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

by Caterina Marconi, Alessandro Pecci, Flavia Palombo, Federica Melazzini, Roberta Bottega, Elena Nardi, Valeria Bozzi, Michela Faleschini, Serena Barozzi, Tania Giangregorio, Pamela Magini, Carlo L. Balduini, Anna Savoia, Marco Seri, Patrizia Noris, and Tommaso Pippucci

Received: March 4, 2022.
Accepted: August 29, 2022.


Publisher’s Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors’ final approval; the final version of the manuscript will then appear in a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.
Exome sequencing in 116 patients with inherited thrombocytopenia that remained of unknown origin after systematic phenotype-driven diagnostic workup

Exome sequencing in inherited thrombocytopenias

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

* These authors equally contributed to the work.
** These authors equally contributed to the work as senior authors.
Corresponding author

Affiliations
1 Department of Medical and Surgical Science, University of Bologna, Bologna, Italy
2 Department of Internal Medicine, University of Pavia, Pavia, Italy
3 Medicina Generale 1, IRCCS Policlinico San Matteo Foundation, Pavia, Italy
4 Institute for Maternal and Child Health – IRCCS Burlo Garofolo, Trieste, Italy
5 Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy
6 Medical Genetics Unit, IRCCS Azienda Ospedaliero Universitaria di Bologna, Policlinico di Sant’Orsola, Bologna, Italy
7 Department of Medical Sciences, University of Trieste, Trieste, Italy

Authors’ contributions

Corresponding Author
Prof. Marco Seri
IRCCS Azienda Ospedaliero Universitaria di Bologna, Policlinico di Sant’Orsola and Department of Medical and Surgical Sciences, University of Bologna, Via Massarenti 9, 40138 Bologna Italy
marco.seri@unibo.it
Tel +39 051 2088421, Fax: +39 051 2088416

Data sharing statement
Original sequencing data cannot be shared due to privacy reasons.

Word Count
Abstract: 200 words
Main text: 3994 words
Figures: 2
Tables: 3
Supplementary Data Files: 1
**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 Ministry of Health to IRCCS Burlo Garofolo (RC 2/18) and to FM (GR-2018-12367235).

**Conflict-of-interest disclosure**

The authors declare no competing financial interests.
Abstract

Inherited thrombocytopenias (ITs) are genetic diseases characterized by low platelet count, possibly associated with congenital defects or predisposition to develop additional conditions. Next generation sequencing has consistently improved our knowledge of ITs, with >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 on 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 puzzled by atypical inheritance, sex-related effects or false negative laboratory results. Finally, ES-based Copy Number Variant (CNV) analysis disclosed an unexpected high prevalence of RUNX1 deletions, predisposing to hematological malignancies. Our findings demonstrate that ES, including CNV 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 bleeding tendency of variable degree. In these disorders, thrombocytopenia can be isolated or associated with additional congenital defects; moreover, some IT are featured by the predisposition to develop additional diseases over time, such as hematological malignancies, bone marrow aplasia or renal failure.¹

Until recent years, the diagnosis of ITs was based on a complex process requiring a multi-step clinical and laboratory characterization of patients and the subsequent resequencing of candidate genes.² The introduction of next generation sequencing (NGS) 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 NGS led to the identification of many novel genes underlying ITs. These advances revealed a picture of wide genetic heterogeneity, with at least 40 genes implicated¹; of note, 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 ITs 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,⁴⁵ has proven effective in the description of novel forms of IT and the identification of causative variants 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 clinical picture, or is not associated with detectable alterations in platelet function.

Here we report the analysis of 116 patients with non-syndromic ITs 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 (Table S1) with familial thrombocytopenia that 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 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 (Figure S1). Details about the different forms of IT diagnosed through the application of the phenotype-driven diagnostic algorithm are reported in Table S3. All the cases recruited to the ES analysis presented apparently non-syndromic forms, with no clinical features additional to thrombocytopenia nor specific platelet function alterations. The study was approved by the Ethic Committee of the IRCCS Policlinico San Matteo Foundation.

All investigated individuals or their legal guardians provided written informed consent for 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 (Supplementary Methods). Sequence data analysis was carried out as already described.\(^10\)

Known IT genes analysis
We evaluated each patient for small variants and Copy number Variants (CNVs) in 43 genes associated with ITs (Table S2), including 21 genes pre-screened as part of the diagnostic algorithm and 22 genes found to be associated with IT in recent years (Table S2).
Single Nucleotide Variants and small InDels were selected according to minor allele frequencies and variant consequences, as detailed in Supplementary Methods, and confirmed by Sanger sequencing in probands and available relatives. Variant classification followed the American College of Medical Genetics (ACMG) guidelines\(^\text{11}\) and the Clin Gen Expert curation panel guidelines for \(\text{RUNX1}\) variants\(^\text{12}\). We evaluated segregation data as recommended by Jarvik and Browning.\(^\text{13}\)

Analysis of Runs of Homozygosity (ROH) was carried out on Family.43 with H\(^3\)M\(^2\)\(^\text{14}\) to identify large exomic homozygous regions. The genomic inbreeding coefficient was calculated as the percentage of the cumulative length of autosomal ROH >1.5 Mb over the overall length of the autosomal genome.

