

Next-generation sequencing for the diagnosis of *MYH9*-RD: predicting pathogenic variants

Loredana Bury¹, Karyn Megy^{2,3}, Jonathan C Stephens^{2,3}, Luigi Grassi^{2,3}, Daniel Greene^{2,3,20}, Nick Gleadall^{2,3}, Karina Althaus^{4,5}, David Allsup⁶, Tadbir K Bariana^{2,3,7}, Mariana Bonduel⁸, Nora V Butta⁹, Peter Collins¹⁰, Nicola Curry¹¹, Sri VV Deevi^{2,3}, Kate Downes^{2,3}, Daniel Duarte^{2,3}, Kim Elliott BSc¹², Emanuela Falcinelli¹, Bruce Furie¹³, David Keeling¹⁴, Michele P Lambert^{15,16}, Rachel Linger^{2,3}, Sarah Mangles¹⁷, Rutendo Mapeta^{2,3}, Carolyn M Millar^{18,19}, Christopher Penkett^{2,3}, David J Perry²⁰, Kathleen E Stirrups^{2,3}, Ernest Turro^{2,3,21}, Sarah K Westbury²², John Wu²³, NIHR BioResource²⁷, Keith Gomez⁶, Kathleen Freson²⁴, Willem H Ouwehand^{2,3,25,26}, *Paolo Gresele, *Ilenia Simeoni^{2,3}

Affiliate Institutions

¹Department of Internal Medicine, Section of Internal and Cardiovascular Medicine, University of Perugia, Perugia, Italy. ²Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK. ³NIHR BioResource - Rare Diseases, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, UK. ⁴Institute for Immunology and Transfusion Medicine, Universitätsmedizin Greifswald Ernst-Moritz-Arndt University Greifswald, Greifswald, Germany. ⁵Center for Clinical Transfusion Medicine Tuebingen, Tuebingen, Germany. ⁶Hull York Medical School, York, UK. ⁷The Katharine Dormandy Haemophilia Centre and Thrombosis Unit, Royal Free London NHS Foundation Trust, London, UK. ⁸Hematology/Oncology Department. Hospital de Pediatría “Prof. Dr. Juan P. Garrahan”, Buenos Aires, Argentina ⁹Servicio de Hematología y Hemoterapia Hospital Universitario La Paz-IDIPaz, Madrid, Spain. ¹⁰Arthur Bloom Haemophilia Centre, Institute of Infection and Immunity, School of Medicine, Cardiff University, UK. ¹¹Oxford Radcliffe Hospitals NHS Trust. ¹²Oxford Haemophilia & Thrombosis Centre, Department of Haematology, Oxford University Hospitals NHS Trust, Churchill Hospital, Oxford and the NIHR BRC, Blood Theme, Oxford Centre for Haematology, Oxford, UK. ¹³Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA. ¹⁴Churchill Hospital, Oxford University Hospitals, UK. ¹⁵Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA. ¹⁶Division of Hematology, Children's Hospital of Philadelphia, Philadelphia, PA. ¹⁷Basingstoke and Hampshire NHS Foundation Trust. ¹⁸Hampshire Hospital NHS Foundation Trust, UK. ¹⁹Centre for Haematology, Hammersmith Campus, Imperial College Academic Health Sciences Centre, Imperial College London, London, UK. ²⁰Department of Haematology, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK. ²¹Medical Research Council

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/humu.23927.

Biostatistics Unit, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Cambridge, UK. ²²School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK. ²³B.C. Children's Hospital, Vancouver, Canada. ²⁴Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, KU Leuven, Leuven, Belgium. ²⁵NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK. ²⁶Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, ²⁷NIHR BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK.

Grant numbers:

WH is supported by: NIHR RG65966, NIHR BioResource - Rare Diseases, British Heart Foundation, European Commission, MRC, NHS Blood and Transplant and Wellcome Trust. PG is supported by: MIUR-FIRB (Protocol # RBFR12W5V5_004), Telethon Foundation Grant (GGP15063) and Fondazione Umberto Veronesi. C.M.M. is supported by the NIHR Imperial College Biomedical Research Centre. NVB is recipient of FIS-Fondos FEDER CP14/00024.

Corresponding Author:

Dr Ilenia Simeoni, Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, United Kingdom, Tel:+44 1223 588902, email: is250@cam.ac.uk

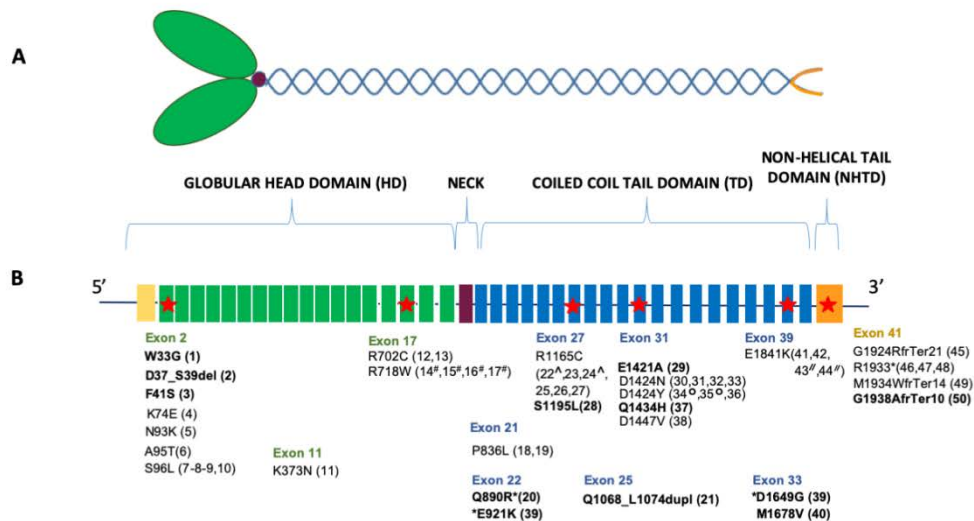
ABSTRACT

The heterogeneous manifestations of the *MYH9*-related disorder (MYH9-RD), characterized by macrothrombocytopenia, Döhle-like inclusion bodies in leukocytes, bleeding of various severity with in some cases ear, eye, kidney and liver involvement make the diagnosis for these patients still challenging in clinical practice. We collected phenotypic data and analysed the genetic variations in more than 3,000 patients with a bleeding or platelet disorder. Patients were enrolled in the BRIDGE-BPD and ThromboGenomics Projects and their samples processed by high throughput sequencing (HTS). We identified 50 patients with a rare variant in *MYH9*. All patients had macrothrombocytes and all except two had thrombocytopenia. Some degree of bleeding

diathesis was reported in 41 patients. Eleven patients presented hearing impairment, three renal failure and two elevated liver enzymes. Among the 28 rare variants identified in *MYH9*, 12 were novel. HTS was instrumental to diagnose 23 patients (46%). Our results confirm the clinical heterogeneity of *MYH9*-RD and show that, in the presence of an unclassified platelet disorder with macrothrombocytes, *MYH9*-RD should always be suspected. A HTS-based strategy is a reliable method to reach a conclusive diagnosis of *MYH9*-RD in clinical practice.

Graphical Abstract

MYH9-related disorder diagnosis is still challenging in clinical practice. We analysed the genetic variations in more than 3,000 patients with a bleeding or platelet disorder and identified 50 patients with a rare variant in MYH9. In the presence of an unclassified platelet disorder with macrothrombocytes, MYH9-RD should always be suspected.



Key words: MYH9-related disorders, variant classification, clinical diagnosis, high throughput next generation sequencing, genomics, ACMG Guidelines

INTRODUCTION

Non-muscle myosin heavy chain 9 related disorder (*MYH9*-RD) is a rare autosomal-dominant syndrome characterized by large/giant platelets and thrombocytopenia associated with the presence of Döhle-like inclusion bodies in neutrophils (Kunishima, *et al* 2003). Clinical manifestations include a mild to moderate bleeding tendency (Orsini, *et al* 2017) and the risk of developing progressive nephropathy, sensorineural deafness, pre-senile cataract or alteration of liver enzymes during infancy or adult life (Balduini, *et al* 2011, Pecci, *et al* 2012, Pecci, *et al* 2018). The disease is caused by heterozygous variants in *MYH9*, the gene coding for the heavy chain of non-muscle myosin of class IIA (NMMHC-IIA), a 1,960 amino acid residue protein involved in platelet cytoskeletal contraction, granule secretion, and in the Rho GTPases and Ca²⁺/calmodulin signaling pathways (Vicente-Manzanares, *et al* 2009). *MYH9* is located on chromosome 22q12-13 and is composed of 41 exons. The coding region from exons 2 to 19 forms the globular head domain (HD), exon 20 for the neck region, and exons 21 to 40 for the coiled-coil tail domain (TD). The final 34 amino acid residues of the C-terminal non-helical tail domain (NHTD) are encoded by exon 41.

