Exome sequencing in 116 patients with inherited thrombocytopenia that remained of unknown origin after systematic phenotype-driven diagnostic workup

Caterina Marconi,[1*] Alessandro Pecci,[2,3*] Flavia Palombo,[1] Federica Melazzini,[2,3] Roberta Bottega,[4] Elena Nardi,[5] Valeria Bozzi,[3] Michela Faleschini,[4] Serena Barozzi,[3] Tania Giangregorio,[4] Pamela Magini,[6] Carlo L. Balduini,[2] Anna Savoia,[4,7] Marco Seri,[1,6] Patrizia Noris[2,3#] and Tommaso Pippucci[6#]

[1]Department of Medical and Surgical Science, University of Bologna, Bologna; [2]Department of Internal Medicine, University of Pavia, Pavia; [3]Medicina Generale 1, IRCCS Policlinico San Matteo Foundation, Pavia; [4]Institute for Maternal and Child Health – IRCCS Burlo Garofolo, Trieste; [5]Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna; [6]UO Genetica Medica, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna and [7]Department of Medical Sciences, University of Trieste, Trieste, Italy

**CM and AP contributed equally as co-first authors.*
#PN and TP contributed equally as co-senior authors.

Correspondence: M. Seri
marco.seri@unibo.it

Received: March 4, 2022.
Accepted: August 29, 2022.
Early view: December 15, 2022.

https://doi.org/10.3324/haematol.2022.280993

Abstract

Inherited thrombocytopenias (IT) are genetic diseases characterized by low platelet count, sometimes associated with congenital defects or a predisposition to develop additional conditions. Next-generation sequencing has substantially improved our knowledge of IT, with more than 40 genes identified so far, but obtaining a molecular diagnosis remains a challenge especially for patients with non-syndromic forms, having no clinical or functional phenotypes that raise suspicion about specific genes. We performed exome sequencing (ES) in a cohort of 116 IT patients (89 families), still undiagnosed after a previously validated phenotype-driven diagnostic algorithm including a targeted analysis of suspected genes. ES achieved a diagnostic yield of 36%, with a gain of 16% over the diagnostic algorithm. This can be explained by genetic heterogeneity and unspecific genotype-phenotype relationships that make the simultaneous analysis of all the genes, enabled by ES, the most reasonable strategy. Furthermore, ES disentangled situations that had been puzzling because of atypical inheritance, sex-related effects or false negative laboratory results. Finally, ES-based copy number variant analysis disclosed an unexpectedly high prevalence of *RUNX1* deletions, predisposing to hematologic malignancies. Our findings demonstrate that ES, including copy number variant analysis, can substantially contribute to the diagnosis of IT and can solve diagnostic problems that would otherwise remain a challenge.

Introduction

Inherited thrombocytopenias (IT) are a heterogeneous group of disorders characterized by low platelet count that can result in a bleeding tendency of variable degree. In these disorders, thrombocytopenia can be isolated or associated with additional congenital defects; moreover, some cases of IT have a predisposition to develop additional diseases over time, such as hematologic malignancies, bone marrow aplasia and renal failure.[1]

Until recently, the diagnosis of IT was based on a complex process requiring a multi-step clinical and laboratory characterization of patients and subsequent resequencing of candidate genes.[2] The introduction of next-generation sequencing revolutionized the diagnostic approach to these disorders, allowing the analysis of virtually all known genes at one time by exome sequencing (ES). Moreover, application of next-generation sequencing led to the identification of many novel genes underlying IT. These advances revealed a picture of wide genetic heterogeneity, with at least 40 genes implicated;[1] it should, however, be noted that only a few genes account for most of the cases, while most genes explain less than 2% cases each.[3] In spite of the recent progress, almost half of patients with familial IT still remain without a definite molecular diagnosis,[3] which would be of key importance for clinical management and counseling.

ES of large numbers of patients, in association with clustering of results according to standardized clinical or functional phenotypes,[4,5] has proven effective in the description of novel forms of IT and the identification of causative vari-

ants in known IT genes, thus achieving a molecular diagnosis for a substantial proportion of patients.[6-9] However, phenotype-based approaches may be ineffective when thrombocytopenia is non-syndromic, with no additional phenotypes contributing to the patient's clinical picture, or is not associated with detectable alterations in platelet function.

Here we report the analysis of 116 patients with non-syndromic IT who had remained without a definite molecular diagnosis after an extensive diagnostic workup. We show the power of the ES approach to improve the diagnostic yield over a phenotype-driven diagnostic algorithm and illustrate the reasons that make ES a strategy of choice in the molecular elucidation of these genetically heterogeneous disorders.

Methods

Patients

We recruited 116 patients among probands and available affected relatives from 89 families (*Online Supplementary Table S1*) with familial thrombocytopenia who remained without a molecular diagnosis after the application of a validated diagnostic algorithm[2] including systematic phenotypic investigation and sequencing of candidate genes among the 21 IT genes defined before the beginning of the study (listed *in Online Supplementary Table S2*). By this approach, 54%of the probands did not have a definite diagnosis, and all those with the availability of DNA samples of suitable quality for ES analysis were enrolled in this study (*Online Supplementary Figure S1*). Details about the different forms of IT diagnosed through the application of the phenotype-driven diagnostic algorithm are reported in *Online Supplementary Table S3*. All the cases recruited into the ES analysis presented apparently non-syndromic forms, with no clinical features additional to thrombocytopenia and no specific platelet function alterations.

The study was approved by the Ethics Committee of the IRCCS Policlinico San Matteo Foundation. All investigated individuals or their legal guardians provided written informed consent to participation in the study, which was conducted in accordance with the Declaration of Helsinki. Individual data were completely de-identified.

Exome sequencing

ES was carried out on DNA from whole blood after enrichment and capture with different strategies (*Online Supplementary Methods*). Sequence data analysis was performed as already described.[10]

Analysis of known inherited thrombocytopenia genes

We evaluated each patient for small variants and copy number variants (CNV) in 43 genes associated with IT (*Online Supplementary Table S2*), including 21 genes prescreened as part of the diagnostic algorithm and 22 genes found to be associated with IT in recent years (*Online Supplementary Table S2*).

Single nucleotide variants (SNV) and small insertions/deletions (indels) were selected according to minor allele frequencies and variant consequences, as detailed in the *Online Supplementary Methods*, and confirmed by Sanger sequencing in probands and available relatives. Variant classification followed the American College of Medical Genetics (ACMG) guidelines[11] and the ClinGen Expert curation panel guidelines for *RUNX1* variants.[12] We evaluated segregation data as recommended by Jarvik and Browning.[13]

Analysis of runs of homozygosity was carried out on Family.43 with H^3M^2 to identify large exomic homozygous regions.[14] The genomic inbreeding coefficient was calculated as the percentage of the cumulative length of autosomal runs of homozygosity >1.5 Mb over the overall length of the autosomal genome.