Exome alignments were used to call and genotype CNVs with Excavator\(^\text{2}\).\(^\text{15}\) Deletions encompassing \(\text{RUNX1}\) were confirmed by real-time PCR and chromosomal microarray analysis (CMA) on 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, healthy or presenting genetic disorders with no hematological involvement.

Qualifying variants were defined as SNVs affecting the canonical transcript of protein-coding sequences (non-synonymous and splice-site variants) with MAF ≤ 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 Supplementary Methods for more details).

To exclude population stratification among cases and controls, we removed from the analysis samples with non-European ancestry, based on self-reported information and outlier samples from a Principal Component Analysis (PCA) carried out on genotypes of our samples together with 2,504 samples from the 1000genomes dataset [https://www.internationalgenome.org/](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 control. Multiple testing control was done applying the false discovery rate criterion proposed by Benjamini and Hochberg choosing a FDR=0.2.\textsuperscript{16}

Variant confirmation by Sanger sequencing was carried out for the top ranking genes presenting a non-corrected p≤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%) (Table S2).

A total of 104 variants fulfilling selection criteria were identified in 60/89 probands (67%). Following 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 Table S4 reports VUS and B/LB variants. Table S6 reports applied ACMG criteria 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 AD or AD/AR forms, while the remaining ones were either heterozygous X-linked variants (6%; 2/32) or heterozygous, compound heterozygous and homozygous variants in AR genes (19%; 6/32).

The majority of P/LP variants (62%; 20/32) impacted genes that could not be previously analysed as not included in the diagnostic algorithm,\textsuperscript{2} and that were already described in our previous publications.\textsuperscript{10,17}\textsuperscript{24} Conversely, 9/31 P/LP variants (29%) were heterozygous variants in IT-associated genes of which the analysis should have been contemplated (Table 1), but that were overlooked due to 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 addi-
tional 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 variant in ITGA2B, flow cytometry failed to detect a decreased expression of glycoprotein complex IIb-IIIa on platelet surface.\(^{26}\) Moreover, \textit{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 1.A) 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 \textit{ABCG8} c.1234C>T, p.Arg412* variant, which was homozygous in II-1, III-1 and III-2 and heterozygous in II-2.

In a second family (Family 41, Figure 1.B), the LP variant \textit{WAS} c.134C>T, pThr45Met on chromosome X, was inherited by two males 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 (120 x10\(^9\)/L). In this situation, the high variability of platelet counts between the two probands (19 and 22 x10\(^9\)/L) and mother (80 x10\(^9\)/L) was attributed to other non-genetic factors, while it is most likely associated to the hemizygous vs. 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 of 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 1.C) we identified the heterozygous c.146delC, p.P50Rfs*86 variant in \textit{GATA1}, on the X chromosome, in the female proband (II-1). In this pedigree, an X-linked inheritance was not suspected due to 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.\textsuperscript{28,29} Details about these three last cases are reported in Supplementary Results.

We then reasoned that not all actual disease-contributing variants may be classified as P/LP, either because in IT-associated genes but not fulfilling ACMG guidelines, or because 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 qualifying variants (defined as predicted protein-altering variants with MAF≤0.0001 in ExAC Non-Finnish Europeans) in 81 unrelated cases and 215 controls of homogeneous Italian ancestry according to PCA. 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 2 of the 18 IT-associated genes mutated in this cohort, namely \textit{ACTN1}\textsuperscript{30} and \textit{ETV6}\textsuperscript{31,32} attained a study-wide significant excess of qualifying variants in probands (p=0.0001 and p=0.0004, respectively). \textit{ACTN1} had 9 qualifying variants in cases and 1 in controls, while \textit{ETV6} had 6 and 0, respectively (Table 2). Almost all \textit{ACTN1} and \textit{ETV6} alleles, including VUS and B/LB variants, were characterized by CADD and GERP scores 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, Table S4) but not all variant alleles could be defined as causative. As supported by our previous work, 7 \textit{ACTN1} and 3 \textit{ETV6} alleles were classified as P/LP,\textsuperscript{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-values≤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 (Table S5). In this context, we suggest \textit{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 vari-
ants (Residual Variation Intolerance Score, RVIS 22.30, ExAC constraint 4.11) (Table 2, Table S5). In two families, relatives were available for the analysis and PREX1 variants were segregating with the disorder in two and three affected members respectively (Figure S2). This gene has a role in regulating aggregation and dense granule secretion of mouse platelets. 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 CNVs, could play a role in disease. By using Excavator2, we identified heterozygous deletions involving RUNX1 in three unrelated families (Table 3, Figure 2). All the deletions were confirmed by RT-PCR, and CMA was used to annotate the genomic boundaries of the alterations (Figure 2.A). 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 confirmed in six affected individuals of Family.72 (Figure 2.B), four affected individuals in Family.44 (Figure 2.C), and two patients in Family.66 (Figure 2.D) and were found to span 900Kb (7 genes), 1.9Mb (18 genes) and 2.7Mb (29 genes), respectively. 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 the enrollment in this study, in accordance to the diagnostic algorithm. Few instances of large RUNX1 intragenic deletions have been described. The main clinical and laboratory features of our patients with heterozygous deletions involving RUNX1 are detailed in Supplementary Results. A history of myeloid neoplasms was reported in Family.72 and Family.44. Interestingly, if both small variants (SNVs and InDels) and CNVs would be included in the enrichment analysis, RUNX1 would rank in the top-ranking positions (5 variants in cases, 1 variants 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% (29 and 3 families with disease-causative SNV and CNV, respectively, out of 89) diagnostic yield. Compared to the phenotype-driven diagnostic algorithm on the same target genes, we estimated in 16% (14/89) the increase of the diagnostic yield attained by ES, including 11 families with SNV and 3 families with full RUNX1 deletions.
**Discussion**

Until a few years ago, the diagnosis of ITs 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. More recently, several groups introduced the use of NGS 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 where causative variants have been identified vary greatly according to different investigations.