101 *MYH9* variants are listed in the Human Gene Mutation Database (HGMD, public version, as of July 2019, Stenson, *et al* 2017): 72 missense/nonsense, 4 splicing substitutions, 17 small deletions/insertions, 8 large deletions/insertions. One very large deletion of 1220 bp involving exon 26 and some cases of somatic or germinal mosaicism have also been described (Gresele, *et al* 2013, Kunishima, *et al* 2005, Kunishima, *et al* 2009).

Genotype-phenotype correlation studies in *MYH9*-RD patients have reported that variants

in the HD are associated with more severe thrombocytopenia and a higher frequency and/or a more rapid progression of nephropathy and deafness than variants in the TD, with the amino acid substitution p.Arg702Cys resulting in the most severe phenotype reported to date (Pecci, *et al* 2014, Pecci, *et al* 2008b, Saposnik, *et al* 2014). However, some exceptions exist: the p.Asp1424His variant which lies in the TD, is also associated with a high risk of developing syndromic manifestations. Moreover, patients carrying variants at the interface between the SH3-like motif and the motor domain (MD) of the HD (SH3/MD interface), present a mild clinical phenotype consisting of mild macrothrombocytopenia and delayed risk of sensorineural deafness (Pecci, *et al* 2014).

The diagnosis of *MYH9*-RD requires skilled laboratory investigations, including the correct assessment of the degree of thrombocytopenia, made difficult by the abnormal size of platelets, the identification of macrothrombocytes, and the determination of the presence of Döhle-like inclusion bodies in neutrophils on a blood smear (Balduini, *et al* 2003). The latter test is performed by May-Grünwald-Giemsa (MGG) staining or through the identification of NMMHC-IIA aggregates by immunofluorescence (Kunishima, *et al* 2003, Pecci, *et al* 2008b), a test which is not available in most of the hematology diagnostic laboratories despite its high sensitivity. Moreover, high variability in the syndromic manifestations can complicate the interpretation of the clinical presentation. The identification of the causal *MYH9* variant in a patient is key to reach a conclusive diagnosis, predict the course of extra-hematological symptoms and consequently implement a personalized clinical monitoring and therapeutic approach (Pecci, *et al* 2008a, Pecci, *et al* 2010a). HTS techniques represent a comprehensive and cost-effective strategy for a precise diagnosis of inherited bleeding, thrombotic and platelet disorders (BPDs) (Simeoni, *et al* 2016, Zhang, *et al* 2016). The efficacy of HTS in patients with

uncharacterized macrothrombocytopenia has been recently demonstrated (Rabbolini, *et al* 2017). Here, we report the patients with a rare *MYH9* variant discovered after genome sequencing of 1,481 subjects enrolled in the BRIDGE-BPD study and 1,550 patients enrolled in the clinical diagnostic ThromboGenomics study (Simeoni, *et al* 2016). We identified 28 causal rare *MYH9* variants in 50 patients (44 index cases), 20 with a known diagnosis of *MYH9*-RD (based on the presence of macrothrombocytopenia, Döhle-like bodies and an extra-haematological phenotype in some of them) with no genetic confirmation, 11 with an already suspected but not confirmed *MYH9*-RD and 19 in whom *MYH9*-RD was not previously suspected despite an expert evaluation of their clinical and laboratory data. We describe the 28 *MYH9* variants identified, 12 of which are novel and classify the variants for pathogenicity and contribution to phenotype. We also describe the phenotypic profiles of this *MYH9*-RD cohort adding new insight into genotype-phenotype correlations and expanding the knowledge on this rare inherited platelet disorder.

METHODS

Patient cohort

Patients gave their written informed consent and were enrolled through two main projects: the NIHR BioResource - Rare Diseases study (specifically, the BRIDGE-BPD project) and the clinical diagnostic ThromboGenomics study. The BRIDGE-BPD project includes patients with rare inherited BPDs of unknown etiology who were screened mainly by genome sequencing and a small subset by exome sequencing. DNA samples from BPD patients with clinical and laboratory phenotypes indicative of a particular molecular etiology were sequenced using the ThromboGenomics HTS test. Inclusion

criteria have been previously described (Simeoni, *et al* 2016, Westbury, *et al* 2015). Ethics authorities and approval numbers are provided in **Supp. Table S1**.

Clinical and laboratory phenotypes

Clinical and laboratory phenotypes were submitted by the referring clinicians as Human Phenotype Ontology (HPO) terms, as previously described (Westbury, *et al* 2015). The severity of bleeding was coded as numerical scores using the MCMDM-1 VWD Bleeding Assessment Tool (http://www1.wfh.org/docs/en/Resources/Assessment_Tools_MCMDM-1VWD.pdf).

Centralized analysis of blood smears for the identification of Döhle-like inclusion bodies in patients with previously not performed or negative detection of Döhle-like inclusion bodies was performed by two independent centres. Blood films obtained from patients and from healthy controls were randomly analysed by two operators blindly. Inclusions in neutrophils were classified as type I, II and III based on their size, shape and pattern of distribution (Kunishima, *et al* 2003, Pecci, *et al* 2008b). Information on hearing impairment, renal and liver dysfunctions were also collected.

Variant prioritisation and assessment

Sequencing results were processed by using a single analysis approach and filtering as previously described (Greene, *et al* 2016, Simeoni, *et al* 2016) (see also Supporting Information). An average of 5 variants per patient remained after the bioinformatics analysis and each of these variants was assessed following the ACMG Guidelines (Richards, *et al* 2015) by a Multi-Disciplinary Team (MDT) composed of clinicians, clinical geneticists, bioinformaticians and clinical scientists. The Congenica™ software (Congenica Ltd., Hinxton, UK) was used to visualise the data and assign pathogenicity

and contribution to phenotype to each variant based on the clinical picture, predicted consequence for the protein, presence in the Human Gene Mutation Database (HGMD (Stenson, *et al* 2017)) and allele frequency in control datasets as the Exome Aggregation Consortium (ExAC) (Karczewski, *et al* 2017) and the genome Aggregation Database (gnomAD) (Lek, *et al* 2016)). The MDT also evaluated the Minor Allele Frequency (MAF) of the variants found in patients with aggregate data from more than 13,000 individuals enrolled in other non-BPD BRIDGE projects. The LRG transcript LRG_567t1 (NM_002473.5, ENST00000216181.10) was used as reference sequence. Variants and their pathogenicity have been deposited in ClinVar under accession numbers SCV000891130 to SCV000891157. They are accessible by searching for the accession number (e.g. SCV000891130) or with the keywords “MYH9 AND NIHR AND BioResource”(<https://www.ncbi.nlm.nih.gov/clinvar/?term=MYH9+AND+NIHR+BioResource>).

RESULTS

Novel *MYH9* variants

3,031 patients were enrolled in the BRIDGE-BPD and ThromboGenomics studies and screened for rare variants in the *MYH9* gene. We found 74 individuals with a variant in the *MYH9* gene, however only 50 patients were considered for this study. The remaining 24 were excluded for the following reasons: 1) the *MYH9* variant was also present in other non-BPD patients; 2) the platelet disorder and/or phenotype was not compatible with *MYH9*-RD (e.g. thrombocytosis, no large platelets); 3) the phenotype was explained by the presence of a causal variant in another gene; 4) the *MYH9* variant was also identified in an unaffected family member. All the patients excluded from this study and

the reasons for their exclusion are shown in detail in **Supp. Table S2**.

In the 50 patients analysed, of whom 44 are index cases, we found 28 *MYH9* variants, namely 21 missense, three frameshifts, two stop gains, one in-frame deletion and one in-frame insertion. The variants identified were positioned in 11 of the 41 exons of the *MYH9* gene (**Figure 1**). Of the 28 variants, 12 are novel as absent from the HGMD database (public version, as of July 2019, Stenson, *et al* 2017), the literature and all other publicly accessible *MYH9*-RD databases at the time of the analysis (**Table 1**). Of the novel variants, three affect the SH3/MD interface of the globular MYH9 head, including a new c.97T>G transversion in exon 2, leading to p.Trp33Gly amino acid change, an in-frame deletion p.Asp37_Ser39del and one missense variant p.Phe41Ser caused by the c.353T>C transition. In silico protein modelling predicts that these three variants may disturb the hydrophobicity of the SH3/MD interface (**Supp. Figure S1**). We also found eight novel variants localized in the coiled-coil domain. These include one missense variant p.Glu921Lys and a nonsense variant p.Gln890Arg*, leading to a premature stop codon causing the formation of a shorter MYH9 protein of 890 amino acids, both in exon 22; one in-frame insertion p.Gln1068_Leu1074dup in exon 25 and five missense variants, p.Ser1195Leu in exon 27, p.Glu1421Ala and p.Gln1434His in exon 31, p.Asp1649Gly and p.Met1678Val in exon 33. In the non-helical tail domain of the protein, we found one further novel variant and a frameshift leading to a premature stop in the protein, p.Gly1938Alafs*10. The read coverage of whole genome sequencing (WGS) and targeted sequencing results for the 12 novel variants (in 11 patients) is shown in **Supp. Figure S2**.

