ORIGINAL ARTICLE

Improving diagnostic precision, care and syndrome definitions using comprehensive next-generation sequencing for the inherited bone marrow failure syndromes

Ibrahim Ghemlas,[1,2,3] Hongbing Li,[1] Bozana Zlateska,[1] Robert Klaassen,[4] Conrad V Fernandez,[5] Rochelle A Yanofsky,[6] John Wu,[7] Yves Pastore,[8] Mariana Silva,[9] Jeff H Lipton,[10] Josee Brossard,[11] Bruno Michon,[12] Sharon Abish,[13] MacGregor Steele,[14] Roona Sinha,[15] Mark Belletrutti,[16] Vicky R Breakey,[17] Lawrence Jardine,[18] Lisa Goodyear,[19] Lillian Sung,[20] Santhosh Dhanraj,[1,21] Emma Reble,[1] Amanda Wagner,[2] Joseph Beyene,[22] Peter Ray,[1,23] Stephen Meyn,[1] Michaela Cada,[2] Yigal Dror[1,2,21]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2015-103270).

For numbered affiliations see end of article.

Correspondence to
Dr Yigal Dror, Division of Hematology/Oncology and Program in Genetics and Genome Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8; yigal.dror@sickkids.ca.

Received 15 May 2015
Accepted 7 June 2015
Published Online First 1 July 2015

ABSTRACT

Background Phenotypic overlap among the inherited bone marrow failure syndromes (IBMFSs) frequently limits the ability to establish a diagnosis based solely on clinical features. >70 IBMFS genes have been identified, which often renders genetic testing prolonged and costly. Since correct diagnosis, treatment and cancer surveillance often depend on identifying the mutated gene, strategies that enable timely genotyping are essential.

Methods To overcome these challenges, we developed a next-generation sequencing assay to analyse a panel of 72 known IBMFS genes. Cases fulfilling the clinical diagnostic criteria of an IBMFS but without identified causal genotypes were included.

Results The assay was validated by detecting 52 variants previously found by Sanger sequencing. A total of 158 patients with unknown mutations were studied. Of 75 patients with known IBMFS categories (eg, Fanconi anaemia), 59% had causal mutations. Among 83 patients with unclassified IBMFSs, we found causal mutations and established the diagnosis in 18% of the patients. The assay detected mutant genes that had not previously been reported to be associated with the patient phenotypes. In other cases, the assay led to amendments of diagnoses. In 20% of genotype cases, the results indicated a cancer surveillance programme.

Conclusions The novel assay is efficient, accurate and has a major impact on patient care.

INTRODUCTION

Inherited bone marrow failure syndromes (IBMFSs) are multisystem disorders with underproductive bone marrow and single-lineage or multilineage cytopenia.[1] Many of the disorders carry a risk of cancer. The term IBMFSs is reserved for disorders that are caused by germline mutations, either inherited or arising de novo with the patient.[2 3] Based on transmission patterns (eg, autosomal dominant (AD), autosomal recessive (AR) or X-linked), segregation of alleles within families and molecular analysis of IBMFS genes, all known IBMFSs appear to be monogenic.[4–6]

The wide range of physical anomalies associated with the IBMFSs help establish a specific diagnosis. However, the substantial phenotypic overlap among the disorders frequently limits the ability to establish a diagnosis based solely on clinical manifestations. Furthermore, physical malformations may be absent or appear later in life.[7] Identifying the specific mutated gene is essential. It helps establish a diagnosis, predict disease course (eg, cancer risk), direct genetic counselling and treatment and select healthy sibling donors for haematopoietic stem cell transplantation (HSCT). Since >70 IBMFS genes have been identified, timely and cost-effective strategies for genetic testing are necessary to provide proper care.

Sanger sequencing often poses significant limitations during the diagnostic evaluation of patients who have or suspected to have IBMFSs. First, there are >25 defined IBMFSs with substantial clinical overlap among as well as between IBMFSs and acquired marrow failure syndromes. Second, individual diseases (eg, Fanconi anaemia (FA), dyskeratosis congenita (DC) and Diamond–Blackfan anaemia (DBA)) can be caused by mutations in multiple genes. Since Sanger sequencing of multiple IBMFS genes is costly and lengthy,[8] it is not feasible in the setting of acute illnesses (eg, presentation with severe aplastic anaemia (SAA)) before urgent treatment decisions are made.

Next-generation sequencing (NGS) generates data on multiple DNA fragments in a single reaction.[9] Although application of NGS gene panels has been reported in several heterogeneous disease groups,[10–14] there are insufficient data about the clinical impact of this genetic approach on disorders with cancer risk with regard to facilitating a diagnosis, delivery of care and counselling of patients and other family members. There is one published study that used an NGS gene panel to study a variety of IBMFSs.[15] In that study that

To cite: Ghemlas I, Li H, Zlateska B, *et al*. *J Med Genet* 2015;**52**:575–584.

included a mixed population of paediatric and adult patients with bone marrow failure (including IBMFSs) and myelodysplastic syndromes (MDSs), the mutation detection rate was only 11%. We hypothesise that an NGS gene panel, which includes all known genes for single and multilineage cytopenias and uses the HaloPlex technology, could form the basis of a comprehensive testing strategy for IBMFSs in order to provide accurate and clinically relevant molecular diagnoses in a timely fashion and at a significantly reduced cost.

We developed an NGS assay to sequence a panel of 72 known IBMFS genes related to disorders with pancytopenia (eg, FA), disorders with predominantly anaemia (eg, DBA), disorders with predominantly neutropenia (eg, Kostmann/severe congenital neutropenia (K/SCN)), disorders with predominately thrombocytopenia (eg, familial thrombocytopenia (FT)) and inherited MDS (eg, Emberger syndrome). We applied the assay to a large population of 158 patients with IBMFSs. We first assessed the ability of the assay to identify causal mutations in patients that had been diagnosed with a specific IBMFS (eg, FA, DBA, K/SCN and FT), but whose genotype had not been identified. We then tested the ability of the assay to identify genotypes and establish a diagnosis in patients with unclassified IBMFSs.

METHODS

Patients

The Canadian Inherited Marrow Failure Registry (CIMFR) is a multicentre study, which was approved by the Institutional Ethics Board of all 17 participating institutions. In Canada, all paediatric patients with IBMFSs are typically treated in one of the CIMFR institutions. This registry is population-based as >90% of the patients in this study are from centres that enrol >80% of the patients at their institutions. Patients have been prospectively enrolled since January 2001 after obtaining written consent. Detailed information was collected at presentation, study entry and periodic follow-up.

Individuals who fulfilled published diagnostic criteria for an IBMFS[5] were recruited at the participating centres of the CIMFR. In short, the eligibility criteria include evidence of chronic bone marrow failure in addition to either family history or physical malformations or presentation earlier than 1 year of age or positive genetic testing. When possible, each case was assigned a specific syndromic diagnosis by the participating centre. Diagnoses were reviewed centrally and if necessary were adjusted based on published diagnostic criteria of specific IBMFSs[2 5] and after discussions with the respective centre. Cases that fulfilled the diagnostic criteria of an IBMFS but did not meet the clinical, laboratory and genetic diagnostic criteria for any known IBMFS[2 5] were defined unclassified IBMFSs.[8] The majority of these patients had undergone extensive genetic testing, which was negative. Patients who presented with SAA and did not respond to immunosuppressive therapy (ie, multilineage severe cytopenia and marrow cellularity <25% at 3 months after starting therapy) were also eligible for the CIMFR as having unclassified IBMFSs since a proportion of such patients would be ultimately diagnosed with IBMFS.

In the present study, 158 patients with classified or unclassified IBMFS without identified causal mutations were included; 155 were enrolled in the CIMFR and 3 were enrolled in an internal bone marrow failure study at the Hospital for Sick Children, Toronto. Of the 158 patients, 69 had prior clinical genetic testing, 72 were not tested and 17 had no available data on previous testing.

NGS IBMFS gene panel assay

We designed an NGS assay for a comprehensive panel of 72 IBMFS genes that had been published as of January 2013 (see online supplementary table S1). DNA was extracted from peripheral blood in most cases. For patients with MDS/acute myeloid leukaemia (AML), skin fibroblasts, marrow fibroblasts or peripheral blood T-cells were used for DNA extraction to minimise detection of somatic changes. We used the HaloPlex Capture Kit (Agilent Technologies, Santa Clara, California, USA) for DNA library preparation and capture according to the manufacturer's instructions. Targeted fragments were amplified while incorporating indexes and generating linear barcoded DNA fragments, and were sequenced on the Illumina HiSeq2500 platform. DNA libraries from 83 patients (first batch) and 85 patients (second batch) were pooled, labelled with different barcodes and sequenced in one lane.

Variant analysis and filtering strategy

The algorithm used to filter non-relevant variants is described in figure 1. To minimise false positive calls, we considered variants as true hits if they had ≥5 positive reads. For heterozygous calls, we also required that positive reads constitute ≥20% of the total reads for the respective nucleotides. However, to study the ability of the assay to detect mutations, we performed Sanger sequencing of any homozygous variants that appeared in ≥2 reads and any heterozygous variants that appeared in ≥20% of the reads. The software programs used to study minor allele frequency (MAF), conservation and potential damage of variants on the protein are listed in table 1.

Variants were defined as 'causal' if they had been reported as disease-causing in public databases (table 1). Novel variants were considered 'most likely causal mutations' if (1) they appeared in allelic dosage that was consistent with the known inheritance mode of the disease, (2) the MAF was <0.001, (3) evolutionary conserved amino acid/s are affected and (4) the variant was considered damaging by at least two of the following prediction software programs: PolyPhen2, SIFT/SIFT-Indel, Provean, MutationTaster and Human Splicing Finder. In this paper, we referred to both previously published mutations and novel mutations as 'causal mutations'.

Previously reported variants were considered 'not causal' if they had been published to be polymorphic. Novel variants that did not fulfil the criteria in the previous paragraph were deemed 'most likely not causal'. Variants that fulfilled most but not all the above criteria remained of unknown significance.

Sanger sequencing

Targeted sequences were analysed after PCR amplification by bi-directional sequencing as previously described.[16]

Statistical analysis and calculation of test sensitivity and precision

Calculation of test sensitivity and precision was performed as previously described.[17]

Sanger sequencing was considered as the reference standard method. Assay sensitivity was determined as $\sum$ true positive/$\sum$ condition position, where 'true positive' was considered when a mutation was observed by both NGS and Sanger sequencing and 'condition positive' was considered in any case where a mutation was observed by Sanger sequencing. Assay precision was determined as $\sum$ true positive/$\sum$ test outcome positive, where 'test outcome positive' was considered as any case with an identified causal mutation by NGS.

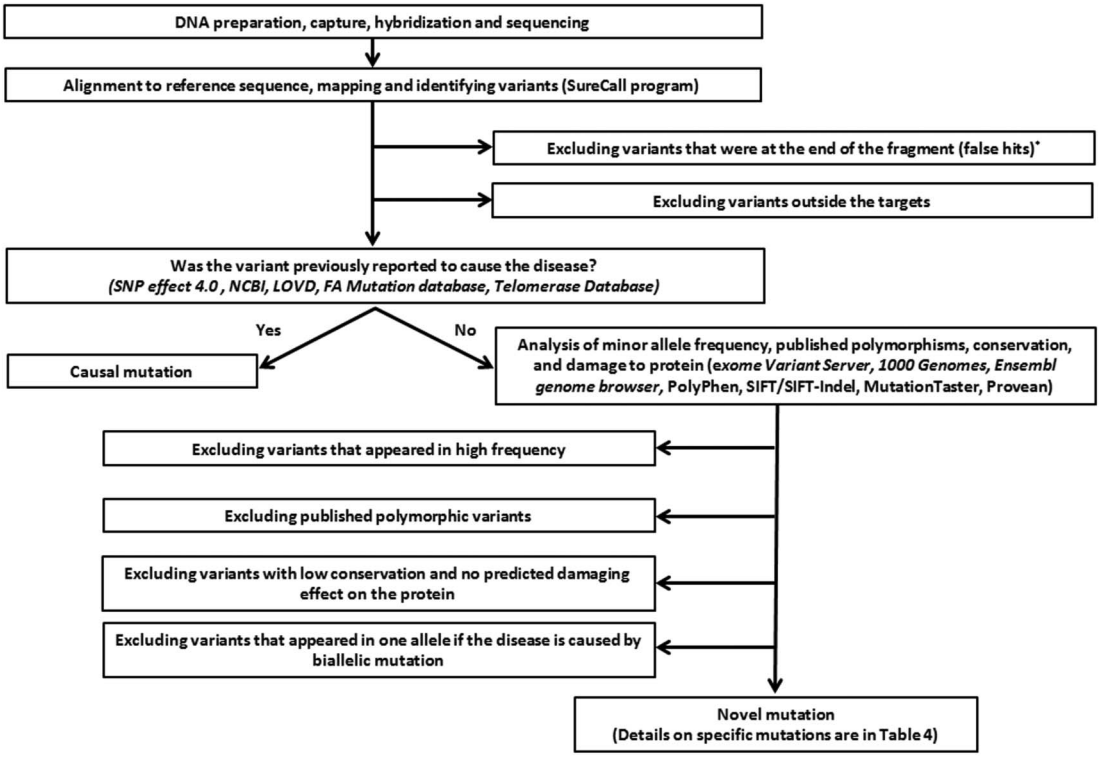


Figure 1 An algorithm summarising analysis and filtering processes. FA, Fanconi anaemia; LOVD, Leiden Open Variation Database; NCBI, National Center for Biotechnology Information; PolyPhen, polymorphism phenotyping; SIFT, sorting intolerant from tolerant.