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.

For this reason, our population is particularly informative to assess the advantages and issues of ES as a complement of 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 the phenotype-based diagnostic approach, we observed a 16% increase in the diagnostic yield attained with ES. This increase can be explained by its unbiased approach to the analysis of protein-coding variation that can be articulated in three major reasons.

First, the genetic heterogeneity of ITs, with more than 40 associated genes, makes the gene-by-gene approach very laborious and complex, a situation that may negatively affect the adherence of the clinicians to the diagnostic algorithm. As an addition in complexity, patients with non-syndromic IT lack straightforward phenotypic features that easily raise diagnostic suspicion towards 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 ITs can confound the process, as we observed in our cohort. Genetic heterogeneity partially reflects also on ES, since the larger the set of genes the higher the chance that candidate regions are not adequately represented (as reported here for GP1BA, GP9 and MPIG6B) and that VUS are found.
Indeed, the interpretation of ES-identified variants was confirmed to be a main challenge. Here, we describe a proportion of cases carrying at least one VUS in known IT genes 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. In our analysis, the availability of previous phenotypic and laboratory characterizations of pedigrees and the prompt access to DNA samples of 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 synergic role for NGS 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. Moreover, a periodic review of NGS data in the light of new discoveries could be useful to refine the classification of variants.

Second, we showed that ES disclose causes of disease that could otherwise be overlooked according to assumptions made on the genetic model prior to the analysis. We observed indeed 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 contribute for a non-negligible proportion of cases in our cohort (3.4%) and therefore we suggest that they should be taken into account in ITs.

Finally, although ES is tailored to detect small variants (i.e. single nucleotide changes and indels), its data can be successfully used also to identify CNVs. Indeed, from 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 which a molecular diagnosis of IT was achieved (>165 families), the relative frequency of FPD-AML due to alterations of RUNX1 is 4.2% and the whole deletions account 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 ITs. In particular, it clearly emerges that only few most frequent genes, including ACTN1, ETV6 and some of the genes that were here pre-screened (MYH9, ANKRD26, GP1BA, GP1BB) are prevalent in ITs. Conversely, a constellation of many minor 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 (Table S2), accounting for seven cases collectively and for maximum two cases each and thereby emphasizing IT genetic heterogeneity.

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 attain convincing evidence for association with ITs in this study. It is worth noticing, in this respect, that if we include CNV and SNV in the variant enrichment study, RUNX1 would be ranking with a top p-value (5 variants in cases, 1 variants 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 ITs since this disorder associates with a strong predisposition to haematological neoplasms. Therefore, once a pathogenic variant in RUNX1 is identified, molecular analysis should be extended to all available family members: all individuals carrying the RUNX1 mutation, including possible subjects with normal blood counts, should receive a proper genetic counselling and be offered an appropriate follow-up, with at least annual evaluations according to re-
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 notice that, in our study, segregation data weaken the role of many top-ranking genes. The application of the method on datasets in which more samples from the same family have ES performed could benefit from a prompt enrichment in properly segregating variants.

It is to mention that our study is focused on Mendelian forms of IT. We explored the possibility of multigenic or incomplete penetrance. However, our data are not sufficient to demonstrate a significant role for oligogenic inheritance or incomplete penetrance occurrences.

In conclusion, our results detailed 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 certainly expected to represent the initial diagnostic approach with the highest probability of success. However, a systematic application of this combined approach in all patients appears expensive and time consuming, also in view of the increasing number of disorders discovered as associated with IT, which therefore requires 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 the phenotypic characterization requires fresh blood samples, it usually implies that patients travel long distances to reach the nearest reference center for the study of these rare diseases. Therefore, we deem 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 through 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 reach 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 in ES at least with the approaches
used here. We also suggest that segregation of candidate variants is evaluated on all available family members to allow a correct variant classification.

Finally, our study disclosed how no novel gene gives major contributions to ITs 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.
References


Table 1. Known IT genes. Pathogenic and likely pathogenic variants identified in the 43 known IT genes. 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