Variant pathogenicity and contribution to phenotype

The MDT assigned pathogenicity and contribution to phenotype to each variant according to the clinical features of each patient following the ACMG Standards and Guidelines (Richards, *et al* 2015) (shown in **Table 1**). The choice of the transcript for variant reporting was based on transcript and protein lengths, and expression in blood cells according to the Blueprint data (Javierre, *et al* 2016). Eleven *MYH9* transcripts are expressed in the different blood cells, but only three of them are protein coding. ENST00000216181 (NM_002473, LRG_567t1) is the longest transcripts (7501 base pairs (bp), corresponding to a protein with the expected 1960 amino acids (aa) length), with an equivalent in the RefSeq database (NM_0024736). This is the most expressed transcript in platelets, while its expression is lower in neutrophils and megakaryocytes and much lower in erythroblasts. Of the two remaining protein coding transcripts, ENST00000401701 is much shorter (789 bp, 218aa) and markedly less expressed; ENST00000456729 is also shorter (449 bp, 103aa) and absent ($\log_2(\text{FPKM}) < 1$) in blood cells (**Supp. Figure S3**). For these reasons, ENST00000216181 (NM_002473) was used for variant reporting. This is also the transcript subsequently selected by LRG (LRG_567t1).

The MDT classified the novel variants as follows: the three variants in exon 2 found in patients 1, 2 and 3, as likely pathogenic (in patient 1) and variant of uncertain significance (VUS) (in patients 2 and 3) with full contribution to phenotype. The stop gain p.Gln890Arg*, in patient 20, was classified as VUS with full contribution to the phenotype. Based on the high impact of the variant on the MYH9 protein causing a premature stop, the variant might be considered to be likely pathogenic. However, we have not been able to perform any functional tests due to the difficulties of recalling the

This article is protected by copyright. All rights reserved.

86 year old patient, thus we remained conservative and classified this variant as VUS. The in-frame insertion Gln1068_Leu1074dup and the frameshift variant, p.Gly1938Alafs*10, were considered VUS and likely pathogenic with full contribution to the phenotype, respectively. The six novel missense variants (present in patients 28-29-37-39-40) identified in the coiled-coil domain of the MYH9 protein were classified as VUSs with full contribution to phenotype. The referring clinicians of these five patients with a VUS variant, were re-contacted to arrange co-segregation studies. Pedigree analysis was possible only for two of these patients. This has confirmed, in patient 28, the absence of the variant in the non-affected mother and in patient 39 the presence of the same variant in the daughter affected with mild thrombocytopenia. The pathogenicity and contribution to phenotype assigned to the remaining non-novel variants are listed in **Table 1**.

For all the variants identified in this study we investigated the evolutionary conservation in the MYH9 protein domains. We found that all pathogenic, likely pathogenic and VUS variants affect highly conserved amino acid residues providing further confidence that the variants identified have an impact on MYH9 protein function and consequently on the patients' phenotypes (**Figure 2**).

Immunofluorescence analysis

At enrolment, the presence of Döhle-like inclusion bodies was reported only in 21 (42%) of the 50 patients analysed. Given that the Döhle-like bodies are reported to be invariably present in *MYH9*-RD patients, at least when analysed by immunofluorescence, we recalled the remaining 29 patients, initially labelled as Döhle-like bodies negative, for a centralized blood smear analysis (**Supp. Table S3**). Of these 29 patients, we obtained a

fresh blood smear from 18 patients. An abnormal neutrophils MYH9 distribution was found in all 18 (100%) patients when analysed by immunofluorescence and in 11 patients (61%) when analysed by the MGG staining, in accordance with previous results (Balduini, *et al* 2011).

Immunofluorescence was performed for all the 11 patients with novel variants, except for those in which the Döhle-like bodies were previously identified (patients 2, 3, 21, 50) and in patients 37 and 39 not available for further analysis. In the remaining five patients, we obtained the following results: in patients 1 and 20, neutrophils had circular to oval shaped cytoplasmic spots that have been classified as type II inclusions. In patients 28 and 40, neutrophils had speckled inclusions and in patient 29 inclusions resembled small dots scattered throughout the cytoplasm inclusions were classified as type III.

In conclusion, 33% of the all the 18 patients re-analyzed had type II inclusions and 67% type III myosin IIA inclusions (**Supp. Table S3**). An example of the altered NMMHC-IIA distribution in neutrophils in patients with a pathogenic variant and VUS is shown in **Figure 3**.

Phenotypic description of the *MYH9*-RD cohort and genotype-phenotype correlation

Our cohort includes 21 males and 29 females from 44 unrelated pedigrees. The median age at diagnosis was 20 years (range 1-76). Over a third (19) of the patients were enrolled with a diagnosis of ‘unclassified platelet disorder’ while the remaining (31) had a suspected (11) or known (20) but not confirmed *MYH9*-RD, based on family history, presence of large/giant platelets, thrombocytopenia, presence of Döhle-like bodies and/or extra-hematological symptoms.

Macrothrombocytes were present in all patients, while thrombocytopenia, with various degrees of severity, was present in all but two patients (17 and 40). The median platelet count was $54 \times 10^9/L$ ($8-220 \times 10^9/L$) from automated measurements and $48.5 \times 10^9/L$ by microscopic assessment, although the latter was only available for 8 patients (**Supp. Table S4**). The mean platelet volume (MPV) values are shown in **Supp. Table S5**. Three cases had a normal MPV when measured by automatic blood cell counting, however macro-thrombocytes were noticed upon examination of their blood smears (Greinacher, *et al* 2017, Kunishima, *et al* 2001a). Haematological and non haematological symptoms are shown in **Figure 4**. Bleeding symptoms, mostly mild mucocutaneous bleeding, were reported in 82% of the patients (41 out of 50). Bleeding scores, calculated by the MCMDM-1 VWD Bleeding Assessment Tool, are shown in **Supp. Table S6**. Of the 29 females enrolled, 11 (38%) had menorrhagia, one of the most common symptoms reported by women with congenital platelet disorders.

Genotype-phenotype correlations were analysed by plotting the seven HPO terms representing the major *MYH9*-RD clinical features against the exons in which both previously described and novel *MYH9* variants were found (**Supp. Figure S4**). We first investigated the correlation between the position of variants in the MYH9 protein and the degree of thrombocytopenia, by dividing patients in two groups according to platelet count below (severe/moderate) or above $50 \times 10^9/L$ (mild). We found that 39% of the patients with severe thrombocytopenia have a variant affecting exons in the HD and 61% of the individuals had instead a variant in the coiled coil domain. Genotype-phenotype correlations were also studied for the extra-hematological manifestations of *MYH9*-RD. Details on how patients were screened for hearing impairment, renal dysfunction and liver enzymes alteration are summarised in **Supp. Table S7**. Nephropathy was reported

in patients 8 and 17, who carry p.Ser96Leu and p.Arg718Trp variants, respectively. Patient 8 also has hearing impairment. However, five other patients (10% of this cohort), two unrelated individuals carrying the same variant, p.Ser96Leu (patients 7 and 9), and three pedigree members of case 17, carrying the p.Arg718Trp, did not present any of these non-hematological features. Hearing impairment was present in 22% of the patients: 8% with variants involving the HD, as expected, and 14% involving the coiled-coil and the NHT domains (Balduini, *et al* 2012, Pecci, *et al* 2014, Pecci, *et al* 2008b). Variants observed in patients with bleeding symptoms were randomly distributed across the MYH9 domains, confirming a lack of genotype-phenotype correlation for the bleeding phenotype (Pecci, *et al* 2014, Saposnik, *et al* 2014). Moreover, no correlation was found even between platelet count and bleeding tendency (**Supp. Figure S5**).

DISCUSSION

The *MYH9*-RD, although rare, is considered the most frequent inherited macrothrombocytopenia. In Italy, where a large active patient registry was established in 2006, *MYH9*-RD has an estimated frequency of 1 in 312,000, representing 12% of the inherited thrombocytopenias (Balduini, *et al* 2012, Pecci, *et al* 2014). The complexity and variability of patients' phenotypes can make the diagnosis of *MYH9*-RD rather challenging, even by skilled clinicians at specialist centres. As a consequence, a significant number of patients with *MYH9*-RD are initially misdiagnosed as immune thrombocytopenic purpura (ITP), and thus subjected to ineffective and potentially harmful treatments, or classified as inherited platelet disorder of unknown origin. In this scenario, HTS techniques may represent a reliable method for the diagnosis of *MYH9*-RD.

The present study represents the first systematic analysis of *MYH9* variants by HTS analysis in a large cohort of patients and controls enrolled from over 100 centres worldwide. Here, we report 50 *MYH9*-RD patients with 28 rare variants in *MYH9* found in a group of 3,031 patients (of whom 764 were classified as having thrombocytopenia) and over 13,000 controls.

