# UNIVERSITÀ DEGLI STUDI DI TRIESTE

## XXXI CICLO DEL DOTTORATO DI RICERCA IN SCIENZE DELLA RIPRODUZIONE E DELLO SVILUPPO

# MOLECULAR DIAGNOSIS OF INHERITED THROMBOCYTOPENIAS USING HIGH THROUGHPUT SEQUENCING

Settore scientifico-disciplinare: **BIO/18 GENETICA**

PH.D STUDENT
**TANIA GIANGREGORIO**

PH.D PROGRAM COORDINATOR
**PROF. PAOLO GASPARINI**

THESIS SUPERVISOR
**PROF. ANNA SAVOIA**

**ANNO ACCADEMICO 2017/2018**

Table of Contents

# ABSTRACT

Inherited thrombocytopenias (ITs) are a heterogenous group of rare genetic disorders characterized by reduced platelet count sometimes combined with bleeding tendency and/or other clinical defects. The molecular diagnosis of ITs is essential to make clinical decision and infer personalized prognosis and risks.

About 30 genes have been identified that harbor mutations responsible for ITs (Balduini et al., 2017). In addition, ITs often show phenotypic overlaps that hamper the correct diagnosis with the traditional diagnostic algorithm based on step-wise specialized investigations.

However, the advent of next generation sequencing has changed the diagnostic approach of diseases characterized by high genetic heterogeneity like ITs.

In order to improve the diagnosis of ITs, we designed a targeted next generation sequencing panel to screen 28 genes associated with ITs (IT-NGS). Ninety-seven consecutive probands with a suspicious of ITs had been sequenced.

The analysis led us to reach a definite diagnosis for 37 probands. In these probands we identified known or novel likely pathogenic mutations causing specific diseases, including monoallelic Bernard Soulier syndrome (N=14), biallelic Bernard Soulier syndrome (N=4), *ACTN1*-related thrombocytopenia (N=4), *MYH9*-related disease (N=7), *ANKRD26*-related thrombocytopenia (N=4), congenital amegakaryocytic thrombocytopenia (N=1), grey platelet syndrome (N=1), Wiskott-Aldrich syndrome (N=1) and the Familial Platelet Disorder with propensity to Acute Myelogenous Leukemia (N=1). In another 43 cases we identified variants of uncertain significance (VUS) whose pathogenic role has to be supported by segregation analysis and in-depth functional studies. Since 17 probands had no potential candidate variant impacting IT-NGS genes, they are eligible for whole exome sequencing (WES) to clone novel genes involved in ITs.

In conclusion, since some IT forms predispose to additional acquired disease during life, an accurate diagnosis is essential to infer personalized prognosis and define proper treatments and follow-up.

Because of clinical and genetic heterogeneity, the molecular diagnosis of ITs represents a lengthy and expensive challenge using conventional technologies.

The use of IT-NGS in clinical practice aided by specific investigations clarifying the role of variant of uncertain significance, overcomes these issues facilitating a definite diagnosis in patients with a suspicious of known ITs forms.

# 1 INTRODUCTION

## 1.1 Platelet biogenesis

Platelets are anucleated small cells involved in hemostasis and blood coagulation (Patel et al., 2005).

During early megakaryopoiesis, hematopoietic stem cells in the osteoblastic niche of the bone marrow expands and differentiates into platelet progenitor cells called megakaryocytes (MKs). Early MKs undergo a tightly regulated process of growth and maturation leading to cellular polyploidia and cytoskeletal reorganization in an internal membrane system.

Finally, mature MKs migrate to the vascular niche of the bone marrow and form cytoplasmic protrusion called proplatelet that are extruded into the bone marrow sinusoids and shed into the blood stream (Machlus and Italiano, 2013).

This process is finely regulated by the hematopoietic growth factor thrombopoietin (THPO) synthetized by the liver (Wendling et al., 1994). THPO binds the c-MPL specific receptor expressed on the membrane of bone marrow MKs triggering several downstream signaling pathways.

## 1.2 Inherited Thrombocytopenias (ITs)

Inherited thrombocytopenias are heterogenous genetic disorders characterized by reduced platelet count sometimes associated with bleeding diathesis or other clinical manifestation. They are caused by mutation affecting a wide variety of genes involved in transcriptional regulation, cytoskeleton organization and transmembrane glycoprotein signaling pathways. All the IT-related genes identified so far, play a synergic role in the megakaryopoiesis process aimed at platelet production (Figure 1).

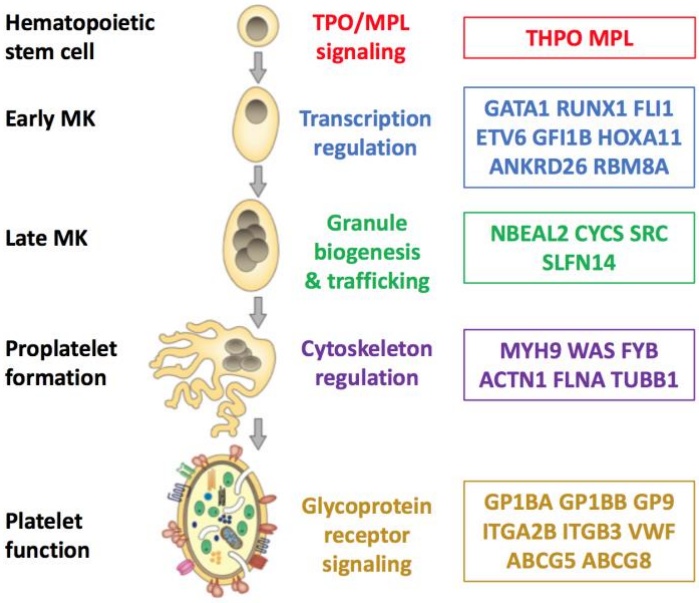


**Figure 1. Representation of megakaryopoiesis process and platelet formation.** Each gene included in IT-NGS design is indicated and categorized according to their effect on megakaryocyte and platelet biology. Abbreviation: MK, megakaryocytes. Cartoon adapted from Lentaigne et al, Blood 2016.

Given the marked heterogeneity of ITs, they can be classified according to several parameters like mode of inheritance, platelets size or associated abnormalities of platelets, red blood cells or leukocytes.

Recently Noris et al have proposed a new classification of ITs based on clinical features (Noris and Pecci, 2017). They divided ITs into 3 categories: forms purely characterized by platelet defect, forms characterized by additional congenital defects and forms with increased susceptibility to develop additional diseases during life like hematological malignancies, bone marrow aplasia or extra-hematological defects (Table 1). This distinction have prognostic implications for ITs patients.

| Disease (abbreviation, OMIM entry) | | Frequency* | Inheritance | Gene | Locus |
|---|---|---|---|---|---|
| **Forms with only thrombocytopenia** | | | | | |
| Bernard-Soulier syndrome (BSS, 231200/153670) | Biallelic | ++++ | AR | GP1BA<br>GPIBB<br>GP9 | 17p13<br>22q11<br>3q21 |
| | Monoallelic | +++ | AD | | |
| Gray platelet syndrome (GPS, 139090) | | ++ | AR | NBEAL2 | 3p21 |
| ACTN1-related thrombocytopenia (ACTN1-RT, 615193) | | ++ | AD | ACTN1 | 14q24 |
| ITGA2B/ITGB3-related thrombocytopenia (ITGA2B/ITGB3-RT, 187800) | | + | AD | ITGA2B<br>ITGB3 | 17q21<br>17q21 |
| TUBB1-related thrombocytopenia | | + | AD | TUBB1 | 20q13 |

| | | | | |
|---|---|---|---|---|
| (TUBB1-RT, 613112) | | | | |
| CYCS-related thrombocytopenia (CYCS-RT or THC4, 612004) | + | AD | CYCS | 7p15 |
| GFI1b-related thrombocytopenia (GFI1b-RT, 187900) | + | AD | GFI1B | 9q34 |
| FYB-related thrombocytopenia (FYB-RT or THC3, 273900) | + | AR | FYB | 5p13.1 |
| SLFN14-related thrombocytopenia (SLFN14-RT or BDPLT20, 616913 ) | + | AD | SLFN14 | 17q12 |
| FLI1-related thrombocytopenia (FLI1-RT or BDPLT21, 617443) | + | AR | FLI1 | 11q24.3 |
| Inherited thrombocytopenia from monoallelic THPO mutation (THPO-RD, NA) | + | AD | THPO | 3q27.1 |
| Von Willebrand disease types 2B (VWD2B, 613554) | ++ | AD | VWF | 12p13 |
| **Forms with additional clinically relevant congenital defects/syndromic forms** | | | | |
| Thrombocytopenia-absent radius syndrome (TAR, 2740009) | +++ | AR | RBM8A | 1q21 |
| Wiskott-Aldrich syndrome (WAS, 301000)<br>X-linked thrombocytopenia (XLT, 313900) | ++++ | XL | WAS | Xp11 |
| FLNA-related thrombocytopenia (FLNA-RT, NA) | + | XL | FLNA | Xq28 |
| GATA1-related disease (GATA1-RD) (Dyserythropoietic anemia with thrombocytopenia - NA, 300367 – X-linked thrombocytopenia with thalassemia – XLTT, 314050) | ++ | XL | GATA1 | Xp11 |
| Thrombocytopenia associated with sitosterolemia (STSL, 210250) | + | AR | ABCG5<br>ABCG8 | 2p21 |
| **Forms with increased risk of acquiring additional illnesses/predisposing forms** | | | | |
| Congenital amegakaryocytic thrombocytopenia (CAMT, 604498) | ++ | AR | MPL | 1p34.2 |
| MYH9-related disease (MYH9-RD, 155100 ) | ++++ | AD | MYH9 | 22q12 |
| Familial platelet disorder with propensity to acute myelogenous leukemia (FPD/AML, 601399) | ++ | AD | RUNX1 | 21q22 |
| ANKRD26-related thrombocytopenia (ANKRD26-RT or THC2, 188000) | ++ | AD | ANKRD26 | 10p12 |
| ETV6-related thrombocytopenia (ETV6-RT or THC5, 616216 ) | + | AD | ETV6 | 12p13 |
| Radioulnar synostosis with amegakaryocytic thrombocytopenia (RUSAT, 605432) | + | AR | HOXA11 | 7p15 |
| Thrombocytopenia 6 (THC6, 616937) | + | AD | SRC | 20q12 |

**Table 1. Inherited thrombocytopenia forms targeted by our IT-NGS panel classified according to their clinical features.** AD, autosomal dominant; AR, autosomal recessive; NA, not available; XL, X-linked. *n°familes reported: ++++, > 100; +++, > 50; ++, ≥ 10; +, < 10.

The traditional diagnostic algorithm used for the differential diagnosis of IT exploits patients medical history and physical examination to distinguish syndromic and predisposing ITs forms, whereas the evaluation of platelet size and peripheral blood smear guides the diagnostic process in non-syndromic forms (Balduini et al., 2013) (Figure 2). However, genetic studies are required to confirm the diagnostic hypothesis definitively.

Of note, approximately 50% of ITs proband do not fit the criteria for any known forms, this suggests that novel forms of ITs have yet to be characterized (Balduini and Noris, 2016).

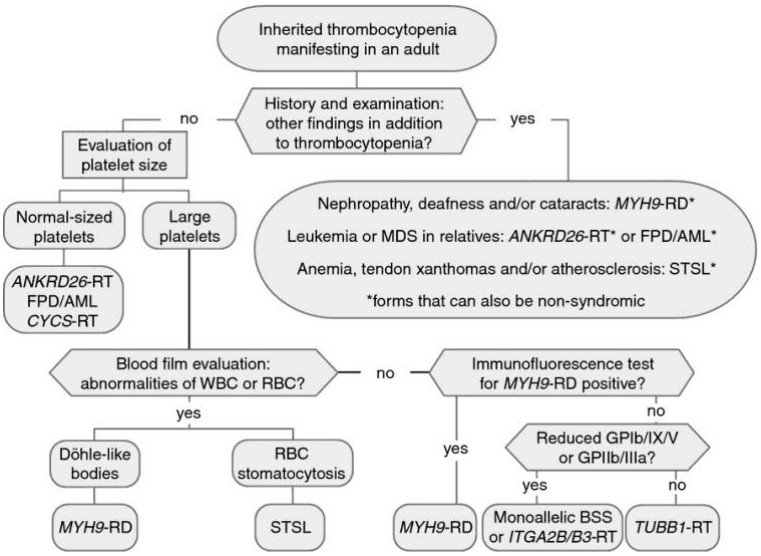


**Figure 2. Traditional diagnostic algorithm used for the differential diagnosis of ITs.** Abbreviations: ANKRD26-RT, ANKRD26-related thrombocytopenia; CYCS-RT, CYCS-related thrombocytopenia; FPD/AML, familial platelet disorder with predisposition to acute myelogenous leukemia; STSL, sitosterolemia; MDS, myelodysplastic syndrome; RBC, red blood cell; WBC, white blood cell.

The next sections describe the more representative IT forms frequently reported (Balduini and Noris, 2016), focusing on genetic aspects, clinical and laboratory features.

### 1.2.1 Bernard-Soulier syndrome

Bernard-Soulier syndrome (BSS) is caused by mutations in the *GP1BA*, *GP1BB* or the *GP9* genes. This genes encode for 3 of the 4 different subunits that compose GPIb-IX-V complex, a platelet membrane glycoprotein that functions as a receptor for von Willebrand factor (VWF). The VWF is a multimeric glycoprotein that mediates platelet adhesion to the vascular endothelium at sites of injury (Denis, 2002).

The presence of large platelet is one of BSS distinctive feature useful for differential diagnosis of ITs (Balduini et al., 2013). To date, two forms of BSS have been described: the most severe biallelic BSS (BSSA1) caused by homozygous or compound heterozygous mutation and the monoallelic BSS (BSSA2), which is a mild autosomal dominant form of IT caused by heterozygous mutations (Savoia et al., 2014).

BSSA1 is characterized by prolonged bleeding time and no platelet agglutination after stimulation with ristocetin, whereas BSSA2 has absent or mild bleeding diathesis and normal or reduced platelet aggregation in response to the addition of various activators.

The most frequent BSSA2 mutation in Italy is the nonsynonymous c.515C>T variant in *GP1BA* gene also called “Bolzano mutation” (Savoia et al., 2001).

Since 42 apparently unrelated families with macrothrombocytopenia and Bolzano mutation share the same haplotype, this mutation may be associated to a founder effect (Noris et al., 2012).

### 1.2.2 *ACTN1*-related thrombocytopenia

*ACTN1*-related thrombocytopenia (*ACTN1*-RT) is an autosomal dominant form of macrothrombocytopenia characterized by no or mild bleeding tendency. It is caused by mutations in the gene *ACTN1* encoding for the non-muscle isoforms of $\alpha$-actinin, an actin-crosslinking protein expressed in megakaryocytes and platelets. This protein has an N-terminal actin-binding domain, four spectrin repeats and a C-terminal calmodulin-like domain (Sjöblom et al., 2008).

To date, the different pathogenetic variants described in literature are mainly amino acid substitutions affecting the N-terminal or C-terminal functional domains (Bottega et al., 2015; Kunishima et al., 2013).

Since immunofluorescence assays in cells overexpressing the mutant forms of ACTN1 show abnormal cytoskeletal organization of the actin filaments, *ACTN1* mutations may have a dominant effect on the actin filament assembly (Bottega et al., 2015; Kunishima et al., 2013).