Exome alignments were used to call and genotype CNV with Excavator2.[15] Deletions encompassing *RUNX1* were confirmed by real-time polymerase chain reaction (RT-PCR) and chromosomal microarray analysis (CMA) on an Agilent 8x60K platform.

Gene-based rare variant enrichment analysis

With the aim of identifying possible major contributing IT genes, in terms of prevalence, we carried out exome-wide collapsing of rare variants by gene to identify those bearing an excess of qualifying variants in IT probands compared to population-matched, unrelated controls. These included subjects of our exome datasets who were healthy or had genetic disorders with no hematologic involvement.

Qualifying variants were defined as SNV affecting the canonical transcript of protein-coding sequences (non-synonymous and splice-site variants) with a minor allele frequency ≤0.0001 in the ExAC non-Finnish European subpopulation. All 20,345 Gencode v19 protein-coding genes that were well-covered in most samples were considered (see the *Online Supplementary Methods* for more details). To exclude population stratification among cases and controls, we removed from the analysis: (i) samples with non-European ancestry, based on self-reported information, and (ii) outlier samples from a principal component analysis carried out on genotypes of our samples together with 2,504 samples from the 1000genomes dataset (*https://www.internationalgenome.org/*).

For each gene under analysis, we modeled the number of subjects with at least one variant as a binomial distribution. We performed a two-proportion pooled test to verify the null hypothesis of equality of proportions of subjects with at least one variant in patients and controls. Multiple testing control was done applying the false discovery rate

criterion proposed by Benjamini and Hochberg choosing a false discovery rate of 0.2.[16]
Variant confirmation by Sanger sequencing was carried out for the top ranking genes with a non-corrected $P \leq 0.0015$. Segregation analysis was performed whenever possible.

Results

First, we wanted to identify variants in IT-associated genes that could be defined as disease-contributing following ACMG criteria for establishing variant pathogenicity.
All the 43 known IT-associated genes, except *GP1BB*, *GP9* and *MPIG6B*, achieved adequate sequence representation with average coverage of 132X (range: 72X-289X) and coverage higher than 20X on an average of 96.2% of targeted bases (range: 83.3% -100%) (*Online Supplementary Table S2*).
A total of 104 variants fulfilling selection criteria were identified in 60/89 probands (67%). According to ACMG criteria, 32 variants were classified as benign (B) or likely benign (LB), 40 were variants of uncertain significance (VUS), and 32 were classified as pathogenic (P) or likely pathogenic (LP). Table 1 reports P/LP variants, while *Online Supplementary Table S4* lists the VUS and B/LB variants. *Online Supplementary Table S6* reports the ACMG criteria applied for the classification of each variant. P/LP variants affected 30 probands and involved 18 genes. Most (75%; 24/32) were heterozygous variants in genes associated with autosomal dominant or autosomal dominant/recessive forms, while the remaining ones were either heterozygous X-linked variants (6%; 2/32) or heterozygous, compound heterozygous and homozygous variants in autosomal recessive genes (19%; 6/32).
The majority of P/LP variants (62%; 20/32) were in genes that would not have been previously analyzed because they were not included in the diagnostic algorithm,[2] and that were already described in our previous publications.[10,17-24] Conversely, 9/31 P/LP variants (29%) were heterozygous variants in IT-associated genes which should have been considered for analysis (Table 1), but that were overlooked because of erroneous interpretation or execution of the diagnostic algorithm: five variants affected genes that should have been suspected according to the clinical and laboratory findings of the patients (Family.9, *CYCS*; Family.21, *ITGB3* and *GP1BA*; Family.26 and Family.27, *RUNX1*) but sequencing analysis was not performed. In four additional cases, laboratory or genetic tests failed to identify the implicated gene (in Family.8 and Family.35 Sanger sequencing missed the variants in *ANKRD26* and *GP1BA*, respectively; in Family.23 the presence of MYH9 protein aggregates in leukocytes was not recognized at the immunofluorescence assay;[25] in Family.17 with a variant in *ITGA2B*, flow cytometry failed to detect decreased expression of glycoprotein complex IIb-IIIa on the platelet surface[26]. Moreover, *in vitro* platelet aggregation in response to collagen, ADP, and arachidonate resulted within the normal range in the proband of this family).
Finally, in three cases the identification of the pathogenic variants had been hindered by non-Mendelian inheritance patterns or gender-related effects and only the unbiased (with respect to suspected inheritance) evaluation of qualifying variants in IT-associated genes allowed their recognition.
In one case (Family.43) (Figure 1A) the unreported consanguinity between both parents (II-1 and II-2) and grandparents (I-1 and I-2) of the probands (III-1 and III-2) resulted in a pseudo-dominant inheritance pattern of the *ABCG8* c.1234C>T, p.Arg412Ter variant, which was homozygous in II-1, III-1 and III-2 and heterozygous in II-2.
In a second family (Family.41) (Figure 1B), the LP variant *WAS* c.134C>T, p.Thr45Met on chromosome X, was inherited by two male siblings (IV-1 and IV-3) from their affected mother (III-4). An X-linked inheritance was initially not suspected, due to a phenocopy in the grandfather (II-1), who was considered to be affected because of a platelet count slightly below the normal threshold ($120x10^9$/L). In this situation, the high variability of platelet counts between the two probands (19 and $22x10^9$/L) and mother ($80x10^9$/L) was attributed to other non-genetic factors, while it is most likely associated with the hemizygous *versus* heterozygous status of the variant. The status of the grandfather, II-1, revised based on the age- and gender-adjusted reference intervals for platelet count,[27] was then classified as non-affected. Unfortunately a DNA sample from II-1 was not available; however, the fact that the other daughter, III-3, does not carry the variant indicates that II-1 is not a variant carrier.
In a third case (Family.40) (Figure 1C) we identified the heterozygous c.146delC, p.Pro50Argfs86Ter variant in *GATA1*, on the X chromosome, in the female proband (II-1). In this pedigree, an X-linked inheritance was not suspected because of a female-to-female transmission (I-2 also affected, but not available for genetic testing). Moreover, mild isolated thrombocytopenia due to *GATA1* pathogenic variants (OMIM #300367) has been rarely reported in heterozygous females.[28,29]
Details about these three last cases are reported in the *Online Supplementary Results*.
We then reasoned that not all actual disease-contributing variants may be classified as P/LP, either because they are present in IT-associated genes but not fulfilling ACMG guidelines, or because they are in “novel” genes. We thus needed a method to highlight genes that had a significant burden of disease-contributing variants regardless of their P/LP classification. To this end, we performed a gene-based rare variant enrichment analysis focused on qual-