In agreement with previous studies, 75% of the variants identified (21 out of 28) are annotated in the most commonly affected *MYH9* exons (Pecci, *et al* 2014, Pecci, *et al* 2008b, Saposnik, *et al* 2014).

We identified 12 novel variants affecting *MYH9* in highly conserved amino acid residues, including eight missense variants (one in a previously described amino acid residue but with a different nucleotide change (Jang, *et al* 2012, Kahr, *et al* 2009), one inframe deletion, one stop gain, one inframe insertion and one frameshift. Patients carrying the eight missense variants did not show extra-hematological symptoms, had a platelet count ranging from 15 to $96 \times 10^9/L$ and three of them (1, 20, 21) had a history of excessive bleeding. The in frame deletion, Asp37_Ser39del, was found in a 31-year-old man with no bleeding symptoms, moderate thrombocytopenia and no extra-hematological symptoms. The new pathogenic stop codon in the coiled-coil domain, Gln890Arg*, leading to a large deletion of 1070 amino acids in the MYH9 protein was found in a 86-year-old man who was originally diagnosed as an “unclassified platelet disorder”, with mild thrombocytopenia ($88 \times 10^9/L$), no extra-hematological manifestations and a pathologic bleeding score due to major bleeding after surgery. We have not been able to test the presence of the truncated MYH9 protein in this patient’s cells, and a non classical distribution of NMMHC-IIA, with just small punctuate clusters (Althaus, *et al* 2009) was observed by IF- and MGG-staining in granulocytes (**Supp. Figure S6**). We also report

This article is protected by copyright. All rights reserved.

the first in-frame insertion, Gln1068_Leu1074dup. The same amino acids were previously described to be involved in an in-frame deletion in two patients (Ishida, *et al* 2013, Saposnik, *et al* 2014). This was found in a young girl who presented with moderate thrombocytopenia ($70 \times 10^9/L$), large platelets and moderate/severe bleeding (Bleeding score 7), similarly to the previously published cases, but with no current extra-haematological symptoms. In contrast, the two patients previously described present several extra-hematological features like hearing loss since childhood, congenital cataracts and mild proteinuria in a 59 year-old woman (Saposnik, *et al* 2014) and end-stage renal disease and bilateral hearing loss in a 27 year-old woman (Ishida, *et al* 2013). The novel frameshift, Gly1938Alafs*10, located in a known mutational hot spot, was found in a patient with severe thrombocytopenia ($16 \times 10^9/L$), large platelets and mild bleeding.

All the variants were discussed in MDT meetings and pathogenicity and contribution to phenotype assigned according to the ACMG Guidelines. The novel variants were labelled as pathogenic or likely pathogenic when supported by strong evidences, including the impact of the variant on the protein, the presence of strong *MYH9*-RD phenotype and/or another *MYH9*-RD feature and, when possible, by pedigree analysis. In all the remaining cases, the novel variants were classified as VUS. Previously reported variants were classified mainly as pathogenic. One variant, initially classified as VUS, was re-classified as benign (patient 71 in **Supp. Table S2**), one variant initially VUS to a likely pathogenic (in patients 18 and 19) and four initially likely pathogenic as pathogenic (in patients 5, 6, 12, and 13).

We have previously shown that HTS technologies can successfully be applied to diagnose inherited bleeding, platelet and thrombotic disorders (Simeoni, *et al* 2016). In This article is protected by copyright. All rights reserved.

the present study, a total of 23 patients, 12/19 initially coded as “unclassified platelet disorder” and 11/11 for whom only a suspicion of *MYH9*-RD was put forward with no conclusive diagnosis, received a molecular diagnosis of *MYH9*-RD because a likely pathogenic or pathogenic variant in *MYH9* was found.

Our data confirm that the presence of Döhle-like bodies is an invariable feature of *MYH9*-RD. Indeed, Döhle-like bodies were found in all 18 patients that were re-analysed by immunofluorescence (in 9 patients with a pathogenic, in 5 patients with a likely pathogenic and in 4 patients with a VUS variant) and in 11/18 by MGG (in 6 patients with a pathogenic, in 4 patients with a likely pathogenic and in a single patient with a VUS variant) bringing the percentage of patients positive for Döhle-like bodies inclusion and with a variant in the *MYH9* to 100%. Interestingly, we noted that 52% (11 out of the 21) of the patients in which Döhle-like bodies were reported at enrollment by MGG staining had a variant in the tail or in the S2 fragment, which are the regions of the *MYH9* protein that, when mutated, are associated with the presence of type I inclusions, the most visible at MGG staining and more easy to identify. Our attempt to identify genotype/phenotype correlations in this cohort of patients generally confirms previously published data (Pecci, et al 2014, Pecci, et al 2008b, Saposnik, et al 2014), although with some exceptions. Our study confirms that variants in the HD are frequently associated with more severe thrombocytopenia and higher risk of other organ involvement contrarily to variants in the TD. In fact, two patients (cases 8 and 10) with severe/moderate thrombocytopenia, kidney disease and hearing impairment had variants in the HD (exon 2 and 17 respectively), while most of the cases with variants in the TD showed mild thrombocytopenia and no extra-haematological organ involvement. Regarding the exceptions, three patients in our series carrying variants in the HD (patients 11, 12 and

13) had only mild thrombocytopenia, very mild or absent bleeding symptoms, and no other extra haematological manifestations, except hearing loss in patient 12. Moreover, four patients in our cohort (cases 28, 30, 33 and 50) with variants in the TD had severe thrombocytopenia ($\leq 20 \times 10^9/L$). Also patients carrying the same variant (p.Ser96Leu and p.Arg718Trp) did not share the same extra-hematological phenotype, showing that the risk of developing deafness or renal failure may be variable among patients carrying the same variant. The risk to develop these phenotypes is known to increase with age. In our cohort we did not observe a clear age-dependency for the development of an extra-hematological phenotype, however, it must be considered that our patients are mostly relatively young. For instance, patients 12 and 13 with the p.Arg702Cys variant are 7 and 11 years old, respectively. Their health care management will take in consideration the high risk of developing extra-hematologic features by age of 40 due to this known pathogenic variant.

Our cohort confirms that the presence of macrothrombocytes is an invariable feature of this disorder while thrombocytopenia, although highly frequent, may be absent (Pecci, *et al* 2014, Saposnik, *et al* 2014). This is the case for patients 17 and 40, with a platelet count of 187 and 220 $\times 10^9/L$, respectively. Patient 17 shares the same variant with her sister (patient 14), her nephew (patient 15) and her mother (patient 16) who have all three mild thrombocytopenia. Unfortunately, we could not investigate further patient 40. Thus, this cohort confirms that a wide platelet count variability is a feature of *MYH9*-RD (Balduini, *et al* 2011). In conclusion, our study expands the number of variants causing *MYH9*-RD, highlights the heterogeneity of the *MYH9*-RD phenotypes and, despite supporting previous correlation studies, shows that exceptions exist in genotype/phenotype correlations. The application of HTS-based strategies revealed to be

a reliable and fast method to reach a conclusive diagnosis of *MYH9*-RD and exclude other thrombocytopenias with potential susceptibility to malignancies and may represent the first line of investigation for this disorder, even after preliminary expert evaluation.

Acknowledgements

We are grateful to patients who accepted to participate in these projects. We would like to thank all the members of the BRIDGE-BPD and ThromboGenomics Consortia for their valuable contributions. In particular, Veerle Labarque from Leuven for recruiting patients for the ThromboGenomics panel, and Dr Romina Petersen for helping with data analysis. We thank Anna Maria Mezzasoma (University of Perugia, Italy) for the help with MGG-staining and analysis of blood smears. We gratefully acknowledge the participation of all NIHR BioResource volunteers, and thank the NIHR BioResource centre and staff for their contribution. We thank the National Institute for Health Research and NHS Blood and Transplant.

Authorship Contribution

L.B. wrote the paper, performed immunofluorescence and provided samples; K.M. reviewed the paper, managed data, set up and oversaw MDT meetings; J.C.S. provided genotyping results and processed samples; L.G. provided BLUEPRINT data; D.G. performed statistical analysis and provided HPO tables; N.G. performed variant conservation analysis; K.A. and L.B. performed immunofluorescence; D.A., T.K.B, M.B., N.V.B., P.C., N.C., K.E., E.F, B.F., D.K., C.M.M., M.P.L, S.M., D.J.P., S.S, S.K.W., K.F., K.G., L.B. and P.G. provided samples and clinical data; S.V.V.D supported the bioinformatics analysis, K.D. managed ThromboGenomics; R.M. performed DNA extraction; D.D processed ThromboGenomics samples, S.P. coordinated the NIHR

This article is protected by copyright. All rights reserved.

BioResource Rare Diseases BPD project; C.P. managed and supervised the WGS pipeline; K.S. managed the NIHR BioResource sequencing pipeline; E.T. managed BRIDGE-BPD data analysis; K.F. and K.G. chaired the MDT meetings; P.G., K.F. and W.H.O. reviewed the paper; I.S. wrote the paper, managed and processed ThromboGenomics samples.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Table 1.