Table 1 Software that were used in this study to evaluate conservation, minor allele frequency and potential damage of variants on splicing and protein structure and function

	Software	Source	Comments
Conservation	1. MutationAssessor 2. Exome Variant Server database	http://mutationassessor.org http://evs.gs.washington.edu/EVS	
Allele frequency	1. Exome Variant Server 2. 1000 Genomes 3. Variant Effect Predictor at the Ensembl genome browser	http://evs.gs.washington.edu/EVS http://browser.1000genomes.org/index.html http://useast.ensembl.org/Homo_sapiens/UserData/UploadVariations?db=core	Known polymorphism and common variants were excluded. We used a minor allele frequency (MAF) cut-off of 0.2% for variants in genes with AD and XL disease inheritance, and 0.5% for variants in genes with AR inheritance
Predict potential damage of amino acid substitution on protein structure and function	1. Polymorphism Phenotyping 2 (PolyPhen2) 2. Sorting Intolerant From Tolerant (SIFT) 3. Sorting Intolerant From Tolerant (SIFT)—Indel 4. Protein Variation Effect Analyzer (Provean) 5. MutationTaster	http://genetics.bwh.harvard.edu/pph2 http://sift.jcvi.org http://sift.bii.a-star.edu.sg/www/SIFT_indels2.html http://provean.jcvi.org http://www.mutationtaster.org	
Predict potential damage on splicing	1. Human Splicing Finder 3.0 2. MutationTaster	http://www.umd.be/HSF http://www.mutationtaster.org	
Reported disease-causing mutations	1. SNP effect 4.0 2. The National Center for Biotechnology Information's ClinVar database 3. the Leiden Open Variation Database (LOVD) 4. Fanconi Anemia Mutation database 5. The Telomerase Database	http://snpeffect.switchlab.org http://www.ncbi.nlm.nih.gov/clinvar http://databases.lovd.nl/shared/genes http://www.rockefeller.edu/fanconi/genes http://telomerase.asu.edu/diseases.html	

RESULTS

Frequencies of variants and causal mutations

The assay covered 456 351 bp. The average gene coverage was 99.12% (see online supplementary table S1). The average read depth was 680×. Also, 91.2% of targeted regions were covered with >100×. Among the 158 patients without known causal mutations, we identified 66 393 variants. After filtering (figure 1), 77 nucleotide-level variants (mutant alleles) in 59 patients were deemed causal (tables 2 and 3; see online supplementary tables S3 and S4). The majority of the mutations (44) have previously been reported; 33 were novel (see online supplementary table S5). Of the novel mutations, 15 were splicing mutations or indels, and 18 were missense mutations. Three of the novel mutations recurred in more than one patient.

Efficiency of variant detection

We evaluated the ability of the assay to detect 53 variants that were found by clinical testing prior to the present study; 40 were

Table 2 Genotyping patients with classified IBMFSs and unknown genes

Clinical diagnosis	Number tested	Number genotyped	Mutated genes	Number of cases with mutations in this gene	Type of mutations in this gene	Number of alleles with this type of mutation
DBA	23	16	*RPS26* (het)	5	Indel/frameshift	2
					Start code lost	2
					Splicing	1
			RPL11 (het)	3	Indel/frameshift	3
			RPS24 (het)	2	Start code lost	1
					Splicing	1
			RPS19 (het)	3	Splicing	1
					Missense	1
					Nonsense	1
			RPL35A (het)	1	Indel/inframe	1
			RPS7 (het)	1	Missense	1
			SBDS (combined)	1	Missense	1
					Splicing	1
FA	12	9	FANCA (hom or combined)	4	Missense	1
					Indel/frameshift	4
					Nonsense	2
					Splicing	1
			FANCA/BRIP1 (hom)	2	Nonsense	4
			FANCE (hom)	1	Missense	2
			FANCQ/ERCC4 (hom)	1	Missense	2
			TINF2 (het)	1	Missense	1
SCN	10	4	*ELANE* (het)	3	Missense	3
			HAX1 (hom)	1	Nonsense	2
CN	7	3	*ELANE* (het)	3	Missense	2
					Indel/frameshift	1
FT	4	4	*MYH9* (het)	1	Missense	1
			ANKRD26 (het)	1	Indel/frameshift	1
			TERT (het)	2	Splicing	2
DC	5	2	*RTEL1* (combined)	1	Missense	1
					Indel/frameshift	1
			TERC (combined)	1	ncRNA	2
TAR	2	2	*RBM8A**	2	5′-UTR	2
					Large deletion*	2
CDA	2	2	*CDAN1* (combined)	2	Missense	4
SDS†	5	1	*SBDS* (combined)	1	Indel/frameshift	1
					Splicing	1
CSA	2	1	*SLC25A38* (hom)	1	Missense	2
Radioulnar dysostosis	1	0				
Reticular dysgenesis	1	0				
CAMT	1	0				
Total	75	44		44		62‡

*Focused analysis of the NGS data at the *RBM8A* gene locus using the SureCall CNV algorithm identified a large deletion on the allele without the mutation (see online supplementary figures 42 and 43). This confirmed a compound heterozygosity state, which is the commonest genotype combination in this disease.
†Four of the patients with SDS had previous SBDS testing by Sanger sequencing in a clinical lab, which was negative. No mutations in other IBMFS genes were identified in these patients. A fifth patient had no prior testing for the SBDS gene; this patient was identified to have two SBDS mutations by the panel.
‡The total number of 62 mutant alleles includes two large deletions that were detected by focused analysis of the NGS data at the *RBM8A* gene using the SureCall CNV algorithm.
AD, autosomal dominant; AR, autosomal recessive; CAMT, congenital amegakaryocytic thrombocytopenia; CDA, congenital dyserythropoietic anaemia; CSA, congenital sideroblastic anaemia; CN, cyclic neutropenia; DBA, Diamond–Blackfan anaemia; DC, dyskeratosis congenita; FA, Fanconi anaemia; FT, familial thrombocytopenia; IBMFSs, inherited bone marrow failure syndromes; ncRNA, non-coding RNA; NGS, next-generation sequencing; SCN, severe congenital neutropenia; SDS, Shwachman–Diamond syndrome; TAR, thrombocytopenia with absent radii.

Table 3 Genotyping of patients with unclassified inherited bone marrow failure syndromes

Clinical phenotype	Number tested	Number genotyped	Genes mutated	Number of patients with mutations in this gene	Mutation type	Number of alleles	Diagnosis based on this study
Unclassified—IBMFS with predominantly neutropenia	14	3	*GATA2* (het)	1	Missense	1	Familial MDS
			WAS (hemi)	1	Missense	1	SCN
			G6PC3 (hom)	1	Indel/frameshift	2	SCN
Unclassified—IBMFS with bilineage or trilineage cytopenia	6	0					
Unclassified—IBMFS with bilineage or trilineage cytopenia	53	9	*TERT* (het)	3	Missense	3	DC
			TERC (het)	1	ncRNA	1	DC
			TINF2 (het)	1	Missense	1	DC
			CXCR4 (het)	1	Missense	1	WHIM syndrome
			RPL5 (het)	1	Indel/frameshift	1	DBA
			MYH9 (het)	1	Missense	1	MYH9- related disorder
			WAS (hemi)	1	Indel/inframe	1	WAS
SAA; no response to immunosuppressive therapy	10	3	*RTEL1* (comb)	1	Missense	2	DC
			TERT (Het)	1	Missense	1	DC
			MASTL (het)	1	Splicing	1	*MASTL*-associated disorder
Total	83	15		15		17	

AD, autosomal dominant; AR, autosomal recessive; DBA, Diamond–Blackfan anaemia; DC, dyskeratosis congenita; eADA, adenosine deaminase; IBMFSs, inherited bone marrow failure syndromes; IgG, immunoglobulin G; IST, immunosuppressive therapy; MDS, myelodysplastic syndrome; SAA, severe aplastic anaemia; SCN, severe congenital neutropenia; TL, telomere length; WAS, Wiskott–Aldrich syndrome; WHIM, Warts–Hypogamaglobulinemia–Infection–Myelokathexis.

polymorphisms in 21 of the 158 subjects in this study, and 13 were causal mutations in 10 other patients on the registry that had been genotyped (see online supplementary table S2). All variants were detected by the NGS assay, except for one polymorphic variant that was not covered; yielding a sensitivity of 98%.

Next, we determined the ability of the assay to detect mutations with sizable number of reads, which we defined as homozygous mutations of ≥5 reads and heterozygous mutations that appear in ≥5 reads and constitute ≥20% of the total reads. All 76 identified such causal mutations were validated by Sanger sequencing, giving a precision of 100% (see online supplementary figure S1–S58). Next, we studied the ability of the NGS assay to detect mutation calls with <5 reads. Among three such calls, two were found true by Sanger sequencing (see online supplementary figures S10 and S22). This suggests that calls with few reads may still be true and require validation.

We encountered several cases where the NGS assay outperformed Sanger sequencing in assessing complex genotypes. For example, the assay enabled determination whether two mutations in *RTEL1* were on the same allele (figure 2A), and whether a mutation in *SBDS* is true and not a contaminating pseudogene sequence (figure 2B).

Genotyping patients with classified IBMFSs

Of the 75 IBMFS patients with clinically classified IBMFSs (table 2), we identified 60 nucleotide-level causal mutations (9 of them were homozygous) in 44 patients (59%) by the NGS assay (see online supplementary table S3). Among patients who had not had previous genetic testing, 66% were genotyped. DBA was the most common IBMFS in the Canadian registry, followed by FA; among these disorders, 70% and 75% were found to have causal mutations by the NGS assay, respectively.

The NGS assay helped establish a precise diagnosis and discriminate between disorders with similar initial phenotypes but different natural histories. For example, one of two patients with FT (see online supplementary table S3, patients 38 and 39) had a mutation in *MYH9*, leading to a specific diagnosis of *MYH9*-associated FT, while the other was classified as having *ANKRD26*-associated FT, based on having a frameshift mutation in the *ANKRD26* gene. This frameshift causes loss of the last 50 amino acids of the protein, a region that is critical for the binding of ANKRD26 to its partner, TRIO, that shares cellular functions with ANKRD26.[18] Importantly, in contrast to *MYH9*-associated FT, *ANKRD26*-associated FT is associated with an increased risk of haematological malignancies[19] and indicates cancer surveillance.[20]

The analysis of two cases of thrombocytopenia with absent radii (TAR) syndrome exemplifies how compound heterozygosity that includes both one allelic deletion and a nucleotide-level mutation can be detected by NGS. Such a compound heterozygosity in *RBM8A* is the commonest cause of TAR syndrome.[6 21] The NGS data indicated a previously reported mutation in 5′-UTR area of *RBM8A* in both patients (figure 3A, see online supplementary table S3, patients 43 and 44). To determine whether the patients have homozygous mutations or compound heterozygosity with a submicroscopic monoallelic deletion, we used the SureCall CNV detection algorithm and our NGS data. Comparison of read numbers along *RBM8A* in these patients with five other subjects (figure 3B) suggested one allelic deletion in both patients. Copy number analysis in one of these cases by Affymetrix SNP6.0 array validated a submicroscopic deletion (figure 3C).

Amendment of diagnoses

The diagnosis of four clinically classified patients (9%) was amended after the results of the NGS gene panel assay became available. The first example is of a mother and son who were clinically diagnosed with non-syndromic FT (see online supplementary table S3, patients 40 and 41). Both were found to be

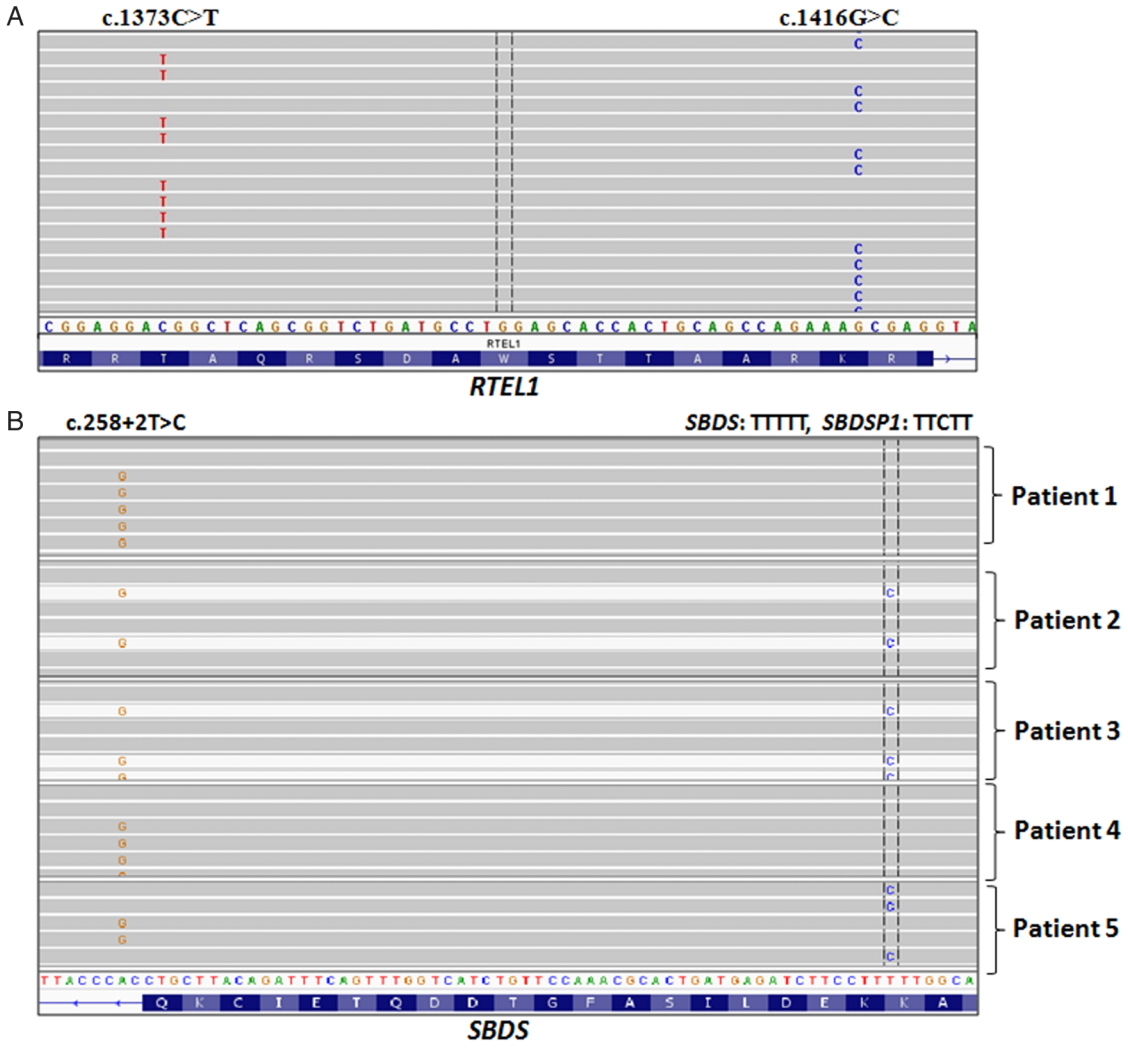


Figure 2 Next-generation sequencing (NGS) inherited bone marrow failure syndromes gene panel assay revealed challenging genotypes. (A) Using BAM files to depict results of massive parallel sequencing, each of the obtained sequences of *RTEL1* (see online supplementary table S4, patient 13) could be visualised separately. Both variants are located closer than the 150 bp size of the oligonucleotides used and it is possible to determine that they are on different alleles (and read). (B) *SBDS* and its pseudogene (*SBDSP1*) share 97% homology. Both were amplified by the designed NGS oligonucleotide. Sequences from several patients with the c.258+2T>C variant are seen. This common SDS variant results from conversion with *SBDSP1*. Based on the difference between *SBDS* sequences upstream to the mutation (TTTTT) and *SBDSP1* (TTCTT), we can discriminate true mutation from a contaminating *SBDSP1* sequence. Patients 1 and 4 have c.258+2T>C variant together with the *SBDS* sequence TTTTT, therefore, those variants are real *SBDS* mutations. Patients 2 and 3 have c.258+2T>C variant together with the SBDSP1 sequence, TTCTT, indicating that these sequences are of *SBDSP1* sequences and not real *SBDS* mutations. Patient 5 has both, c.258+2T>C variant and TTCTT, but they exist in different fragments, indicating that both are real *SBDS* mutations, and that they exist in different alleles.

heterozygous for a *TERT* mutation c.2383-15T>C (figure 4A), which was predicted to disrupt the binding site of splicing factor SRp40 and break the adjacent splicing site (see online supplementary table S5, patient 9). This mutation appeared in another unrelated patient in our registry, who had aplastic anaemia, very short telomeres in the range that is characteristic to DC and response to androgen therapy (figure 4B, see online supplementary table S2, patient 9). This mutation is very rare in the general population. Based on this information, the diagnosis was changed to DC.