ifying variants (defined as predicted protein-altering variants with minor allele frequency ≤0.0001 in ExAC non-Finnish Europeans) in 81 unrelated cases and 215 controls of homogeneous Italian ancestry according to principal component analysis. A total of 6,320 genes with at least one qualifying variant in either cases or controls and fulfilling thresholds for adequate coverage were left for statistical analysis. Only two of the 18 IT-associated genes mutated in this cohort, namely *ACTN1*[30] and *ETV6*,[31,32] attained a study-wide significant excess of qualifying variants in probands (P=0.0001 and P=0.0004, respectively). *ACTN1* had nine qualifying variants in cases and one in controls,

Table 1. Pathogenic and likely pathogenic variants identified in the 43 known inherited thrombocytopenia genes.

Family	N of tested members (affected carriers/healthy carriers)	Gene	Variant	Status	Variant Class
AUTOSOMAL DOMINANT INHERITANCE					
Family 1	4 (4/na)	***ACTN1***	NM_001130004.1: c.2305G>A ; p.(Glu769Lys)	Het	LP
Family 2	3 (3/na)	***ACTN1***	NM_001130004.1: c.673G>A ; p.(Glu225Lys)	Het	LP
Family 3	4 (4/na)	***ACTN1***	NM_001130004.1: c.384G>C ; p.(Trp128Cys)	Het	LP
Family 4	3 (3/na)	***ACTN1***	NM_001130004.1: c.136C>T ; p.(Arg46Trp)	Het	LP
Family 5	3 (3/na)	***ACTN1***	NM_001130004.1: c.136C>T ; p.(Arg46Trp)	Het	LP
Family 6	3 (2/0)	***ACTN1***	NM_001130004.1: c.982G>A ; p.(Val328Met)	Het	LP
Family 7	1 (1/0)	***ACTN1***	NM_001130004.1: c.698C>T ; p.(Pro233Leu)	Het	LP
Family 8	2 (2/na)	*ANKRD26*	NM_014915.2: c.-125T>C ; p.?	Het	LP
Family 9	11 (7/0)	*CYCS*	NM_018947.5: c.145T>C ; p.(Tyr49His)	Het	LP
Family 11	2 (2/na)	***ETV6***	NM_001987.4: c.641C>T ; p.(Pro214Leu)	Het	LP
Family 12	2 (2/na)	***ETV6***	NM_001987.4: c.1105C>T ; p.(Arg369Trp)	Het	LP
Family 13	2 (2/na)	***ETV6***	NM_001987.4: c.1105C>T ; p.(Arg369Trp)	Het	LP
Family 17	1 (1/na)	*ITGA2B*	NM_000419.4: c.3076C>T ; p.(Arg1026Trp)	Het	LP
Family 21	1 (1/na)	*ITGB3*	NM_000212.2: c.1768A>G ; p.(Thr590Ala)	Het	LP
Family 16	5 (4/0)	***GFI1B***	NM_004188.6: c.648+5G>A ; p.?	Het	LP
Family 23	4 (4/na)	*MYH9*	NM_002473.5: c.121T>A ; p.(Phe41Ile)	Het	LP
Family 26	1 (1/na)	*RUNX1*	NM_001754.4: c.351+1G>A ; p.?	Het	LP
Family 27	3 (3/na)	*RUNX1*	NM_001754.4: c.578T>A ; p.(Ile193Asn)	Het	LP
Family 28	3 (3/na)	***SLFN14***	NM_001129820.1: c.667C>T ; p.(Arg223Trp)	Het	LP
Family 29	3 (1/0)	***SRC***	NM_198291.2: c.1579G>A ; p.(Glu527Lys)	Het	LP
AUTOSOMAL DOMINANT/RECESSIVE INHERITANCE					
Family 21	1 (1/na)	*GP1BA*	NM_000173.6: c.104del ; p.(Lys35ArgfsTer4)	Het	LP*
Family 35	1 (1/na)	*GP1BA*	NM_000173.6: c.169A>G ; p.(Asn57Asp)	Het	LP
Family 18	2 (2/na)	***THPO***	NM_000460.4: c.91C>T ; p.(Arg31Ter)	Het	P
Family 39	2 (2/na)	***THPO***	NM_000460.4: c.91C>T ; p.(Arg31Ter)	Het	P
X-LINKED INHERITANCE					
Family 40	1 (1/na)	*GATA1*	NM_002049.3: c.149del ; p.(Pro50ArgfsTer87)	Het	LP
Family 41	6 (3/0)	*WAS*	NM_000377.2: c.134C>T ; p.(Thr45Met)	Het	LP
AUTOSOMAL RECESSIVE INHERITANCE					
Family 43	4 (3/1)	*ABCG8*	NM_022437.2: c.1234C>T ; p.(Arg412Ter)	Hom	P
Family 81	1 (1/na)	***GNE***	NM_001128227.2: c.1768G>A ; p.(Gly590Arg)	Hom	LP
Family 82	1 (1/na)	***GNE***	NM_001128227.2: c.1427G>T ; p.(Arg476Met)	Hom	LP
Family 50	1 (1/na)	***GNE***	NM_001128227.2: c.98A>G ; p.(Glu33Gly)	Het	LP*
Family 51	4 (2/2)	***PTPRJ***	NM_002843.3: c.97-2A>G ; p.? / NM_002843.3: c.1875del ; p.(Ser627AlafsTer8)	Compound Het	P/P

In bold, genes not pre-screened in the diagnostic algorithm. *Variants considered not explicative of the phenotype. na: not applicable; Het: heterozygous; Hom: homozygous; P: pathogenic; LP: likely pathogenic.