List of variants with nucleotide change position, protein alteration and allele frequency. Cases labelled from 1 to 50. Ethnicity indicated when known. MTD outcome is shown per each variant as pathogenicity and contribution to phenotype. ACMG= American College of Medical Genetics; CADD=Combined Annotation Dependent Depletion; MDT= Multi-Disciplinary Team; EUR= European; SAS= South Asian; AFR= African; VUS= Variant of Uncertain Significance; ExAC= Exome Aggregation Consortium; gnomAD = Genome Aggregation Database.

Figures

Figure 1. Schematic representation of the heavy chain A of non-muscle myosin class IIA (NMMHC-IIA) and variants position.

A. Schematic representation of NMMHC-IIA protein. Non-muscle myosin II A shows a hexameric structure consisting of two heavy chains, namely NMMHC-IIA, and two pairs of light chains. Each heavy chain includes a N-terminal globular head domain (HD), a neck region which binds the light chains, and a C-terminal α -helical coiled-coil tail

domain (TD), which ends with a non-helical tail domain (NHTD) involved in the subcellular localization of the protein. The HD includes four subdomains: the N-terminal SRC-Homology 3 like motif (SH3), the upper and lower 50kDa subdomains, that together form the motor domain (MD), and the converter subdomain.

In green the globular HD, in violet the neck domain and in blue the coiled coil TD with the NHTD at the 3'UTR in orange. **B. Affected exons and variants identified.** The most affected exons are highlighted with red stars. The novel variants are shown in bold and the number in brackets is the patient ID number. Colors reflect protein domains. All variants described were confirmed by Sanger sequencing. The * indicates the two mutations identified in the same patient (39). #*,^,°,// represent members of the same family.

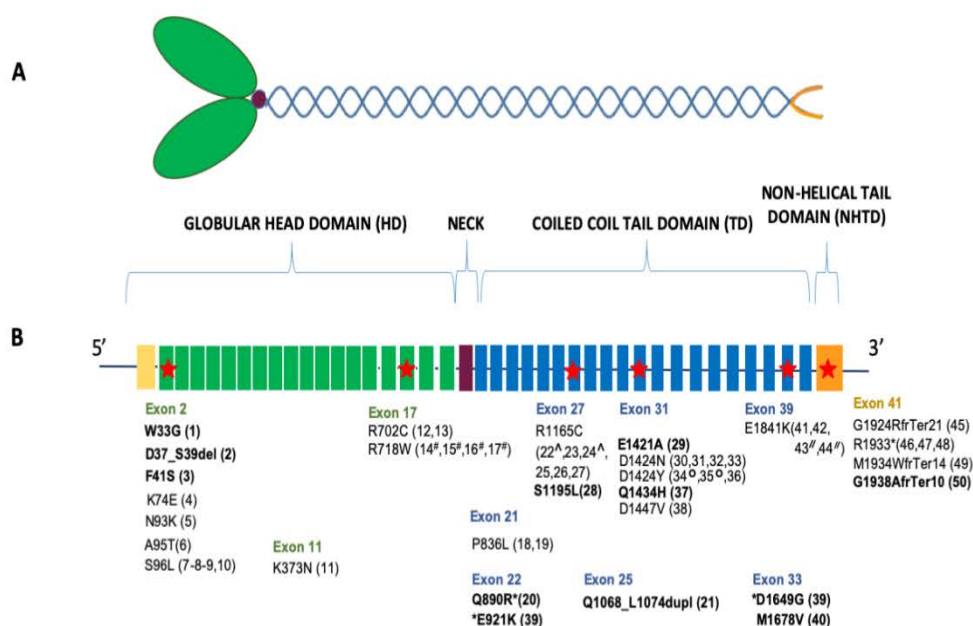


Figure 2. Evolutionary conservation variant analysis. From the outer to the inner circle. **MYH9 protein domains:** in green the N-terminal globular head domain (HD), in purple the neck domain, in blue the C-terminal α -helical coiled-coil tail domain (TD) and in orange the 3'UTR. **Evolutionary conserved regions in the MYH9 protein** in grey. All the pathogenic, likely pathogenic and VUS variants affect highly conserved amino acid residues. **Variant minor allele frequency (MAF) in gnomAD database** is represented by green bars. Smaller is the green bar lower is the allele frequency. **Variants present in ClinVar and LOVD** are represented by blue bars. The height of each blue bar represents the number of patients previously described with the same variant. **Variants in this cohort previously seen in the literature** include 'clearly pathogenic' and 'likely pathogenic' variants, in red and orange, respectively. **Novel variants in this cohort** include 'likely pathogenic' variants and variants of uncertain significance (VUS) in orange and black, respectively. UTR= Untranslated region.

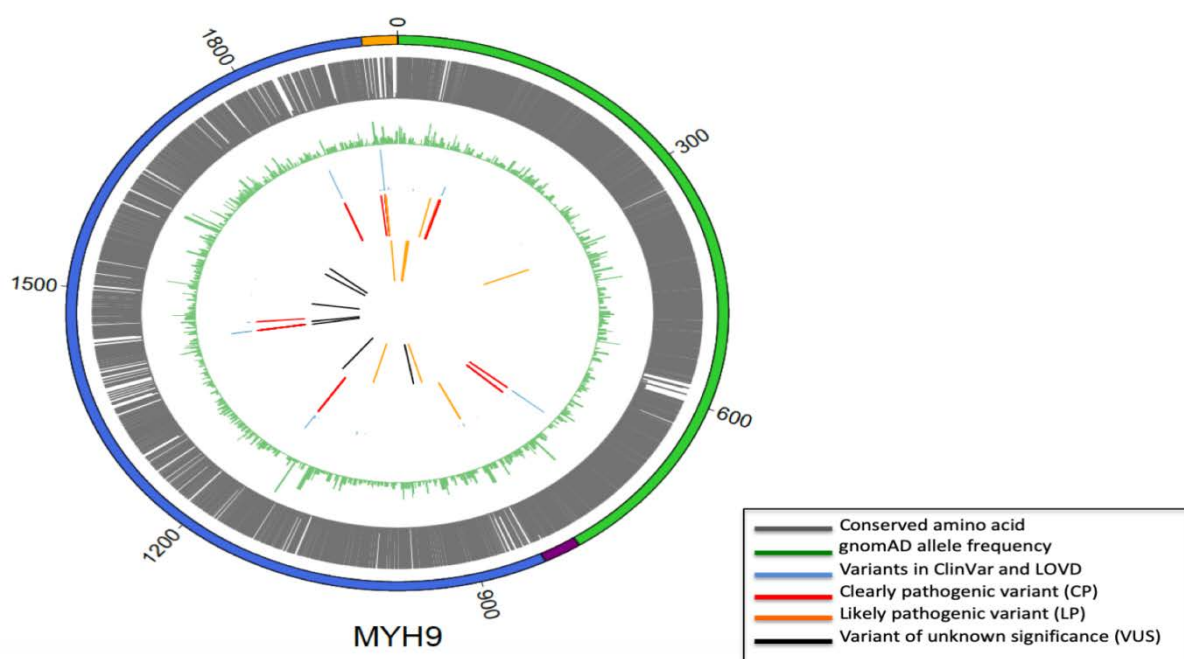


Figure 3. Döhle-like inclusion bodies localization by NMMHC-IIA immunofluorescence or MGG staining. Light microscopy and immunofluorescence analyses of granulocytes in a healthy control (control), in patients (32 for immunofluorescence, 43 for light microscopy) with a pathogenic variant (pathogenic) and in three patients (28, 29 and 40) with a variant of uncertain significance (VUS). The analysis was performed by two independent centres: Panels I-V show results obtained by centre 1; Panels VI-X show results obtained by centre 2. Both centres used rabbit anti-human NMMHCIIA Ab followed by Alexa-Fluor 488-conjugated secondary antibody. Results between the two centers were highly comparable. The patient's sample in which a pathogenic variant was identified shows circular to oval shaped cytoplasmic punctuate spots, classified as type II inclusions (panels II and VII). Patients' samples in which VUSs were identified show a speckled staining (panels III and VIII and panels V and X, respectively), and many small dots scattered throughout the cytoplasm (panels IV and IX) classified as type III inclusions.

Panels XI-XV show May Grunwald-Giemsa staining. Panels XII and XV show the presence of Döhle-like bodies (arrowhead) in patients' samples with a pathogenic variant (XII) and a VUS (XV). NMMHCIIA= non-muscle myosin of class IIA. VUS= Variant of Uncertain Significance.