<table>
<thead>
<tr>
<th>Family_id</th>
<th>N° tested members (affected carriers/healthy carriers)</th>
<th>Gene</th>
<th>Variant</th>
<th>Status</th>
<th>Variant Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTOSOMAL DOMINANT INHERITANCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family.1</td>
<td>4 (4/na)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.2305G&gt;A; p.(Glu769Lys)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.2</td>
<td>3 (3/na)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.673G&gt;A; p.(Glu225Lys)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.3</td>
<td>4 (4/na)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.384G&gt;C; p.(Trp128Cys)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.4</td>
<td>3 (3/na)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.136C&gt;T; p.(Arg46Trp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.5</td>
<td>3 (3/na)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.136C&gt;T; p.(Arg46Trp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.6</td>
<td>3 (2/0)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.982G&gt;A; p.(Val328Met)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.7</td>
<td>1 (1)</td>
<td>ACTN1</td>
<td>NM_001130004.1: c.698C&gt;T; p.(Pro233Leu)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.8</td>
<td>2 (2/na)</td>
<td>ANKR D26</td>
<td>NM_014915.2: c.-125T&gt;C; p.?</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.9</td>
<td>11 (7/0)</td>
<td>CYCS</td>
<td>NM_018947.5: c.145T&gt;C; p.(Tyr49His)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.11</td>
<td>2 (2/na)</td>
<td>ETV6</td>
<td>NM_001987.4: c.641C&gt;T; p.(Pro214Leu)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.12</td>
<td>2 (2/na)</td>
<td>ETV6</td>
<td>NM_001987.4: c.1105C&gt;T; p.(Arg369Trp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.13</td>
<td>2 (2/na)</td>
<td>ETV6</td>
<td>NM_001987.4: c.1105C&gt;T; p.(Arg369Trp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.17</td>
<td>1 (1/na)</td>
<td>ITG A2B</td>
<td>NM_000419.4: c.3076C&gt;T; p.(Arg1026Trp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.21</td>
<td>1 (1/na)</td>
<td>ITGB3</td>
<td>NM_000212.2: c.1768A&gt;G; p.(Thr590Ala)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.16</td>
<td>5 (4/0)</td>
<td>GFI1B</td>
<td>NM_004188.6: c.648+5G&gt;A; p.?</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.23</td>
<td>4 (4/na)</td>
<td>MY H9</td>
<td>NM_002473.5: c.121T&gt;A; p.(Phe41Ile)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.26</td>
<td>1 (1/na)</td>
<td>RUNX1</td>
<td>NM_001754.4: c.351+1G&gt;A; p.?</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.27</td>
<td>3 (3/na)</td>
<td>RUNX1</td>
<td>NM_001754.4: c.578T&gt;A; p.(Ile193Asn)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.28</td>
<td>3 (3/na)</td>
<td>SLFN14</td>
<td>NM_001129820.1: c.667C&gt;T; p.(Arg223Trp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.29</td>
<td>3 (1/0)</td>
<td>SRC</td>
<td>NM_198291.2: c.1579G&gt;A; p.(Glu527Lys)</td>
<td>Het</td>
<td>P</td>
</tr>
<tr>
<td>AUTOSOMAL DOMINANT/RECESSIVE INHERITANCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family.21</td>
<td>1 (1/na)</td>
<td>GP1BA</td>
<td>NM_000173.6: c.104del; p.(Lys35 ArgfsTer4)</td>
<td>Het</td>
<td>LP*</td>
</tr>
<tr>
<td>Family.35</td>
<td>1 (1/na)</td>
<td>GP1BA</td>
<td>NM_000173.6: c.169A&gt;G; p.(Asn57Asp)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.18</td>
<td>2 (2/na)</td>
<td>T H P O</td>
<td>NM_000460.4: c.91C&gt;T; p.(Arg31Ter)</td>
<td>Het</td>
<td>P</td>
</tr>
<tr>
<td>Family.39</td>
<td>2 (2/na)</td>
<td>T H P O</td>
<td>NM_000460.4: c.91C&gt;T; p.(Arg31Ter)</td>
<td>Het</td>
<td>P</td>
</tr>
<tr>
<td>X-LINKED INHERITANCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family.40</td>
<td>1 (1/na)</td>
<td>GATA1</td>
<td>NM_002049.3: c.149del; p.(Pro50ArgfsTer87)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>Family.41</td>
<td>6 (3/0)</td>
<td>WAS</td>
<td>NM_000377.2: c.134C&gt;T; p.(Thr45Met)</td>
<td>Het</td>
<td>LP</td>
</tr>
<tr>
<td>AUTOSOMAL RECESSIVE INHERITANCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family.43</td>
<td>4 (3/1)</td>
<td>ABCG8</td>
<td>NM_022437.2: c.1234C&gt;T; p.(Arg412Ter)</td>
<td>Hom</td>
<td>P</td>
</tr>
<tr>
<td>Family.81</td>
<td>1 (1/na)</td>
<td>GNE</td>
<td>NM_001128227.2: c.1768G&gt;A; p.(Gly590Arg)</td>
<td>Hom</td>
<td>LP</td>
</tr>
<tr>
<td>Family.82</td>
<td>1 (1/na)</td>
<td>GNE</td>
<td>NM_001128227.2: c.1427G&gt;T; p.(Arg476Met)</td>
<td>Hom</td>
<td>LP</td>
</tr>
<tr>
<td>Family.50</td>
<td>1 (1/na)</td>
<td>GNE</td>
<td>NM_001128227.2: c.98A&gt;G; p.(Glu33Gly)</td>
<td>Het</td>
<td>LP*</td>
</tr>
<tr>
<td>Family.51</td>
<td>4 (2/2)</td>
<td>PTP R J</td>
<td>NM_002843.3: c.97-2A&gt;G; p.?/NM_002843.3: c.1875del; p.(Ser627AlafsTer8)</td>
<td>Compound Het</td>
<td>P/P</td>
</tr>
</tbody>
</table>
Table 2. Gene-based rare variant collapsing analysis. 24 genes showing an uncorrected p value<=0.0015 are reported. RVIS, Residual Variation Intolerance Score. *genes with significant enrichment after multiple test correction. **significant score