The third patient was diagnosed clinically with FA based on haematological findings noticed at the age of 12.5 months, non-haematological features and increased chromosome fragility with hypersensitivity to mitomycin C and diepoxybutane (see online supplementary table S3, patient 25). Using the NGS assay, we found a mutation in *TINF2* (c.734C>A) (figure 4C), which was previously reported in a patient with aplastic anaemia. Accordingly, the diagnosis was amended to DC. Varying degree of chromosomal instability has been reported in DC,[22–25] but not to the degree found in this patient. Telomere length measurement was not available before the patient died.

The fourth patient was diagnosed with DBA based on severe anaemia, reticulocytopenia and markedly reduced marrow erythrocytes (see online supplementary table S3, patient 16). Marrow cellularity was reduced for the patient's age (70%), and moderate neutropenia (0.69×10^9/L) was registered once. The patient failed to respond to steroids. We found two mutations in *SBDS* (figure 4D): c.258+2T>C, which is the most common *SBDS* mutation, and c.127G>T, which is predicted to replace valine with leucine. Importantly, the substitution of G at the second last nucleotide of exon 1 (c.127) is also predicted to break the adjacent splice donor site. Sequencing of samples from the patient's parents showed that each parent was heterozygous for one of these mutations (figure 4E), confirming the

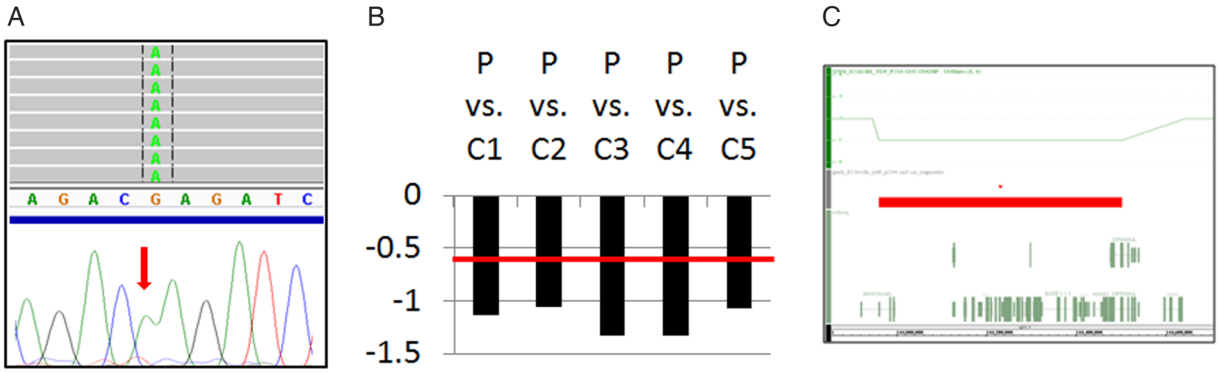


Figure 3 Revealing deletions in the *RBM8A* gene. (A) A previously described missense mutation in the *RBM8A* gene: c.-21G>A was detected by the NGS data (upper panel) and validated by Sanger sequencing (lower panel). Only a mutant allele was detected, which suggests either a homozygous mutation or a nucleotide-level mutation that is accompanied by a deletion on the other allele. (B) Copy number analysis was detected in the *RBM8A* gene. Patient read numbers in the gene region obtained from the NGS data were compared with read numbers from five subjects by the SureCall software. One copy number deletion was found as indicated by log2 of <0.6 when patient calls are compared with each of the other controls. (C) One copy number deletion in the *RBM8A* gene was validated by the Affymetrix SNP6.0 array.

biallelic nature of the maturations in the child. Sequencing of reverse transcription PCR products using three different pairs of primers repeatedly showed multiple cDNA products consisting with alternative splicing by both mutations. The overall SBDS protein was reduced to 60% of the level in normal samples (figure 4F). Based on the above, a diagnosis of SDS was deemed more likely than DBA.

Identifying mutations and establishing diagnoses in patients with unclassified IBMFSs

Eighty-three patients with unclassified IBMFSs were studied. These patients posed major diagnostic dilemmas and frequently underwent multiple testing over many years. We identified 17 causal mutations and established the specific IBMFS diagnosis in 15 patients (18%) (table 3, see online supplementary figure S59).

Three of the successfully genotyped unclassified patients with IBMFS had predominantly neutropenia (see online supplementary table S4, patients 1–3). Two patients had mutations in known K/SCN genes: *WAS* and *G6PC3*. However, in one case the mutated gene, *GATA2*, was not known as a K/SCN gene,[26–28] and biased testing for K/SCN-related genes would not have identified the causal genotype.

Nine of the successfully genotyped unclassified patients with IBMFS had multilineage cytopenia (see online supplementary table S4, patients 4–12). Five of these patients were ultimately diagnosed with DC due to mutations in telomere maintenance-associated genes (see online supplementary table S4, patients 4–8). The haematological phenotype of these patients with DC varied from predominantly anaemia to moderate aplastic anaemia and SAA. Interestingly, three other patients with pancytopenia were ultimately diagnosed with predominantly neutropenia syndrome (Warts–Hypogamaglobulinemia–Infection–Myelokathexis (see online supplementary table S4, patient 9), predominantly erythrocytopenia syndrome (DBA, see online supplementary table S4, patient 10) or predominantly thrombocytopenia syndrome (*MYH9*-related FT) (see online supplementary table S4, patient 11). These three patients would not have been genotyped if only pancytopenia-related genes had been sequenced.

Three genotyped unclassified patients with IBMFS belonged to a group of 10 patients who had SAA and no response to immunosuppressive therapy (see online supplementary table S4, patients 13–15). Two patients had mutations in pancytopenia-associated genes, *RTEL1* and *TERT*. Surprisingly, one patient had a mutation in microtubule-associated serine/threonine kinase-like (*MASTL*), which would not have been normally tested in cases of SAA, as so far it has been associated only with FT.

Changing indications for implementation of cancer surveillance programme

In 20% of genotype cases, the results indicated a cancer surveillance programme and proper family counselling. One example is a patient with chronic severe neutropenia and marrow monosomy 7 (see online supplementary table S4, patient 1), who was found to have a mutation in *GATA2*. Another example is a patient with chronic moderate neutropenia (see online supplementary table S4, patient 2), solitary kidney and hypocellular bone marrow, who was found to have an activating *WAS* mutation.[29]

Expanding syndromes' phenotype

Our study expanded the known phenotype of two syndromes. Neutropenia has not been reported as a feature of *MYH9*-associated FT.[30] One of the patients with unclassified IBMFSs presented with early-onset thrombocytopenia and neutropenia, which persisted at moderate level (see online supplementary table S4, patient 11). The successful genotyping suggests that mutations in *MYH9* may also cause neutropenia. The second syndrome is *MASTL*-associated disorder, which thus far has been linked only to non-syndromic FT, but the results in our study suggest that it can also be associated with SAA.

Cost consideration

We compared the cost of NGS assay to the cost of clinical testing for 30 patients with IBMFS enrolled at one of the CIMFR institutions (the Hospital for Sick Children, Toronto) between December 2010 and December 2013. Also, 21/30 had clinical genetic testing, averaged 5.95 tests/patient and US $4643/patient. These costs did not include expenses of DNA extraction and shipping to designated laboratories. The cost of NGS averaged $470/patient. This included reagents, sequencing service, bioinformatics and salary. In case of urgent testing without sequencing batching, the maximum price for NGS testing would be $2605/patient. This cost did not include a profit charge in the NGS cost, while this charge was

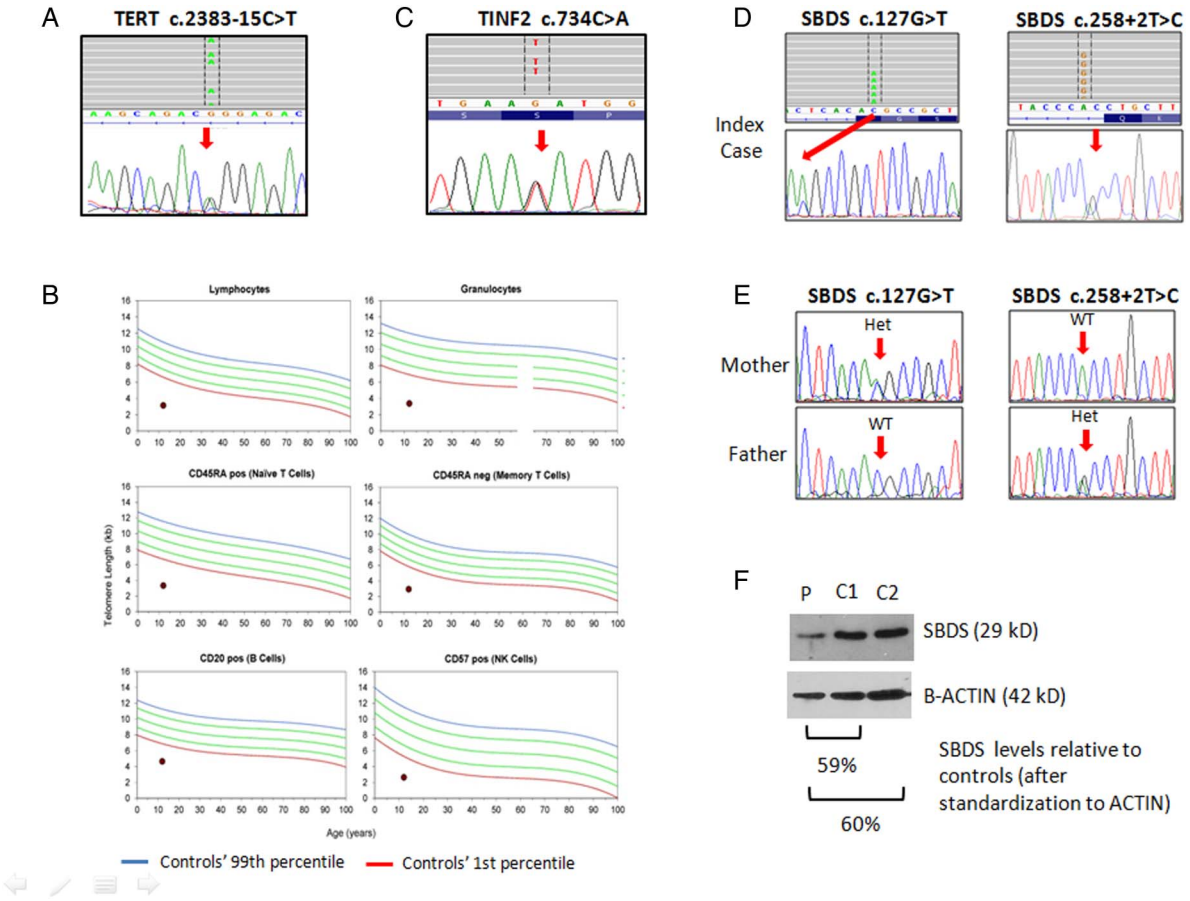


Figure 4 Mutations that lead to amendment of diagnosis. (A) Heterozygous mutation in *TERT*, c.2383-15C>T, was identified in a mother and son who were clinically diagnosed as autosomal-dominant non-syndromic familial thrombocytopenia (see also online supplementary table S3, patient 40). Bone marrow biopsy at the age of 17 years showed cellularity of 70%. The upper panel shows the nucleotide-level sequence by the BAM next-generation sequencing (NGS) file, which is the DNA reverse-complement counterpart. The lower panel shows the results of Sanger sequencing. (B) Telomere length in various peripheral blood cell lineages from a with *TERT* mutations c.2383-15C>T in the study was measured by flow fluorescent in situ hybridisation. The dots represent the patient values. The lines represent percentiles of telomere length among 391 healthy control persons ranging from birth to 100 years of age done in Repeat Diagnostic Laboratory (Vancouver, Canada). (C) *TINF2* mutation in a patient who was clinically diagnosed with Fanconi anaemia (FA). The upper panel shows the nucleotide-level sequence as found by NGS. The lower panel shows the results of Sanger sequencing. The patient underwent haplo-identical haematopoietic stem cell transplantation from her parent with FA-adjusted conditioning, but did not engraft. The patient had also a novel germline *RUNX1* variant in the marrow fibroblast DNA that was tested (c.383T>C, V128A), which might have contributed to her disease course. (D) *SBDS* mutations in a patient that was diagnosed clinically with Diamond–Blackfan anaemia. The upper panel shows the nucleotide sequence as found by next-generation sequencing. The middle panel shows the results of Sanger sequencing. (E) Sanger sequencing of *SBDS* in parents' samples. (F) for western blotting of the SBDS protein in the patient with SBDS mutations: c.127G>T/SBDS c.258+2T>C (P) and two controls (C1, C2) (C, control; P, patient).

incorporated in the service cost by the clinical laboratories. It is anticipated that rapid genotyping will not only reduce cost of genetic testing but also the cost of frequent clinic visits after presentation and other diagnostic ancillary laboratory tests such as skeletal survey, telomere length, adenosine deaminase levels, chromosome fragility test and pancreatic function tests.

DISCUSSION

We evaluated the effectiveness of a new comprehensive NGS IBMFS gene panel assay on a large cohort of patients with IBMFSs and showed that it detects mutations with high sensitivity and precision. The test assisted in establishing a diagnosis in difficult cases and amended diagnoses that have been established solely on clinical basis. It is rapid, cost-effective and yields high-positive hits compared with the typical diagnostic odyssey that many of these patients with IBMFS encounter.