### 1.2.3 Congenital amegakaryocytic thrombocytopenia

Homozygous or compound heterozygous mutations in *MPL* gene cause the congenital amegakaryocytic thrombocytopenia (CAMT), an autosomal recessive disease.

To date, forty-five different mutations of *MPL* have been identified mainly in the first five exons (Ballmaier and Germeshausen, 2009).

The *MPL* gene encodes for the receptor of thrombopoietin, an hematopoietic growth factor crucial for megakaryopoiesis and platelets production (Freedman and Estrov, 1990).

CAMT is characterized by severe thrombocytopenia at birth (less than 50.000 platelet/µl) leading to hemorrhagic manifestation. The bone marrow aspiration of CAMT patients shows a reduction or absence of megakaryocytes. Moreover, TPO plasma level are high in affected patients (Ballmaier et al., 2001; Muraoka et al., 1997). These findings are helpful for differential diagnosis of CAMT.

Since the disease usually progresses to bone marrow aplasia within the first years of life, affected individuals require hematopoietic stem/progenitor cell (HSPC) transplantation.

### 1.2.4 *MYH9*-related disease

*MYH9*-related disease (*MYH9*-RD) is an autosomal-dominant disorder caused by mutations in *MYH9* gene. This gene encoded for the human non-muscle myosin heavy chain IIA (NMMHC-IIA) expressed in fibroblasts, endothelial cells, and macrophages (Kunishima et al., 2001; Seri et al., 2000, 2003).

NMMHC-IIA has two different domains: the globular head domain (HD) at the N-terminus, which includes the actin-binding site and the ATP hydrolysis region involved in the mechanical translocation of myosin along actin filament, and the tail domain (TD) at the C-terminus, crucial for heavy chain dimerization and the assembly of myosin filaments.

More than 80 different mutations related to *MYH9*-RD have been identified; mainly non-synonymous mutations affecting the head or tail domain (Pecci et al., 2018).

However, in the 70% of cases the mutations hit only 6 different residues of NMMHC-IIA: Ser96 and Arg702 in the HD, Arg1165, Asp1424 and Glu1841 in the coiled-coil region of TD, and Arg1933 in the non-helical region of TD (Balduini et al., 2011a).

All *MYH9*-RD affected cases show macrothrombocytopenia related to a variable degree of bleeding tendency and neutrophils inclusion bodies (Balduini et al., 2011a; Kunishima and Saito, 2010).

However, these hematological features are often associated with extra-hematological defect that may occur in childhood or adult life, like severe deafness and/or presenile cataract and/or alteration of liver enzymes and/or glomerulonephritis evolving to end-stage kidney failure (Pecci et al., 2012).

Genotype-phenotype studies showed that the risk and the severity of extra-hematological manifestations are predicted by specific *MYH9* mutations. Overall, mutations affecting the HD are associated with a more severe clinical evolution than those involving the TD (Pecci et al., 2014a).

### 1.2.5 Wiskott-Aldrich syndrome and X-linked thrombocytopenia

The Wiskott-Aldrich syndrome (WAS) and the X-linked thrombocytopenia (XLT) are congenital ITs both caused by mutations in *WAS* gene located on chromosome X (Aldrich et al., 1954; Canales and Mauer, 1967). *WAS* gene encodes for the WAS protein, a protein regulating the cytoskeleton organization expressed in hemopoietic cells.

Male cases with WAS or XLT have small-sized platelet and moderate or very severe platelet count reduction. In addition, patients with WAS also have a severe immunodeficiency, recurrent infections and an increased risk of lymphoproliferative

disorders generally leading to early death. Allogeneic HSPC transplantation can be resolutive for WAS patients, but the absence of fully matched donors it is often associated with an increased mortality.

The differences between XLT and WAS clinical phenotypes may be explained by the type of *WAS* mutations identified respectively (Jin et al., 2004).

Indeed, XLT is typically due to missense or splice-site mutations whereas WAS is generally caused by complex mutations, like non-sense and frameshift mutations, which are associated to most severe clinical phenotypes.

### 1.2.6 Inherited thrombocytopenias with increased risk for myeloid malignancies

*ANKRD26*-Related Thrombocytopenia (*ANKRD26*-RT), Familial Platelet Disorder with propensity to Acute Myelogenous Leukemia (FPD/AML), and *ETV6*-Related Thrombocytopenia (*ETV6*-RT) are autosomal dominant platelet disorders caused by mutations of *ANKRD26*, *RUNX1* and *ETV6* genes respectively .

These forms are characterized by moderate thrombocytopenia with normal-sized platelets and mild bleeding tendency (Melazzini et al., 2016a; Noris et al., 2014).

Since these forms share an increased risk of developing hematological myeloid malignancies (8%, 40%, and 20% for *ANKRD26*-RT, FPD/AML, and *ETV6*-RT, respectively), the 2016 WHO revision of hematological neoplasms clustered them in the group of “myeloid neoplasms with germline predisposition and preexisting platelet disorder” (Arber et al., 2016).

*ANKRD26* gene encodes for a protein with N-terminal ankyrin repeats involved in protein-protein interactions. So far, all the *ANKRD26*-RT related mutations identified hit a highly-conserved stretch in the 5' untranslated region (5'UTR); this region contains RUNX1 binding site (Pippucci et al., 2011).

RUNX1 is a transcription factor, encoded by *RUNX1* gene, that regulates the expression of multiple hematopoiesis-specific genes, in particular it downregulates ANKRD26 expression in the late stages of megakaryopoiesis (Bluteau et al., 2014).

Even *ETV6* gene encodes for a transcription factor. The product of this gene contains two functional domains: a N-terminal pointed domain that is involved in protein-protein interactions with itself and other proteins, and a C-terminal DNA-binding domain.
Gene knockout studies in mice suggest that these transcription factors are required for hematopoiesis, however the pathogenic mechanism that underlies ANKRD26-RT FPD/AML and ETV6-RT remains unclear (Okuda et al., 1996; Wang et al., 1998).

## 1.3 Next generation sequencing

The advent of Next Generation Sequencing (NGS) technologies has radically changed the field of human genetic, accelerating the discoveries of new genes involved in rare genetic disease (Bamshad et al., 2011). Of note, 14 novel ITs forms have been discovered since 2010 thanks to NGS (Balduini and Noris, 2016).
In addition, thanks to the reduction of NGS platform costs, these strategies have been recently introduced in the clinical practice for diagnostic propose.
The capture of specific genomic regions, generally the whole-exome or pre-defined genes, in combination with the high throughput sequencing of DNA allows the rapid identification of all the genomic variants of an individual within the genomic target analyzed. In this context the major limitation is the ability to discriminate the true disease-causing mutation, among all the rare, possibly disease-linked variants.
This task is often hampered by the intrinsic limitations of NGS technologies such as the presence of false positives and false negatives in the process of variant calling as well as the presence of systematic errors for particular types of mutations such as indels and repetitions (Shendure and Ji, 2008).
Moreover, the variants identified may have no connection with the disease as the region harboring the causal variant could be outside the genomic target investigated.
Since whole-exome analysis and even targeted sequencing of pre-defined genes often reveal private non-synonymous variants not reported in genomic database, the interpretation of the molecular impact of these variants is a non-trivial task (Richards et al., 2015).

For the interpretation of missense change effect have been developed a variety of in silico tools based on one or a combination of these criteria:

- evolutionary conservation of an amino acid or nucleotide
- location and context within the protein sequence
- biochemical consequences of the amino acid substitution

The more common in silico tools used in clinical laboratories are PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/index.shtml), SIFT (http://sift.jcvi.org/www/SIFT_enst_submit.html), Mutation Taster (http://www.mutationtaster.org), Mutation Assessor (http://mutationassessor.org/r3/) and the Combined Annotation Dependent Depletion (CADD, https://cadd.gs.washington.edu/).

Mutation Assessor tool assesses the functional impact of amino-acid substitutions in proteins considering only the evolutionary conservation of the affected amino acid in protein homologs.

In contrast, Polyphen2, SIFT and Mutation Taster tools evaluate the impact of an amino acid substitution on the structure and function of a human protein using physical and comparative considerations whereas CADD integrates multiple annotations like conservation metrics, functional genomic data, transcript information and protein-level scores.

As each tool has his own strengths and limits, the use of multiple predictive tools for the variant interpretation is advised. However, they should be used carefully in the assessment of variant pathogenicity since they provides only predictions (Richards et al., 2015). Thus, only well-established in vitro or in vivo functional studies can provide further evidences essential to support the damaging effect of putative disease-causing mutations.

# 2 AIMS OF THE STUDY

The molecular diagnosis of ITs is pivotal to make clinical decision and infer personalized prognosis and risks. However, since ITs are clinically and genetically heterogeneous, their molecular diagnosis is a lengthy and expensive process using conventional technologies.

To address this limitations we implemented an ITs diagnostic approach based on the application of Ion Torrent targeted sequencing technology. This strategy has been used for the simultaneous analysis of a panel of 28 genes associated with ITs.

We assessed the pathogenicity of the candidate variants identified in a case series of 97 proband with a suspicious of IT in order to define the disease-causing mutations in this families.

# 3 MATERIAL AND METHODS

## 3.1 Case series

Ninety-seven Italian index cases with a suspicious of ITs have been recruited in this nationwide multicentric study (50 male and 47 female).

All subjects gave informed consent to the research in accordance with Declaration of Helsinki. Genomic DNA was extracted from peripheral blood samples using standard methods.

## 3.2 IT panel design

We designed an NGS panel covering 28 ITs related genes selected from literature revision (IT-NGS, Table 2) with Ion Ampliseq Designer tool (https://www.ampliseq.com).

IT-NGS panel exploited 788 primer pairs, equally split in two primer pool, for the libraries preparation.

The capture design includes all the protein-coding exons, the flanking intronic regions and the 5' untranslated regions of *ANKRD26* gene. Overall, the target size is of 148.8 kb.

The fraction of covered bases per gene with IT-NGS platform is reported in Table 2.

| Gene (RefSeq) | Coverage (%) |
|---|---|
| RBM8A (NM_005105.3) | 70.07 |
| MPL (NM_005373.2) | 99.76 |
| ABCG5 (NM_022436.2) | 100 |
| ABCG8 (NM_022437.2) | 99.7 |
| NBEAL2 (NM_015175.2) | 99.52 |
| GP9 (NM_000174.3) | 100 |
| CYCS (NM_018947.5) | 100 |
| HOXA11 (NM_005523.5) | 100 |
| ANKRD26 (NM_014915.2) | 76.62 |
| FLI1 (NM_002017.4) | 100 |
| VWF (NM_000552.3) | 97.12 |
| ACTN1 (NM_001102.3) | 100 |

| Gene (RefSeq) | Coverage (%) |
|---|---|
| GP1BA (NM_000173.5) | 98.22 |
| TUBB1 (NM_030773.3) | 90.99 |
| RUNX1 (NM_001001890.2) | 98.41 |
| MHY9 (NM_002473.4) | 92.17 |
| GP1BB (NM_000407.4) | 97.95 |
| WAS (NM_000377.2) | 94.86 |
| GATA1 (NM_002049.3) | 98.89 |
| FLNA (NM_001110556.1) | 99.97 |
| ETV6 (NM_001987.4) | 100 |
| THPO (NM_000460.3) | 100 |
| GFI1B (NM_004188.5) | 100 |
| SLFN14 (NM_001129820.1) | 99.9 |

| | | | | |
|---|---|---|---|---|
| ITGB3 (NM_000212.2) | 95.31 | | FYB (NM_001465.4) | 100 |
| ITGA2B (NM_000419.3) | 97.81 | | SRC (NM_005417.4) | 98.9 |

**Table 2. List of genes included in IT-NGS design.** For each gene is reported the percentage of bases covered by Ion AmpliSeq Designer software.

## 3.3 Library preparation and sequencing

We set up DNA target amplification reactions using Ion Ampliseq library kit 2.0 (Thermo Fisher). Briefly, 10 ng of genomic DNA were amplified using IT-NGS primer pools and Ion AmpliSeq™ HiFi Mix.

After multiplex PCR reaction, different barcode adapters were ligated to each library in order to sequence multiple libraries on a single chip ( 5 and 11 libraries per 316™ and 318™ Ion Chip respectively) .

Barcoded libraries have been purified following manufacturers' instruction and quantified using KAPA Library Quantification Kit for Ion Torrent platforms.

We diluted libraries to 10 pM and performed the emulsion PCR, with the Ion OneTouch™ Instrument, for the generation of template positive Ion Sphere Particles (ISPs) containing clonally amplified amplicon. The template-positive ISPs were enriched with the Ion OneTouch ™ ES Instrument.

Finally, enriched ISPs were sequenced with the Ion Personal Genome Machine (PGM™) system and Ion 316™ or 318™ Chip (Life Technologies) according to the manufacturers' procedures.

## 3.4 Sequencing data analysis

Raw data from Personal Genome Machine (PGM) were processed using the Torrent Suite Software analysis pipeline.

The raw unmapped reads were aligned to human hg19 reference genome with the Torrent Mapping Aligment program (TMAP).

Variant detection and annotation were performed using the Variant Caller Plugin and Annovar software respectively.

Single nucleotide variants (SNV) and small insertion or deletion (indels) were annotated with allele frequency in the GnomAD Aggregation Consortium (http://gnomad.broadinstitute.org), presence in human gene mutation databases (Human Gene Mutation Database - HGMD, ClinVar), InterVar clinical interpretation and various bioinformatic tools evaluating deleterious effect on the protein and/or evolutionary conservation metrics like SIFT, Polyphen, Mutation Taster, Mutation Assessor and CADD. For synonymous changes, and intronic variants we run Human Splicing Finder tool (HSF, http://www.umd.be/HSF3/HSF.shtml) in order to identify mutations possibly leading to splicing defects.

We filtered out all the variants with a minor allele frequency (MAF) more than 0.01 in GnomAD database and synonymous changes or intronic variants with probably no impact on splicing as result of HSF prediction.

After filtering, the output candidate variants were checked by Sanger sequencing using the BigDye Terminator cycle Sequencing V3.1 (Applied Biosystems) and 3500 Dx Genetic Analyzer Instrument (Applied Biosystems). Likewise, segregation analysis of variants was performed in available family members.

The true positive variants, confirmed by Sanger sequencing gold standard, have been classified in 3 different categories following the American College of Medical Genetics and Genomics guidelines:

a. known pathogenic mutations previously described in the literature;
b. likely pathogenetic variants;
c. missense variants of uncertain significance (VUS)

In order to determine whether VUS could be regarded as pathogenic, we evaluated their potential effect on protein function (Table 2S) using either word-based or numerical bioinformatic software (SIFT, Polyphen, Mutation Taster, Mutation Assessor and CADD). To compare predictions across the different programs, an effect was considered as "deleterious", when the output was "deleterious" (SIFT), "probably damaging" (PolyPhen-2), "possibly damaging" (PolyPhen-2), "disease-causing" (MutationTaster), "high" (MutationAssessor) and "medium" (MutationAssessor), or as "tolerated", when the output was "tolerated" (SIFT), "benign" (PolyPhen-2), "polymorfism" (MutationTaster), "low" (MutationAssessor), or "neutral" (MutationAssessor). For the numerical tools CADD we set a threshold of 15 for deleteriousness (the threshold on

deleteriousness suggested by developers is between 10 and 20). We considered the concordance between this five bioinformatic tools for the prioritization of variant of uncertain significance.