<table>
<thead>
<tr>
<th>Ranking position</th>
<th>Gene</th>
<th>Uncorrected p value</th>
<th>Cases with qualifying variants</th>
<th>Cases with qualifying variants retained after segregation studies</th>
<th>Controls with qualifying variants</th>
<th>ExAC constraint z score</th>
<th>RVIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACTN1*</td>
<td>5.20E-06</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>3.82**</td>
<td>3.43</td>
</tr>
<tr>
<td>2</td>
<td>ETV6*</td>
<td>4.10E-05</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>2.2</td>
<td>20.53</td>
</tr>
<tr>
<td>3</td>
<td>PXDN</td>
<td>1.60E-04</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>2.72</td>
<td>11.47</td>
</tr>
<tr>
<td>3</td>
<td>YARS</td>
<td>1.60E-04</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1.15</td>
<td>17.20</td>
</tr>
<tr>
<td>4</td>
<td>PLA2G4C</td>
<td>3.20E-04</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>-0.63</td>
<td>91.06</td>
</tr>
<tr>
<td>5</td>
<td>TNS1</td>
<td>4.28E-04</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>-0.64</td>
<td>97.39</td>
</tr>
<tr>
<td>6</td>
<td>ADAR</td>
<td>6.70E-04</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3.01</td>
<td>2.20</td>
</tr>
<tr>
<td>6</td>
<td>ANKRD55</td>
<td>6.70E-04</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0.34</td>
<td>17.26</td>
</tr>
<tr>
<td>6</td>
<td>CLEC16A</td>
<td>6.70E-04</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1.44</td>
<td>25.02</td>
</tr>
<tr>
<td>6</td>
<td>GNAS</td>
<td>6.70E-04</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>4.34**</td>
<td>51.28</td>
</tr>
<tr>
<td>6</td>
<td>LY9</td>
<td>6.70E-04</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>-1.14</td>
<td>76.18</td>
</tr>
<tr>
<td>6</td>
<td>NECTIN1</td>
<td>6.70E-04</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>na</td>
<td>12.08</td>
</tr>
<tr>
<td>6</td>
<td>NNPAP1</td>
<td>6.70E-04</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>-1.39</td>
<td>78.58</td>
</tr>
<tr>
<td>6</td>
<td>PLBD1</td>
<td>6.70E-04</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0.25</td>
<td>61.52</td>
</tr>
<tr>
<td>6</td>
<td>PREX1</td>
<td>6.70E-04</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4.11**</td>
<td>22.30</td>
</tr>
<tr>
<td>7</td>
<td>CC2D2A</td>
<td>1.20E-03</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>-2.06</td>
<td>92.94</td>
</tr>
<tr>
<td>7</td>
<td>CCP110</td>
<td>1.20E-03</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0.13</td>
<td>24.49</td>
</tr>
<tr>
<td>7</td>
<td>GSG1L2</td>
<td>1.20E-03</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>7</td>
<td>LAMM4</td>
<td>1.20E-03</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>-0.67</td>
<td>15.11</td>
</tr>
<tr>
<td>7</td>
<td>PHLP1</td>
<td>1.20E-03</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2.46</td>
<td>7.51</td>
</tr>
<tr>
<td>7</td>
<td>PTH2R</td>
<td>1.20E-03</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>-1.72</td>
<td>54.83</td>
</tr>
<tr>
<td>7</td>
<td>ZNF394</td>
<td>1.20E-03</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>36.39</td>
</tr>
<tr>
<td>8</td>
<td>INPP5D</td>
<td>1.45E-03</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3.65**</td>
<td>15.51</td>
</tr>
<tr>
<td>8</td>
<td>MGA</td>
<td>1.45E-03</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>-1.31</td>
<td>2.51</td>
</tr>
</tbody>
</table>
Table 3. Deletions involving *RUNX1* identified through analysis of exome data with Excavator2. Annotation of breakpoints identified by chromosomal microarray analysis are reported.

<table>
<thead>
<tr>
<th>Family ID</th>
<th>N° tested members (affected carriers/healthy carriers)</th>
<th>Size (Kb)</th>
<th>Average LOG ratio</th>
<th>Genes</th>
<th>ISCN nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family.44</td>
<td>9 (7/0)</td>
<td>1911.5</td>
<td>-0.88</td>
<td>ITSN1, ATP5O, LINC00649, LOC101928126, MRPS6, SLC5A3, LINC00310, KCNE2, SMIM11A, C21orf140, KCNE1, RCAN1, CLIC6, LINC00160, LINC01426, <strong>RUNX1</strong>, RUNX1-IT1, LOC100506403</td>
<td>arr[GRCh37] 21q22.11q22.12(35171289_37082807)x1</td>
</tr>
<tr>
<td>Family.66</td>
<td>2 (2/0)</td>
<td>2718.8</td>
<td>-0.913</td>
<td>LINC00649, LINC01928126, SLC5A3, MRPS6, LINC00310, KCNE2, SMIM11A, C21orf140, KCNE1, RCAN1, CLIC6, LINC00160, LINC01426, <strong>RUNX1</strong>, RUNX1-IT1, LOC100506403, MIR802, PPP1R2P2, LOC101928269, LINC01436, SETD4, LOC100133286, CBR1, CBR3-AS1, CBR3, DOPEY2, MORC3, CHAF1B, CLDN14</td>
<td>arr[GRCh37] 21q22.11q22.13(35289266_38008029)x1</td>
</tr>
<tr>
<td>Family.72</td>
<td>9 (5/0)</td>
<td>885.9</td>
<td>-0.903</td>
<td>RCAN1, CLIC6, LINC00160, LINC01426, <strong>RUNX1</strong>, RUNX1-IT1, LOC100506403</td>
<td>arr[GRCh37] 21q22.12(35888934_36774802)x1</td>
</tr>
</tbody>
</table>
**Figure legends**