REFERENCES

- Althaus, K., Greinacher, A. (2009) MYH9-related platelet disorders. *Semin Thromb Hemost*, **35**, 189-203.
- Arrondel, C., Vodovar, N., Knebelmann, B., Grunfeld, J.P., Gubler, M.C., Antignac, C. & Heidet, L. (2002) Expression of the nonmuscle myosin heavy chain IIA in the human kidney and screening for MYH9 mutations in Epstein and Fechtner syndromes. *J Am Soc Nephrol*, **13**, 65-74.
- Balduini, C.L., Cattaneo, M., Fabris, F., Gresele, P., Iolascon, A., Pulcinelli, F.M., Savoia, A. & Italian Gruppo di Studio delle, P. (2003) Inherited thrombocytopenias: a proposed diagnostic algorithm from the Italian Gruppo di Studio delle Piastrine. *Haematologica*, **88**, 582-592.
- Balduini, C.L., Pecci, A. & Noris, P. (2012) Inherited thrombocytopenias: the evolving spectrum. *Hamostaseologie*, **32**, 259-270.
- Balduini, C.L., Pecci, A. & Savoia, A. (2011) Recent advances in the understanding and management of MYH9-related inherited thrombocytopenias. *Br J Haematol*, **154**, 161-174.
- Greene, D., BioResource, N., Richardson, S. & Turro, E. (2016) Phenotype Similarity Regression for Identifying the Genetic Determinants of Rare Diseases. *Am J Hum Genet*, **98**, 490-499.
- Greinacher, A., Pecci, A., Kunishima, S., Althaus, K., Nurden, P., Balduini, C.L. & Bakchoul, T. (2017) Diagnosis of inherited platelet disorders on a blood smear: a tool to facilitate worldwide diagnosis of platelet disorders. *J Thromb Haemost*, **15**, 1511-1521.
- Gresele, P., De Rocco, D., Bury, L., Fierro, T., Mezzasoma, A.M., Pecci, A. & Savoia, A. (2013) Apparent genotype-phenotype mismatch in a patient with MYH9-related disease: when the exception proves the rule. *Thromb Haemost*, **110**, 618-620.
- Heath, K.E., Campos-Barros, A., Toren, A., Rozenfeld-Granot, G., Carlsson, L.E., Savage, J., Denison, J.C., Gregory, M.C., White, J.G., Barker, D.F., Greinacher, A., Epstein, C.J., Glucksman, M.J. & Martignetti, J.A. (2001) Nonmuscle myosin heavy chain IIA mutations define a spectrum of autosomal dominant macrothrombocytopenias: May-Hegglin anomaly and Fechtner, Sebastian, Epstein, and Alport-like syndromes. *Am J Hum Genet*, **69**, 1033-1045.
- Ishida, M., Mori, Y., Ota, N., Inaba, T. & Kunishima, S. (2013) Association of a novel in-frame deletion mutation of the MYH9 gene with end-stage renal failure: case report and review of the literature. *Clin Nephrol*, **80**, 218-222.
- Jang, M.J., Park, H.J., Chong, S.Y., Huh, J.Y., Kim, I.H., Jang, J.H., Kim, H.J. & Oh, D. (2012) A Trp33Arg mutation at exon 1 of the MYH9 gene in a Korean patient with May-Hegglin anomaly. *Yonsei Med J*, **53**, 662-666.

- Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., Burden, F., Farrow, S., Cutler, A.J., Rehnstrom, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., Consortium, B., Stunnenberg, H.G., Todd, J.A., Zerbino, D.R., Stegle, O., Ouwehand, W.H., Frontini, M., Wallace, C., Spivakov, M. & Fraser, P. (2016) Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, **167**, 1369-1384 e1319.
- Kahr, W.H., Savoia, A., Pluthero, F.G., Li, L., Christensen, H., De Rocco, D., Traivaree, C., Butchart, S.E., Curtin, J., Stollar, E.J., Forman-Kay, J.D. & Blanchette, V.S. (2009) Megakaryocyte and platelet abnormalities in a patient with a W33C mutation in the conserved SH3-like domain of myosin heavy chain IIA. *Thromb Haemost*, **102**, 1241-1250.
- Kanematsu, T., Suzuki, N., Yoshida, T., Kishimoto, M., Aoki, T., Ogawa, M., Kagami, Y., Kiyoi, H., Matsushita, T. & Kunishima, S. (2016) A case of MYH9 disorders caused by a novel mutation (p.K74E). *Ann Hematol*, **95**, 161-163.
- Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K.E., Cummings, B.B., Birnbaum, D., The Exome Aggregation, C., Daly, M.J. & MacArthur, D.G. (2017) The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*, **45**, D840-D845.
- Kunishima, S., Kojima, T., Matsushita, T., Tanaka, T., Tsurusawa, M., Furukawa, Y., Nakamura, Y., Okamura, T., Amemiya, N., Nakayama, T., Kamiya, T. & Saito, H. (2001a) Mutations in the NMMHC-A gene cause autosomal dominant macrothrombocytopenia with leukocyte inclusions (May-Hegglin anomaly/Sebastian syndrome). *Blood*, **97**, 1147-1149.
- Kunishima, S., Matsushita, T., Yoshihara, T., Nakase, Y., Yokoi, K., Hamaguchi, M. & Saito, H. (2005) First description of somatic mosaicism in MYH9 disorders. *Br J Haematol*, **128**, 360-365.
- Kunishima, S., Matsushita, T., Kojima, T., Amemiya, N., Choi, Y.M., Hosaka, N., Inoue, M., Jung, Y., Mamiya, S., Matsumoto, K., Miyajima, Y., Zhang, G., Ruan, C., Saito, K., Song, K.S., Yoon, H.J., Kamiya, T. & Saito, H. (2001b) Identification of six novel MYH9 mutations and genotype-phenotype relationships in autosomal dominant macrothrombocytopenia with leukocyte inclusions. *J Hum Genet*, **46**, 722-729.
- Kunishima, S., Matsushita, T., Kojima, T., Sako, M., Kimura, F., Jo, E.K., Inoue, C., Kamiya, T. & Saito, H. (2003) Immunofluorescence analysis of neutrophil nonmuscle myosin heavy chain-A in MYH9 disorders: association of subcellular localization with MYH9 mutations. *Lab Invest*, **83**, 115-122.
- Kunishima, S., Takaki, K., Ito, Y. & Saito, H. (2009) Germinal mosaicism in MYH9 disorders: a family with two affected siblings of normal parents. *Br J Haematol*, **145**, 260-262.

- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G. & Exome Aggregation, C. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285-291.
- Neveling, K., Feenstra, I., Gilissen, C., Hoefsloot, L.H., Kamsteeg, E.J., Mensenkamp, A.R., Rodenburg, R.J., Yntema, H.G., Spruijt, L., Vermeer, S., Rinne, T., van Gassen, K.L., Bodmer, D., Lugtenberg, D., de Reuver, R., Buijsman, W., Derks, R.C., Wieskamp, N., van den Heuvel, B., Ligtenberg, M.J., Kremer, H., Koolen, D.A., van de Warrenburg, B.P., Cremers, F.P., Marcelis, C.L., Smeitink, J.A., Wortmann, S.B., van Zelst-Stams, W.A., Veltman, J.A., Brunner, H.G., Scheffer, H. & Nelen, M.R. (2013) A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat*, **34**, 1721-1726.
- Orsini, S., Noris, P., Bury, L., Heller, P.G., Santoro, C., Kadir, R.A., Butta, N.C., Falcinelli, E., Cid, A.R., Fabris, F., Fouassier, M., Miyazaki, K., Lozano, M.L., Zuniga, P., Flaujac, C., Podda, G.M., Bermejo, N., Favier, R., Henskens, Y., De Maistre, E., De Candia, E., Mumford, A.D., Ozdemir, G.N., Eker, I., Nurden, P., Bayart, S., Lambert, M.P., Bussel, J., Zieger, B., Tosetto, A., Melazzini, F., Glembotsky, A.C., Pecci, A., Cattaneo, M., Schlegel, N., Gresele, P., European Hematology Association - Scientific Working Group on, t. & platelet function, d. (2017) Bleeding risk of surgery and its prevention in patients with inherited platelet disorders. *Haematologica*, **102**, 1192-1203.
- Pecci, A., Biino, G., Fierro, T., Bozzi, V., Mezzasoma, A., Noris, P., Ramenghi, U., Loffredo, G., Fabris, F., Momi, S., Magrini, U., Pirastu, M., Savoia, A., Balduini, C., Gresele, P. & Italian Registry for, M.Y.H.r.d. (2012) Alteration of liver enzymes is a feature of the MYH9-related disease syndrome. *PLoS One*, **7**, e35986.
- Pecci, A., Granata, A., Fiore, C.E. & Balduini, C.L. (2008a) Renin-angiotensin system blockade is effective in reducing proteinuria of patients with progressive nephropathy caused by MYH9 mutations (Fechtner-Epstein syndrome). *Nephrol Dial Transplant*, **23**, 2690-2692.
- Pecci, A., Gresele, P., Klersy, C., Savoia, A., Noris, P., Fierro, T., Bozzi, V., Mezzasoma,

A.M., Melazzini, F. & Balduini, C.L. (2010a) Eltrombopag for the treatment of the inherited thrombocytopenia deriving from MYH9 mutations. *Blood*, **116**, 5832-5837.