Moreover, in order to predict the presence of copy number variation (CNV) from high throughput sequencing data we implemented an algorithm exploiting amplicon read depth.

First we excluded from the analysis all the amplicons that did not achieve a median coverage of 30X across all sample, we then performed a two-step normalization for each library to account for sample and sequencing run variations.

In the first step we computed an intra-sample normalization. In details, we calculated the ratio between the reads count per amplicon and the global mean read count of each library obtained considering all the amplicons within autosomal chromosomes.

In the second step, we performed an inter-sample normalization determined dividing the intra-sample normalization ratio of a specific amplicon by the mean of the intra-sample normalization ratios obtained for the same amplicon from 10 control libraries.

Inter-sample normalization ratios lower than 0.7 indicate the putative presence of heterozygous deletions while ratios higher than 1.3 suggest possible duplications.

# 4 RESULTS AND DISCUSSION

## 4.1 Spectrum of variants identified by IT-NGS

IT-NGS screening allowed us to detect rare (MAF<0.01) variants (N=122) in 80 of the 97 unrelated probands with a suspected diagnosis of inherited thrombocytopenia.

Since in the remaining 17 affected individuals no potential disease-causing mutations were detected, their platelet defects are likely due to mutations in genes not included in the IT-NGS design.

Of the 122 filtered alterations, 6 were homozygous variants of the *GP1BB* (N=2), *GP9* (N=2), *NBEAL2* (N=1), or *TUBB1* (N=1) genes, 4 hemizygous variants of the X-linked *FLNA* (N=2), *GATA* (N=1) or *WAS* (N=1) genes in four males and 112 heterozygous variants.

Overall, they affect the coding regions or the flanking "ag"/"gt" canonical splicing dinucleotides of 24 different genes, as well as the 5'-UTR of the *ANKRD26* gene. Synonymous and intronic variants occurring outside the acceptor and donor splice sites with probably no impact on splicing, as result of HSF prediction, were excluded from this study.

Since 10 alterations were detected in more than one patient (Table 3), 100 were the different variants identified (Suppl. Table 1).

| Gene | Genomic Variation | Allele count / Tot allele | | Allele frequency | | p-value |
|---|---|---|---|---|---|---|
| | | our cohort | GnomAD NFE | our cohort | GnomAD NFE | |
| ABCG5 (NM_022436) | c.1864A>G (p.Met622Val) | 3/194 | 887/126068 | 0.0155 | 0.00704 | 0.16 |
| | c.293C>G (p.Ala98Gly) | 2/194 | 209/96148 | 0.0103 | 0.00217 | 0.068 |
| ITGA2B (NM_000419) | c.2602G>A (p.Val868Met) | 3/194 | 342/89426 | 0.0155 | 0.00382 | 0.040 |
| VWF (NM_000552) | c.5191T>A (p.Ser1731Thr) | 2/194 | 136/129190 | 0.0103 | 0.00105 | 0.019 |
| | c.1781C>G (p.Ala594Gly) | 2/194 | 18/76526 | 0.0103 | 0.000235 | 0.0012 |

| | | | | | | |
|---|---|---|---|---|---|---|
| TUBB1 (NM_030773) | c.326G>A (p.Gly109Glu) | 5^/194 | 209/129178 | 0.0206 | 0.00162 | 1.99e-05 |
| **MPL (NM_005373)** | **c.1327G>C (p.Gly443Arg)** | **2/194** | **0/56406** | **0.0103** | **0** | **1.17e-05** |
| **MYH9 (NM_002473)** | **c.4535C>T (p.Ser1512Phe)** | **2/194** | **0/~251294** | **0.0103** | **0** | **5.92e-07** |
| | **c.5521G>A (p.Glu1841Lys)** | **2/194** | **0/~250896** | **0.0103** | **0** | **5.94e-07** |
| **GP1BA (NM_000173)** | **c.515C>T (p.Ala172Val)** | **10/194** | **0/~249014** | **0.0515** | **0** | **2.2e-16** |

**Table 3. Variants identified in more than one probands**. Comparisons of allele frequencies between our IT case series and controls from GnomAD Non-Finnish European population using Fisher exact test. Five alleles are significantly ($p<0.001$) over-represented in our case series. In bold are reported the variants with known pathogenic effect significantly over-represented in ITs cohort compared to controls as expected.
^5 alleles detected in 4 probands (three heterozygous and one homozygous).

Among the 100 unique variants, 11 were deemed as deleterious (Table 4 and 5), as they were stop gain (N=1 novel), frameshift (N=3 novel; N=1 known), start lost (N=1 novel), splicing (N=1 novel) mutations, or known nucleotide substitutions affecting the 5' untranslated region of *ANKRD26* gene (N=4 known). The 89 left were amino acids substitutions (Figure 3). Of these, 14 were pathogenic as they had previously been associated with ITs, and 75 were novel or described in public databases but with no reported ITs association. This relatively high number of potential pathogenic missense variants is consistent with the spectrum of mutations in ITs, which are mainly autosomal dominant diseases due to amino acid substitutions.

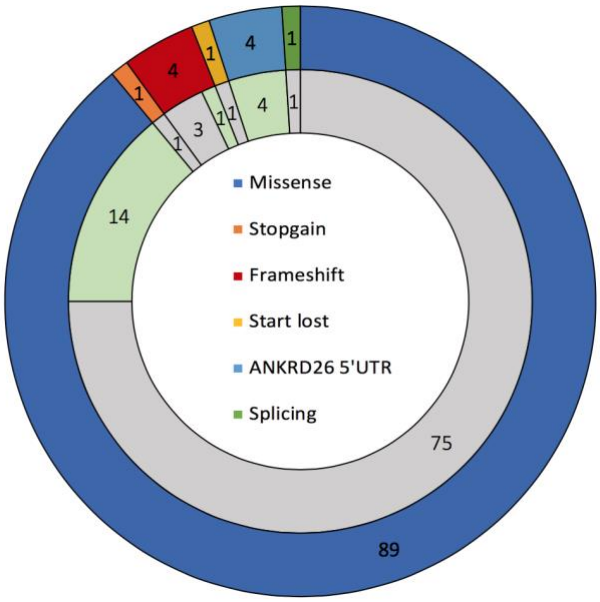


**Figure 3. Frequency of the different types of variants identified.** For each category (outer track), the inner track shows the number of novel variants (grey) vs known deleterious variants previously reported in literature as IT-related (green).

In order to ascertain whether the 75 amino acid substitutions never associated with IT could be regarded as pathogenic, we evaluated their potential effect on protein function using five predictive tools described in Material and Methods section: SIFT, Polyphen, Mutation Taster, Mutation Assessor and CADD (Suppl. Table 1). All these programs confer a "deleterious" effect to 17 of the 75 amino acid substitutions. On the contrary, 13 missense variants were evaluated as "tolerated" by all the five tools. For the remaining 45 variants, outputs were not always concordant, being four (N=17), three (N=8), two (N=8), or one (N=12) the in silico tools that predict a deleterious effect on protein function.

Since these predictive software take in account different parameters, we are not surprised to find discordant prediction. For this reason, functional assays should be developed to assess whether mutant proteins have a pathogenic effect.

Considering different aspects, such as type of mutation, predictive tools, model of inheritance and phenotypic features observed in probands we made a definite diagnosis in 37 (38%) of the 97 probands included in this studies, as described below in more detail.

## 4.2 Molecular diagnosis in families with known IT-causing mutations

In 30 index cases, we identified IT-causing mutations reported in the literature (Table 4). Overall we recognized 27 families with the monoallelic or biallelic forms of BSS, *ACTN1*-RT, *MYH9*-RD, *ANKRD26*-RD, or CAMT.

In three probands the identification of a mutated *MPL* allele alone did not allowed us to reach a definite diagnosis, since CAMT is an autosomal recessive disease (Table 4).

| Proband | Gene | Genomic variation | Status | Protein effect | Disease | Literature citation |
|---|---|---|---|---|---|---|
| IT1 - IT10 | GP1BA (NM_000173) | c.515C>T | het | p.Ala172Val | BSSA2 | Savoia (2001) Balduini (2009) Noris (2012) |
| IT11 | | c.104delA | het | p.Lys35Argfs*4 | BSSA2 | Li (1996) |
| IT12 | GP1BB (NM_000407) | c.179T>C | het | p.Leu60Pro | BSSA2 | Ferrari (2018) |
| IT13 | GP9 (NM_000174) | c.182A>G | hom | p.Asn61Ser | BSSA1 | Wright (1993) Kanda (2017) |
| IT14 | | c.284A>G | hom | p.Tyr95Cys | BSSA1 | Savoia (2011) |
| IT15 | ACTN1 (NM_001102) | c.313G>A | het | p. Val105Ile | ACTN1-RT | Kunishima (2013) |
| IT16 | | c.673G>A | het | p.Glu225Lys | ACTN1-RT | Kunishima (2013) |
| IT17 | | c. 2212C>T | het | p.Arg738Trp | ACTN1-RT | Kunishima (2013) |
| IT18 | | c.2255G>A | het | p.Arg752Gln | ACTN1-RT | Kunishima (2013) |
| IT19 | MYH9 (NM_002473) | c.279C>G | het | p.ASn93Lys | MYH9-RD | Seri (2000) |
| IT20 | | c.2680G>A | het | Glu894Lys | MYH9-RD | Saposnik (2014) |
| IT21 IT22 | | c.5521G>A | het | p.Glu1841Lys | MYH9-RD | Seri (2000) Ruhoy (2016) Cechova (2018) |
| IT23 | ANKRD26 (NM_014915) | c.-116C>T | het | - | THC2 | Pippucci (2011) Greene (2017) Perez Botero (2018) |
| IT24 | | c.-118C>T | het | - | THC2 | Pippucci (2011) Marquez (2014) Greene (2017) |
| IT25 | | c.-126T>G | het | - | THC2 | Noris (2011) |

| | | | | | | |
|---|---|---|---|---|---|---|
| IT26 | | c.-128G>A | het | - | THC2 | Pippucci (2011)<br>Greene (2017)<br>Zaninetti (2017) |
| IT27 | MPL (NM_005373) | c.1904C>T<br>c.1210G>A | compound het | p.Pro635Leu<br>p.Gly404Arg^ | CAMT | Oudenrijn (2000)<br>Tonelli (2000)<br>Tijssen (2008) |
| IT28<br>IT29 | | c.1327G>C | het | p.Gly443Arg | CAMT ? | Savoia (2007) |
| IT30 | | c.304C>T | het | p.Arg102Cys | CAMT ? | Steele (2005) |

**Table 4.** List of the ITs-related mutation reported in the literature and identified in our cohort.
^ p.Gly404Arg is a novel variant detected in trans with the p.Pro635Leu in MPL gene which is reported in literature as CAMT disease-causing mutation.

### 4.2.1 Bernard-Soulier syndrome

In 10 families we identified the heterozygous c.515C>T (p.Ala172Val) substitution of the *GP1BA* gene. This mutation, which is relatively frequent in the Italian population, is associated with BSSA2, the autosomal dominant form of Bernard-Soulier syndrome.
Consistent with a founder effect (Noris et al., 2012), all these patients shared the same haplotype at the *GP1BA* locus defined by the c.482C>T substitution (p.T161M) and 3 repeats of the variable number of tandem repeat polymorphism (allele B) (data not shown). In 5 families, the segregation analysis showed that the mutation was transmitted to all the affected family members, further supporting its pathogenic effect.
In recent years, different heterozygous mutations not only of *GP1BA* but also of the *GP1BB* gene have been correlated with mild macrothrombocytopenia (Sivapalaratnam et al., 2017), this suggests that the prevalence of monoallelic BSS could be underestimated.
Indeed, we identified two additional BSSA2 families, the former carrying the c.104delA (p.Lys35Argfs*4) deletion of *GP1BA* and the latter with the c.179T>C (p.Leu60Pro) substitution of *GP1BB*. Interestingly, the p.Leu60Pro variant was detected in other 5 apparently unrelated Italian families: 2 were previously investigated by our group but not included in this IT case series and 3 were recently reported in the literature (Ferrari et al., 2018). Therefore, the p.Leu60Pro variant of *GP1BB* could be regarded as the second more frequent cause of monoallelic BSS in the Italian population.

In the context of BSS, we also recognized two families with BSSA1, the severe autosomal recessive form of BSS. We found that the affected individuals were homozygous for the c.182A>G (p.Asn61Ser) or c.284A>G (p.Tyr95Cys) missense mutations of the *GP9* gene. Molecular data are consistent with the clinical features of the disease, with mean platelet number of 45 x$10^9$/L, giant platelet and defect of platelet aggregation after stimulation with ristocetin.

### 4.2.2 ACTN1-related thrombocytopenia and MYH9-related disease

In 8 families the thrombocytopenia was associated with defects of proteins playing a critical role in cytoskeletal organization, such as $\alpha$-actinin 1 and non-muscle myosin heavy chain IIA which are encoded by *ACTN1* and *MYH9* genes, respectively. The seven distinct mutations identified in these two genes are all missense variants affecting well-known functional domains of each protein (Kunishima et al., 2013; Pecci et al., 2018).

Probands (IT15-IT18) carrying p.Val105Ile, p.Glu225Lys, p.Arg738Trp or Arg752Gln mutations of the *ACTN1* gene had the *ACTN1*-RT. Consistent with mild phenotype of this form of IT, their thrombocytopenia ranges from 90 to 155x$10^9$/L and their platelets are large in size (Faleschini et al., 2018).

In contrast to *ACTN1*-RT, the expressivity is variable in *MYH9*-RD. This disease shows strong genotype-phenotype correlations as the risk of developing extra-hematological non-congenital manifestations is due to specific mutations of *MYH9* (Pecci et al., 2014). Proband IT19 was a 2 months old infants with 30x$10^9$/L platelets at birth, carrying p.Asn93Lys de novo mutation of the *MYH9* gene. This mutation affects the head domain of non-muscle myosin IIA and is associated with high risk of deafness but low risk of kidney damage or cataract .

On the contrary, in the two unrelated probands (IT21 and IT22) carrying p.Glu1841Lys substitution, the thrombocytopenia (40-50x$10^9$ platelets/L) is expected to remain the only feature of the disease throughout their life with very low risk of non-congenital defects (Pecci et al., 2014).

For the p.Glu894Lys mutation identified in IT20 proband there are no genotype-phenotype correlation data (only two families reported in Saposnik et al., 2014).

However, since this mutation hits the coil-coiled tail of the non-muscle myosin IIA it is likely associated with a mild phenotype characterized by thrombocytopenia and low risk of extra-hematological features, like p.Glu1841Lys and the other mutations affecting this domain. Thus, the identification of the mutations affecting *MYH9* gene is important not only to establish a correct diagnosis but also to set a personalized follow-up program for the patients.

### 4.2.3 ANKRD26-related thrombocytopenia

In four family we identified one (c.-116C>T, c.-118C>T, c.-126T>G, and c.-128G>A) of the 13 nucleotide substitutions identified in the 5'-UTR of *ANKRD26* gene (Noris et al., 2011; Pippucci et al., 2011). Considering that these mutations are associated with increased susceptibility to hematological malignancies, these patients will receive a personalized follow up in order to monitor the possible onset of leukemia.