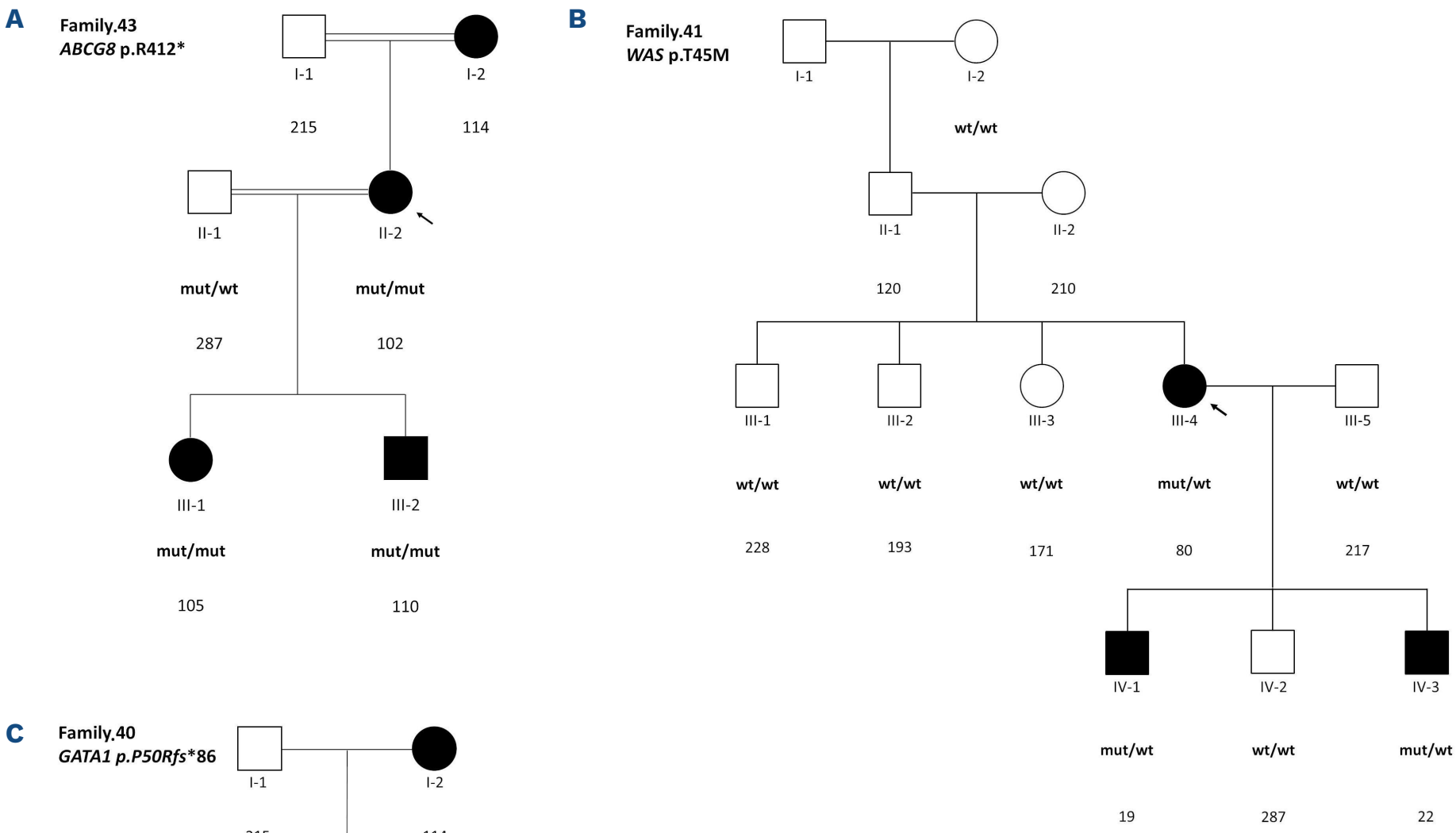

Figure 1. Pedigrees of families with inherited thrombocytopenia. (A-C) Families with pathogenic variants in *ABCG8* (A), *WAS* (B) and *GATA1* (C). Double lines indicate consanguineous unions. Platelet count (x10^9/L) and genotype are reported when available. wt: wt allele; mut: mutated allele.

while *ETV6* had six and zero, respectively (Table 2). Almost all *ACTN1* and *ETV6* alleles, including VUS and B/LB variants, were characterized by Combined Annotation Dependent Depletion (CADD) and Genomic Evolutionary Rate Profiling (GERP) scores as suggestive of a deleterious effect (i.e., CADD >20, GERP RS>4) as well as an ultra-low frequency in the general population (Table 1, *Online Supplementary Table S4*) but not all variant alleles could be defined as causative. As supported by our previous work, seven *ACTN1* and three *ETV6* alleles were classified as P/LP,[17,18,22] while the other alleles were VUS or B/LB. No novel genes attained statistical significance.

We also asked whether disease-contributing genes could have top-ranking *P* values although not reaching statistical significance. We chose the highest 25 genes, i.e. showing a non-corrected *P* value ≤0.0015 (Table 2), finding no additional IT-associated gene. To understand whether novel genes could be present in this list, we performed segregation analysis of the variants in all the available relatives and collected the genes in which we found variants that segregated according to the disease (*Online Supplementary Table S5*). In this context, we suggest *PREX1* as a potentially promising candidate, having no variants in controls and three variants in cases, all with scores indicative of a deleterious effect and of an intolerance to missense variants (Residual Variation Intolerance Score, RVIS 22.30, ExAC constraint 4.11) (Table 2, *Online Supplementary Table S5*).[33] In two families, relatives were available for the analysis and *PREX1* variants were segregating with the disorder in two and three affected members, respectively (*Online Supplementary Figure S2*). This gene has a role in regulating aggregation and dense granule secretion of mouse platelets.[34] However, further genetic and/or functional data are needed to prove that this is an IT-associated gene.

We then argued that different types of variation, namely CNV, could play a role in disease. By using Excavator2,[15] we identified heterozygous deletions involving *RUNX1* in three unrelated families (Table 3, Figure 2). All the deletions were confirmed by RT-PCR, and chromosomal microarray analysis was used to annotate the genomic boundaries of the alterations (Figure 2A). In all the cases, *RUNX1* is completely included within the breakpoints and is therefore present in a single copy in the carriers. The deletions were

Table 2. Gene-based rare variant collapsing analysis.