Our assay was effective in identifying causal mutations in 59% of the classified cases and led to amendment of clinical diagnoses in 9% of the genotyped cases. This exemplifies the potential pitfalls of targeting a specific diagnosis in patients with IBMFSs, even when the clinical features are highly suggestive of that disorder. These pitfalls can be overcome only by a comprehensive panel testing of all IBMFS genes that associated with single and/or multilineage cytopenias, and not by panel assays that are restricted to genes associated with one disease/phenotype.

The results of this study also underscore the advantage of comprehensive testing in unclassified IBMFSs. We genotyped and consequently established a diagnosis in 18% of these cases. Our study uncovered atypical presentations of patients with specific genotypes (eg, certain *MASTL* mutations in SAA) that allow expansion of clinical definitions of syndromes and refinement of their diagnostic criteria. Although the number of patients with SAA that did not respond to immunosuppressive therapy that tested herein was only 10, the proportion of successfully genotyped patients appears to be higher (30%) than that in previous reports (<5%).[31–33] This is the first report of a patient with

SAA who was tested and found positive for mutations in *MASTL*. *MASTL*-associated IBMFS is AD and is characterised by moderate thrombocytopenia. Platelets are of normal size and function. Haemoglobin levels and neutrophils were reported to be normal. Marrow megakaryocytes are typically reduced.[34–36]

Our study shows that the NGS assay can have a major impact on clinical care. For example, the amendment of clinical diagnosis of FA to DC in one of the patients indicates a need for substantially more intensive HSCT preparatory regimen to achieve successful engraftment. Further, patients with IBMFS with similar presentations were found to have syndromes that carry markedly different cancer risk. For examples, patients with isolated neutropenia were classified as having a *GATA2*-associated disorder (very high risk of leukaemia, no risk of carcinomas) or a *CXCR4*-associated disorder (low or potentially absence of a risk of leukaemia, risk of mucocutaneous carcinoma). Also, four patients with a clinical diagnosis of non-syndromic FT were accurately categorised as having either *ANKRD26*-associated FT and DC (substantial risk of MDS/AML) or *MYH9*-related FT (no risk of MDS/AML).

Not all patients were genotyped by the NGS assay. This might be due to incomplete target coverage (~1%), exclusion of deep intronic areas, large indels, inability of bioinformatics tools to determine whether certain rare variant are causal and/or incomplete knowledge of all the IBMFS genes. Small indels and promoter mutations are captured by our panel design. In a proportion of the patients with FA, one allele of an early haematopoietic stem cells/progenitor undergoes spontaneous genetic correction and the respective developing precursors lose the increased chromosome fragility phenotype. This results in a mixture of healthy and FA cells and peripheral blood cell mosacism on chromosome fragility testing. In compound heterozygous cases, NGS will be able to detect the aberrant reads from non-corrected alleles. In most cases with homozygous mutations and mosacism, functional correction results in a sequence that is not identical to the wild-type one and the non-corrected mutant allele will still be identified by NGS. Genetic counselling should always be recommended, and the above limitations should be mentioned when results are disclosed to patients. Similar to reports of Sanger sequencing results, novel variants may be reported as likely positive or likely negative. Newly discovered genes can be incorporated in the panel as they are discovered by determining the precise coordinates of the fragments to be sequenced and designing oligonucleotides as described in online supplementary table S1. Hence, periodic repeat testing by updated gene panels may result in successful genotyping. Our assay can serve as a screening test before applying gene discovery methods such as exome sequencing.

In summary, our novel NGS IBMFS gene panel assay is a rapid, accurate and cost-effective strategy to genetically investigate patients with IBMFSs. The correct classification of IBMFs by NGS facilitates the more accurate medical management of these complex conditions. Therefore, we propose that NGS gene panels be considered as the first-line clinical molecular diagnostic test when the list of potentially mutated genes includes multiple candidates; this applies to the majority of patients with IBMFSs. Similar strategies may also be applied to other groups of genetic disorders with variable disease expression and presentation.

Author affiliations
[1]Program in Genetics and Genome Biology, Research Institute, Toronto, Ontario, Canada
[2]Marrow Failure and Myelodysplasia Program, Division of Hematology/Oncology, Department of Paediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada
[3]King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia
[4]Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada
[5]IWK Health Centre, Halifax, Nova Scotia, Canada
[6]CancerCare Manitoba, Winnipeg, Manitoba, Canada
[7]British Columbia Children's Hospital, Vancouver, British Columbia, Canada
[8]Hôpital Ste. Justine, Montréal, Québec, Canada
[9]Queen's University, Kingston, Ontario, Canada
[10]Princess Margaret Hospital, Toronto, Ontario, Canada
[11]Centre hospitalier universitaire, Sherbrooke, Quebec, Canada
[12]Centre Hospital University Quebec-Pav CHUL, Sainte-Foy, Quebec, Canada
[13]Montreal Children's Hospital, Montreal, Québec, Canada
[14]Alberta Children's Hospital, Calgary, Alberta, Canada
[15]University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[16]Stollery Children's Hospital, University of Alberta, Edmonton, Alberta, Canada
[17]McMaster Children's Hospital, McMaster University, Hamilton, Ontario, Canada
[18]Children's Hospital at London Health Sciences Centre, London, Ontario, Canada
[19]Janeway Child Health Centre, St. John's, Newfoundland, Canada
[20]Population Health Sciences, Research Institute, The Hospital For Sick Children, Toronto, Ontario, Canada
[21]Faculty of Medicine, Institute of Medical Sciences, University of Toronto, Toronto, Ontario, Canada
[22]Program in Population Genomics, Department of Clinical Epidemiology & Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
[23]Molecular Genetic Laboratory, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

Contributors IG and HL contributed equally to the paper and should be considered co-first authors. IG performed research, analysed data and wrote the paper. HL performed research, analysed data and wrote the manuscript. BZ collected and analysed data. RK, CVF, RAY, JW, YP, MSi, JHL, JBr, BM, SA, MSt, RS, MB, VRB, LJ, LG and MC are study co-investigators, contributed vital data and review and/or edited the paper. LS is a study investigator, involved in study design, interpretation of results and edited the manuscript. SD and ER performed research, analysed data and interpreted results. AW evaluated and contributed clinical and genetic data. JBe is a study co-investigator, contributed analytical tools. PR and SM analysed data, interpreted results and wrote the manuscript. YD designed research, oversaw the project, analysed data and wrote the paper.

Funding This work was supported by grants from C17 Canadian Research Network and Candlelighters Canada, from the Nicola Kids' Triathlon Fund and from the Canadian Institute of Health Research (funding reference 102528).

Competing interests None declared.

Ethics approval The Hospital for Sick Children Research Ethics Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Dror Y, Freedman F. Inherited bone marrow failure disorders. In: Churchill Livingstone EE, ed. *Hoffman's Textbook of Hematology: Principles and Practice*. 6th edn. Elsevier Health Sciences, 2012:307–49.
2 Alter BP. Inherited bone marrow failure syndromes. In: Nathan DG, Orkin SH, Ginsberg D, Look AT, eds. *Hematology of Infancy and Childhood*. Philadelphia: W.B. Saunders, 2003:280–365.
3 Dokal I, Vulliamy T. Inherited aplastic anaemias/bone marrow failure syndromes. *Blood Rev* 2008;22:141–53.
4 Shimamura A, Alter BP. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev* 2010;24:101–22.
5 Tsangaris E, Klaassen R, Fernandez CV, Yanofsky R, Shereck E, Champagne J, Silva M, Lipton JH, Brossard J, Michon B, Abish S, Steele M, Ali K, Dower N, Athale U, Jardine L, Hand JP, Odame I, Canning P, Allen C, Carcao M, Beyene J, Roifman CM, Dror Y. Genetic analysis of inherited bone marrow failure syndromes from one prospective, comprehensive and population-based cohort and identification of novel mutations. *J Med Genet* 2011;48:618–28.
6 Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA, Jolley JD, Cvejic A, Kostadima M, Bertone P, Breuning MH, Debili N, Deloukas P, Favier R, Fiedler J, Hobbs CM, Huang N, Hurles ME, Kiddle G, Krapels I, Nurden P, Ruivenkamp CA, Sambrook JG, Smith K, Stemple DL, Strauss G, Thys C, van Geet C, Newbury-Ecob R, Ouwehand WH, Ghevaert C. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 2012;44:435–9.
7 Alter BP. Diagnosis, genetics, and management of inherited bone marrow failure syndromes. *Hematology Am Soc Hematol Educ Program* 2007;2007:29–39.
8 Teo JT, Klaassen R, Fernandez CV, Yanofsky R, Wu J, Champagne J, Silva M, Lipton JH, Brossard J, Samson Y, Abish S, Steele M, Ali K, Athale U, Jardine L, Hand JP, Tsangaris E, Odame I, Beyene J, Dror Y. Clinical and genetic analysis of unclassifiable inherited bone marrow failure syndromes. *Pediatrics* 2008;122:e139–148.

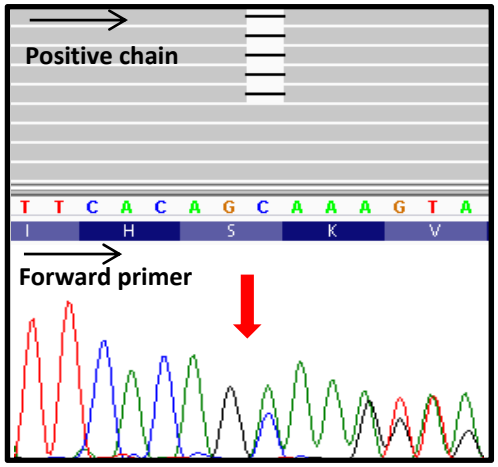
9 Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prev Res (Phila)* 2012;5:887–900.
10 Amstutz U, Andrey-Zurcher G, Suciu D, Jaggi R, Haberle J, Largiader CR. Sequence capture and next-generation resequencing of multiple tagged nucleic acid samples for mutation screening of urea cycle disorders. *Clin Chem* 2011;57:102–11.
11 Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Med* 2009;1:100.
12 Meder B, Haas J, Keller A, Heid C, Just S, Borries A, Boisguerin V, Scharfenberger-Schmeer M, Stahler P, Beier M, Weichenhan D, Strom TM, Pfeufer A, Korn B, Katus HA, Rottbauer W. Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Cir Cardiovasc Genet* 2011;4:110–22.
13 DaRe JT, Vasta V, Penn J, Tran NT, Hahn SH. Targeted exome sequencing for mitochondrial disorders reveals high genetic heterogeneity. *BMC Med Genet* 2013;14:118.
14 Berg JS, Evans JP, Leigh MW, Omran H, Bizon C, Mane K, Knowles MR, Weck KE, Zariwala MA. Next generation massively parallel sequencing of targeted exomes to identify genetic mutations in primary ciliary dyskinesia: implications for application to clinical testing. *Genet Med* 2011;13:218–29.
15 Zhang MY, Keel SB, Walsh T, Lee MK, Gulsuner S, Watts AC, Pritchard CC, Salipante SJ, Jeng MR, Hofmann I, Williams DA, Fleming MD, Abkowitz JL, King MC, Shimamura A. Genomic analysis of bone marrow failure and myelodysplastic syndromes reveals phenotypic and diagnostic complexity. *Haematologica* 2015;100:42–8.
16 Majeed F, Jadko S, Freedman MH, Dror Y. Mutation analysis of SBDS in pediatric acute myeloblastic leukemia. *Pediatr Blood Cancer* 2005;45:920–4.
17 Nicoll D, Detmer W. Current medical diagnosis & treatment. In: Tierney LM Jr, McPhee SJ, Papadekis MA, eds. *Basic Principles of Diagnostic Test Use and Interpretation*, 36th edn. Stamford, CT: Appleton & Lange, 1997:1–16.
18 Liu XF, Bera TK, Kahue C, Escobar T, Fei Z, Raciti GA, Pastan I. ANKRD26 and its interacting partners TRIO, GPS2, HMMR and DIPA regulate adipogenesis in 3T3-L1 cells. *PLoS One* 2012;7:e38130.
19 Noris P, Favier R, Alessi MC, Geddis AE, Kunishima S, Heller PG, Giordano P, Niederhoffer KY, Bussel JB, Podda GM, Vianelli N, Kersseboom R, Pecci A, Gnan C, Marconi C, Auvrignon A, Cohen W, Yu JC, Iguchi A, Miller Imahiyerobo A, Boehlen F, Ghalloussi D, De Rocco D, Magini P, Civaschi E, Biino G, Seri M, Savoia A, Balduini CL. ANKRD26-related thrombocytopenia and myeloid malignancies. *Blood* 2013;122:1987–9.
20 Balduini CL, Pecci A, Noris P. Diagnosis and management of inherited thrombocytopenias. *Semin Thromb Hemost* 2013;39:161–71.
21 Klopocki E, Schulze H, Strauss G, Ott CE, Hall J, Trotier F, Fleischhauer S, Greenhalgh L, Newbury-Ecob RA, Neumann LM, Habenicht R, Konig R, Seemanova E, Megarbane A, Ropers HH, Ullmann R, Horn D, Mundlos S. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am J Hum Genet* 2007;80:232–40.
22 DeBauche DM, Pai GS, Stanley WS. Enhanced G2 chromatid radiosensitivity in dyskeratosis congenita fibroblasts. *Am J Hum Genet* 1990;46:350–7.
23 Deng Z, Glousker G, Molczan A, Fox AJ, Lamm N, Dheekollu J, Weizman OE, Schertzer M, Wang Z, Vladimirova O, Schug J, Aker M, Londono-Vallejo A, Kaestner KH, Lieberman PM, Tzfati Y. Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. *Proc Natl Acad Sci USA* 2013;110:E3408–16.
24 Murnane JP. Telomere dysfunction and chromosome instability. *Mutat Res* 2012;730:28–36.
25 Sabatier L, Ricoul M, Pottier G, Murnane JP. The loss of a single telomere can result in instability of multiple chromosomes in a human tumor cell line. *Mol Cancer Res* 2005;3:139–50.
26 Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, Frucht DM, Vinh DC, Auth RD, Freeman AF, Olivier KN, Uzel G, Zerbe CS, Spalding C, Pittaluga S, Raffeld M, Kuhns DB, Ding L, Paulson ML, Marciano BE, Gea-Banacloche JC, Orange JS, Cuellar-Rodriguez J, Hickstein DD, Holland SM. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* 2011;118:2653–5.
27 Ostergaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, Dafou D, Kilo T, Smithson S, Lunt P, Murday VA, Hodgson S, Keenan R, Pilz DT, Martinez-Corral I, Makinen T, Mortimer PS, Jeffery S, Trembath RC, Mansour S. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* 2011;43:929–31.
28 Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, Babic M, Lin M, Carmagnac A, Lee YK, Kok CH, Gagliardi L, Friend KL, Ekert PG, Butcher CM, Brown AL, Lewis ID, To LB, Timms AE, Storek J, Moore S, Altree M, Escher R, Bardy PG, Suthers GK, D'Andrea RJ, Horwitz MS, Scott HS. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* 2011;43:1012–17.
29 Ancliff PJ, Blundell MP, Cory GO, Calle Y, Worth A, Kempski H, Burns S, Jones GE, Sinclair J, Kinnon C, Hann IM, Gale RE, Linch DC, Thrasher AJ. Two novel activating mutations in the Wiskott-Aldrich syndrome protein result in congenital neutropenia. *Blood* 2006;108:2182–9.
30 Seri M, Cusano R, Gangarossa S, Caridi G, Bordo D, Lo Nigro C, Ghiggeri GM, Ravazzolo R, Savino M, Del Vecchio M, d'Apolito M, Iolascon A, Zelante LL, Savoia A, Balduini CL, Noris P, Magrini U, Belletti S, Heath KE, Babcock M, Glucksman MJ, Aliprandis E, Bizzaro N, Desnick RJ, Martignetti JA. Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Hegglin/Fechtner Syndrome Consortium. *Nat Genet* 2000;26:103–5.
31 Du HY, Mason PJ, Bessler M, Wilson DB. TINF2 mutations in children with severe aplastic anemia. *Pediatr Blood Cancer* 2009;52:687.
32 Pigullo S, Pavesi E, Dianzani I, Santamaria G, Svahn J, Risso M, Van Lint MT, Pillon M, Iori AP, Longoni D, Ramenghi U, Lanciotti M, Dufour C. NOLA1 gene mutations in acquired aplastic anemia. *Pediatr Blood Cancer* 2009;52:376–8.
33 Field JJ, Mason PJ, An P, Kasai Y, McLellan M, Jaeger S, Barnes YJ, King AA, Bessler M, Wilson DB. Low frequency of telomerase RNA mutations among children with aplastic anemia or myelodysplastic syndrome. *J Pediatr Hematol Oncol* 2006;28:450–3.
34 Drachman JG, Jarvik GP, Mehaffey MG. Autosomal dominant thrombocytopenia: incomplete megakaryocyte differentiation and linkage to human chromosome 10. *Blood* 2000;96:118–25.
35 Gandhi MJ, Cummings CL, Drachman JG. FLJ14813 missense mutation: a candidate for autosomal dominant thrombocytopenia on human chromosome 10. *Hum Hered* 2003;55:66–70.
36 Johnson HJ, Gandhi MJ, Shafizadeh E, Langer NB, Pierce EL, Paw BH, Gilligan DM, Drachman JG. In vivo inactivation of MASTL kinase results in thrombocytopenia. *Exp Hematol* 2009;37:901–8.