### 4.2.4 Known IT-related mutations in MPL gene

Patient IT27 was a compound heterozygote for the c.1904C>T (p.Gly404Arg) and c.1210G>A (p.Pro635Leu) variants of *MPL*, the gene encoding for thrombopoietin receptor. The paternally inherited p.Pro635Leu substitution is a known mutation whereas the maternally transmitted p.Gly404Arg variant is novel. All the predictive tools classified p.Gly404Arg as deleterious (Suppl. Table 1). Moreover, the trans condition of the two variants supports his possible deleterious effect. Consistent with a diagnosis of CAMT, the three-years old boy had severe thrombocytopenia and no megakaryocytes in his bone marrow.

In the same gene, we also identified two mutations reported in the literature (p.Gly443Arg or p.Arg102Cys) in three probands (IT28 - IT30; Table 4). Consistent with their reported pathogenic effect, these variants are extremely rare: in GnomAD database Gly443Arg has a MAF of 7.15e-06 whereas p.Arg102Cys has a MAF of 1.22e-05 (Suppl. Table 1).

Since they were heterozygous, we cannot exclude the presence of a second mutant *MPL* allele not detected by IT-NGS. However, the phenotype of these patients is not ascribable to CAMT, being their thrombocytopenia moderate and their megakaryocytes – when bone marrow aspiration was available for analysis – normal in number and morphology. Moreover, segregation analysis showed that the two affected siblings of IT30 family were heterozygous for p.Arg102Cys, which was inherited from the healthy mother.

Although we should regard probands IT28-IT30 as asymptomatic carriers of CAMT and hypothesize that their thrombocytopenia is caused by mutations in different genes, it would be interesting to assess whether haploinsufficiency or dominant negative effect of their mutant MPL receptor could interfere with the physiological processes of megakaryopoiesis and platelet production in specific genomic or environmental context.

For instance, in family IT30 the affected children may have inherited an “asymptomatic” hypomorphic variant of *MPL* from their healthy father, lowering but not abolishing, as in CAMT, the activity of the thrombopoietin receptor.

For instance, in family IT30 the affected siblings could have inherited from their healthy father an “asymptomatic” hypomorphic variant of *MPL* that reduce but not abolish as in CAMT, the activity of the thrombopoietin receptor.

Alternatively, we can hypothesize that a 50% reduction of the thrombopoietin receptor activity combined with the effect of variants in other IT genes could impair platelet production in a digenic or oligogenic model of inheritance. For instance, proband IT28 is heterozygous for missense variants of *FLNA* (p.Ala2235Gly) and *NBEAL2* (p.Thr2525Met) genes, which being classified as deleterious or likely deleterious by the in silico analyses could further contribute to impair the platelet biogenesis process.

## 4.3 Molecular diagnosis in families with novel likely pathogenic mutations

In 11 index cases, we detected novel likely pathogenic variants in one of the IT genes (Table 5), whose potential effect on the protein function is discussed below.

| Proband | Gene | Genomic variation | Status | Protein effect | Evidence for pathogenicity |
|---|---|---|---|---|---|
| IT31 | GP1BA (NM_000173) | c.169A>G | het | p.Asn57Asp | Deleterious (4/5) + SE1 |
| IT32 | GP1BB (NM_000407) | c.1A>C | het | p.Met1? | start loss |
| IT33 | | c.528_550del | hom | p.Arg177Serfs*124 | frameshift deletion |
| IT34 | | c.347T>C | apparently hom | p.Leu116Pro | Deleterious (3/5) |
| IT35<br>IT36 | MYH9 (NM_002473) | c.4535C>T | het | p.Ser1512Phe | Deleterious (5/5) + SE2 |
| IT37 | | c.4563C>A | het | p.His1521Gln | Deleterious (5/5 ) + SE1 |
| IT38 | NBEAL2 (NM_015175) | c.6212G>C | hom | p.Arg2071Pro | Deleterious (5/5) |
| IT39 | WAS (NM_000377) | c.680dupG | hem | p.Ser228Leufs*10 | frameshift insertion |
| IT40 | RUNX1 (NM_001754) | c.614-2A>G | het | - | splicing |
| IT41 | ACTN1 (NM_001102) | c.2210C>G | het | p.Thr737Ser | Deleterious (2/5) + SE1 *but excluded by segregation and functional studies |
| IT42 | TUBB1 (NM_030773) | c.165C>A | het | p.Tyr55* | nonsense |
| IT43<br>IT44<br>IT45<br>IT46 | | c.326G>A | het<br>het<br>hom<br>het | p.Gly109Glu | Deleterious (5/5) + significant overrepresentation in our cohort |
| IT47 | | c.742G>A | het | p.Ala248Thr | Deleterious (4/5) + localization in $\alpha/\beta$ intradimer interface |

**Table 5**. **ITs patients with novel likely pathogenic variants.** Abbreviation: SE1 (supporting evidence 1) → novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before, SE2 (supporting evidence 2)→ cosegregation with disease in multiple affected family members in a gene known to cause the disease

### 4.3.1 Frequent forms of Inherited Thrombocytopenia

According to the relatively high prevalence of Bernard-Soulier Syndrome and *MYH9*-related disease in the Italian IT cohort (Balduini and Noris, 2016), we identified novel variants affecting their causative genes.

In two families (IT31 and IT32), we detected a heterozygous mutation of *GP1BB* (c.1A>C/p.Met1?) or *GP1BA* (c.169A>G/p.Asn57Asp). Whereas the start loss variant is regarded as deleterious, the p.Asn57Asp variant has a CADD score of 23.5 and is predicted to be pathogenic by three word-based tools (Suppl. Table 1). Moreover, it affects the same residue of p.Asn57His substitution which was previously reported in two unrelated families with BSSA2 (Berndt and Andrews, 2011; Vettore et al., 2008).

The presence of this missense change at an amino acid residue previously reported as mutated support the likely pathogenic effect of p.Asn57Asp variant. Moreover, laboratory and clinical data are consistent with a diagnosis of monoallelic BSS for IT31 proband.

The IT33 affected individual has a novel homozygous mutation (c.528_550del/p.Arg177Serfs*124) in the *GP1BB* gene, which leads to a diagnosis of biallelic BSS. Interestingly, this deletion is the first alteration that destroys the cytoplasmic region (residues 178-206) of the relative glycoprotein. In collaboration with Dr. Alessandro Pecci (IRCCS San Matteo, Pavia) further investigations are in progress to determine the biochemical and physiological consequence of this mutation on megakaryopoiesis and platelet production.

In the same gene, we identified another novel apparently homozygous variant (c.347T>C/p.Leu116Pro) inherited from the father but not from the mother of proband IT34, a 1 year old girl. Multiple sequence alignments showed that p.Leu116Pro is highly conserved in mammalians (Figure 4A).

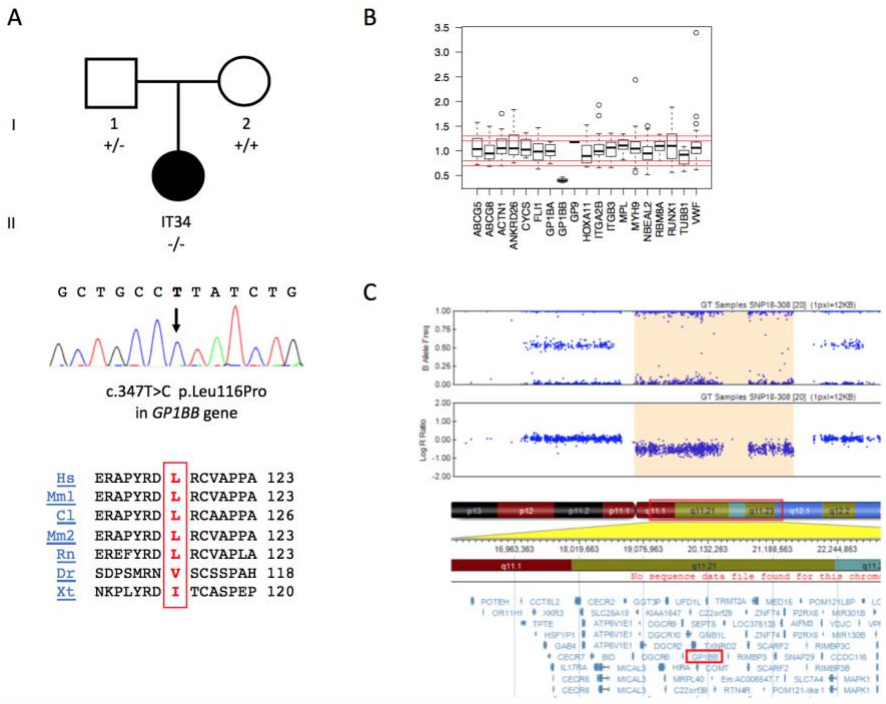


**Figure 4.** A) Pedigree of IT34 family and direct sequencing of PCR products showing the p.Leu116Pro mutations in GP1BB. Alignment of *GP1BB* orthologs from different species is shown. The mutated residue is boxed and highlighted in red. Hs, H.sapiens (NP_000398), Mm1, M.mulatta (XP_001105321), Cl, C.lupus (XP_005636645), Mm2, M.musculus (NP_034457), Rn, R.norvegicus (NP_446382), Dr, D.rerio (NP_001138284) and Xt, X.tropicalis (XP_002931927) at https://www.ncbi.nlm.nih.gov/homologene/30972.
B) Box plots representing the distribution of the inter-sample normalization ratio of the amplicons covering 19 IT-NGS autosomal genes in IT34 proband. A median between 0.7 and 1.3 is indicative of two copy of the gene whereas a median below 0.7 suggest the presence of one copy of the gene. The circles show outlier values. C) SNP array analysis showing the presence of 2.6 Mb hemizygous deletion of chromosome 22q11.2 including GP1BB.

Since clinical features, as well as the ristocetin response, of this affected individual are consistent with a diagnosis of biallelic BSS, to determine whether the apparent homozygosity is due to a de novo or maternally inherited deletion we carried out copy number variation analysis (CNV) from high throughput sequencing data using the algorithm we implemented described in Material and Methods section.

The amplicons covering the *GP1BB* gene had a median inter-sample normalization ratio lower than 0.7, this suggests the presence of a single copy of *GP1BB* gene in IT34 proband (Figure 4C). On the contrary, the medians of the other autosomal genes reported in figure fall within the normal range (0.7-1.3) so IT24 had two copies of these genes.

In addition, this hypothesis was confirmed by SNP array analysis which identified a 2.6 Mb hemizygous deletion of chromosome 22q11.2 including *GP1BB*. Deletions of this region cause DiGeorge syndrome, a congenital malformation and neuropsychiatric

disorder. Affected individuals have mild to serious clinical features often including congenital heart disease, immunodeficiency, autoimmune disease, palatal abnormalities, hypocalcemia, thyroid disease, renal anomalies, skeletal anomalies characteristic facial features and thrombocytopenia (Morrow et al., 2018).

Although the spectrum of the *MYH9* mutations is relatively limited (more than 80% of patients have mutations affecting only 7 protein residues (Balduini et al., 2011); we identified two novel amino acid substitutions, c.4535C>T (p.Ser1512Phe) and c.4563C>A (p.His1521Gln) which cosegregate with thrombocytopenia in multiple affected family members (Figure 5). They are not reported in GnomAD and classified as deleterious by all the in silico predictive programs (Suppl. Table 1). Multiple sequence alignment showed that Ser1512 and His1521 are highly conserved between different species (Figure 5).

Moreover, the former has been identified in two apparently unrelated families (IT35 and IT36) whereas the latter affects an amino acid residue where a different missense change determined to be pathogenic has been seen before (Ghemlas et al., 2015).

In the cases carrying these two variants the immunofluorescence analysis showed the presence of Döhle-like inclusions in their neutrophils confirming the diagnosis of *MYH9*-RD.

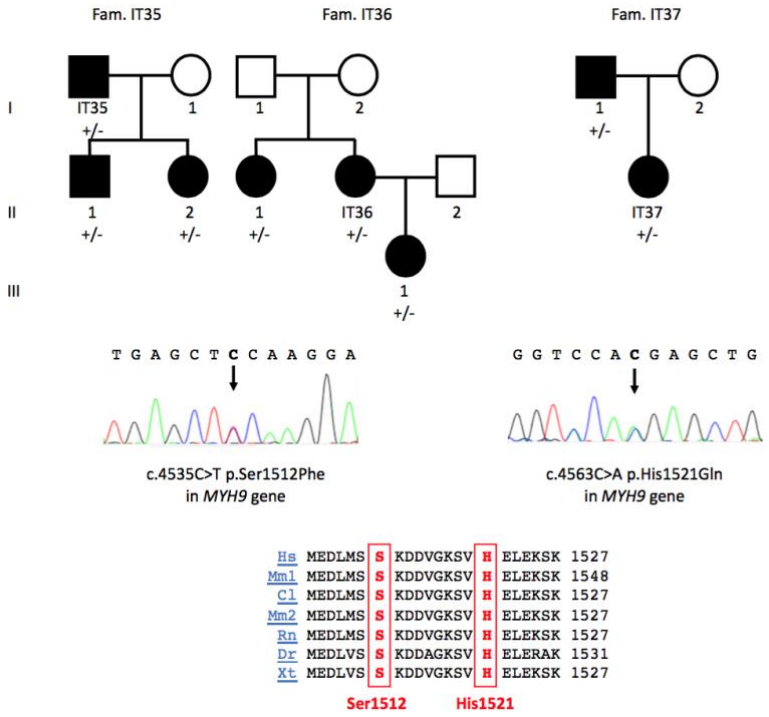


**Figure 5.** Pedigrees of families with novel mutation in MYH9 gene and direct sequencing of PCR products showing the respective mutations. For the p.Ser1521Phe and p.His1521Gln alignment of *MYH9* orthologs from different species is shown. Hs, H. sapiens (NP_002464), Mm1, M.mulatta (XP_015005719), Cl, C. lupus (NP_001104237), Mm2, M. musculus (NP_071855), Rn, R. norvegicus (NP_001292806), Dr, D. rerio (NP_001091647) and Xt, X. tropicalis (XP_017949131) https://www.ncbi.nlm.nih.gov/homologene/137255.

### 4.3.2 Rare forms of Inherited Thrombocytopenia

In three families we identified variants in genes *NBEAL2*, *WAS*, and *RUNX1*, whose mutations are responsible for very rare forms of IT, such as grey platelet syndrome, Wiskott-Aldrich syndrome, and platelet familial disorder with predisposition to acute myeloid leukemia, respectively.

**Family IT38.** In patient IT38, we identified a homozygous variant (c.6212G>C/p.Arg2071Pro) of the *NBEAL2* gene. Considering that, in addition to its homozygous status, all the predictive tools evaluated this substitution as deleterious and that is not present in GnomAD, we diagnosed this individual as affected by GPS. According to this conclusion, electron microscopy showed that the patient's platelets were "grey" for complete absence of alpha-granules (Bottega et al., 2017).

**Family IT39**. In a newborn male (IT39) with a suspected form of IT we identified a de novo hemizygous duplication c.680dupG (p.Ser228Leufs*10) of the X-linked *WAS* gene, leading to a suspicion diagnosis of WAS. Although the disease is one of the few forms of thrombocytopenia characterized by small-sized platelets, a useful feature in the differential diagnosis of IT, the patient's mean platelet volume was in the normal range. Of note, IT39 was also heterozygous for the c.4221_4222del (p.Ser1410Glnfs*14) frameshift mutation of the *NBEAL2* gene transmitted from the father. This variant could have compensated platelet size defects considering that the "asymptomatic" carriers of mutations in this gene have an increased platelet diameter but not thrombocytopenia (Bottega et al., 2013).