Ranking position	Gene	Uncorrected *P* value	Cases with qualifying variants	Cases with qualifying variants retained after segregation studies	Controls with qualifying variants	ExAC constraint z score	RVIS
1	*ACTN1**	5.20E-06	9	8	1	3,82**	3.43
2	*ETV6**	4.10E-05	6	3	0	2.2	20.53
3	*PXDN*	1.60E-04	5	4	0	2.72	11.47
3	*YARS*	1.60E-04	5	3	0	1.15	17.20
4	*PLA2G4C*	3.20E-04	6	1	1	-0.63	91.06
5	*TNS1*	4.28E-04	8	5	3	-0.64	97.39
6	*ADAR*	6.70E-04	4	1	0	3.01	2.20
6	*ANKRD55*	6.70E-04	4	3	0	0.34	17.26
6	*CLEC16A*	6.70E-04	4	1	0	1.44	25.02
6	*GNAS*	6.70E-04	4	2	0	4,34**	51.28
6	*LY9*	6.70E-04	4	2	0	-1.14	76.18
6	*NECTIN1*	6.70E-04	4	2	0	na	41.08
6	*NPAP1*	6.70E-04	4	2	0	-1.39	78.58
6	*PLBD1*	6.70E-04	4	2	0	0.25	61.52
6	*PREX1*	6.70E-04	4	4	0	4,11**	22.30
7	*CC2D2A*	1.20E-03	5	4	1	-2.06	92.94
7	*CCP110*	1.20E-03	5	2	1	0.13	24.49
7	*GSG1L2*	1.20E-03	5	0	1	na	na
7	*LAMA4*	1.20E-03	5	3	1	-0.67	15.11
7	*PHLPP1*	1.20E-03	5	2	1	2.46	7.51
7	*PTH2R*	1.20E-03	5	3	1	-1.72	54.83
7	*ZNF394*	1.20E-03	5	5	1	1	36.39
8	*INPP5D*	1.45E-03	7	2	3	3,65**	15.51
8	*MGA*	1.45E-03	7	5	3	-1.31	2.51

Twenty-four genes showing an uncorrected *P* value ≤0.0015 are reported. *Genes with significant enrichment after multiple test correction. **Significant score. RVIS: residual variation intolerance score.

confirmed in seven, two and five affected individuals in Family.44 (Figure 2B), Family.66 (Figure 2C) and Family.72 (Figure 2D) respectively, in which they spanned 1.9 Mb (18 genes), 2.7 Mb (29 genes) and 900 Kb (7 genes). *RUNX1* haploinsufficiency causes an autosomal-dominant IT with predisposition to myeloid malignancies (FPD-AML, OMIM #601399). Of note, Sanger sequencing of the whole *RUNX1* gene had been performed in all the three probands before enrollment in this study, in accordance with the diagnostic algorithm. Few instances of large *RUNX1* intragenic deletions have been described.[35] The main clinical and laboratory features of our patients with heterozygous deletions involving *RUNX1* are detailed in the *Online Supplementary Results*. A history of myeloid neoplasms was reported in Family.72 and Family.44. Interestingly, if both small variants (SNV and indels) and CNV were included in the enrichment analysis, *RUNX1* would be among the top-ranking positions (5 variants in cases, 1 variant in controls; uncorrected *P*=0.00125).

Overall, ES on all 89 families followed by a targeted analysis of 43 IT-associated genes achieved a 36% diagnostic yield (29 and 3 families with disease-causative SNV and CNV, respectively, out of 89). Compared to the phenotype-driven diagnostic algorithm on the same target genes, we estimated the increase of the diagnostic yield attained by ES as 16% (14/89), including 11 families with SNV and 3 families with full *RUNX1* deletions.

Discussion

Until a few years ago, the diagnosis of IT was based on a multi-step clinical and laboratory characterization of patients and screening of candidate genes. In our experience, this approach made it possible to identify the causative genetic defect in just under 50% of cases.[2] More recently, several groups introduced the use of next-generation sequencing techniques for a single-step, parallel sequencing of all the known genes associated with IT as a more effective, easier, and faster diagnostic approach. The results in terms of proportion of cases for which causative variants have been identified vary greatly according to differ-

ent investigations.[9,36-39]

Differently from previous studies, here we applied ES to a cohort of IT patients who had remained without a definite diagnosis after the application of a systematic, well-defined diagnostic workup based on phenotype characterization and screening of candidate genes.[2] For this reason, our population is particularly informative to assess the advantages and issues of ES as a complement to the traditional approach in the diagnosis of these disorders.

We report a 36% diagnostic rate in a large cohort of patients with non-syndromic IT. If we consider the same target genes as those in the phenotype-based diagnostic approach, we observed a 16% increase in the diagnostic yield attained with ES. This increase can be explained by the unbiased approach of ES to the analysis of protein-coding variations that overcomes three major problems. First, the genetic heterogeneity of IT, with more than 40 associated genes, makes the gene-by-gene approach very