Supplemental Figures

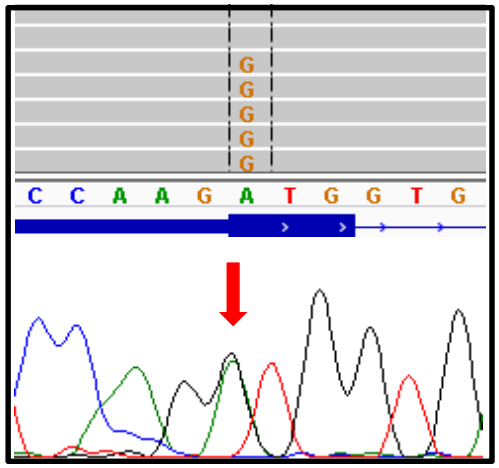
Improving diagnostic precision, care and syndrome definitions using comprehensive next generation sequencing for the inherited bone marrow failure syndromes

Ibrahim Ghemlas, Hongbing Li, Bozana Zlateska, Robert Klaassen, Conrad V Fernandez, Rochelle A Yanofsky, John Wu, Yves Pastore, Mariana Silva, Jeff H Lipton, Josse Brossard, Michon Bruno, Sharon Abish, MaCregor Steele, Roona Sinha, Mark Belltrutti, Vicky Breaky, Lawrence Jardine, Lisa Goodyear, Lillian Sung, Santhosh Dhanraj, Emma Reble, Amanda Wagner, Joseph Beyene, Peter Ray, Stephen Meyn, Michaela Cada, Yigal Dror

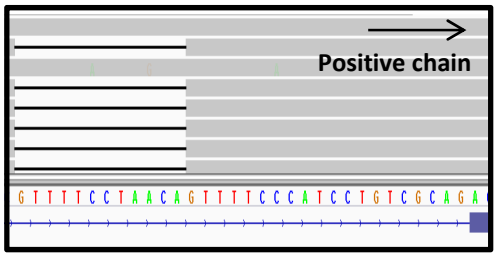
Figures 1-58 show the mutation areas from the next generation sequencing BAM files and information about Sanger sequencing validation. Please note that in the BAM files only a proportion of the reads are shown (about 7-8 of an average of 680 reads/patient).



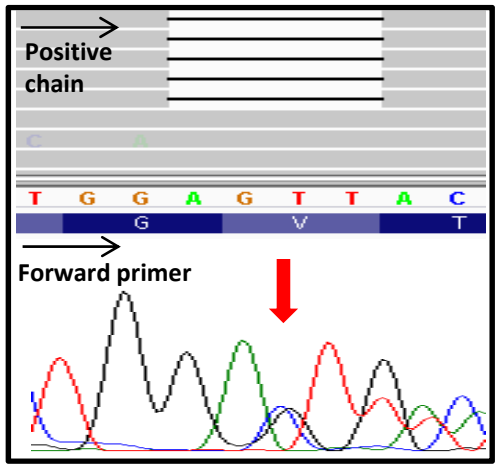
1. Patient 1 in in Supplemental Table 3. *RPS26* c.243delC



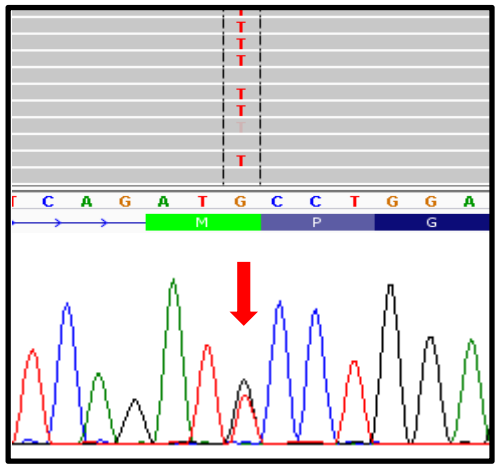
2. Patient 2 in Supplemental Table 3. *RPS26* c.1A>G



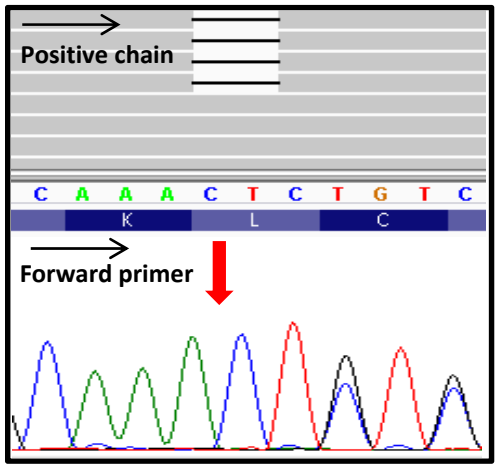
3. Patient 3 in Supplemental Table 3. *RPS26* c.4-32_21delGTTTTCCTAACA (Mutation was Validated in a clinical lab)



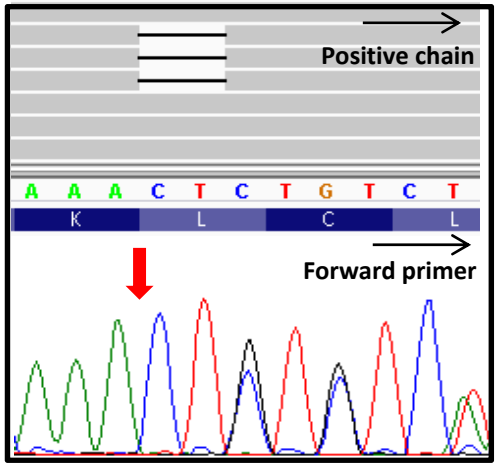
4. Patient 4 in Supplemental Table 3. *RPS19* c.10_13delAGTT



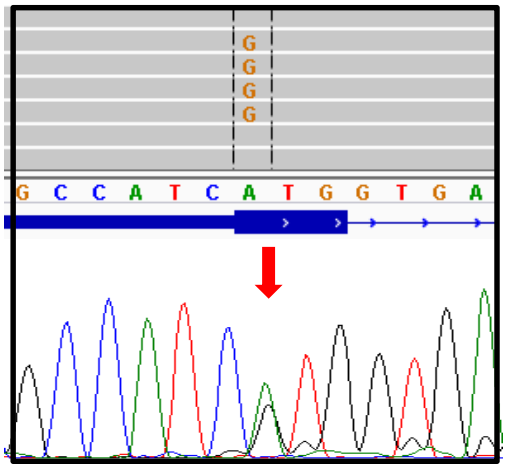
5. Patient 5 in Supplemental Table 3. *RPS19* c.3G>T



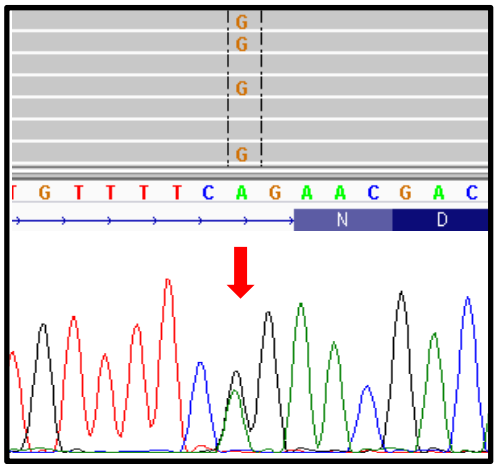
6. Patient 1 in Supplemental Table 3. *RPL11* c.60_61delCT



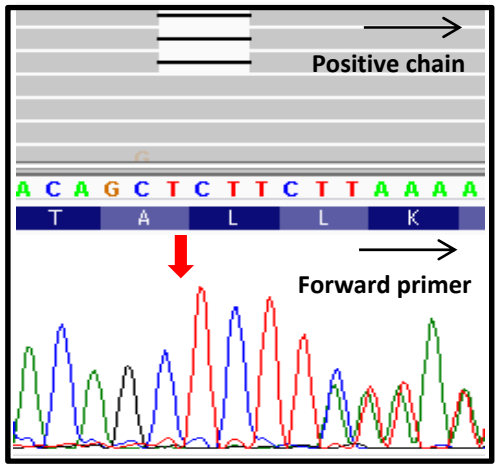
7. Patient 7 in Supplemental Table 3.
RPL11 c.60_61delCT



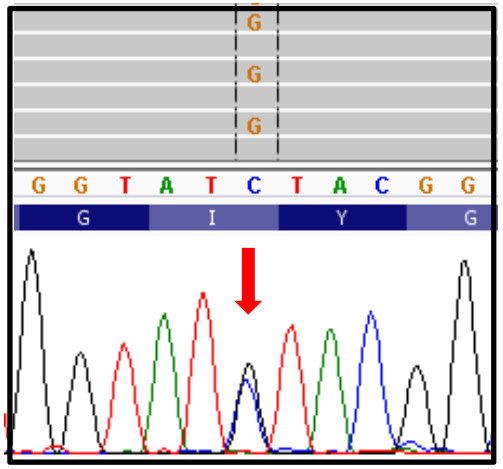
8. Patient 8 in Supplemental Table 3.
RPS24 c.1A>G



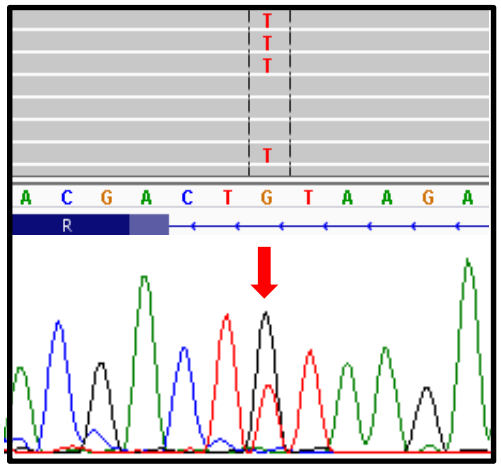
9. Patient 9 in Supplemental Table 3.
RPS24 c.4-2A>G



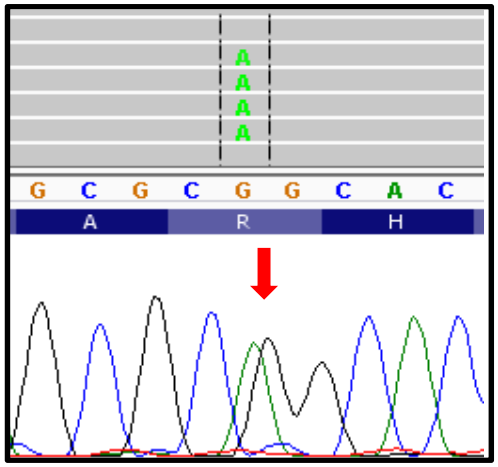
10. Patient 10 in Supplemental Table 3.
RPL35A c.78_80delTCT



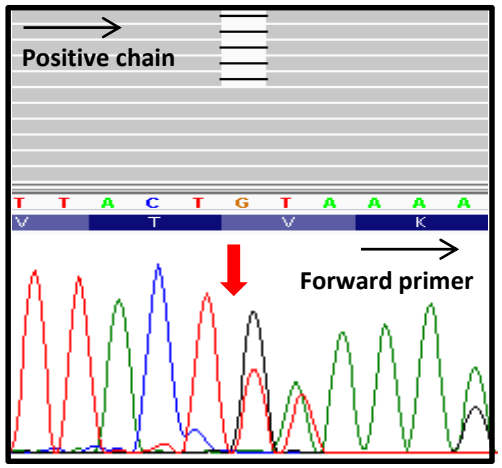
11. Patient 11 in Supplemental Table 3.
RPL11 c.372C>G



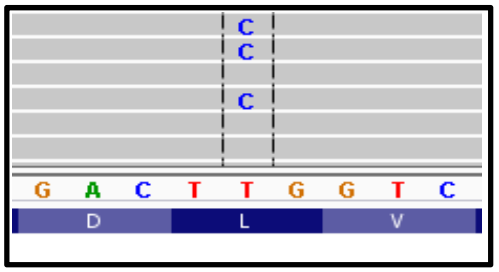
12. Patient 12 in Supplemental Table 3.
RPS29 c.63-3C>A