- Pecci, A., Klersy, C., Gresele, P., Lee, K.J., De Rocco, D., Bozzi, V., Russo, G., Heller, P.G., Loffredo, G., Ballmaier, M., Fabris, F., Beggiato, E., Kahr, W.H., Pujol-Moix, N., Platokouki, H., Van Geet, C., Noris, P., Yerram, P., Hermans, C., Gerber, B., Economou, M., De Groot, M., Zieger, B., De Candia, E., Fraticelli, V., Kersseboom, R., Piccoli, G.B., Zimmermann, S., Fierro, T., Glembotsky, A.C., Vianello, F., Zaninetti, C., Nicchia, E., Guthner, C., Baronci, C., Seri, M., Knight, P.J., Balduini, C.L. & Savoia, A. (2014) MYH9-related disease: a novel prognostic model to predict the clinical evolution of the disease based on genotype-phenotype correlations. *Hum Mutat*, **35**, 236-247.
- Pecci, A., Ma, X., Savoia, A. & Adelstein, R.S. (2018) MYH9: Structure, functions and role of non-muscle myosin IIA in human disease. *Gene*, **664**, 152-167.
- Pecci, A., Panza, E., De Rocco, D., Pujol-Moix, N., Giroto, G., Podda, L., Paparo, C., Bozzi, V., Pastore, A., Balduini, C.L., Seri, M. & Savoia, A. (2010b) MYH9 related disease: four novel mutations of the tail domain of myosin-9 correlating with a mild clinical phenotype. *Eur J Haematol*, **84**, 291-297.
- Pecci, A., Panza, E., Pujol-Moix, N., Klersy, C., Di Bari, F., Bozzi, V., Gresele, P., Lethagen, S., Fabris, F., Dufour, C., Granata, A., Doubek, M., Pecoraro, C., Koivisto, P.A., Heller, P.G., Iolascon, A., Alvisi, P., Schwabe, D., De Candia, E., Rocca, B., Russo, U., Ramenghi, U., Noris, P., Seri, M., Balduini, C.L. & Savoia, A. (2008b) Position of nonmuscle myosin heavy chain IIA (NMMHC-IIA) mutations predicts the natural history of MYH9-related disease. *Hum Mutat*, **29**, 409-417.
- Rabbolini, D.J., Chun, Y., Latimer, M., Kunishima, S., Fixter, K., Valecha, B., Tan, P., Chew, L.P., Kile, B.T., Burt, R., Radhakrishnan, K., Bird, R., Ockelford, P., Gabrielli, S., Chen, Q., Stevenson, W.S., Ward, C.M. & Morel-Kopp, M.C. (2017) Diagnosis and treatment of MYH9-RD in an Australasian cohort with thrombocytopenia. *Platelets*, 1-8.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L. & Committee, A.L.Q.A. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, **17**, 405-424.
- Saposnik, B., Binard, S., Fenneteau, O., Nurden, A., Nurden, P., Hurtaud-Roux, M.F., Schlegel, N. & French, M.Y.H.n. (2014) Mutation spectrum and genotype-phenotype correlations in a large French cohort of MYH9-Related Disorders. *Mol Genet Genomic Med*, **2**, 297-312.
- Savoia, A., Germeshausen, M., De Rocco, D., Henschel, B., Kratz, C.P., Kuhlen, M., Rath, B., Steuhl, K.P., Wermes, C. & Ballmaier, M. (2010) MYH9-related

disease: Report on five German families and description of a novel mutation. *Ann Hematol*, **89**, 1057-1059.

- Seri, M., Cusano, R., Gangarossa, S., Caridi, G., Bordo, D., Lo Nigro, C., Ghiggeri, G.M., Ravazzolo, R., Savino, M., Del Vecchio, M., d'Apolito, M., Iolascon, A., Zelante, L.L., Savoia, A., Balduini, C.L., Noris, P., Magrini, U., Belletti, S., Heath, K.E., Babcock, M., Glucksman, M.J., Aliprandis, E., Bizzaro, N., Desnick, R.J. & Martignetti, J.A. (2000) Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Hegglin/Fechtner Syndrome Consortium. *Nat Genet*, **26**, 103-105.
- Simeoni, I., Stephens, J.C., Hu, F., Deevi, S.V., Megy, K., Bariana, T.K., Lentaigne, C., Schulman, S., Sivapalaratnam, S., Vries, M.J., Westbury, S.K., Greene, D., Papadia, S., Alessi, M.C., Attwood, A.P., Ballmaier, M., Baynam, G., Bermejo, E., Bertoli, M., Bray, P.F., Bury, L., Cattaneo, M., Collins, P., Daugherty, L.C., Favier, R., French, D.L., Furie, B., Gattens, M., Germeshausen, M., Ghevaert, C., Goodeve, A.C., Guerrero, J.A., Hampshire, D.J., Hart, D.P., Heemskerk, J.W., Henskens, Y.M., Hill, M., Hogg, N., Jolley, J.D., Kahr, W.H., Kelly, A.M., Kerr, R., Kostadima, M., Kunishima, S., Lambert, M.P., Liesner, R., Lopez, J.A., Mapeta, R.P., Mathias, M., Millar, C.M., Nathwani, A., Neerman-Arbez, M., Nurden, A.T., Nurden, P., Othman, M., Peerlinck, K., Perry, D.J., Poudel, P., Reitsma, P., Rondina, M.T., Smethurst, P.A., Stevenson, W., Szkotak, A., Tuna, S., van Geet, C., Whitehorn, D., Wilcox, D.A., Zhang, B., Revel-Vilk, S., Gresele, P., Bellissimo, D.B., Penkett, C.J., Laffan, M.A., Mumford, A.D., Rendon, A., Gomez, K., Freson, K., Ouwehand, W.H. & Turro, E. (2016) A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood*, **127**, 2791-2803.
- Stenson, P.D., Mort, M., Ball, E.V., Evans K., Hayden M., Heywood S., Hussain M., Phillips A.D. & Cooper, D.N. (2017) The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*, **136**, 665-677.
- Vicente-Manzanares, M., Ma, X., Adelstein, R.S. & Horwitz, A.R. (2009) Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nat Rev Mol Cell Biol*, **10**, 778-790.
- Westbury, S.K., Turro, E., Greene, D., Lentaigne, C., Kelly, A.M., Bariana, T.K., Simeoni, I., Pillois, X., Attwood, A., Austin, S., Jansen, S.B., Bakchoul, T., Crisp-Hihn, A., Erber, W.N., Favier, R., Foad, N., Gattens, M., Jolley, J.D., Liesner, R., Meacham, S., Millar, C.M., Nurden, A.T., Peerlinck, K., Perry, D.J., Poudel, P., Schulman, S., Schulze, H., Stephens, J.C., Furie, B., Robinson, P.N., van Geet, C., Rendon, A., Gomez, K., Laffan, M.A., Lambert, M.P., Nurden, P., Ouwehand, W.H., Richardson, S., Mumford, A.D., Freson, K. & Consortium, B.-B. (2015) Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Med*, **7**, 36.

Zhang, J., Barbaro, P., Guo, Y., Alodaib, A., Li, J., Gold, W., Ades, L., Keating, B.J., Xu, X., Teo, J., Hakonarson, H. & Christodoulou, J. (2016) Utility of next-generation sequencing technologies for the efficient genetic resolution of haematological disorders. *Clin Genet*, **89**, 163-172.

Table I - MYH9 variants

Patient	Ethnicity	Chromosomal position (NG_011884.2)	Transcript alteration (NM_002473.5)	Protein alteration (NP_002464.1)	MYH9 domain affected	Variant type	CADD score	AF (ExAC and gnomAD)	Pathogenicity and contribution to phenotype	ACMG Evidence	Diagnosis
Novel MYH9 variants											
1	EUR	22:43879T>G	c.97T>G	p.(Trp33Gly)	HD	missense	27.0	Not present	Likely pathogenic Full	PM2, PM5, PP3, PP4, A	<u>New diagnosis</u> MYH9-RD not suspected
2	EUR	22:43890_43898del	c.108_116delCGACAAGAG	p.(Asp37_Ser39del)	HD	inframe deletion	N/A	Not present	VUS Full	PM2, PM4, PP4, A	Known MYH9-RD Patient with no genetic confirmation
3	EUR	22:43904T>C	c.122T>C	p.(Phe41Ser)	HD	missense	28.7	Not present	VUS Full	PM6, PM2, PP3, PP4, A	No diagnosis Assumed de novo varia