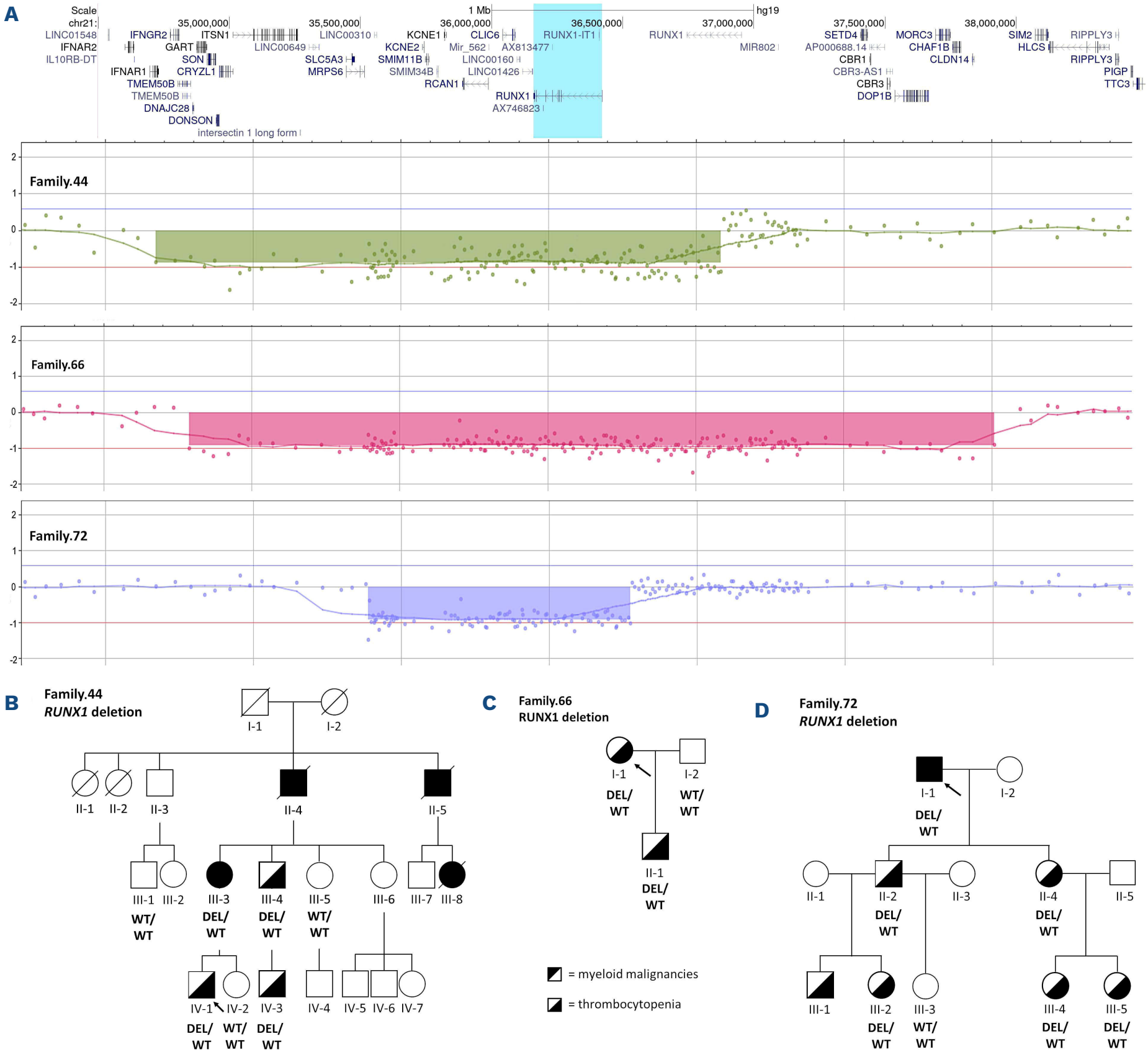

Figure 2. *RUNX1* deletions. (A) *RUNX1* deletions identified by exome sequencing were confirmed by chromosomal microarray analysis. Profiles from probands of Families 44, 66 and 72 (top to bottom panels) are shown. Highlighting indicates deleted regions. The top panel was produced by the UCSC Genome Browser (*https://genome.ucsc.edu/*) and shows the genomic positions and all genes included in the region. Segregation of deletions was evaluated by real-time polymerase chain reaction in Family.44 (B), Family.66 (C) and Family.72 (D). Genotype is reported when available.

Table 3. Deletions involving *RUNX1* identified through analysis of exome data with Excavator2. Annotation of breakpoints identified by chromosomal microarray analysis are reported.

Family ID	N of tested members (affected carriers/ healthy carriers)	Size (Kb)	Average LOG ratio	Genes	ISCN nomenclature
Family 44	9 (7/0)	1911.5	-0.88	*ITSN1, ATP5O, LINC00649, LOC101928126, MRPS6, SLC5A3, LINC00310, KCNE2, SMIM11A, C21orf140, KCNE1, RCAN1, CLIC6, LINC00160, LINC01426,* ***RUNX1****, RUNX1-IT1, LOC100506403*	arr[GRCh37] 21q22.11q22.12(35171289_37082807)x1
Family 66	2 (2/0)	2718.8	-0.913	*LINC00649, LOC101928126, SLC5A3, MRPS6, LINC00310, KCNE2, SMIM11A, C21orf140, KCNE1, RCAN1, CLIC6, LINC00160, LINC01426,* ***RUNX1****, RUNX1-IT1, LOC100506403, MIR802, PPP1R2P2, LOC101928269, LINC01436, SETD4, LOC100133286, CBR1, CBR3-AS1, CBR3, DOPEY2, MORC3, CHAF1B, CLDN14*	arr[GRCh37] 21q22.11q22.13(35289266_38008029)x1
Family 72	9 (5/0)	885.9	-0.903	*RCAN1, CLIC6, LINC00160, LINC01426,* ***RUNX1****, RUNX1-IT1, LOC100506403*	arr[GRCh37] 21q22.12(35888934_36774802)x1

ISCN: International System for Human Cytogenomic Nomenclature.

laborious and complex, which may negatively affect the adherence of clinicians to the diagnostic algorithm. As an additional complexity, patients with non-syndromic IT lack straightforward phenotypic features that easily raise diagnostic suspicion about specific genes, making the simultaneous analysis of all the genes the most reasonable strategy. Finally, false negative results from laboratory assays exploited in the diagnosis of IT can confound the process, as we observed in our cohort. Genetic heterogeneity also partially reflects on ES, since the larger the set of genes the higher the chance that candidate regions are not adequately represented (as reported here for *GP1BB*, *GP9* and *MPIG6B*) and that VUS are found.

Indeed, the interpretation of ES-identified variants was confirmed to be a major challenge. Here, we found that the proportion of cases carrying at least one VUS in known IT genes was as high as 29% (26/89). Previously reported data range from 13 to 50%, depending on criteria for selecting patients to be analyzed, sequencing techniques, and bioinformatic processing of data.[9,36-39] In our analysis, the availability of previous phenotypic and laboratory characterizations of pedigrees and the prompt access to DNA samples from patients' relatives was essential to define the pathogenic or non-pathogenic role of variants in many cases. In particular, segregation analysis was determinant for downgrading 14 variants from P/LP to VUS and 13 variants from VUS to B/LB, as well as for upgrading one VUS to LP.

This emphasizes the synergistic role of next-generation sequencing and accurate phenotype description of pedigrees for the improvement of the diagnostic process for IT. In particular, such an approach would significantly improve the interpretation of variants, thereby reducing the number of VUS, and overcome the pitfalls of the traditional diagnostic workup. In this framework, interaction between specialists and discussion of cases in multidisciplinary teams including geneticists, hematologists and laboratory experts appears to be the most proficient strategy.[25] Moreover, a periodic review of next-generation sequencing data in the light of new discoveries could be useful to refine the classification of variants.

Second, we showed that ES discloses causes of disease that could be otherwise overlooked according to assumptions made on the genetic model prior to the analysis. In-

deed, we observed how instances of pseudo-dominance (Family.43), presence of phenocopies (Family.41) and unexpected female-to-female X-linked transmission (Family.40) can hinder a diagnosis during a process in which the mode of inheritance is incorrectly assumed to drive the selection of the suspected genes. Of note, these events, although occurring rarely, collectively contributed for a non-negligible proportion of cases in our cohort (3.4%) and we therefore suggest that they should be taken into account in IT.