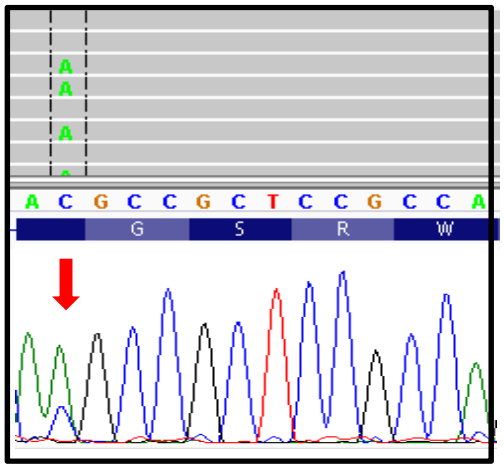
13. Patient 13 in Supplemental Table 3. *RPS19* c.185G>A



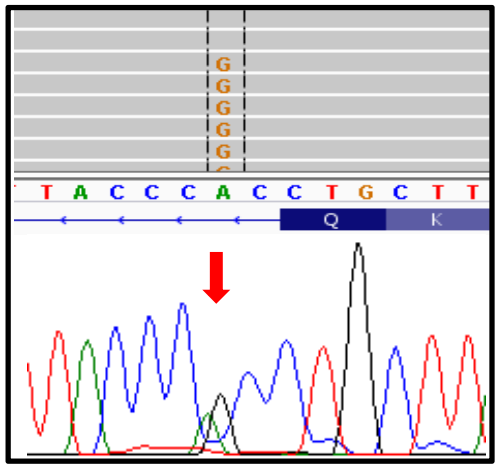
14. Patient 14 in Supplemental Table 3. *RPS19* c.16delG



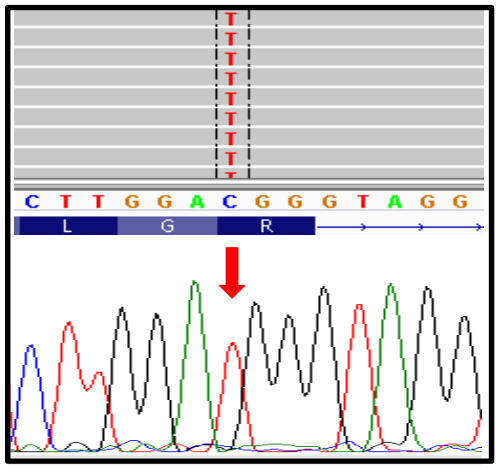
15. Patient 15 in Supplemental Table 3. *RPS7* c.398T>C. (Mutation was Validated in a clinical lab)



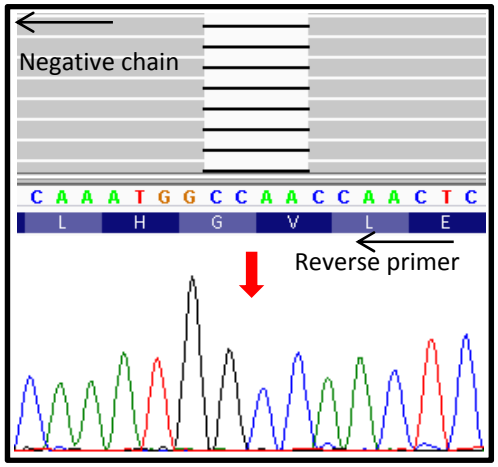
16a. Patient 16 in Supplemental Table 3. *SBDS* c.127G>T



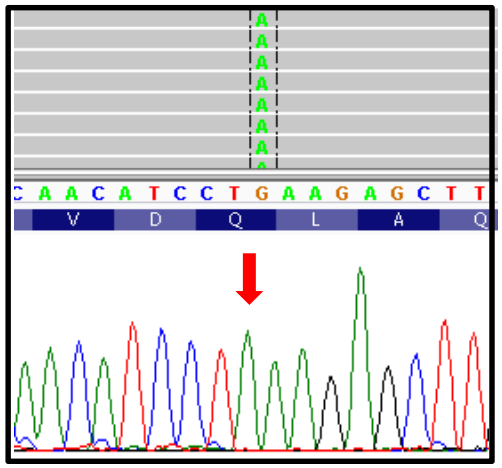
16b. Patient 16 in Supplemental Table 3. *SBDS* c.258+2T>C



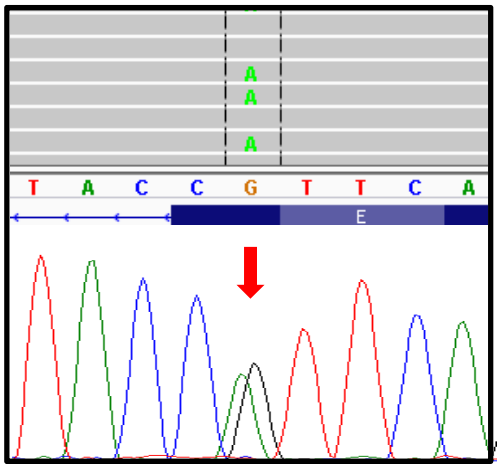
17. Patient 17 in Supplemental Table 3. *FANCE* c.1111C>T (hom)



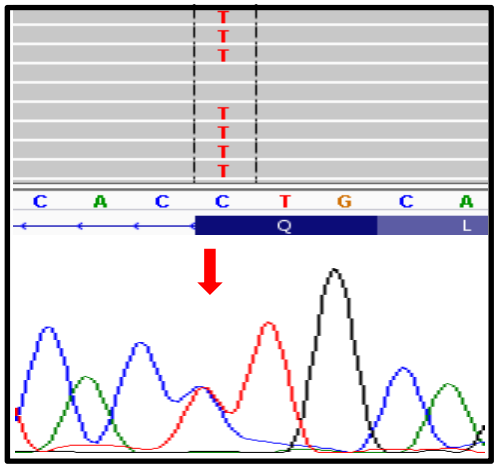
18. Patient 18 in Supplemental Table 3. *FANCA* c.1115_1118delTTGG (hom)



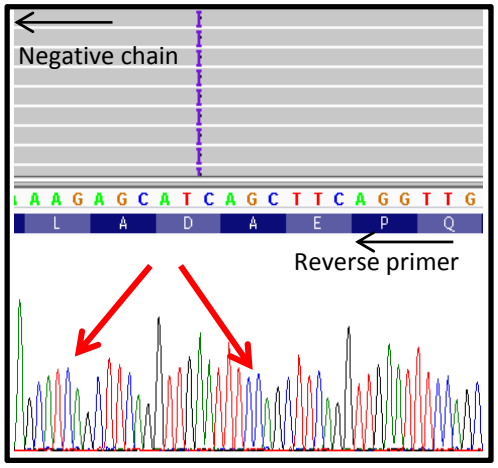
19. Patient 19 in Supplemental Table 3. *FANCA* c.1645C>T (hom)



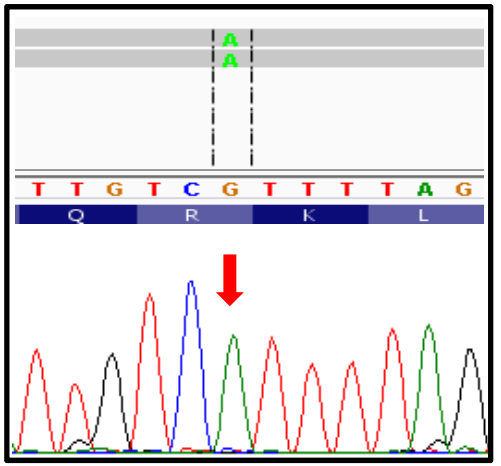
20a. Patient 20 in Supplemental Table 3. *FANCA* c.2851C>T



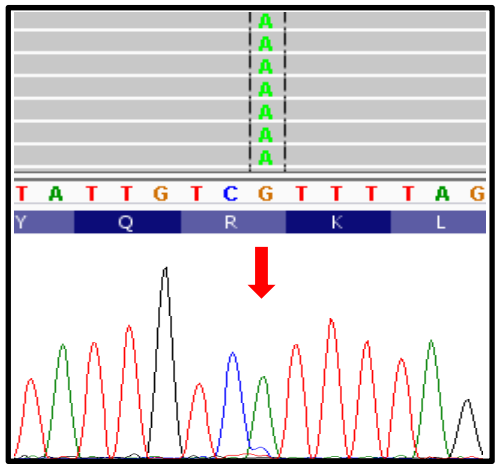
20b. Patient 20 in Supplemental Table 3. *FANCA* c.1470G>A



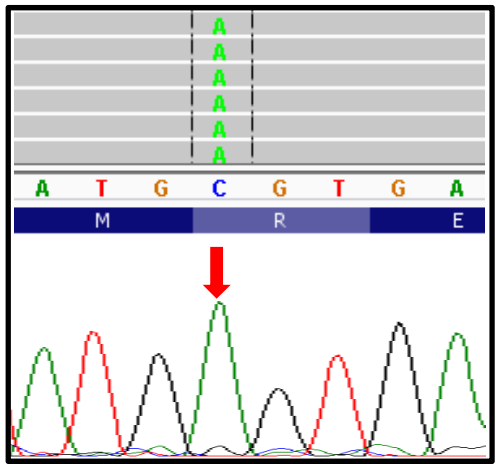
21. Patient 21 in Supplemental Table 3. *FANCA* c.2830_2831InsGAAATTCAACCT GAAGCTG (hom)



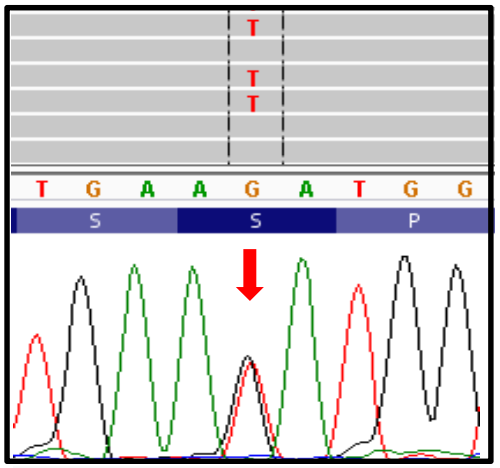
22. Patient 22 in Supplemental Table 3. *FANCJ/BRIP1* c.2392C>T (hom). See Results Section for more information about this homozygous mutation with 2 reads .



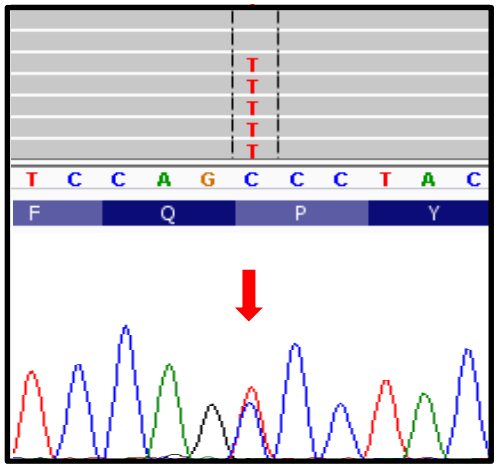
23. Patient 23 in Supplemental Table 3. *FANCJ/BRIP1* c.2392C>T (hom)



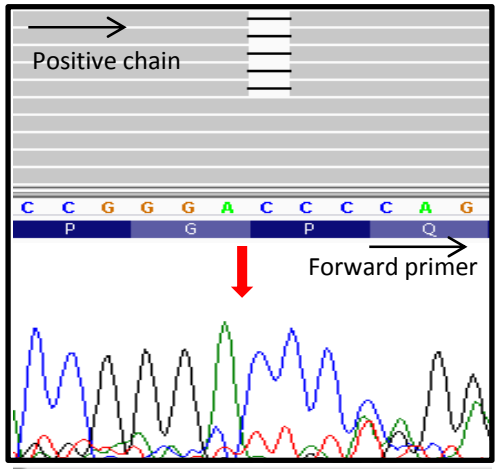
24. Patient 24 in Supplemental Table 3. *ERCC4* c.2065C>A (hom)



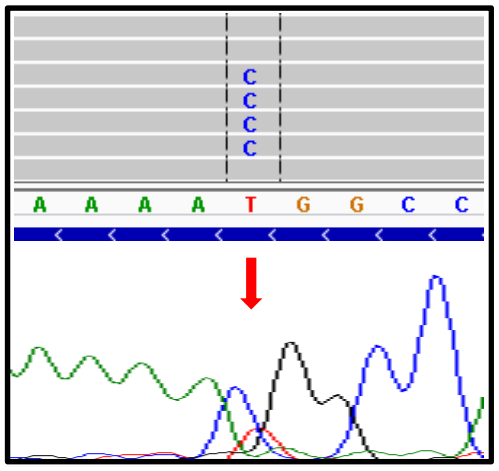
25. Patient 25 in Supplemental Table 3. *TINF2* c.734C>A



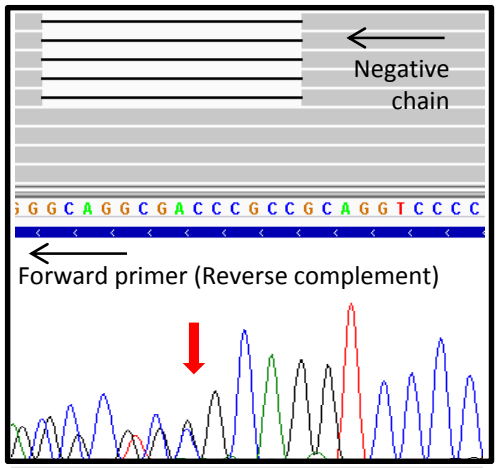
26a. Patient 26 in Supplemental Table 3. RTEL1 c.49C>T



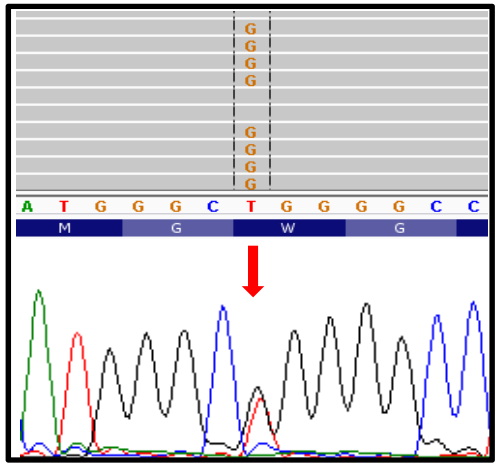
26b. Patient 26 in Supplemental Table 3. *RTEL1* c.3442delC