20	EUR	22:9199 8del	c.2668delC	p.(Gln890 Arg*)	TD	stop- gain	N/A	Not pre sent	VUS [□] Full	PM2, PVS1, PP4, A	No diag nosi s		
39	EUR	22:9209 1G>A	c.2761G>A	p.(Glu921 Lys)	TD	mis sen se	33	1:1 00, 000 (Ex AC and gno mA D)	VUS Full	PM2, A, B	No diag nosi s		
21	EUR	22:9610 6_96126 dup	c.3202_3222dupCA GGCCAGATCGCG GAGCTC	p.(Gln106 8_Leu107 4dup)	TD	infr ame inse rtio n	N/A	Not pre sent	VUS Full	PM2, PM4, PP4, A, B	Known MYH 9- RD Patie nt with no genet ic confir matio n		
28	EUR	22:9804 1C>T	c.3584C>T	p.(Ser119 5Leu)	TD	mis sen se	23 .1	Not pre sent	VUS Full	PM2, A	No diag nosi s		
29	EUR	22:1009 51A>C	c.4262A>C	p.(Glu142 1Ala)	TD	mis sen se	28 .5	Not pre sent	VUS Full	PM2, PP3, A	No diag nosi s		
37	AFR	22:1009 91G>C	c.4302G>C	p.(Gln143 4His)	TD	mis sen se	23 .5	1:1 0,0 00 (Ex AC) ; 1:2 50, 000	VUS Full	PM2	No diag nosi s		

									(gnomAD)			
39	EUR	22:106186A>G	c.4946A>G	p.(Asp1649Gly)	TD	missense	32	Not present	VUS	Full	PM2, PP3, PP1, A, B	No diagnosis
40	N/A	22:106272A>G	c.5032A>G	p.(Met1678Val)	TD	missense	18.4	Not present	VUS	Uncertain	PM2, A	No diagnosis
50	N/A	22:110276del	c.5808delG	p.(Gly1938Alafs*10)	NH TD	frameshift	N/A	Not present	Likely pathogenic	Full	PM2, #PM4, PM1, PP4, A	Known MYH9-RD Patient with no genetic confirmation

Known MYH9 variants

4	EUR	22:44002A>G	c.220A>G	p.(Lys74Glu) (Kanematsu, <i>et al</i> 2016)	HD	missense	23.50	Not present	Likely pathogenic	Full	PS4, supporting, PM2, PP4, PP3, A, B	Known MYH9-RD patient
5	EUR	22:44061C>G	c.279C>G	p.(Asn93Lys) (Seri, <i>et al</i> 2000)	HD	missense	25	Not present	**Pathogenic	Full	PS4, PS3, PM1, PM2, PP4, PP3, A	Known MYH9-RD patient
6	EUR	22:44065G>A	c.283G>A	p.(Ala95Thr) (Kunishima, <i>et al</i> 2001b)	HD	missense	29.70	Not present	**Pathogenic	Full	PS4, PM6, PM2, PM1, PP3, A	<u>New diagnosis</u> MYH9-RD suspected

14 [#] , 15 [#] , 16 [#] , 17 [#]	EUR ¹⁴ ,15,16,1 7	22:8708 2C>T	c.2152C>T	p.(Arg718 Trp) (Pecci, <i>et al</i> 2008b)	HD	mis sen se	34	Not pre sent in ExA C; 1:2 50, 000 (gn om AD)	Patho genic Full	PS4, PM2, PP1_s trong, PP4, PP3, A ¹⁷ ,B ¹ 4,15,16 ,17	<u>New diag nosi s</u> for all patie nts MYH 9-RD suspe cted
18,1 9	EUR ¹⁸ ,19	22:9136 1C>T	c.2507C>T	p.(Pro836 Lys) (Neveling, <i>et al</i> 2013)	TD	mis sen se	34	Not pre sent	*Likel y patho genic Full	PS4, PM2, PP4, PP3, A ¹⁸	<u>New diag nosi s</u> for both patie nts Patie nts 18 (MYH 9-RD suspe cted); 19 (not suspe cted)
22 [^] , 23, 24 [^] , 25, 26,2 7	EUR ²² ,23,24,2 6,27 N/A ²⁵	22:9795 0C>T	c.3493C>T	p.(Arg116 5Cys) (Seri, <i>et al</i> 2000)	TD	mis sen se	34	Not pre sent	Patho genic Full	PS4, PM2, PP1_s trong, PP4, PP3, A ²⁵ , B ²²⁻²⁴	<u>New diag nosi s</u> for patie nts 23,2 5,26, 27 Patie nts 23 and 25 (MYH 9-RD suspe cted); 26 and 27 (not suspe cted); 22 and 24 (kno wn MYH 9- RD)

30,31,32,33	Iraqi ³⁰ , N/A ³¹ , SAS ³² , Hispanic ³³	22:100959G>A	c.4270G>A	p.(Asp1424Asn) (Kunishima, <i>et al</i> 2001a)	TD	mis sense	33	Not present	Pathogenic Full	PS4, PP1_strong, PM2, PP3, A ^{31,33}	<u>New diagnosis</u> for all patients Patients 31, 32, 33 (MYH9-RD suspected); 30 (not suspected)
34□, 35□, 36	EUR ³⁴ , ,35,36	22:100959G>T	c.4270G>T	p.(Asp1424Tyr) (Kunishima, <i>et al</i> 2001b)	TD	mis sense	32	Not present	Pathogenic Full	PS4, PP1_strong, PM2, PP3, A, B ³⁴⁻³⁵	<u>New diagnosis</u> for patient 36 (MYH9-RD suspected); 34 and 35 (known MYH9-RD)
38	EUR	22:101029A>T	c.4340A>T	p.(Asp1447Val) (Pecci, <i>et al</i> 2008b)	TD	mis sense	31	Not present	Likely pathogenic Full	PS4, PM2, PP3, A, B	Known MYH9-RD patient with no genetic confirmation
41,42,43□,	EUR ⁴² , ,44,45	22:108545G>A	c.5521G>A	p.(Glu1841Lys)	TD	mis sense	34	Not present	Pathogenic	PS4, PP1_strong,	Known MYH9-

44	□	Filipino 43			(Seri, <i>et al</i> 2000)	se	sent	Full	PM2, PP4, PP3, A ⁴¹⁻⁴⁴ , B ^{43,44}	RD Patie nts with no genet ic confir matio n		
45		EUR	22:1102 38_1102 47del	c.5770_5779delGGG GACCTGC	p.(Gly1924 Argfs*21) (Pecci, <i>et al</i> 2010b)	NH TD	fram eshi ft	36	1:5 0,0 00 (Ex AC) ; not pre sent in gno mAD	Patho genic Full	PS4, #PM4, PM2, PM1, PP4, A	<u>New diag nosi s</u> MYH 9-RD not sus pected
46,4 7,48		N/A	22:1102 65C>T	c.5797C>T	p.(Arg193 3*) (Seri, <i>et al</i> 2000)	NH TD	stop - gain	50	1:1 00, 000 (Ex AC) ; 1:2 50, 000 (gn om AD)	Patho genic Full	PS4, PM2, PM1, PP4, A ⁴⁶⁻⁴⁸ B ⁴⁷	Known MYH 9- RD Patie nts with no genet ic confir matio n
49		EUR	22:1102 68del	c.5800delA	p.(Met193 4Trpfs*14) (Savoia, <i>et al</i> 2010)	NH TD	fram eshi ft	34	Not pre sent	Likely patho genic Full	PS4, #PM4, PM2, PM1, PP4, A	Known MYH 9- RD Patie nt with no genet ic confir matio n

ACMG Guidelines Evidences: **PVS1 (very strong evidence):** Null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where LOF is a known mechanism of disease; **PS1 (strong evidence):** same amino acid change as a previously published pathogenic variant regardless of nucleotide change; **PS4 (strong evidence):** The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls; **PS3 (strong evidence):** Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product; **PM2 (moderate evidence):** Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium; **PM1 (moderate evidence):** Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation; **PM4 (moderate evidence):** Protein length changes as a result of in-frame deletions/insertions in a non repeat

region or stop-loss variants; **PM6** (supporting evidence): Assumed de novo, but without confirmation of paternity and maternity; **PP2** (supporting evidence): Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease; **PP3** (supporting evidence): Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.); ; **PP1** (supporting evidence): Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease; **PP4** (supporting evidence): Patient's phenotype or family history is highly specific for a disease with a single genetic etiology; **#PM4** used instead of PVS1 in case of protein termination within the last 50 amino acids; **A**: presence of at least one main *MYH9*-RD feature in addition to macrothrombocytes (Dohle-like bodies, hearing impairment, nephropathy, alteration of liver enzymes); **B**: family history of *MYH9*-RD.

The ACMG evidence PP3 is used only in case of all the following evidences are present: highly conserved nucleotide, highly conserved amino acid, effect on protein, alignment with GVDG of C65 and deleterious prediction in SIFT, Mutation Taster and PolyPhen.

Superscript numbers in the 'Ethnicity' and in the 'ACMG Evidences' columns represent the patients. *Variant initially considered as 'VUS' but then changed into 'likely pathogenic'; ** Variants initially labelled as 'likely pathogenic' and subsequently re-classified as 'pathogenic'; variant labelled as VUS, independently by the ACMG evidences, due to the impossibility to perform functional tests.

Members of the same family are represented with the following symbols: #, ^, □, □

HD= head domain; TD=coiled-coil tail domain; NHTD = non-helical tail domain.