Finally, although ES is tailored to detect small variants (i.e., single nucleotide changes and indels), its data can be successfully used to identify CNV as well. Indeed, based on ES data we detected large deletions encompassing *RUNX1* in three cases (3.4%). Notably, this specific gene was suspected during the application of the diagnostic algorithm and its analysis by PCR and Sanger sequencing was correctly requested, but this technology could not identify these whole gene deletions. This finding is important as, if we consider all the IT probands available to us for whom a molecular diagnosis of IT was achieved (>165 families), the relative frequency of FPD-AML due to alterations of *RUNX1* was 4.2% and whole deletions accounted for 43% of FPD-AML cases. Thus, deletions involving *RUNX1* appear to be a relatively common cause of FPD-AML that may have been overlooked so far.

The gene-based variant enrichment analysis allowed us to identify *ACTN1* and *ETV6* as main contributing genes in this cohort. No further IT-associated gene, nor any “novel” gene, reached the study-wise statistical significance, confirming a picture of vast genetic heterogeneity for the genetic landscape of IT. It clearly emerged that only the few most frequent genes, including *ACTN1*, *ETV6* and some of the genes that were pre-screened here (*MYH9*, *ANKRD26*, *GP1BA*, *GP1BB*) are prevalent in IT. Conversely, a constellation of many other genes, each accountable for substantially less than 2% of cases, must be searched for variants in a diagnostic setting. Accordingly, we observed that apart from *ACTN1* and *ETV6*, the only two genes attaining study-wise statistical significance in the enrichment analysis, clinically relevant variation was dispersed across five (*GNE*, *PTPRJ*, *SLFN14*, *SRC* and *THPO*) of the 22 genes not in the pre-screening (*Online Supplementary Table S2*), accounting collectively for seven cases and for a maximum of two cases each, thereby emphasizing once again the genetic heterogeneity of IT.

Similarly, ”novel” genes (e.g., *PREX1*) may be present among those with top-ranking *P* values in the enrichment analysis, but their prevalence was too low to provide convincing evidence for an association with IT in this study. It is worth noting, in this respect, that if CNV and SNV had been included in the variant enrichment study, *RUNX1* would have ranked with a top *P* value (5 variants in cases, 1 variant in controls; uncorrected *P*=0.00125). This supports an important role for *RUNX1* in terms of prevalence and further highlights the importance of a comprehensive analysis of CNV and SNV, especially for this gene. We emphasize the importance of recognizing FPD-AML among IT since this disorder associates with a strong predisposition to hematologic neoplasms. Therefore, once a pathogenic variant in *RUNX1* has been identified, molecular analysis should be extended to all available family members: all individuals carrying the *RUNX1* mutation, including possible subjects with normal blood counts,[6] should receive proper genetic counseling and be offered an appropriate follow-up, with at least annual evaluations according to recent recommendations.[40] We expect that the application of the gene-based variant enrichment method presented here to larger cohorts might lead to the identification of new IT genes. As a limitation, we note that, in our study, segregation data weakened the role of many top-ranking genes. The application of the method on datasets in which more samples from the same family undergo ES could benefit from prompt enrichment in properly segregating variants.

It should be mentioned that our study was focused on Mendelian forms of IT. We explored the possibility of multigenic or incomplete penetrance, but our data are not sufficient to demonstrate a significant role for these mechanisms.

In conclusion, our results show how the application of an unbiased genomic approach to IT, inclusive of CNV evaluation, substantially increased the diagnostic rate in patients who remained undiagnosed after a thorough phenotype-driven investigation. The combined execution of both ES and accurate clinical-laboratory characterization in all patients with IT is expected to be the initial diagnostic approach with the highest probability of success. However, systematic application of this combined approach in all patients would be expensive and time-consuming, also in view of the increasing number of disorders being discovered as associated with IT, therefore necessitating an increasingly complex diagnostic workup for phenotypic characterization. Moreover, the study of IT patients needs specialized skills that often are not available locally: given that phenotypic characterization requires fresh blood samples, this usually means that patients have to travel long distances to reach the nearest reference center for the study of these rare diseases. Therefore, we consider it reasonable to propose that ES, which is becoming more and more economically convenient and can be performed on shipped samples, represents the initial investigation. The diagnosis indicated by ES should be confirmed though the study of the patient's clinical and laboratory phenotype and family history, in order to provide a correct interpretation of genetic variants. If this approach does not culminate in a diagnosis, then a complete phenotypic characterization needs to be

performed at a center with specific expertise in the diagnosis of IT, especially considering the disorders associated with genes not completely covered by ES. We also suggest that segregation of candidate variants is evaluated on all available family members to allow correct classification of variants.

Finally, our study disclosed that no novel genes make major contributions to IT, in terms of prevalence, thereby suggesting the need for larger, collaborative studies to identify the genes associated with the almost 50% cases with unknown molecular causes.

Disclosures

No conflicts of interest to disclose.

Contributions

CM, AP, AS, CLB, PN, TP, and MS conceived the study; FP and TP curated the data; CM, EN, and TP were responsible for the formal analysis; AP, AS, CLB, FM, PN and MS acquired funding; CM, AP, FM, FP, RB, EN, MF, SB, TG, VB, PM, PN, TP, and MS performed the investigations; EN and TP were responsible for the methodology; FP and TP were responsible for the software; AP, TP, and MS supervised the study; CM, AP, TP and MS wrote the original draft of the manuscript; and all the authors reviewed and edited the subsequent versions.

Funding

This work was supported by grants GGP13082 from Telethon Foundation (to MS, PN, and AS), by grant IG 2018-21974 from the Italian Association of Cancer Research (to AS), from the IRCCS Policlinico San Matteo Foundation (to AP) and from the Ministry of Health to IRCCS Burlo Garofolo (RC 2/18) and to FM (GR-2018-12367235).

Data-sharing statement

Original sequencing data cannot be shared for reasons of privacy.