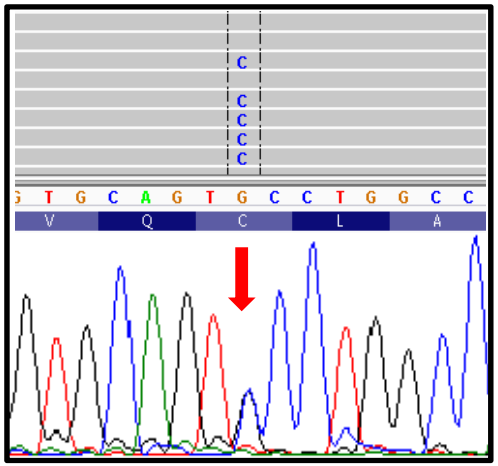
27a. Patient 27 in Supplemental Table 3. *TERC* c.37A>G



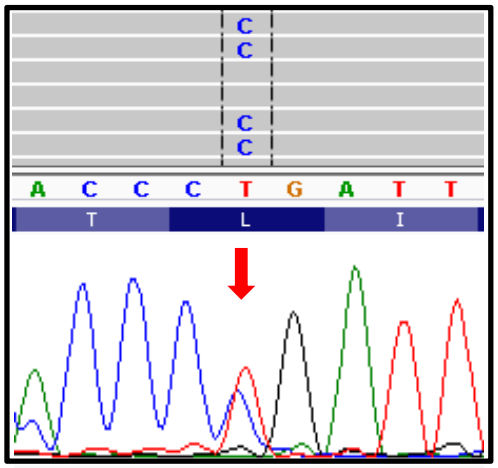
27b. Patient 27 in Supplemental Table 3.
TERC c.216_229del



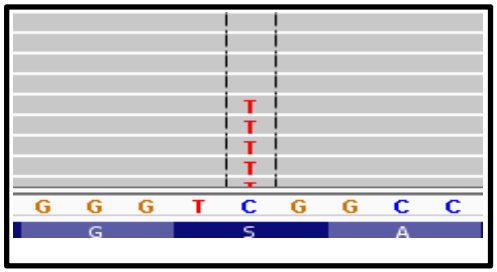
28. Patient 28 in Supplemental Table 3.
ELANE c. 466T>G



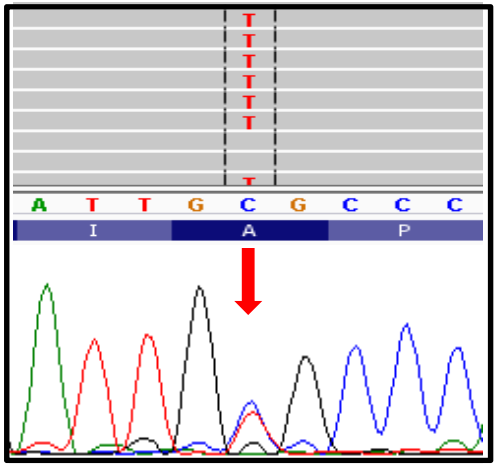
29. Patient 29 in Supplemental Table 3.
ELANE c.452G>C



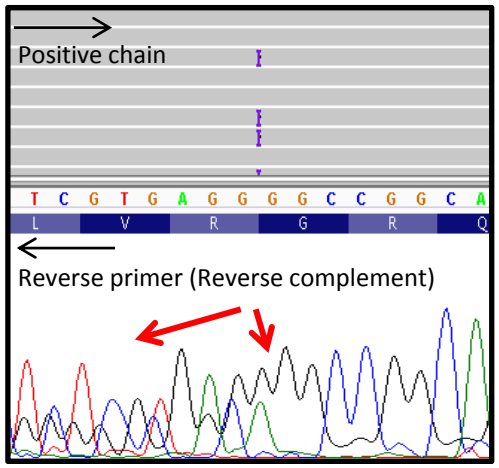
30. Patient 30 in Supplemental Table 3.
ELANE c.176T>C



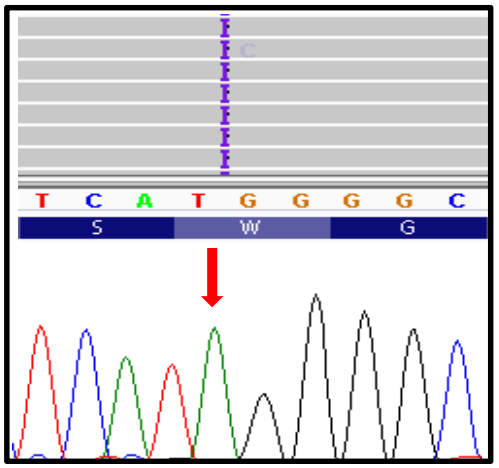
31. Patient 31 in Supplemental Table 3.
ELANE c.377C>T (Mutation was Validated in a clinical lab)



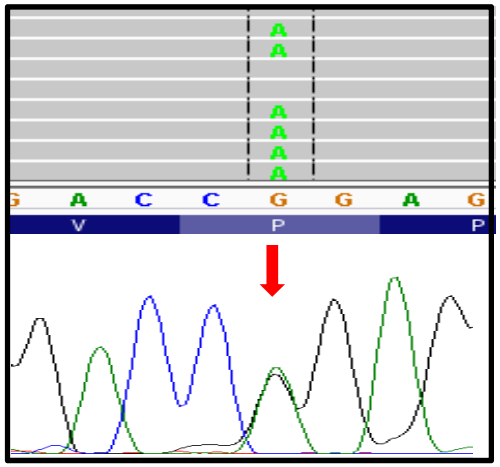
32. Patient 32 in Supplemental Table 3.
ELANE c.182C>T



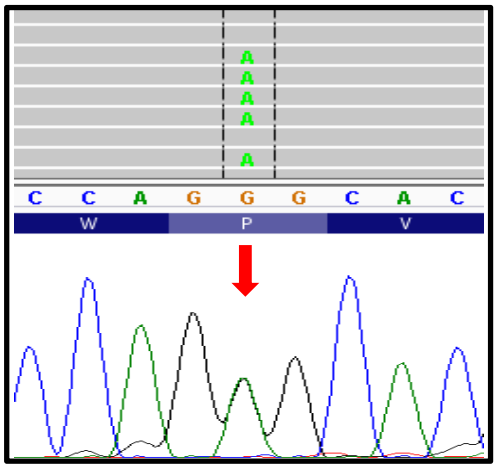
33. Patient 33 in Supplemental Table 3.
ELANE c.574_581dupGGCCGGCA



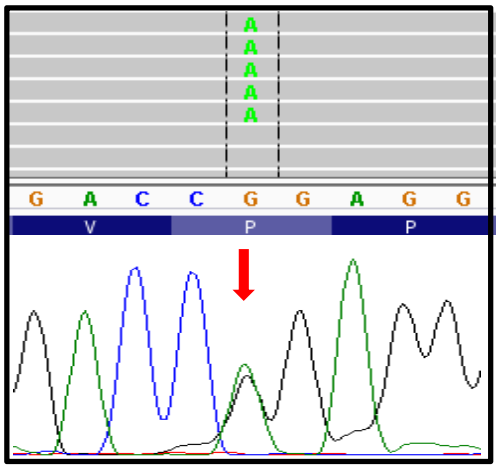
34. Patient 34 in Supplemental Table 3.
HAX1 c.131insA (hom)



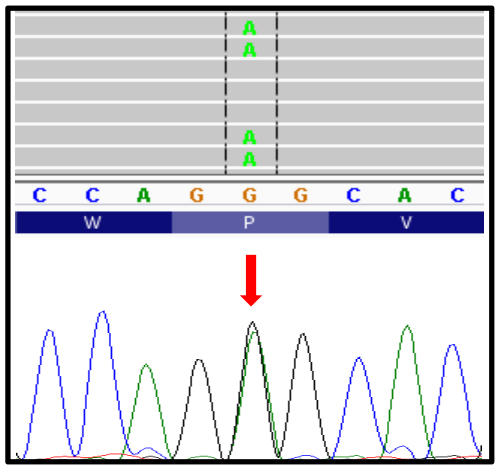
35a. Patient 35 in Supplemental Table 3.
CDAN1 c.2015C>T



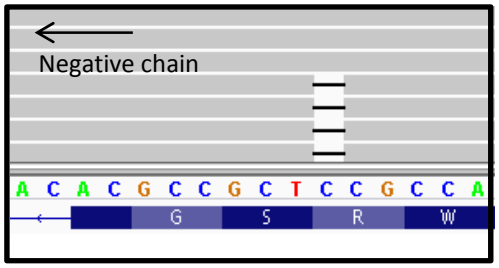
35b. Patient 35 in Supplemental Table 3.
CDAN1 c.2081C>T



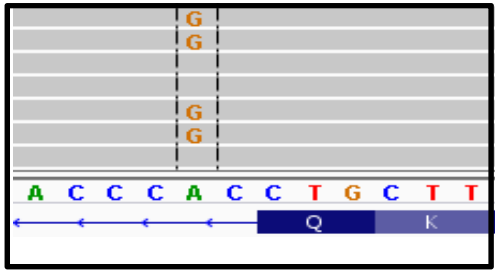
36a. Patient 36 in Supplemental Table 3.
CDAN1 c.2015C>T



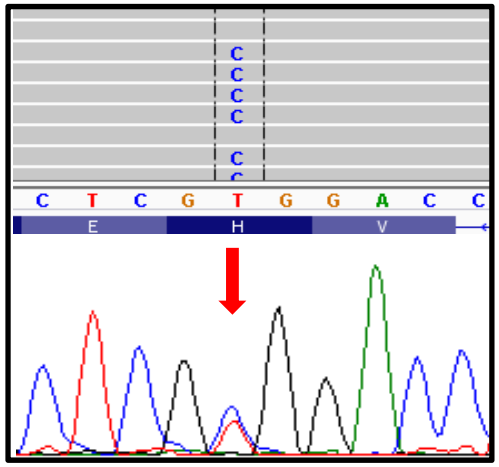
36b. Patient 36 in Supplemental Table 3.
CDAN1 c.2081C>T



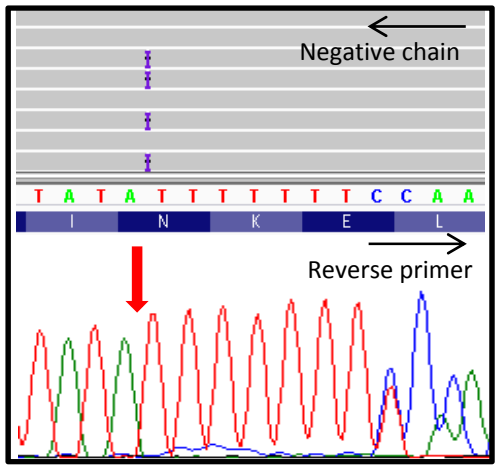
37a. Patient 37 in Supplemental Table 3. *SBDS* c.120delG (Mutation was Validated in a clinical lab)



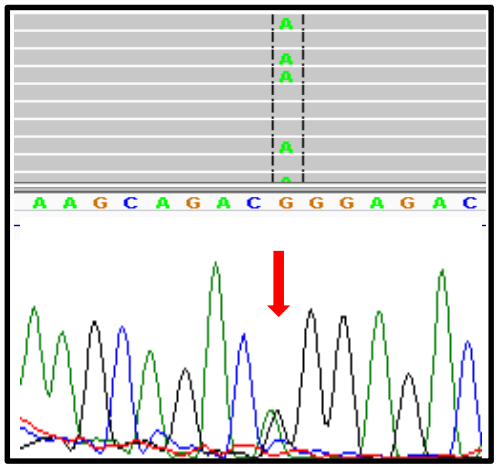
37b. Patient 37 in Supplemental Table 3. *SBDS* c.258+2T>C (Mutation was Validated in a clinical lab)



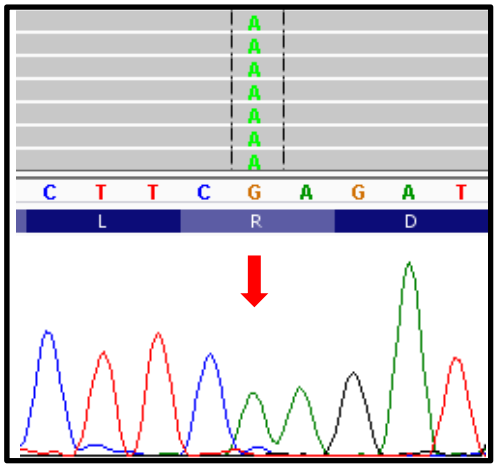
38. Patient 38 in Supplemental Table 3. *MYH9* c.4562A>G



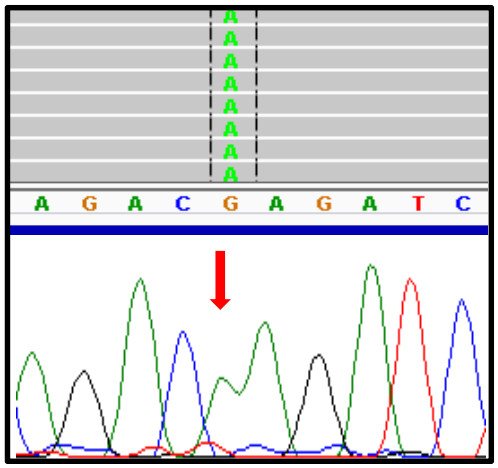
39. Patient 39 in Supplemental Table 3. *ANKRD26* c.4976dupA



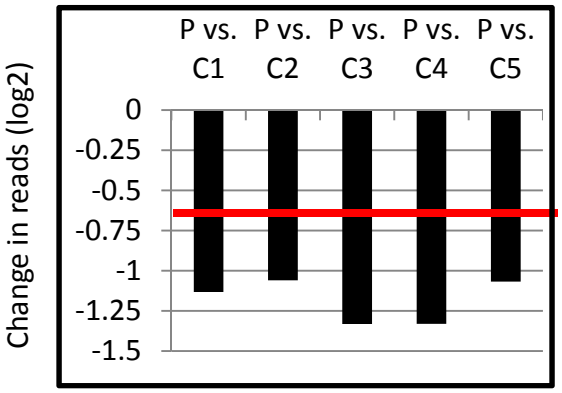
40. Patient 40 in Supplemental Table 3. *TERT* c.2383-15C>T. The NGS gene panel detected the same mutation in the affected mother of this patient (Supplemental Fig 3, Patient 42)



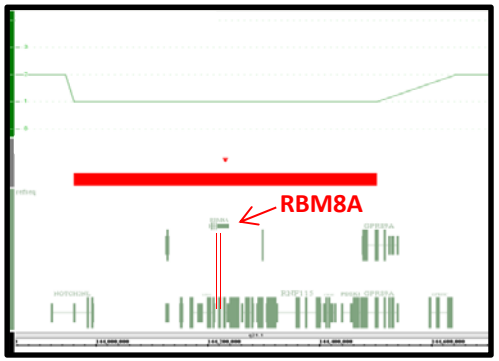
41. Patient 42 in in Supplemental Table 3. *SLC25A38* c.560G>A



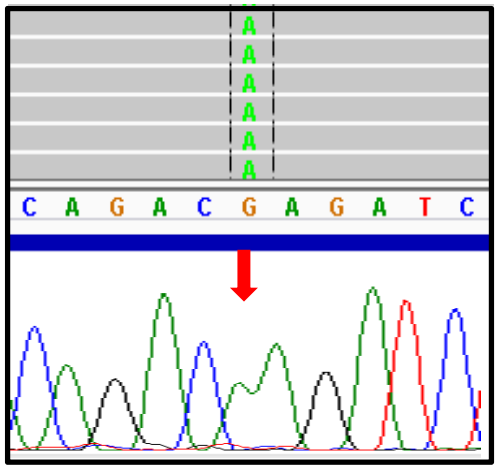
42a. Patient 43 in in Supplemental Table 3. *RBM8A* c.-21G>A (compound het)



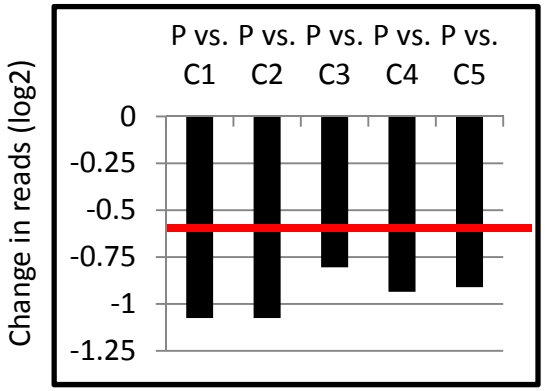
42b. Patient 43 in in Supplemental Table 3. *RBM8A* one copy number deletion as indicated by log2< -0.6 when patient calls are compared to controls (compound het) (C, control; P, patient.)