**Figure 1.** Families carrying pathogenic variants in *ABCG8* (A), *WAS* (B) and *GATA1* (C). Double link lines indicate consanguineous unions. Platelet count ($\times 10^9$/L) and genotype are reported when available. wt=wt allele; mut=mutated allele.

**Figure 2.** *RUNX1* deletions. A) *RUNX1* deletions identified by exome sequencing were confirmed by CMA. Profiles from probands of families 44, 66 and 72 (top to bottom panels) are shown. Highlighting indicates deleted regions. The top panel is 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 PCR in Family.44 (B), Family.66 (C) and Family.72 (D). Genotype is reported when available.
Supplementary Methods

Exome sequencing

DNA was extracted from whole blood using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Exome enrichment and capture was carried out with the BGI Exome kit (BGI TECH SOLUTIONS, Hong Kong) in cases. Regarding control samples: 150 samples were captured with the BGI exome kit (BGI TECH SOLUTIONS), 139 with the Nextera DNA Library Preparation Kit, 18 with the TruSeq Exome Kit (Illumina), 139 with the SeqCap EZ Exome + UTR Kit or the SeqCap EZ MedExome Kit (Roche, Basel, Switzerland). All samples were sequenced as 100/150 bp paired-end reads on Illumina sequencing platforms. Single nucleotide variants (SNVs) and small insertions/deletions (InDels) were annotated with the Ensembl tool Variant Effect Predictor v.76 on Gencode v19 transcripts against the GRCh37 reference genome.

Qualifying variants criteria for known IT genes analysis

Criteria for qualifying variants selection for known IT genes analysis are the following:

1. Minor Allele Frequencies (MAF) from the ExAC (http://exac.broadinstitute.org/)\(^1\). Non-Finnish European sub-population and across our in-house exomes dataset
   ≤ 0.01 for genes with biallelic or X-linked mechanisms;
   ≤ 0.001 for genes with monoallelic mode of inheritance;
2. Variant consequence. For ANKRD26, only variants in the 5’UTR were taken into account, while for all other genes, variants predicted to impact on the protein-coding sequence (nonsynonymous, splice-site and small insertions/deletions) were considered.

Classification following ACMG guidelines was done exploiting the VarSome utility (https://varsome.com/)\(^2\).

Copy Number Variants (CNVs) identification through EXCAVATOR2

We used EXCAVATOR2\(^3\) as follows. A window of 50Kb was used and samples from female and male individuals were analyzed separately. Each group was analyzed in the “pooled” mode, against 20 sex-paired healthy controls. CNVs called by Excavator2 and overlapping for at least 1 bp any of the 43 known IT genes were retained as candidate and confirmed by real-time PCR through the Universal Probe Library (UPL) system (Roche), using the TaqMan Universal PCR Master Mix (ThermoFisher Scientific). Primer pairs were designed (Assay Design Center, http://lifescience.roche.com/) to amplify a genomic region mapping within each CNV (sequences available on request). The DNA of two healthy individuals was processed along with patients’ DNA. Reactions were performed in triplicate for each primer pair, simultaneously amplifying the RPPH1 gene as diploid reference (TaqMan CopyNumber Reference Assay; ThermoFisher Scientific). The ΔΔCt method was applied to qPCR data to obtain a relative quantification of the copy number of analyzed genomic regions.

The SurePrint G3 Unrestricted CGH ISCA 8x60K platform (Agilent Technologies), with a 60 kb median probes spacing, was used to confirm RUNX1 deletions. Scanned images were analyzed by the Cytogenomics v.5.0 software, and aberrations were called with the ADM1 algorithm with a sensitivity threshold of 6.0.

Variants selection in the gene-based collapsing analysis

Well covered genes were defined as having 80% of the target regions covered by ≥10 reads in at least 80% of the cohort (both cases and controls).

Sample selection in the gene-based collapsing analysis

The Principal Component Analysis (PCA) was done exploiting the EIGENSTRAT software\(^4\) and restricted to 70,000 HapMap SNP sites with MAF >= 5% and residing within or near exons (+/-100 bp). Non-European individuals were defined as those whose value on one of the top two axes of variation was at least 5 standard deviation from the mean values of 909 Non-Finnish European individuals (including in-house samples and individuals from TSI, GBR, IBS and CEU populations of the 1000 Genomes).
Statistical data analyses were performed in R3.4.2 (https://www.R-project.org/).