References

1. Pecci A, Balduini CL. Inherited thrombocytopenias: an updated guide for clinicians. Blood Rev. 2021;48:100784.
2. Pecci A. Diagnosis and treatment of inherited thrombocytopenias. Clin Genet. 2016;89(2):141-153.
3. Balduini CL, Melazzini F, Pecci A. Inherited thrombocytopenias - recent advances in clinical and molecular aspects. Platelets. 2017;28(1):3-13.
4. Watson SP, Lowe GC, Lordkipanidzé M, Morgan NV; GAPP Consortium. Genotyping and phenotyping of platelet function disorders [published correction appears in J Thromb Haemost. 2013;11(9):1790]. J Thromb Haemost. 2013;11(Suppl 1):351-363.
5. Westbury SK, Turro E, Greene D, et al. BRIDGE-BPD Consortium. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. Genome Med. 2015;7(1):36.
6. Stockley J, Morgan NV, Bem D, et al. UK Genotyping and Phenotyping of Platelets Study Group. Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects. Blood. 2013;122(25):4090-4093.
7. Fletcher SJ, Son B, Lowe GC, et al. SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet secretion defects. J Clin Invest 2015;125(9):3600-3605.
8. Turro E, Greene D, Wijgaerts A, et al. A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. Sci Transl Med. 2016;8(328):328ra30.
9. Johnson B, Lowe GC, Futterer J, et al. UK GAPP Study Group. Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects. Haematologica. 2016;101(10):1170-1179.
10. Marconi C, Di Buduo CA, Barozzi S, et al. SLFN14-related thrombocytopenia: identification within a large series of patients with inherited thrombocytopenia. Thromb Haemost. 2016;115(5):1076-1079.
11. Richards S, Aziz N, Bale S, et al. ACMG Laboratory Quality Assurance Committee. 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. 2015;17(5):405-424.
12. Luo X, Feurstein S, Mohan S, et al. ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. Blood Adv. 2019;3(20):2962-2979.
13. Jarvik GP, Browning BL. Consideration of cosegregation in the pathogenicity classification of genomic variants. Am J Hum Genet. 2016;98(6):1077-1081.
14. Magi A, Tattini L, Palombo F, et al. H3M2: detection of runs of homozygosity from whole-exome sequencing data. Bioinformatics. 2014;30(20):2852-2859.
15. D'Aurizio R, Pippucci T, Tattini L, Giusti B, Pellegrini M, Magi A. Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2. Nucleic Acids Res. 2016;44(20):e154.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol 1995;57(1):289-300.
17. Bottega R, Marconi C, Faleschini M, et al. ACTN1-related thrombocytopenia: identification of novel families for phenotypic characterization. Blood. 2015;125(5):869-872.
18. Melazzini F, Palombo F, Balduini A, et al. Clinical and pathogenic features of ETV6-related thrombocytopenia with predisposition to acute lymphoblastic leukemia. Haematologica. 2016;101(11):1333-1342.
19. De Rocco D, Cerqua C, Goffrini P, et al. Mutations of cytochrome c identified in patients with thrombocytopenia THC4 affect both apoptosis and cellular bioenergetics. Biochim Biophys Acta. 2014;1842(2):269-274.
20. De Rocco D, Melazzini F, Marconi C, et al. Mutations of RUNX1 in families with inherited thrombocytopenia. Am J Hematol. 2017;92(6):E86-E88.
21. Noris P, Marconi C, De Rocco D, et al. A new form of inherited

thrombocytopenia due to monoallelic loss of function mutation in the thrombopoietin gene. Br J Haematol. 2018;181(5):698-701.
22. Faleschini M, Melazzini F, Marconi C, et al. ACTN1 mutations lead to a benign form of platelet macrocytosis not always associated with thrombocytopenia. Br J Haematol. 2018;183(2):276-288.
23. Marconi C, Di Buduo CA, LeVine K, et al. Loss-of-function mutations in PTPRJ cause a new form of inherited thrombocytopenia. Blood. 2019;133(12):1346-1357.
24. Barozzi S, Di Buduo CA, Marconi C, et al. Pathogenetic and clinical study of a patient with thrombocytopenia due to the p.E527K gain-of-function variant of SRC. Haematologica. 2021;106(3):918-922.
25. Greinacher A, Pecci A, Kunishima S. Diagnosis of inherited platelet disorders on a blood smear: a tool to facilitate worldwide diagnosis of platelet disorders. J Thromb Haemost. 2017;15(7):1511-1521.
26. Nurden AT, Pillois X, Fiore M, Heilig R, Nurden P. Glanzmann thrombasthenia-like syndromes associated with macrothrombocytopenias and mutations in the genes encoding the αIIbβ3 integrin. Semin Thromb Hemost. 2011;37(6):698-706.
27. Zaninetti C, Biino G, Noris P, Melazzini F, Civaschi E, Balduini CL. Personalized reference intervals for platelet count reduce the number of subjects with unexplained thrombocytopenia. Haematologica. 2015;100(9):e338-340.
28. Millikan PD, Balamohan SM, Raskind WH, Kacena MA. Inherited thrombocytopenia due to GATA-1 mutations. Semin Thromb Hemost. 2011;37(6):682-689.
29. Chou ST, Kacena MA, Weiss MJ, Raskind WH. GATA1-related X-linked cytopenia. 2006 Nov 22. Updated 2017. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2019.
30. Kunishima S, Okuno Y, Yoshida K, et al. ACTN1 mutations cause congenital macrothrombocytopenia. Am J Hum Genet. 2013;92(3):431-438.
31. Zhang MY, Churpek JE, Keel SB, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. Nat Genet. 2015;47(2):180-185.
32. Noetzli L, Lo RW, Lee-Sherick AB, et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. Nat Genet. 2015;47(5):535-538.
33. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013;9(8):e1003709.
34. Qian F, Le Breton GC, Chen J, et al. Role for the guanine nucleotide exchange factor phosphatidylinositol-3,4,5-trisphosphate-dependent rac exchanger 1 in platelet secretion and aggregation. Arterioscler Thromb Vasc Biol. 2012;32(3):768-777.
35. Galera P, Dulau-Florea A, Calvo KR. Inherited thrombocytopenia and platelet disorders with germline predisposition to myeloid neoplasia. Int J Lab Hematol. 2019;41 (Suppl 1):131-141.
36. Downes K, Megy K, Duarte D, et al. Diagnostic high-throughput sequencing of 2,396 patients with bleeding, thrombotic and platelet disorders. Blood. 2019;134(23):2082-2091.
37. Bastida JM, Lozano ML, Benito R, et al. Introducing high-throughput sequencing into mainstream genetic diagnosis practice in inherited platelet disorders. Haematologica. 2018;103(1):148-162
38. Leinøe E, Zetterberg E, Kinalis S, et al. Application of whole-exome sequencing to direct the specific functional testing and diagnosis of rare inherited bleeding disorders in patients from the Öresund Region, Scandinavia. Br J Haematol. 2017;179(2):308-322.
39. Romasko EJ, Devkota B, Biswas S, et al. Utility and limitations of exome sequencing in the molecular diagnosis of pediatric inherited platelet disorders. Am J Hematol. 2018;93(1):8-16.
40. Godley LA, Shimamura A. Genetic predisposition to hematologic malignancies: management and surveillance. Blood. 2017;130(4):424-432.