**Family IT40.** Proband IT40 was a carrier for the splicing c.614-2A>G mutation of *RUNX1* gene. The DNA sample from the father, who died of lymphoma, was not available for segregation analysis and the mutation was not inherited by his healthy mother.
Preliminary RNA studies suggested that this variant enhances exon 7 skipping when compared to control (data not shown). Considering that mutations of *RUNX1*, like those of *ANKRD26*, are associated with an increased risk of hematological malignancies, the proband will undergo dedicate protocols to ascertain early signs of leukemia evolution.

### 4.3.3 Functional study of *ACTN1* p.Thr737Ser variant

In IT41, we detected a heterozygous c.2210C>G (p.Thr737Ser) variant of the *ACTN1* gene. Although this substitution was classified as deleterious by only two predictive tools (MutationTaster and CADD; Suppl. Table 1), residue Thr737 is also affected by a different mutation previously reported in literature (p.Thr737Asn), whose pathogenic effect is supported by functional studies (Bottega et al., 2015). To assess the pathogenic role of this variant we performed an immunofluorescence assay in cells overexpressing the mutant form of *ACTN1*. The analysis showed that p.Thr737Ser does not affect the cytoskeleton organization of cells, unlike p.Thr737Asn, indicating that this amino acid substitution is likely to have no effect on platelet production (Figure 6). Then, we were able to perform segregation analysis showing that the variant was inherited from the

mother, who was not thrombocytopenic, leading us to definitely exclude p.Thr737Ser as a pathogenic variant of *ACTN1* (Faleschini et al., 2018).

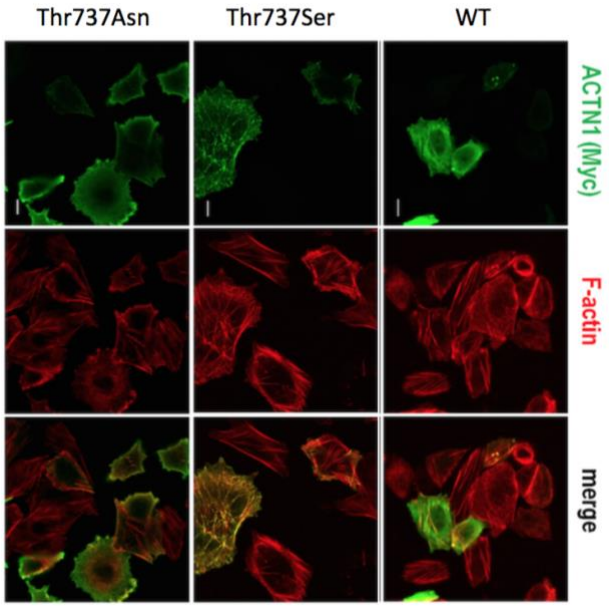


Figure 6. Immunofluorescence analysis of CHO cell line transiently transfected with Myc-tagged wild type (top panel) or mutant (lower panels) *ACTN1* cDNAs. The subcellular localization of the over-expressed α-actinin1 (green) was examined using a c-myc antibody while the actin filaments were stained with phalloidin (red). The cells shown are representative of three independent experiments. Scale bar, 10 µm Picture modified from the research originally published in Faleschini et al., 2018.

## 4.4 Potential role of *TUBB1* variants

*TUBB1* is a gene encoding tubulin-beta 1, a cytoskeleton component that plays a pivotal role in proplatelet formation and platelet release. Its role in IT is not clearly defined, as few are the mutations so far identified in association with reduced platelet count (Fiore et al., 2017; Kunishima et al., 2009, 2014). However, we identified three potential variants in TUBB1 gene that could explain IT in six families.

In the affected individual IT42, we detected a heterozygous nonsense mutation (c.165C>A/p.Tyr55*) (Table 5). Like the known p.Gln423* mutation (Fiore et al., 2017), the p.Tyr55* variant could be a loss-of-function allele leading to haploinsufficiency. However, its pathogenic effect should be further investigated by segregation analysis and functional studies to determine whether 50% expression level of the wild-type protein is enough to guarantee a platelet production above $150x10^9$ platelets/L.

In the same gene, we identified the c.326G>A (p.Gly109Glu) variant in 4 individuals (IT43-IT46; Table 5). Although this allele is classified as deleterious by all the

bioinformatic tools (Suppl. Table 1), it is relatively frequent in Non-Finnish European controls (MAF: 0.00162) where homozygous individuals are expected to be 1:400,000. Despite this, the p.Gly109Glu allele is significantly ($p<0.001$) overrepresented in our case series (Table 3) and one of the four patients (IT45) is homozygous for the variant. Consistent with our data supporting a potential pathogenic effect of p.Gly109Glu, this substitution is enlisted in the HGMD as associated with reduced platelet count (Auer et al., 2014).

The third variant we identified in *TUBB1* is c.742G>A (p.Ala248Thr), a very rare substitution (MAF $3.231 \times 10^{-5}$) classified as deleterious by all the in silico predictions.

Multiple sequence alignment showed that p.Ala248 residue is highly conserved in different species (Figure 7).

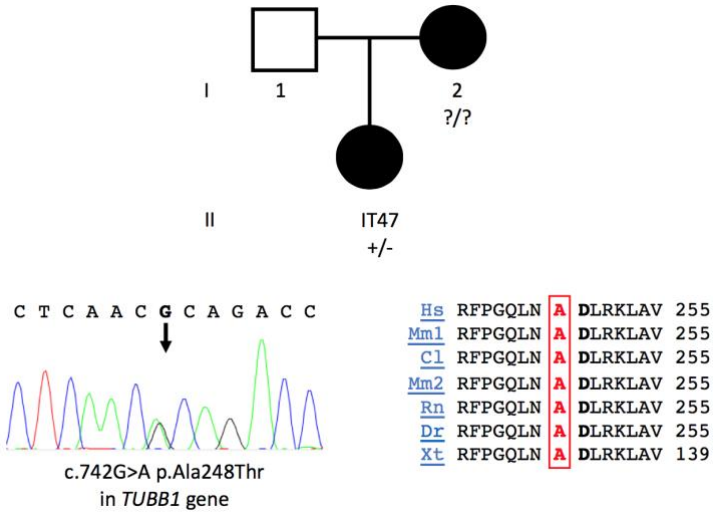


**Figure 7.** Pedigree of IT47 proband and direct sequencing of PCR products showing the p.Ala248Thr mutations in TUBB1. For the p.Ala248 residue alignment of TUBB1 orthologs from different species is shown. The mutated residue is boxed and highlighted in red. Hs, H. sapiens (NP_110400), Mm1, M.mulatta (XP_001082345), Cl, C. lupus (XP_005635298), Mm2, M. musculus (NP_001074440), Rn, R. norvegicus (XP_003749686), Dr, D. rerio (XP_005172182) and Xt, X. tropicalis ( XP_002941758) https://www.ncbi.nlm.nih.gov/homologene/69474.

As mentioned above, in addition to p.Gln423*, few are the known mutations of *TUBB1* associated with thrombocytopenia: Asp249Asn, which causes canine form of macrothrombocytopenia (Davis et al., 2008), p.Phe260Ser and p.Arg318Trp, both identified in two Japanese families (Kunishima et al., 2009, 2014). Of note, residues Ala248, Asp249, Phe260 and Arg318 are all located near the interface of the alpha- and beta-tubulin subunits, suggesting that disruption of the interdimer structure is likely to impair microfilaments assembly leading to defective proplatelet.

In order to confirm whether the p.Gly109Glu and p.Ala248Thr potential pathogenic variants play a role in platelet production we have been setting up studies to determine the effect of this substitution on protein function. Segregation analysis showed that in one family, the p.Gly109Glu variant cosegregates in three affected individuals.

## 4.5 Variants of uncertain significance

As mentioned above, 122 (100 different) variants were identified in 80 probands. Of these alleles, we regarded as pathogenic, likely or potential pathogenic 47 variants (32 different) in 46 affected individuals of our cohort (Tables 4 and 5). Of the remaining 75 alleles (68 different) of uncertain significance, 28 were present as a second or third variant in the 46 individuals just mentioned (Figure 8).


**Figure 8. Classification of probands sequenced with IT-NGS.** For each category in light blue is shown the distribution of variants split in known pathogenic (KPV), likely Pathogenic (LPV) or uncertain significance (VUS) .

Abbreviation → KD: probands with known pathogenic variants; MPL ?: probands heterozygous for MPL known mutations; LD: probands with likely pathogenic variants; TUBB1 ?: probands potentially affected by TUBB1-RT; NDD: proband with no definite diagnosis.

^ the box contain the novel MPL mutation p.Gly404Arg we detected in trans with p.Pro635Leu known mutation of MPL. * the frameshift heterozygous variant p.Ser1410Glnfs*14 in the NBEAL2 gene was added to the variant of uncertain significance.

In patients with confirmed molecular diagnosis (N=37) these additional variants could modulate the expressivity of the thrombocytopenia. On the contrary, the additional

variants detected in the patients with a “undefined” diagnosis (N=9), as they carry monoallelic alleles of *MPL* or variants of *TUBB1*, could have a determinant effect in causing thrombocytopenia.

In 34 individuals, we did not find any strong candidate variant (Figure 8). However, NGS analysis detected 47 missense alleles, with 2 and 9 index cases carrying three and two of these variants, respectively.

Between the variants classified as “deleterious” or “likely deleterious” by 5 or 4 bioinformatic tools (Suppl. Table 1), there are potential disease-causing mutations in *ETV6* and *GP1BA* genes. Conversely, uncertain is the role of *HOXA11* and *FLNA* variants in thrombocytopenia, as well as that of the numerous amino acid substitutions we identified in the *ABCG5*, *ABCG8*, *ITGA2B*, *ITGB3*, *GP1BA*, *GP9* and *VWF* genes, as reported below.

### 4.5.1 Missense variants predicted as deleterious

**ETV6**. In *ETV6*, we identified the c.1040A>C (p.Gln347Pro) substitution in proband IT93. It affects the DNA binding domain of transcription factor *ETV6*, where most of the *ETV6* mutations localize (Melazzini et al., 2016b; Poggi et al., 2017; Topka et al., 2015), suggesting that p.Gln347Pro is likely to be pathogenic (Figure 9).

Using a reporter gene (luciferase) under the control of *ETV6* target promoter (MMP-3), we have been testing the activity of luciferase after transfection of this mutant form, as well as of c.65G>A (p.Ser22Asn), another *ETV6* variant with conflicting pathogenic prediction (deleterious for two bioinformatic tools, Suppl. Table 1). Of note, since mutations of *ETV6* cause a mild form of thrombocytopenia with an increased risk to hematological malignancies, it is pivotal for the correct management of patients to distinguish between pathogenic and neutral *ETV6* variants.

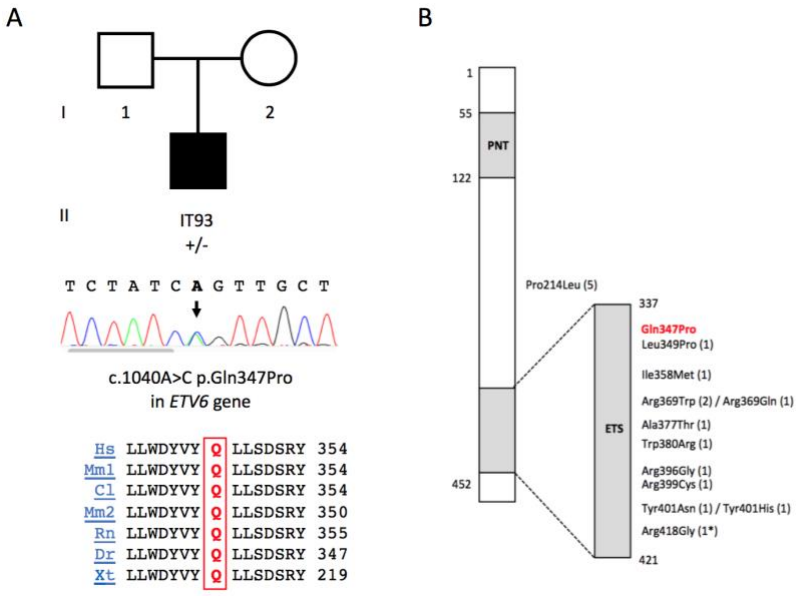


**Figure 9.** A) Pedigree of IT93 proband and direct sequencing of PCR products showing the p.Gln347Pro mutations in ETV6. B) Localization of 12 mutations reported so far in ETV6. The novel mutations here reported is in red.

**GP1BA.** Two additional strong candidate variants, c.1619G>C (p.Trp540Ser) and c.1974C>G (p.His658Gln) of *GP1BA* and *SLFN14*, respectively have been identified in proband IT79. The former has been reported in 5 out of 248948 alleles (MAF 2.0 x $10^{-5}$) whereas the latter is a novel variant not reported in GnomAD. The patient has a mild thrombocytopenia (124 x $10^{6}$/L platelet) with large platelet (MPV: 10.2) that is ascribable to a potential diagnosis of the monoallelic form of BSS and not to alterations of *SLFN14*. Indeed, the thrombocytopenia associated with mutations of *SLFN14*, which has been described in only four families, is characterized by severe thrombocytopenia, defective platelet ATP secretion and an increased risk of bleeding tendency, which is absent in monoallelic BSS. Moreover, p.His658Gln hits a residue far from the small region where the *SFLN14* mutations (p.Lys208Glu, p.Lys219Asn, p.Val220Asp, and p.Arg223Trp) identified so far localize (Marconi et al., 2016).

**FLNA.** In probands IT76 and IT80 two variants predicted as deleterious or likely deleterious have been identified in *FLNA*, a X-linked gene whose mutations are responsible for wide spectrum of rare brain, heart and muscle developmental diseases, including the periventricular nodular heterotopia. Although it is not a strong candidate for IT, the *FLNA* gene has been included in the NGS design because periventricular

nodular heterotopia (X-linked dominant disease) has sometimes been associated with thrombocytopenia and a novel missense variant has been identified in an isolated thrombocytopenia (Nurden et al., 2011). In addition to reduced platelet count, our patients do not have other clinical manifestations, so the interpretation of the variants effect is not trivial. Moreover, all the thrombocytopenic individuals with *FLNA* mutations reported in literature were females but one of our probands was male with healthy heterozygous mother, this further complicate the evaluation of the role of these variants in thrombocytopenia.

**HOXA11.** Two of the variants predicted as deleterious affect *HOXA11* (Suppl. Table 1), a transcription factor that regulates morphogenesis and hematopoiesis. Whereas c.347A>G (p.His116Arg) is a novel variant not reported in public database, the c.396G>C (p.Arg132Ser) missense is relatively frequent having a MAF of $1 \times 10^{-3}$ in the European non-Finnish population.

Moreover, the variant p.His116Arg was inherited in patient IT55 together with a substitution in *ACTN1* gene (c.1822C>T/p.Arg608Trp) similarly predicted as deleterious but which does not cosegregate with thrombocytopenia.

Despite their potential pathogenic effect, the *HOXA11* variants are not responsible for thrombocytopenia, as the only mutation reported so far in this gene is associated with severe platelet reduction evolving to aplastic anemia and radio-ulnar synostosis, all features not observed in families IT55 and IT60.