Supplementary Results

**Pseudo-dominant inheritance of ABCG8 variant**

In Family 43, where thrombocytopenia segregates in an apparent autosomal-dominant fashion (Figure 1A), no candidate variants were found to segregate in a dominant manner. Conversely, we found that the proband II-2 (Figure 1A) and both her affected children (III-1 and III-2) were homozygous for the ABCG8 c.1234C>T, p.Arg412* variant, reported as pathogenic in ClinVar (RCV000005260.5). The apparently unrelated proband's healthy husband (II-1) was found to be heterozygous for the same variant. We identified a large ROH (about 11 Mb) encompassing ABCG8 in II-2 and III-1. We estimated a coefficient of inbreeding of 0.028 for II-2 and 0.029 for III-1, compatible with relatedness between II-1 and II-2 of cousins of 2nd/3rd degree as well as between I-1 and I-2, who, unfortunately, were not available for genetic testing.

Bi-allelic ABCG8 mutations cause sitosterolemia (OMIM #210250), a rare disorder characterized by increased plasma and tissue levels of plant sterols due to intestinal hyperabsorption and poor bile excretion. Clinical spectrum of sitosterolemia is extremely variable: possible features include thrombocytopenia with large platelets, hemolytic anemia, tendon and tuberous xanthomas, hypercholesterolemia, and premature atherosclerosis; asymptomatic individuals have been reported.\(^5\)\(^6\) Our patients were a 51-year-old female and her 16- and 13-year-old children presenting with mild thrombocytopenia (90-100 x10^9/L) with large platelets as the only clinical manifestation of the disease. Apart platelet macrocytosis, platelet morphology was normal in the three affected individuals. Platelet aggregation in response to ADP, collagen, ristocetin and arachidonic acid was tested in II-2 and III-1 and resulted within the normal range.

**WAS mutation inherited from heterozygous female**

In Family 41 (Figure 1B), we identified the WAS c.134C>T, p.Thr45Met variant, reported in ClinVar (VCV00001123.1) as causative of X-linked Thrombocytopenia (XLT, OMIM #313900), the allelic and mild variant of Wiskott-Aldrich Syndrome. The p.Thr45Met variant is present in hemizygous state in the two affected males IV-1 and IV-3, presenting severe thrombocytopenia (platelet count 19 and 22 x10^9/L, respectively) and is inherited from their mother III-4 who has a higher platelet count (80 x10^9/L). X chromosome inactivation test of III-4 showed a balanced profile (data not shown). The father of III-4 (II-1), having a platelet count below the normal threshold (120 x10^9/L), was initially considered to be affected and for these reasons, X-linked inheritance was not suspected in this family. However, based on the age- and gender-adjusted reference intervals for platelet count,\(^7\) platelet count of II-1 actually falls within the normal range (112-361 x10^9/L). DNA sample was not available for II-1 and II-2, and it was therefore not possible to evaluate their carrier status. However, the fact that the other daughter, III-3, does not carry the variant indicates that II-1 is not a variant carrier. Patients III-4, IV-1 and IV-3 have no other relevant clinical features in addition to thrombocytopenia. Analysis of platelet size on peripheral blood smears according to a previously reported method\(^8\) revealed that thrombocytopenia is associated with small platelets in individual IV-3 (mean platelet diameter, MPD, 2.10 \(\mu\)m), whereas III-4 and IV-1 presented normal platelet size (MPD 2.55 and 2.51 \(\mu\)m, respectively). Platelet aggregation in response to ADP, collagen, ristocetin and arachidonic acid was investigated in subject III-4 and resulted normal. Flow cytometry analysis showed that the patients III-4, IV-1 and IV-3 had normal expression of platelet glycoprotein complexes GPIb-IX-V, GPIIb-IIIa, and GP1a-IIIa on the platelet surface.

As reported in the literature, female carriers of WAS mutations are usually healthy because of a skewed inactivation of the mutated X chromosome\(^9,10\) and the few affected heterozygous females have total inactivation of the wild type allele.\(^11\)\(^-\)\(^13\) Of note, in our female patient (III-4) thrombocytopenia is much less severe than in the affected males of the family, suggesting a mild effect of this variant in non-skewed female carriers.

**GATA1 mutation in heterozygous female**

In the female proband of Family 40 (Figure 1C), we identified the heterozygous c.146delC, p.P50Rfs*86 variant in GATA1, a gene mapping to the X chromosome, for which we observed a balanced inactivation (data not shown). GATA1 mutations cause a spectrum of disorders from either dyserythropoietic anemia or hemolytic anemia resembling beta-thalassemia (OMIM #314050) in males to mild isolated thrombocytopenia rarely reported in heterozygous females.\(^14\,\)\(^15\)
Our proband is a 16-year-old female presenting with isolated thrombocytopenia (platelet count 110 x10^9/L), mild platelet macrocytosis, mild alpha-granule defect, and slight reduction of bone marrow megakaryocytes. Platelet aggregation studies showed normal response to ADP, collagen, and ristocetin. Her mother also has a lifelong history of mild thrombocytopenia, but was not available for genetic analysis to confirm the status of carrier.

To the best of our knowledge, few cases of heterozygous females affected with GATA1-associated disorders have been described.\textsuperscript{16-19} In only one case, a non-skewed X chromosome inactivation was demonstrated.\textsuperscript{20}

**RUNX1 haploinsufficiency caused by heterozygous intragenic deletions involving RUNX1**

Table S7 and S8 summarize the main clinical and laboratory findings of the investigated patients with heterozygous intragenic deletions involving RUNX1. The details about the deletions and family pedigrees are reported in Figure 2.

