransferrinemia                                         | Recessive        |

### Table 2

Gene burden analysis results for genes surpassing exome-wide significance.

| Gene     | Full Gene Name                               | cMAF Affected | # Cases | cMAF Unaffected | # Controls | P-Value          |
|----------|---------------------------------------------|---------------|---------|-----------------|------------|------------------|
| CFB      | complement factor B                         | 0.023         | 4       | 0.000           | 0          | 3.03x10^-8       |
| DNASE2   | deoxyribonuclease 2, lysosomal              | 0.041         | 7       | 0.004           | 5          | 3.53x10^-7       |
| FGFR1    | fibroblast growth factor receptor-like 1    | 0.041         | 7       | 0.004           | 5          | 3.53x10^-7       |
| NEURL3   | neuralized E3 ubiquitin protein ligase 3    | 0.116         | 20      | 0.034           | 45         | 7.05x10^-7       |
| CCL5     | C-C motif chemokine ligand 5               | 0.017         | 3       | 0.000           | 0          | 1.62x10^-6       |
| ILSRA    | interleukin 5 receptor subunit alpha        | 0.017         | 3       | 0.000           | 0          | 1.62x10^-6       |
| KCNK18   | potassium two pore domain channel subfamily K member 18 | 0.017   | 3       | 0.000           | 0          | 1.62x10^-6       |
| TBCCD1   | TBCC domain containing 1                    | 0.017         | 3       | 0.000           | 0          | 1.62x10^-6       |
| TBL1X    | transducin beta like 1X-linked              | 0.029         | 5       | 0.002           | 3          | 6.11x10^-6       |
| VWF      | von Willebrand factor                       | 0.111         | 19      | 0.036           | 47         | 7.31x10^-6       |

### Table 3

Relevant results of gene-set burden analyses of HMB patients.

| GO Term/Kegg Pathway                  | Freq Cases | Freq Unaffected | OR  | P-Value   |
|--------------------------------------|------------|-----------------|-----|-----------|
| oxygen transporter activity           | 0.047      | 0.003           | 14.13 | 2.5x10^-10 |
| hemoglobin complex                    | 0.047      | 0.004           | 12.1 | 1.6x10^-9  |
| platelet degranulation                | 0.139      | 0.048           | 3.205 | 7.2x10^-7  |
| positive regulation of erythrocyte differentiation | 0.023 | 0.001           | 20.74 | 7.4x10^-7  |
| platelet alpha granule lumen         | 0.105      | 0.031           | 3.658 | 1.2x10^-6  |
FIGURE LEGENDS:

Figure 1. QQ plot of the p-values for common SNPs. 3 SNPs in FERMT2 surpassed genome-wide significance.

Figure 2. Manhattan plot of the p-values for common SNPs. 3 SNPs in FERMT2 surpassed genome-wide significance (red line), while SNPs in EBAG9, TTC18 and TCN1 surpassed exome-wide significance (blue line).
Figure 1

Expected $-\log_{10}(p)$ vs. Observed $-\log_{10}(p)$

FERMT2