### 4.5.2 IT genes frequently hit by missense variants

Numerous substitutions were frequently identified in genes of various length. In the following sections we discussed the putative effect of these variants.

**ABCG5 and ABCG8.** Sitosterolemia is a recessive disease characterized by high level of plant sterols caused by mutations of *ABCG5* and *ABCG8*. These two genes were included in the IT-NGS panel because sitosterolemia is associated with hematological abnormalities, including macrothrombocytopenia.

In our cohort of patients, 10 heterozygous variants of *ABCG5* or *ABCG8* were identified in 13 probands. Eight variants were classified as benign or likely benign, including p.Arg164Gln of *ABCG8* that - even if deleterious for bioinformatic tools - is classified as benign by both InterVar and ClinVar (data not shown). Although the remaining two are enlisted in HGMD as sitosterolemia causing mutation (p.Arg184His of *ABCG8*) or as variant associated with low platelet count (p.Ala98Gly of ABCG5, Ali et al., 2016), probably they are not the cause of thrombocytopenia as in families with sitosterolemia the heterozygous cases are not thrombocytopenic. Consistent with the hypothesis of p.Ala98Gly neutral effect, the frequency of this variant detected in more than one heterozygous probands of our case series, is comparable to that observed in controls (Table 3).

**ITGA2B and ITGB3.** We also identified 7 variants of the *ITGA2B* and *ITGB3* genes in 9 probands. These genes encode for the GPIIb and GPIIIa integrin subunits of the fibrinogen receptor and when mutated cause Glanzmann thrombasthenia, an autosomal recessive bleeding disorder with normal platelet count. However, there are few gain-of-function variants, mainly affecting the integrin residues located between the transmembrane and cytoplasmic domain, that are responsible for an autosomal dominant form of macrothrombocytopenia without platelet dysfunction. None of the variants we identified have been reported in association with Glanzmann thrombasthenia or other platelet defects. Although in silico predictions suggest a potential deleterious effect for two of them (p.Asp459Asn of *ITGA2B* and p.Arg622His of *ITGB3*), all the variants identified affect the extracellular domain where no IT variant has been reported so far. Consistent with this conclusion, the p.Val868Met of *ITGA2B*, which was identified in three heterozygous individuals, is not significantly overrepresented in our thrombocytopenic case series (Table 3).

**GP1BA and GP9**. Excluding the 7 missense variants of the BSS genes we regarded as pathogenic, likely or potential pathogenic (Tables 2 and 3, and that identified in patient IT79 discussed above), 6 heterozygous variants of the *GP1BA* (N=4) and *GP9* (N=2) genes, respectively were identified in six affected individuals. All of them have been

classified as tolerated by three to five bioinformatic tools. However, considering that, as mentioned above, the monoallelic form of BSS is relatively frequent not in Italy but also in other populations (Sivapalaratnam et al., 2017), their potential effect will be tested in an in vitro assay we are setting up.

**VWF**. The *VWF* gene encodes for the von Willebrand factor (VWF), a crucial player of platelet adhesion to the sub-endothelium after vascular injury. The *VWF* mutations cause the von Willebrand disease (VWD), a genetic bleeding disease classified into different types based on qualitative or quantitative defects of VWF. Since the "classical" form of type 2B VWD is associated with moderate or severe thrombocytopenia, the *VWF* gene has been included in our IT target NGS design. The screening allowed us to identify numerous (N=11) variants in 13 probands, six of which are enlisted in HGMD as pathogenic (N=4) or potential disease-causing (N=2) variants. However, although three additional variants are predicted as potentially deleterious, none of these affect the A1 (residues 1270-1480) domain where the mutations associated with low platelet count are detected (Bellissimo et al., 2012). Finally, p.Ser1731Thr and p.Ala594Gly are not significantly ($p<0.001$) over-represented in our case series though they were identified in more than one proband (Table 3). Altogether, these elements suggest that the variants identified in *VWF* are not directly involved in reduced platelet count although they could worsen the phenotype predisposing carriers to increased risk of bleeding tendency.

# 5 CONCLUSION

Targeted-NGS is a useful screening strategy that allows multiplexing and simultaneous analysis of many genes at once at relatively low costs per patient. For diseases characterized by high level of genetic heterogeneity, such as ITs, NGS has become a helpful approach to avoid the investigations proposed by the IT diagnostic algorithm aimed at identifying potential candidate genes for mutational screening (Balduini et al., 2013).

Using the IT-NGS, we analyzed 97 consecutive probands with a diagnostic suspicion of IT. In 80 (82%) we detected variants in one or more of the 28 IT genes included in the IT-NGS design. In 27 families (28%), the identification of IT-causing mutations previously reported in literature allowed us to reach a conclusive molecular diagnosis. In another 10 (10%) index cases, we detected variants we classified as "deleterious" because of a series of evidence: (i) they were nonsense, start loss, frameshift, and splicing mutations; (ii) the amino acid change hit a residue previously reported to be affected by a different substitution; (iii) unrelated families shared the same variant which cosegregates with thrombocytopenia (iiii) the model pattern of inheritance and the gene mutated were compatible with the phenotype observed in the affected individuals. Using these strictly conditions, a definite clinical and molecular diagnosis was achieved in 38% of families without any previous investigations.

The majority (44%) of the probands carrying one or more variants did not receive any definite diagnosis. Among these, three patients were heterozygous for disease-causing mutations of *MPL* and therefore regarded as "asymptomatic" carriers of CAMT. Although the prevalence of CAMT is unknown, it is unlikely that 3 out of 97 individuals are heterozygous, as the disease is very rare (less than 100 cases reported so far). For this reason, we hypothesize that an alteration in another gene could be responsible for the disease in a digenic model of inheritance. However, since the parents of this probands are healthy, we cannot exclude the presence – for instance in the MPL regulatory regions - of a second allele in trans with the known mutation that could exert hypomorphic effect.

Moreover, five families were heterozygous or homozygous for variants of *TUBB1*, including a premature stop codon, one novel amino acid substitution affecting the tubulin interface domain, or a relatively common variant that is overrepresented in our IT cohort when compared to GnomAD Non-Finnish European population. Although the role of TUBB1 in thrombocytopenia is not clearly defined (at present, 3 families reported), these *TUBB1* variants could be involved in platelet number reduction.

In another two probands, two amino acid substitutions of the *ETV6* (p.Gln347Pro) and *GP1BA* (p.Trp540Ser) genes were regarded as strong candidate mutations. Both were classified as deleterious by the in silico tools. Consistent with the bioinformatic predictions, the former affects the DNA binding domain of the transcription factor where the majority of the mutations have been identified whereas the latter is compatible with diagnostic suspicion of monoallelic BSS.

Despite all these considerations, the pathogenic role of the variants identified in these 10 thrombocytopenic individuals should be supported by additional investigations.

First of all, it would be important to carry out segregation analysis, which is often limited to few family members. Then, we should perform functional studies, which are usually difficult to implement and time consuming. To this end, we have set up some assays to study *ACTN1*, *ETV6*, and *ANKRD26* variants. Thanks to these studies we were able to classify the p.Thr737Ser substitution of *ACTN1* as a “neutral” change though it hits the same residue where a different IT-causing substitution has been reported before (Bottega et al, 2015).

Gene reporter assay are in progress to determine the role of the p.Gln347Pro variant of *ETV6* gene. Our laboratory is also setting up an in vitro analysis to study whether the amino acid substitutions of the *GP1BA*, *GP1BB*, and *GP9* genes are compatible with the expression of the GP1b/IX/V receptor on cellular surface and whether the receptor, once expressed, is able to bind the von Willebrand factor. We are confident that segregation and functional studies in these families will allow us to ascertain the cause of thrombocytopenia in a further 5-10% of cases.

Combining the IT-NGS screening with functional assays for the strongest candidate mutations, we could be able to make a definite diagnosis of IT in 42-47% of the probands and their families enrolled in this study. This percentage is less than that reported in our previous reports (50%), where the majority of patients were enrolled at

a single clinical center (IRCCS San Matteo, Pavia) by clinicians with a great experience in ITs. In our consecutive case series of 97 probands, almost 30% of patients have been enrolled in different centers, indicating the need to define reference institutions for the management of patients with rare disease.

At present, we do not have evidence for candidate mutations in 49 of the 97 probands. Of them, 17 (18%) are eligible for whole exome sequencing, as the IT-NGS has not revealed any potential candidate variant. In the other 32 probands, we identified one or two missense variants of uncertain significance.

Promising in terms of pathogenic effect are the different heterozygous variants of the *GP1BA* and *GP9* genes we identified in 7 unrelated individuals. Although they were classified as "benign/likely benign" by the predictive tools, we cannot exclude a potential role in the relative carriers whose clinical observations are compatible with a diagnosis of monoallelic BSS.

This IT form is relatively frequent in the Italian population, accounting the p.Ala172Val mutation of *GP1BA* for almost 10% of the IT characterized so far at molecular level. In addition, a significant association between rare monoallelic variants in *GP1BB* and thrombocytopenia has recently been reported analyzing data from a collection of more than 1000 genome-sequenced patients with a rare bleeding and/or platelet disorder (Sivapalaratnam et al., 2017).

Therefore, since monoallelic BSS seems to be relatively frequent, the seven amino acid changes of *GP1BA* and *GP9* have been included in the study aimed at determining the impact of the BSS variants on expression and activity of the von Willebrand receptor complex.

The variants we detected in the 32 probands without a diagnosis also affect genes, such as *HOXA11*, *FLNA*, *ABCG5* or *ABCG8*, whose mutations are associated with clinical features that are different from that observed in our patients, or *ITGA2B* and *ITGB3*, whose variants associated with thrombocytopenia are all localized in specific domains of the respective proteins.

According to these data we excluded that the variants identified in these genes were causative of thrombocytopenia. However, few are the thrombocytopenic cases with mutations of *HOXA11*, *FLNA*, *ITGA2B* and *ITGB3* described, moreover the thrombocytopenia has not been well characterized in carriers of *ABCG5* and *ABCG8*

mutations, so further work is necessary to establish the molecular consequences of these variants definitively.

Although in some probands functional studies will allow us to discriminate between "pathogenic" and "neutral" variants, approximately 50% of the index cases would remain without a diagnosis. This is consistent with data from literature (Balduini et al., 2013) and suggests that other genes, not identified yet, are implicated in ITs.

Indeed, in collaboration with Dr Marco Seri (University of Bologna) and Carlo Balduini (IRCCS San Matteo) we performed whole exome analysis in 85 families from a previous case series of probands who remained without a diagnosis despite application of the IT diagnostic algorithm and mutational screening of candidate genes (Marconi et al., 2016). In addition to few pathogenic variants in known IT genes (Marconi personal communication), the analysis allowed us to identify only one novel IT gene involved in a recessive form of severe thrombocytopenia (Marconi et al., 2018). The limited number of cases resolved by exome sequencing leads us to consider whether in many of the families left without a diagnosis the thrombocytopenia is an oligogenic trait. It could be caused by segregation and combination of few rare and/or relatively common variants not only of the 28 IT genes analyzed but also of other potential novel IT genes not included in IT-NGS design. Considering that ITs are usually transmitted as an autosomal dominant disease, in an oligogenic model the variants are expected to be hypomorphic, making it difficult to demonstrate their effect on protein function. Similar, even the segregation analysis would be complicated by incomplete penetrance in some family members though they are carriers of some variants. Only association studies will be determinant to define the genetic factors in these families, studies that will be possible only when large case series of patients are collected.

In conclusion, our results demonstrate that implementing IT-NGS data with specific investigations aimed at determining the role of VUS on protein function, we could outperform the traditional diagnostic algorithm, accelerating a definite diagnosis in patients with clinical suspicion of IT and selecting a suitable cohort of cases for novel investigations on ITs.

An accurate diagnosis of IT is fundamental for proper patient management. Since during life a considerable proportion of IT patients are at risk of developing other acquired diseases (i.e. bone marrow aplasia, haematological malignancies, nephropathy leading

to end-stage renal disease), patients with misdiagnosed thrombocytopenia will not undergo a proper follow-up. Moreover, considering that effective treatments are now available for many forms of IT, diagnostic mistakes not only cause the administration of potentially harmful therapies, such as steroids, immunoglobulins and splenectomy that fail to raise the platelet count to a normal level, but also prevent patients from receiving the adequate treatments.