**Family 44.** The proband (IV-1) was a 16-year-old boy referred for mild thrombocytopenia with easy bruising. Platelet count was stably around 100 x10^9/L with normal MPV. Examination of peripheral blood smears showed no abnormalities of platelet size and morphology. In vitro platelet aggregation demonstrated defective response to ADP and collagen, and normal response to ristocetin (Table S7). Platelet expression of GPIb-IX-V and GPIIb-IIIa was normal. The proband’s 46-year-old mother (III-4) reported a history of mild thrombocytopenia since childhood and was diagnosed with myelodysplastic syndrome with multilineage dysplasia at the age of 44. Of note, she reported a history of leukemia (not otherwise specified) in her paternal lineage (Figure 2C). Even the proband’s maternal uncle (III-4) and a cousin (IV-3) presented a mild reduction in platelet count with normal platelet size and morphology and no significant bleeding tendency.

**Family 66.** The proband (I-1) was a 45-year-old female referred for mild chronic thrombocytopenia (platelet count around 100 x10^9/L) and easy bruising as the only bleeding manifestation. She reported a slight worsening of thrombocytopenia during her only pregnancy; however, the delivery proceeded without bleeding complications. Examination of peripheral blood smears did not reveal alterations of platelet size and morphology. The study of in vitro platelet aggregation showed normal response to ADP, collagen, and ristocetin; flow cytometry revealed normal platelet surface expression of GPIb-IX-V and GPIIb-IIIa. The proband’s son (II-1) presented thrombocytopenia since birth and reported easy bruising as the only bleeding manifestation.

**Family 72.** The proband (I-1) was a 56-year-old male referred to our institution for a lifelong history of thrombocytopenia and easy bruising as the only bleeding manifestation. Platelet count ranged from 50 to 70 x10^9/L with normal mean platelet volume (MPV); no further alterations of blood cell counts were present. At examination of peripheral blood smears, no citomorphological abnormalities were observed, in particular platelet size and morphology were normal. The proband’s 39-year-old son (II-2), his 38-year-old daughter (II-4), and 4 of his 5 grandchildren presented thrombocytopenia. Subject II-2 reported only very mild easy bruising; conversely, his sister II-4 had a lifelong history of easy bruising, frequent epistaxis, gum bleeding, and menorrhagia requiring hormonal therapy and iron supplementation. The proband’s grandchildren presented no bleeding tendency (n=1) or only mild bleeding tendency consisting in easy bruising and/or prolonged bleeding after minor wounds (n=3). All the investigated proband’s relatives had normal platelet size and morphology (Table S7). In vitro platelet aggregation was studied in 5 patients (Table S8). In summary, four patients showed defective response to ADP (n=4 to ADP 5 mM, and n=3 to ADP 20 mM as well); three of them presented also reduced aggregation after stimulation with collagen 4 and 20 µg/mL. One patient showed normal aggregation with all the tested agonists (Table S8). Flow cytometry analysis of platelets showed normal expression of the glycoprotein (GP) complexes Ib-IX-V and IIb-IIIa in all the 5 investigated patients (Table S7).

At age 59 years, the proband I-1 began to present progressive worsening of thrombocytopenia and of bleeding tendency, hyporegenerative anemia, and monocytosis; he was diagnosed with chronic myelomonocytic leukemia. He was treated with chemotherapy and allogeneic hematopoietic stem cell transplantation (HSCT). CMML relapsed a few months after HSCT and the patient died from progression of the hematological disease. No hematological malignancies have been reported in the other family members.
Supplementary References


Supplementary Tables as Excel Files

Online Supplementary Table S1. Main features of the study population.

Online Supplementary Table S2. Known IT genes included in this study. *Sub-optimal coverage for variant calling.

Online Supplementary Table S3. Specific diagnoses made in 116 probands through the application of the phenotype-driven diagnostic algorithm (reference 2) to 255 unrelated probands with inherited thrombocytopenia.

Online Supplementary Table S4. Known IT genes. VUS and (likely) benign variants identified in the 43 known IT genes.

Online Supplementary Table S5. Gene-based enrichment analysis: qualifying variants. Qualifying variants detected in genes with a p value <= 0.0015. Only variants with segregation compatible with the disorder and found in families with no disease causative variants in known IT genes are reported. Variants in the genes ACTN1 and ETV6 are reported in Table 1 and Table S3.

Online Supplementary Table S6. ACMG scoring details for variants in Table 1 and Table S3.

Online Supplementary Table S7. Main clinical and laboratory features of 13 FPD-AML patients with RUNX1 haploinsufficiency caused by heterozygous intragenic deletions involving RUNX1.

Online Supplementary Table S8. In vitro platelet aggregation in 7 FPD-AML patients carrying heterozygous intragenic deletions involving RUNX1, maximal extent (percentages).
Supplementary Figures

Figure S1. Flow-chart summarizing the total number of investigated probands, the results of the application of the phenotype-driven diagnostic algorithm, the recruitment and the results of the exome sequencing (ES) analysis.
Figure S2 Families carrying candidate variants in *PREX1*. Platelet count (x10⁹/L) and genotype are reported when available. wt=wt allele, mut=mutated allele.