SUPPLEMENTAL METHODS

DNA preparation

Genomic DNA was extracted from blood using the QIAamp DNA Blood Midi Kit (Catalog number 51183, Qiagen) and from the buccal sample per manufacturer's instructions (Catalog number 33900, Norgen Biotek).

Microsatellite genotyping

PCR was performed using Herculase II Fusion DNA Polymerase (Catalog number 600672, Agilent Technologies). The genotypes were called by running the PCR product on a 20% Novex TBE gel (Catalog number EC63155BOX, Invitrogen).

Sanger Sequencing

PCR was performed using Herculase II Fusion DNA Polymerase (Catalog number 600672, Agilent Technologies). PCR primers are listed in Supplemental Table 1. The PCR product was purified using the QIAquick PCR purification kit (Catalog Number 28106, Qiagen) and Sanger Sequencing was performed at the University of Michigan Sequencing Core.

Linkage Analysis

DNA samples from the thirteen individuals highlighted with a green arrow in Figure 1A were subjected to SNP genotyping on the Infinium HumanCoreExome v24.1 BeadChip (Illumina), which yielded 547,644 genotyped SNPs, 531,413 of which were autosomal. After filtering out the SNPs with MAF ≤ 0.02 (n=287,379), per SNP genotyping missingness rate ≥ 0.01 (n=16,140), Hardy-Weinberg Equilibrium (HWE) p-value < 1E-5 (n=0), and Mendelian transmission errors (n=120), there were 227,774 SNPs left in the final set for analysis. Linkage analysis was performed on multiple marker sets: (1) All SNPs (n=227,143); (2) a pruned SNP set to eliminate the SNPs in high linkage disequilibrium (LD) (r^2 = 0.2 (n=11,614) and r^2 = 0.01 (n=5,315)); and (3) a panel of SNPs solely based on

a spacing of one SNP for every 300 kb (n=4,665). An autosomal dominant inheritance model was applied for each of the marker sets in *MERLIN*[1] in a parametric model with the following parameters: disease allele frequency of 0.0001 and 0.001, full and incomplete (0.99) penetrance, and phenocopy rates of 0.0001 and 0.001. Frequency calculations required for this analysis were also done in multiple ways: (1) directly from the 13 subjects, assuming unrelatedness; (2) using the Maximum Likelihood Model (MLE) algorithm implemented in MERLIN, which adjusts for relatedness; (3) taken from the HapMap data (European ancestry (CEU), n=60), which required marker subsets that overlaps with the HapMap set. In total, we looked at 40 different analysis combinations (note that the different disease allele frequencies tested don't change the results; therefore only 0.0001 was used in the analyses).

Whole-Exome Sequencing and analysis

Genomic DNA only from the family member III-4 (marked with a blue rectangle in Figure 1A) was sonicated and subjected to exome capture using a NimbleGen SeqCap EZ Exome Enrichment Kit v3.0 (Roche, Madison, WI, USA) and to 100 base pair paired-end sequencing on an Illumina HiSeq 2000 at the University of Michigan Sequencing Core (UMSC). 100-bp paired-end reads were aligned to 1000 Genomes build 37 (GRCh37/hg19) with *BWA* version 0.5.9[2], and duplicates were removed using *PICARD* version 1.74.[3] This sample was one of 733 samples sequenced at the UMSC for which pooled variant calling was performed using GATK version 3.3[4] (n=11,595,739 variants). GATK's VQSR tool was utilized to filter for high quality variants (n=939,202), which were annotated with ANNOVAR (2016 Feb version)[5]. The functional variants (missense, nonsense, splice, stop-loss, frame-shift) (n=208,293) with a minor allele frequency (MAF) of 5% or less, or absent in the 1000 Genomes[6], ExAC databases[7], and the European subpopulation of the NHLBI Exome Sequencing Project (ESP)[8] were kept in (n=186,887). To remove any technical bias in the analysis pipeline, the variants were further subjected to a 10% frequency cut-off in the remaining 732

samples (individuals with no hematologic disease) in the pooled call (n=183,206). Rare variants present in this individual (genotypes = ./. (missing), 0/1 (heterozygous), 1/1 (alternative homozygous)) within the linkage intervals were considered. Because congenital thrombocytopenia is a very rare disease, variants with an allele frequency of >0.1% that are present in public databases or in the other 732 samples sequenced at UMSC were filtered. Variant(s) passing all these filtered were assessed for their pathogenicity using several tools including SIFT[9], PolyPhen2[10], and CONDEL[11]. The remaining variant(s) were confirmed by Sanger sequencing.