# 6 APPENDIX

| Gene | Chr | Genomic variation | Protein effect | gnomAD exome ALL | SIFT pred | Polyphen2 HVAR pred | MTaster pred | MAssessor pred | CADD phred | Our prediction | Cod | n alleles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ABCG5 (NM_022436)** | chr2 | c.T20C | p.L7S | NA | D | B | N | M | 9.495 | D (3/5) | IT95 | 1 |
| **ABCG5 (NM_022436)** | chr2 | c.C293G | p.A98G | 0.0024 | D | D | D | M | 32 | D (5/5) | IT2<br>IT10 | 2 |
| **ABCG5 (NM_022436)** | chr2 | c.A1864G | p.M622V | 0.0054 | T | B | N | N | 0.001 | T (5/5) | IT60<br>IT52<br>IT45 | 3 |
| **ABCG8 (NM_022437)** | chr2 | c.C419T | p.S140L | 4.47E-05 | D | B | D | L | 26 | D (3/5) | IT90 | 1 |
| **ABCG8 (NM_022437)** | chr2 | c.G491A | p.R164Q | 6.00E-04 | D | P | D | L | 27.9 | D (4/5) | IT62 | 1 |
| **ABCG8 (NM_022437)** | chr2 | c.G551A | p.R184H | 1.64E-05 | D | B | D | L | 24.2 | D (3/5) | IT12 | 1 |
| **ABCG8 (NM_022437)** | chr2 | c.A1201T | p.T401S | 0.0015 | T | B | N | L | 14.02 | T (5/5) | IT8 | 1 |
| **ABCG8 (NM_022437)** | chr2 | c.T1208C | p.I403T | 4.08E-06 | T | B | D | L | 11.06 | T (4/5) | IT82 | 1 |
| **ABCG8 (NM_022437)** | chr2 | c.T1837C | p.Y613H | 4.87E-05 | T | B | N | L | 1.594 | T (5/5) | IT2 | 1 |
| **ABCG8 (NM_022437)** | chr2 | c.G1924A | p.A642T | 0.001 | T | B | N | L | 11.67 | T (5/5) | IT71 | 1 |
| **ACTN1 (NM_001102)** | chr14 | c.G2255A | p.R752Q | 1.22E-05 | D | D | D | M | 35 | D (5/5) | IT18 | 1 |
| **ACTN1 (NM_001102)** | chr14 | c.C2212T | p.R738W | NA | D | D | D | H | 34 | D (5/5) | IT17 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ACTN1 (NM_001102)** | chr14 | c.C2210G | p.T737S | NA | T | B | D | L | 22.6 | T (3/5) | IT41 | 1 |
| **ACTN1 (NM_001102)** | chr14 | c.C1822T | p.R608W | 8.13E-06 | D | D | D | L | 34 | D (4/5) | IT55 | 1 |
| **ACTN1 (NM_001102)** | chr14 | c.G673A | p.E225K | 4.06E-06 | D | D | D | M | 35 | D (5/5) | IT16 | 1 |
| **ACTN1 (NM_001102)** | chr14 | c.G313A | p.V105I | NA | D | D | D | L | 34 | D (4/5) | IT15 | 1 |
| **ANKRD26 (NM_014915)** | chr10 | c.-116C>T | - | NA | NA | NA | NA | NA | NA | NA | IT23 | 1 |
| **ANKRD26 (NM_014915)** | chr10 | c.-118C>T | - | NA | NA | NA | NA | NA | NA | NA | IT24 | 1 |
| **ANKRD26 (NM_014915)** | chr10 | c.-126T>G | - | NA | NA | NA | NA | NA | NA | NA | IT25 | 1 |
| **ANKRD26 (NM_014915)** | chr10 | c.-128G>A | - | NA | NA | NA | NA | NA | NA | NA | IT26 | 1 |
| **ETV6 (NM_001987)** | chr12 | c.G65A | p.S22N | 8.13E-06 | T | B | D | L | 15.37 | T (3/5) | IT82 | 1 |
| **ETV6 (NM_001987)** | chr12 | c.A1040C | p.Q347P | NA | D | D | D | M | 27.4 | D (5/5) | IT93 | 1 |
| **FLI1 (NM_002017)** | chr11 | c.G203T | p.R68L | 3.00E-04 | T | P | D | L | 34 | D (3/5) | IT54 | 1 |
| **FLNA (NM_001456)** | chrX | c.G1120A | p.V374M | 6.72E-05 | D | D | D | H | 25.7 | D (5/5) | IT80 | 1 |
| **FLNA (NM_001456)** | chrX | c.A2027G | p.K676R | 2.00E-04 | T | B | D | L | 11.36 | T (4/5) | IT68 | 1 |
| **FLNA (NM_001456)** | chrX | c.G3574A | p.E1192K | NA | T | B | D | M | 18.01 | D (3/5) | IT66 | 1 |
| **FLNA (NM_001456)** | chrX | c.C5227T | p.P1743S | 7.00E-04 | T | B | N | N | 0.087 | T (5/5) | IT84 | 1 |
| **FLNA (NM_001456)** | chrX | c.A6326G | p.N2109S | 1.00E-04 | T | B | D | N | 0.075 | T (4/5) | IT68 | 1 |
| **FLNA (NM_001456)** | chrX | c.G6340A | p.D2114N | 5.60E-06 | D | P | D | M | 24.1 | D (5/5) | IT76 | 1 |
| **FLNA** | chrX | c.C6704G | p.A2235G | 2.27E-05 | D | D | D | H | 24.1 | D (5/5) | IT28 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **(NM_001456)** | | | | | | | | | | | | |
| **FLNA (NM_001456)** | chrX | c.G7604A | p.C2535Y | 6.72E-05 | D | B | N | N | 12.04 | T (4/5) | IT42 | 1 |
| **FYB (NM_001465)** | chr5 | c.C410T | p.P137L | 9.00E-04 | T | B | D | M | 6.799 | T (3/5) | IT40 | 1 |
| **GATA1 (NM_002049)** | chrX | c.T884C | p.L295P | 5.74E-06 | D | P | D | L | 27.8 | D (4/5) | IT23 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.103delA | p.K35Rfs*4 | NA | NA | NA | NA | NA | NA | NA | IT11 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.A169G | p.N57D | NA | D | D | N | H | 23.5 | D (4/5) | IT31 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.C206T | p.P69L | 0.0018 | D | B | N | L | 10.82 | T (4/5) | IT26 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.C515T | p.A172V | NA | D | P | A | L | 26.1 | D (4/5) | IT1-IT10 | 10 |
| **GP1BA (NM_000173)** | chr17 | c.G701A | p.R234H | 8.12E-06 | T | B | N | N | 13.56 | T (5/5) | IT48 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.G775C | p.V259L | 1.62E-05 | T | B | N | L | 15.99 | T (4/5) | IT72 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.G1619C | p.W540S | 2.44E-05 | D | D | D | L | 23 | D (4/5) | IT79 | 1 |
| **GP1BA (NM_000173)** | chr17 | c.C1271T | p.P424L | 7.48E-05 | D | B | N | N | 6.839 | T (4/5) | IT72 | 1 |
| **GP1BB (NM_000407)** | chr22 | c.A1C | p.M1L | NA | D | B | D | . | 10.46 | NA | IT32 | 1 |
| **GP1BB (NM_000407)** | chr22 | c.T179C | p.L60P | NA | D | D | D | H | 26.8 | D (5/5) | IT12 | 1 |
| **GP1BB (NM_000407)** | chr22 | c.T347C | p.L116P | NA | T | D | D | L | 25.4 | D (3/5) | IT34 | apparently 2 |
| **GP1BB (NM_000407)** | chr22 | c.528_550del | p.R177Sfs*123 | | | | | | | NA | IT33 | 2 |
| **GP9 (NM_000174)** | chr3 | c.A182G | p.N61S | 5.00E-04 | D | D | A | H | 23.4 | D (5/5) | IT13 | 2 |
| **GP9 (NM_000174)** | chr3 | c.C368T | p.P123L | 9.00E-04 | D | B | N | L | 18.53 | T (3/5) | IT85 | 1 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **GP9 (NM_000174)** | chr3 | c.G434A | p.R145H | 3.00E-04 | T | B | N | N | 0.01 | T (5/5) | IT41 | 1 |
| **GP9 (NM_000174)** | chr3 | c.A284G | p.Y95C | NA | D | D | D | H | 26.4 | D (5/5) | IT14 | 2 |
| **HOXA11 (NM_005523)** | chr7 | c.A347G | p.H116R | NA | D | B | D | M | 17.97 | D (4/5) | IT55 | 1 |
| **HOXA11 (NM_005523)** | chr7 | c.G396C | p.R132S | 4.00E-04 | D | P | D | H | 25.4 | D (5/5) | IT60 | 1 |
| **HOXA11 (NM_005523)** | chr7 | c.G737T | p.C246F | NA | D | D | D | N | 32 | D (4/5) | IT1 | 1 |
| **ITGA2B (NM_000419)** | chr17 | c.G235C | p.E79Q | NA | T | B | N | M | 0.049 | T (4/5) | IT17 | 1 |
| **ITGA2B (NM_000419)** | chr17 | c.G1097A | p.R366Q | 8.51E-06 | T | B | N | L | 12.57 | T (5/5) | IT45 | 1 |
| **ITGA2B (NM_000419)** | chr17 | c.G1375A | p.D459N | 5.28E-05 | D | D | D | L | 34 | D (4/5) | IT65 | 1 |
| **ITGA2B (NM_000419)** | chr17 | c.G1846A | p.V616M | 9.00E-04 | T | B | N | L | 16.3 | T (4/5) | IT51 | 1 |
| **ITGA2B (NM_000419)** | chr17 | c.G1945T | p.V649L | 0.0011 | T | B | N | L | 0.925 | T (5/5) | IT94 | 1 |
| **ITGA2B (NM_000419)** | chr17 | c.G2602A | p.V868M | 0.0023 | D | B | N | L | 19.45 | T (3/5) | IT38<br>IT77<br>IT92 | 3 |
| **ITGB3 (NM_000212)** | chr17 | c.G1985A | p.R662H | 2.84E-05 | T | P | D | M | 34 | D (4/5) | IT88 | 1 |
| **MPL (NM_005373)** | chr1 | c.C304T | p.R102C | 1.22E-05 | T | D | D | M | 33 | D (4/5) | IT30 | 1 |
| **MPL (NM_005373)** | chr1 | c.C1904T | p.P635L | 4.06E-06 | D | D | A | M | 26.1 | D (5/5) | IT27 | 1 |
| **MPL (NM_005373)** | chr1 | c.G1864A | p.A622T | NA | D | B | N | M | 12.66 | D (3/5) | IT96 | 1 |
| **MPL (NM_005373)** | chr1 | c.G1327C | p.G443R | 7.15E-06 | D | P | N | L | 33 | D (3/5) | IT28<br>IT29 | 2 |
| **MPL (NM_005373)** | chr1 | c.G1210A | p.G404R | 4.07E-06 | D | D | D | M | 27.7 | D (5/5) | IT27 | 2 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **MYH9 (NM_002473)** | chr22 | c.C279G | p.N93K | NA | D | D | A | H | 25 | D (5/5) | IT19 | 1 |
| **MYH9 (NM_002473)** | chr22 | c.G2680A | p.E894K | 4.09E-06 | D | P | D | H | 34 | D (5/5) | IT22 | 1 |
| **MYH9 (NM_002473)** | chr22 | c.G3838A | p.V1280M | 5.76E-05 | T | P | D | L | 23.2 | D (3/5) | IT64 | 1 |
| **MYH9 (NM_002473)** | chr22 | c.C4535T | p.S1512F | NA | D | D | D | M | 25.7 | D (5/5) | IT35 IT36 | 1 |
| **MYH9 (NM_002473)** | chr22 | c.C4563A | p.H1521Q | NA | D | D | D | M | 23 | D (5/5) | IT37 | 1 |
| **MYH9 (NM_002473)** | chr22 | c.G5143A | p.G1715S | 0.0015 | T | B | D | N | 14.95 | T (4/5) | IT13 | 1 |
| **MYH9 (NM_002473)** | chr22 | c.G5521A | p.E1841K | NA | D | D | A | H | 34 | D (5/5) | IT20 IT21 | 2 |
| **NBEAL2 (NM_015175)** | chr3 | c.G137A | p.R46Q | 3.25E-05 | T | D | D | M | 27 | D (4/5) | IT6 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.C386T | p.T129M | 1.00E-04 | T | B | D | L | 21 | T (3/5) | IT48 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.G467A | p.R156H | 0.0014 | T | B | D | N | 17.85 | T (3/5) | IT27 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.G1871A | p.R624Q | 1.00E-04 | T | B | D | N | 23.5 | T (3/5) | IT95 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.4221_4222del | p.S1410Qfs*14 | NA | NA | NA | NA | NA | NA | NA | IT39 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.C5660T | p.P1887L | 6.92E-05 | T | B | N | N | 13.32 | T (5/5) | IT48 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.G6212C | p.R2071P | NA | D | P | D | H | 32 | D (5/5) | IT38 | 2 |
| **NBEAL2 (NM_015175)** | chr3 | c.G6866A | p.R2289Q | 0.0028 | D | B | D | M | 29.8 | D (4/5) | IT26 | 1 |
| **NBEAL2 (NM_015175)** | chr3 | c.C7574T | p.T2525M | 8.16E-06 | D | P | D | M | 31 | D (5/5) | IT28 | 1 |
| **RBM8A (NM_005105)** | chr1 | c.T297G | p.I99M | NA | D | D | D | H | 24.1 | D (5/5) | IT63 | 1 |
| **RUNX1** | chr21 | c.614-2A>G | | NA | . | . | D | . | 24.1 | NA | IT40 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **(NM_001754)** | | | | | | | | | | | | |
| **SLFN14 (NM_001129820)** | chr17 | c.T515C | p.V172A | 0.0043 | T | B | N | N | 0.001 | T (5/5) | IT90 | 1 |
| **SLFN14 (NM_001129820)** | chr17 | c.C1974G | p.H658Q | NA | D | D | N | M | 25.8 | D (4/5) | IT79 | 1 |
| **TUBB1 (NM_030773)** | chr20 | c.C165A | p.Y55X | 2.44E-05 | . | . | A | . | 35 | NA | IT42 | 1 |
| **TUBB1 (NM_030773)** | chr20 | c.G326A | p.G109E | 9.00E-04 | D | D | D | H | 24.1 | D (5/5) | IT43<br>IT44<br>IT45<br>IT46 | 5 |
| **TUBB1 (NM_030773)** | chr20 | c.G742A | p.A248T | 3.23E-05 | D | P | D | L | 25.8 | D (4/5) | IT47 | 1 |
| **VWF (NM_000552)** | chr12 | c.G1325A | p.R442H | 2.00E-04 | T | P | D | M | 25.9 | D (4/5) | IT87 | 1 |
| **VWF (NM_000552)** | chr12 | c.C1781G | p.A594G | 6.86E-05 | T | P | D | M | 20.8 | D (4/5) | IT19<br>IT29 | 2 |
| **VWF (NM_000552)** | chr12 | c.C2878T | p.R960W | 1.00E-04 | T | P | D | M | 34 | D (4/5) | IT97 | 1 |
| **VWF (NM_000552)** | chr12 | c.G3290A | p.C1097Y | NA | D | P | D | M | 28.6 | D (5/5) | IT43 | 1 |
| **VWF (NM_000552)** | chr12 | c.C3797T | p.P1266L | 8.00E-04 | T | D | D | M | 18.14 | D (4/5) | IT74 | 1 |
| **VWF (NM_000552)** | chr12 | c.C4613T | p.T1538M | 4.07E-05 | T | B | N | M | 10.91 | T (4/5) | IT52 | 1 |
| **VWF (NM_000552)** | chr12 | c.T5191A | p.S1731T | 0.0016 | D | P | D | M | 24.5 | D (5/5) | IT12<br>IT91 | 2 |
| **VWF (NM_000552)** | chr12 | c.A6721G | p.K2241E | NA | T | B | N | L | 0.01 | T (5/5) | IT58 | 1 |
| **VWF (NM_000552)** | chr12 | c.C7493A | p.A2498D | 1.22E-05 | D | D | D | M | 25.8 | D (5/5) | IT19 | 1 |
| **VWF (NM_000552)** | chr12 | c.C7940T | p.T2647M | 0.0036 | T | B | N | L | 16.78 | T (4/5) | IT1 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **VWF (NM_000552)** | chr12 | c.G8156T | p.C2719F | 4.08E-06 | D | D | D | M | 29.4 | D (5/5) | IT44 | 1 |
| **WAS (NM_000377)** | chrX | c.G413A | p.R138Q | 5.00E-04 | T | B | N | N | 9.73 | T (5/5) | IT82 | 1 |
| **WAS (NM_000377)** | chrX | c.680dupG | p.S228Lfs*10 | NA | NA | NA | NA | NA | NA | NA | IT39 | 1 |

**Suppl. Table 1** List of the 100 different variants detected by IT-NGS. Prediction abbreviations: SIFT (D: Deleterious - sift≤0.05; T: tolerated - sift>0.05), PolyPhen 2 HVar (D: Probably damaging - pp2_hvar ≥0.909; P: possibly damaging - 0.447≤pp2_hvar≤0.909; B: benign pp2_hvar≤0.446), MutationTaster (A: disease_causing_automatic; D: disease_causing; N: polymorphism; P: polymorphism_automatic), MutationAssessor (H: high; M: medium; L: low; N: neutral H/M means functional and L/N means non-functional). NA: not available. Our prediction considers the concordance between the five bioinformatic tools used for the analysis of the variants.

# 7 BIBLIOGRAPHY


Aldrich, R.A., Steinberg, A.G., and Campbell, D.C. (1954). Pedigree demonstrating a sex-linked recessive condition characterized by draining ears, eczematoid dermatitis and bloody diarrhea. Pediatrics *13*, 133–139.

Ali, S., Ghosh, K., Daly, M.E., Hampshire, D.J., Makris, M., Ghosh, M., Mukherjee, L., Bhattacharya, M., and Shetty, S. (2016). Congenital macrothrombocytopenia is a heterogeneous disorder in India. Haemoph. Off. J. World Fed. Hemoph. *22*, 570–582.

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127*, 2391–2405.

Auer, P.L., Teumer, A., Schick, U., O'Shaughnessy, A., Lo, K.S., Chami, N., Carlson, C., de Denus, S., Dubé, M.-P., Haessler, J., et al. (2014). Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat. Genet. *46*, 629–634.

Balduini, C.L., and Noris, P. (2016). Innovation in the field of thrombocytopenias: achievements since the beginning of the century and promises for the future. Haematologica *101*, 2–4.

Balduini, C.L., Pecci, A., and Savoia, A. (2011a). Recent advances in the understanding and management of MYH9-related inherited thrombocytopenias. Br. J. Haematol. *154*, 161–174.

Balduini, C.L., Pecci, A., and Savoia, A. (2011b). Recent advances in the understanding and management of MYH9-related inherited thrombocytopenias. Br. J. Haematol. *154*, 161–174.

Balduini, C.L., Savoia, A., and Seri, M. (2013). Inherited thrombocytopenias frequently diagnosed in adults. J. Thromb. Haemost. JTH *11*, 1006–1019.

Balduini, C.L., Melazzini, F., and Pecci, A. (2017). Inherited thrombocytopenias-recent advances in clinical and molecular aspects. Platelets *28*, 3–13.

Ballmaier, M., and Germeshausen, M. (2009). Advances in the understanding of congenital amegakaryocytic thrombocytopenia. Br. J. Haematol. *146*, 3–16.

Ballmaier, M., Germeshausen, M., Schulze, H., Cherkaoui, K., Lang, S., Gaudig, A., Krukemeier, S., Eilers, M., Strauss, G., and Welte, K. (2001). c-mpl mutations are the cause of congenital amegakaryocytic thrombocytopenia. Blood *97*, 139–146.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.

Bellissimo, D.B., Christopherson, P.A., Flood, V.H., Gill, J.C., Friedman, K.D., Haberichter, S.L., Shapiro, A.D., Abshire, T.C., Leissinger, C., Hoots, W.K., et al. (2012). VWF mutations and new sequence variations identified in healthy controls are more frequent in the African-American population. Blood *119*, 2135–2140.

Berndt, M.C., and Andrews, R.K. (2011). Bernard-Soulier syndrome. Haematologica *96*, 355–359.

Bluteau, D., Balduini, A., Balayn, N., Currao, M., Nurden, P., Deswarte, C., Leverger, G., Noris, P., Perrotta, S., Solary, E., et al. (2014). Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation. J. Clin. Invest. *124*, 580–591.

Bottega, R., Pecci, A., De Candia, E., Pujol-Moix, N., Heller, P.G., Noris, P., De Rocco, D., Podda, G.M., Glembotsky, A.C., Cattaneo, M., et al. (2013). Correlation between platelet phenotype and NBEAL2 genotype in patients with congenital thrombocytopenia and α-granule deficiency. Haematologica *98*, 868–874.

Bottega, R., Marconi, C., Faleschini, M., Baj, G., Cagioni, C., Pecci, A., Pippucci, T., Ramenghi, U., Pardini, S., Ngu, L., et al. (2015). ACTN1-related thrombocytopenia: identification of novel families for phenotypic characterization. Blood *125*, 869–872.

Bottega, R., Nicchia, E., Alfano, C., Glembotsky, A.C., Pastore, A., Bertaggia-Calderara, D., Bisig, B., Duchosal, M.A., Arbesú, G., Alberio, L., et al. (2017). Gray platelet syndrome: Novel mutations of the NBEAL2 gene. Am. J. Hematol. *92*, E20–E22.

Canales, M.L., and Mauer, A.M. (1967). Sex-linked hereditary thrombocytopenia as a variant of Wiskott-Aldrich syndrome. N. Engl. J. Med. *277*, 899–901.

Davis, B., Toivio-Kinnucan, M., Schuller, S., and Boudreaux, M.K. (2008). Mutation in beta1-tubulin correlates with macrothrombocytopenia in Cavalier King Charles Spaniels. J. Vet. Intern. Med. *22*, 540–545.

Denis, C.V. (2002). Molecular and cellular biology of von Willebrand factor. Int. J. Hematol. *75*, 3–8.

Faleschini, M., Melazzini, F., Marconi, C., Giangregorio, T., Pippucci, T., Cigalini, E., Pecci, A., Bottega, R., Ramenghi, U., Siitonen, T., et al. (2018). ACTN1 mutations lead to a benign form of platelet macrocytosis not always associated with thrombocytopenia. Br. J. Haematol. *183*, 276–288.

Ferrari, S., Lombardi, A.M., Cortella, I., Businaro, M.A., Bertomoro, A., Di Pasquale, I., and Fabris, F. (2018). New heterozygous variant in GP1BB gene is responsible for an inherited form of macrothrombocytopenia. Br. J. Haematol.

Fiore, M., Goulas, C., and Pillois, X. (2017). A new mutation in TUBB1 associated with thrombocytopenia confirms that C-terminal part of β1-tubulin plays a role in microtubule assembly. Clin. Genet. *91*, 924–926.

Freedman, M.H., and Estrov, Z. (1990). Congenital amegakaryocytic thrombocytopenia: an intrinsic hematopoietic stem cell defect. Am. J. Pediatr. Hematol. Oncol. *12*, 225–230.

Ghemlas, I., Li, H., Zlateska, B., Klaassen, R., Fernandez, C.V., Yanofsky, R.A., Wu, J., Pastore, Y., Silva, M., Lipton, J.H., et al. (2015). Improving diagnostic precision, care and syndrome definitions using comprehensive next-generation sequencing for the inherited bone marrow failure syndromes. J. Med. Genet. *52*, 575–584.

Jin, Y., Mazza, C., Christie, J.R., Giliani, S., Fiorini, M., Mella, P., Gandellini, F., Stewart, D.M., Zhu, Q., Nelson, D.L., et al. (2004). Mutations of the Wiskott-Aldrich Syndrome Protein (WASP): hotspots, effect on transcription, and translation and phenotype/genotype correlation. Blood *104*, 4010–4019.

Kunishima, S., and Saito, H. (2010). Advances in the understanding of MYH9 disorders. Curr. Opin. Hematol. *17*, 405–410.

Kunishima, S., Kojima, T., Matsushita, T., Tanaka, T., Tsurusawa, M., Furukawa, Y., Nakamura, Y., Okamura, T., Amemiya, N., Nakayama, T., et al. (2001). Mutations in the NMMHC-A gene cause autosomal dominant macrothrombocytopenia with leukocyte inclusions (May-Hegglin anomaly/Sebastian syndrome). Blood *97*, 1147–1149.

Kunishima, S., Kobayashi, R., Itoh, T.J., Hamaguchi, M., and Saito, H. (2009). Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. Blood *113*, 458–461.

Kunishima, S., Okuno, Y., Yoshida, K., Shiraishi, Y., Sanada, M., Muramatsu, H., Chiba, K., Tanaka, H., Miyazaki, K., Sakai, M., et al. (2013). ACTN1 mutations cause congenital macrothrombocytopenia. Am. J. Hum. Genet. *92*, 431–438.

Kunishima, S., Nishimura, S., Suzuki, H., Imaizumi, M., and Saito, H. (2014). TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia. Eur. J. Haematol. *92*, 276–282.

Machlus, K.R., and Italiano, J.E. (2013). The incredible journey: From megakaryocyte development to platelet formation. J. Cell Biol. *201*, 785–796.

Marconi, C., Di Buduo, C.A., Barozzi, S., Palombo, F., Pardini, S., Zaninetti, C., Pippucci, T., Noris, P., Balduini, A., Seri, M., et al. (2016). SLFN14-related thrombocytopenia: identification within a large series of patients with inherited thrombocytopenia. Thromb. Haemost. *115*, 1076–1079.

Marconi, C., Di Buduo, C.A., LeVine, K., Barozzi, S., Faleschini, M., Bozzi, V., Palombo, F., McKinstry, S., Lassandro, G., Giordano, P., et al. (2018). A new form of inherited thrombocytopenia caused by loss-of-function mutations in PTPRJ. Blood.

Melazzini, F., Palombo, F., Balduini, A., De Rocco, D., Marconi, C., Noris, P., Gnan, C., Pippucci, T., Bozzi, V., Faleschini, M., et al. (2016a). Clinical and pathogenic features of ETV6-related thrombocytopenia with predisposition to acute lymphoblastic leukemia. Haematologica *101*, 1333–1342.

Melazzini, F., Palombo, F., Balduini, A., De Rocco, D., Marconi, C., Noris, P., Gnan, C., Pippucci, T., Bozzi, V., Faleschini, M., et al. (2016b). Clinical and pathogenic features of ETV6-related thrombocytopenia with predisposition to acute lymphoblastic leukemia. Haematologica *101*, 1333–1342.

Morrow, B.E., McDonald-McGinn, D.M., Emanuel, B.S., Vermeesch, J.R., and Scambler, P.J. (2018). Molecular genetics of 22q11.2 deletion syndrome. Am. J. Med. Genet. A. *176*, 2070–2081.

Muraoka, K., Ishii, E., Tsuji, K., Yamamoto, S., Yamaguchi, H., Hara, T., Koga, H., Nakahata, T., and Miyazaki, S. (1997). Defective response to thrombopoietin and impaired expression of c-mpl mRNA of bone marrow cells in congenital amegakaryocytic thrombocytopenia. Br. J. Haematol. *96*, 287–292.

Noris, P., and Pecci, A. (2017). Hereditary thrombocytopenias: a growing list of disorders. Hematol. Am. Soc. Hematol. Educ. Program *2017*, 385–399.

Noris, P., Perrotta, S., Seri, M., Pecci, A., Gnan, C., Loffredo, G., Pujol-Moix, N., Zecca, M., Scognamiglio, F., De Rocco, D., et al. (2011). Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. Blood *117*, 6673–6680.

Noris, P., Perrotta, S., Bottega, R., Pecci, A., Melazzini, F., Civaschi, E., Russo, S., Magrin, S., Loffredo, G., Di Salvo, V., et al. (2012). Clinical and laboratory features of 103 patients from 42 Italian families with inherited thrombocytopenia derived from the monoallelic Ala156Val mutation of GPIbα (Bolzano mutation). Haematologica *97*, 82–88.

Noris, P., Biino, G., Pecci, A., Civaschi, E., Savoia, A., Seri, M., Melazzini, F., Loffredo, G., Russo, G., Bozzi, V., et al. (2014). Platelet diameters in inherited thrombocytopenias: analysis of 376 patients with all known disorders. Blood *124*, e4–e10.

Nurden, P., Debili, N., Coupry, I., Bryckaert, M., Youlyouz-Marfak, I., Solé, G., Pons, A.-C., Berrou, E., Adam, F., Kauskot, A., et al. (2011). Thrombocytopenia resulting from mutations in filamin A can be expressed as an isolated syndrome. Blood *118*, 5928–5937.

Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321–330.

Patel, S.R., Hartwig, J.H., and Italiano, J.E. (2005). The biogenesis of platelets from megakaryocyte proplatelets. J. Clin. Invest. *115*, 3348–3354.

Pecci, A., Biino, G., Fierro, T., Bozzi, V., Mezzasoma, A., Noris, P., Ramenghi, U., Loffredo, G., Fabris, F., Momi, S., et al. (2012). Alteration of liver enzymes is a feature of the MYH9-related disease syndrome. PloS One *7*, e35986.

Pecci, A., Klersy, C., Gresele, P., Lee, K.J.D., De Rocco, D., Bozzi, V., Russo, G., Heller, P.G., Loffredo, G., Ballmaier, M., et al. (2014a). 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., Klersy, C., Gresele, P., Lee, K.J.D., De Rocco, D., Bozzi, V., Russo, G., Heller, P.G., Loffredo, G., Ballmaier, M., et al. (2014b). 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., and Adelstein, R.S. (2018). MYH9: Structure, functions and role of non-muscle myosin IIA in human disease. Gene *664*, 152–167.

Pippucci, T., Savoia, A., Perrotta, S., Pujol-Moix, N., Noris, P., Castegnaro, G., Pecci, A., Gnan, C., Punzo, F., Marconi, C., et al. (2011). Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. Am. J. Hum. Genet. *88*, 115–120.

Poggi, M., Canault, M., Favier, M., Turro, E., Saultier, P., Ghalloussi, D., Baccini, V., Vidal, L., Mezzapesa, A., Chelghoum, N., et al. (2017). Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. Haematologica *102*, 282–294.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (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. Off. J. Am. Coll. Med. Genet. *17*, 405–424.

Saposnik, B., Binard, S., Fenneteau, O., Nurden, A., Nurden, P., Hurtaud-Roux, M.-F., Schlegel, N., and French MYH9 networka (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., Balduini, C.L., Savino, M., Noris, P., Del Vecchio, M., Perrotta, S., Belletti, S., Poggi, null, and Iolascon, A. (2001). Autosomal dominant macrothrombocytopenia in Italy is most frequently a type of heterozygous Bernard-Soulier syndrome. Blood *97*, 1330–1335.

Savoia, A., Kunishima, S., De Rocco, D., Zieger, B., Rand, M.L., Pujol-Moix, N., Caliskan, U., Tokgoz, H., Pecci, A., Noris, P., et al. (2014). Spectrum of the mutations in Bernard-Soulier syndrome. Hum. Mutat. *35*, 1033–1045.

Seri, M., Cusano, R., Gangarossa, S., Caridi, G., Bordo, D., Lo Nigro, C., Ghiggeri, G.M., Ravazzolo, R., Savino, M., Del Vecchio, M., et al. (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.

Seri, M., Pecci, A., Di Bari, F., Cusano, R., Savino, M., Panza, E., Nigro, A., Noris, P., Gangarossa, S., Rocca, B., et al. (2003). MYH9-related disease: May-Hegglin anomaly, Sebastian syndrome, Fechtner syndrome, and Epstein syndrome are not distinct entities but represent a variable expression of a single illness. Medicine (Baltimore) *82*, 203–215.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat. Biotechnol. *26*, 1135–1145.

Sivapalaratnam, S., Westbury, S.K., Stephens, J.C., Greene, D., Downes, K., Kelly, A.M., Lentaigne, C., Astle, W.J., Huizinga, E.G., Nurden, P., et al. (2017). Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia. Blood *129*, 520–524.

Sjöblom, B., Salmazo, A., and Djinović-Carugo, K. (2008). Alpha-actinin structure and regulation. Cell. Mol. Life Sci. CMLS *65*, 2688–2701.

Topka, S., Vijai, J., Walsh, M.F., Jacobs, L., Maria, A., Villano, D., Gaddam, P., Wu, G., McGee, R.B., Quinn, E., et al. (2015). Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. PLoS Genet. *11*, e1005262.

Vettore, S., Scandellari, R., Moro, S., Lombardi, A.M., Scapin, M., Randi, M.L., and Fabris, F. (2008). Novel point mutation in a leucine-rich repeat of the GPIbalpha chain of the platelet von Willebrand factor receptor, GPIb/IX/V, resulting in an inherited dominant form of Bernard-Soulier syndrome affecting two unrelated families: the N41H variant. Haematologica *93*, 1743–1747.

Wang, L.C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F., Alt, F.W., Gilliland, D.G., Golub, T.R., and Orkin, S.H. (1998). The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. Genes Dev. *12*, 2392–2402.

Wendling, F., Maraskovsky, E., Debili, N., Florindo, C., Teepe, M., Titeux, M., Methia, N., Breton-Gorius, J., Cosman, D., and Vainchenker, W. (1994). cMpl ligand is a humoral regulator of megakaryocytopoiesis. Nature *369*, 571–574.