Imputation and Haplotype Analysis

In a previous report[12], the disease haplotype at the *ITGA2B* locus was defined using a set of 11 SNPs. In order to determine if the *ITGA2B* mutation occurs on a different haplotype in our family, our chromosome 17 genotype data were imputed and phased on the Michigan Imputation Server[13] using the 1000 Genome Phase 3 (EUR) panel as the reference. Eight of the eleven variants were present in the final set, which did not include the *ITGA2B* mutation.

In order to confirm that the candidate risk allele segregates with the disease haplotype within this family, we added in the *ITGA2B* mutation genotypes obtained from Sanger Sequencing to the list of eight variants above, and performed in silico within-family phasing using *MACH* package v1.0.16[14].

SUPPLEMENTAL TABLES

MS F1	CCATCCTGCACAACAAAGTG
MS R1	AAAAATAAAAGTTACCAACCCTACTCT
ANKRD26 5'UTR F	ACCGACATTTGTTTGACAG
ANKRD26 5'UTR R	GCAGCTTTGTGGATCTTG
ANKRD26 PCR F	ACAGTGGGCTTCATGTTCTG
ANKRD26 PCR R	AACAGCTCCAACCCAAAGC

Table S1. Primers used in this study.

	Affected/Non-affected	Platelet count (x10^3/µl)	MPV (femtoliters)
II-1	Affected	N/A	N/A
II-2*	Affected	22	10.2
II-3	Affected	N/A	N/A
II-4	Affected	N/A	N/A
III-2	Affected	N/A	N/A
III-3*	Non-affected	255	8.4
III-4*	Affected	43	N/A
III-5	Non-affected	N/A	N/A
III-6	Affected	N/A	N/A
III-7	Non-affected	N/A	N/A
IV-4	Non-affected	N/A	10.1
IV-5	Affected	N/A	9.5
IV-6*	Affected	43	8.5

Table S2. Platelet counts and Mean platelet volumes (MPV) from available family members. *Results obtained from a clinical lab. MPV was available for 2 additional family members who had platelet preps analyzed on our laboratory Advia120 hematology system (Siemens). Platelet count is not available for the latter individuals because platelet preps (not whole blood) were used for analysis. N/A: results not available.

REFERENCES FOR THE SUPPLEMENTAL METHODS

1. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics.* Jan 2002;30(1):97-101.
2. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://arxiv.org/abs/1303.3997.
3. PICARD version 1.74. http://broadinstitute.github.io/picard.
4. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research.* Sep 2010;20(9):1297-1303.
5. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research.* Sep 2010;38(16):e164.
6. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* Oct 01 2015;526(7571):68-74.
7. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* Aug 18 2016;536(7616):285-291.
8. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. http://evs.gs.washington.edu/EVS.
9. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome research.* May 2001;11(5):863-874.
10. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods.* Apr 2010;7(4):248-249.
11. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American journal of human genetics.* Apr 08 2011;88(4):440-449.
12. Kunishima S, Kashiwagi H, Otsu M, et al. Heterozygous ITGA2B R995W mutation inducing constitutive activation of the alphaIIbbeta3 receptor affects proplatelet formation and causes congenital macrothrombocytopenia. *Blood.* May 19 2011;117(20):5479-5484.
13. Michigan Imputation Server. https://imputationserver.sph.umich.edu/.
14. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology.* Dec 2010;34(8):816-834.