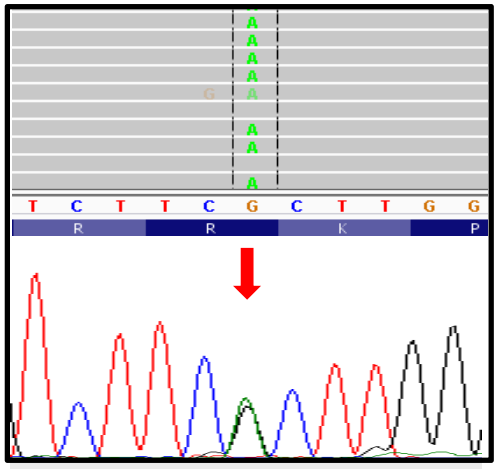
42c. Patient 43 in Supplemental Table 3. One copy number deletion that included *RBM8A* was found by Affymetrix SNP6.0 array. The red arrow points to the location of *RBM8A*. The red vertical lines show its approximate borders .



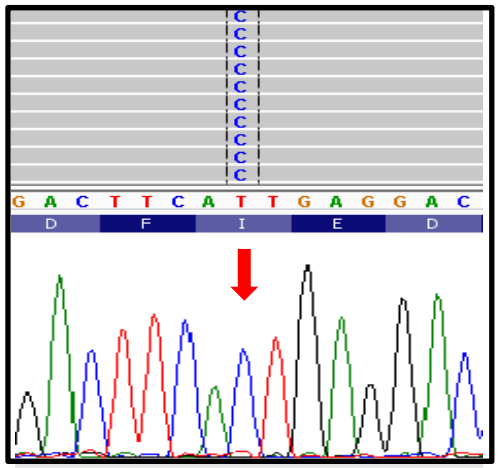
43a. Patient 44 in in Supplemental Table 3. *RBM8A* c.-21G>A (compound het)



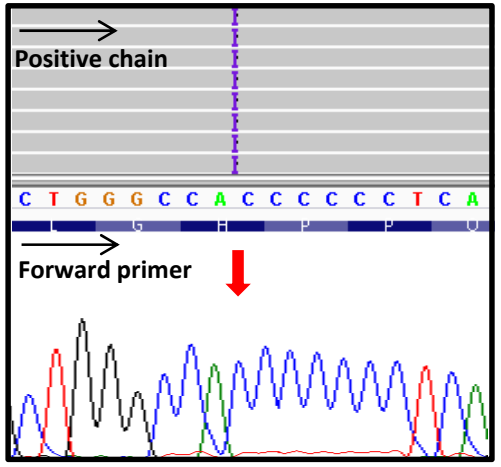
43b. Patient 44 in in Supplemental Table 3. *RBM8A* one copy deletion as indicated by log2< -0.6 when patient calls are compared to controls(compound het) (C, control; P, patient.)



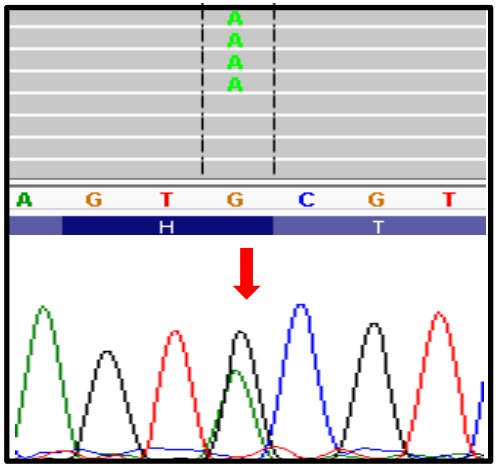
44. Patient 1 in Supplemental Table 4. *GATA2* c.1009C>T



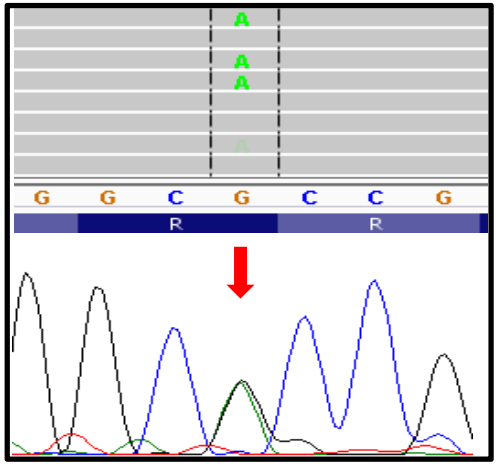
45. Patient 2 in Supplemental Table 4. *WAS* c.881T>C (hemizygous)



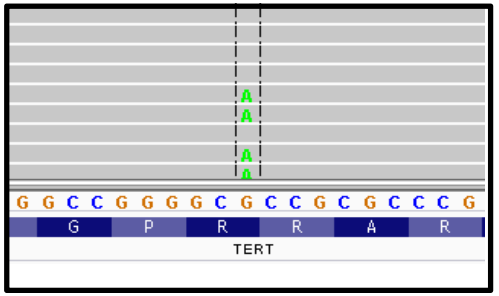
46. Patient 3 in Supplemental Table 4. *G6PC3* c.911dupC (hom)



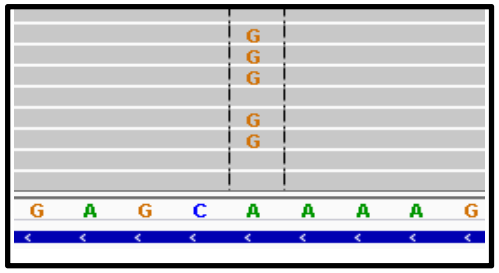
47. Patient 4 in Supplemental Table 4. *TERT* c.1234C>T



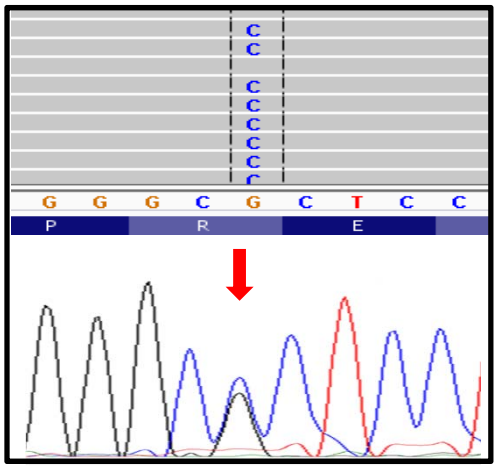
48. Patient 5 in Supplemental Table 4. *TERT* c. 2014C>T



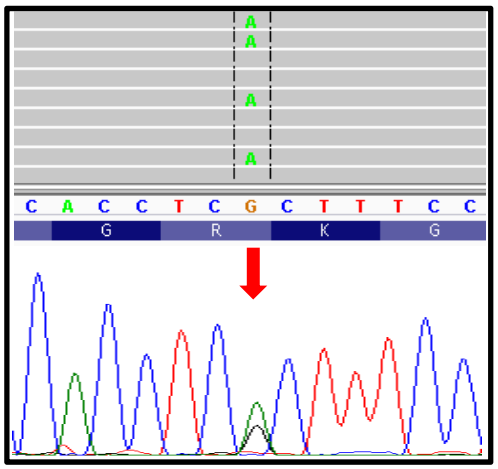
49. Patient 6 in Supplemental Table 4. *TERT* c. 2014C>T (Mutation was Validated in a clinical lab)



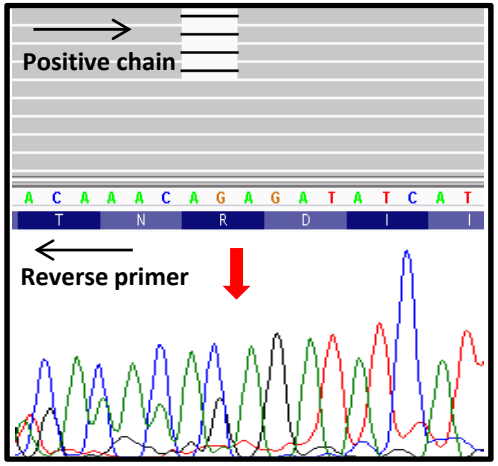
50. Patient 7 in Supplemental Table 4. *TERC* n.83T>C (Mutation was Validated in a clinical lab)



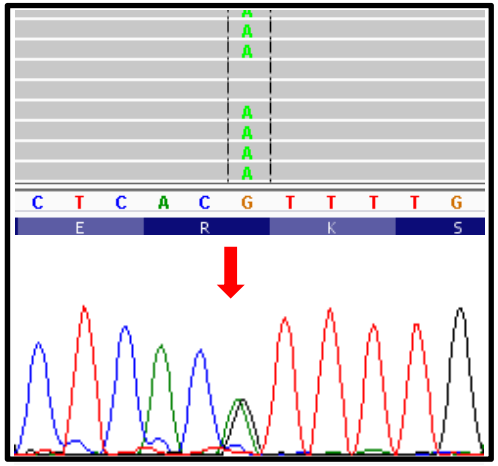
51. Patient 8 in Supplemental Table 4. *TINF2* c. 844C>G



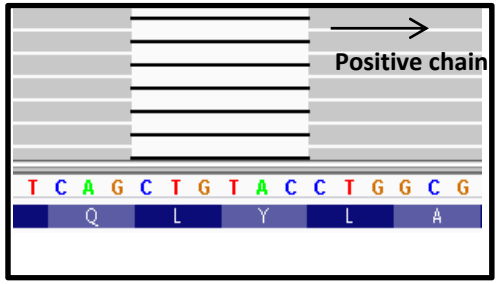
52. Patient 9 in Supplemental Table 4. *CXCR4* c.1000C>T



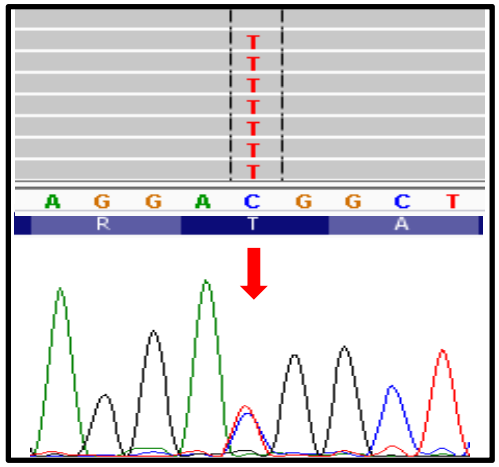
53. Patient 10 in Supplemental Table 4. *RPL5* c.174_175delAG



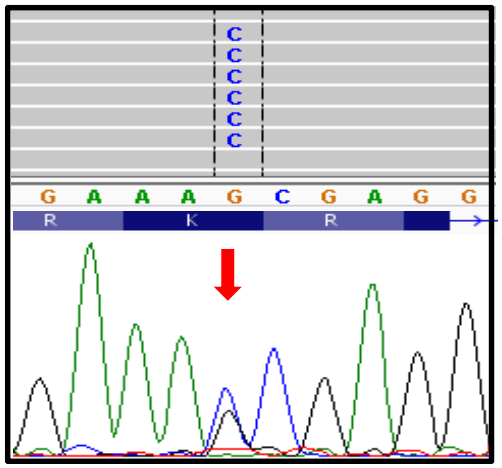
54. Patient 11 in Supplemental Table 4. *MYH9* c.3493C>T



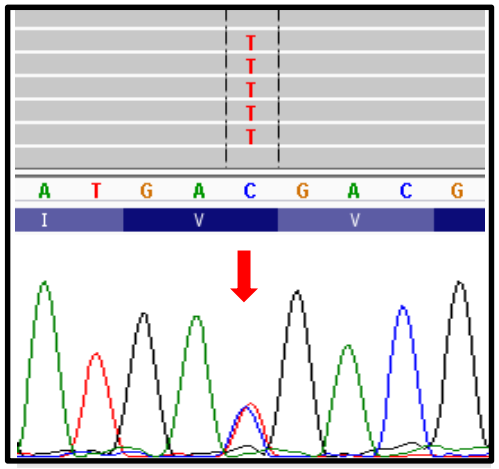
55. Patient 12 in Supplemental Table 4. *WAS* c.157_162delCTGTAC (hemizygous) (Mutation was Validated in a clinical lab)



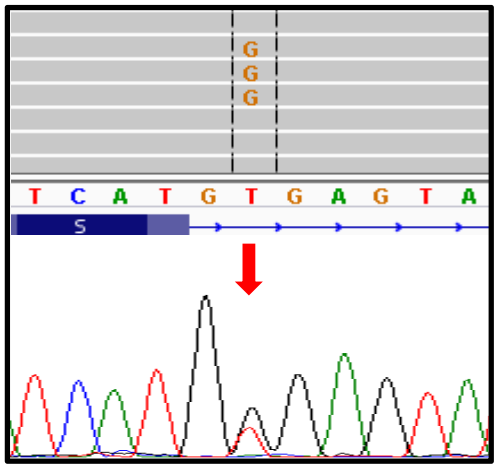
56a. Patient 13 in Supplemental Table 4.
RTEL1 c.1373C>T



56b. Patient 13 in Supplemental Table 4.
RTEL1 c.1416G>C

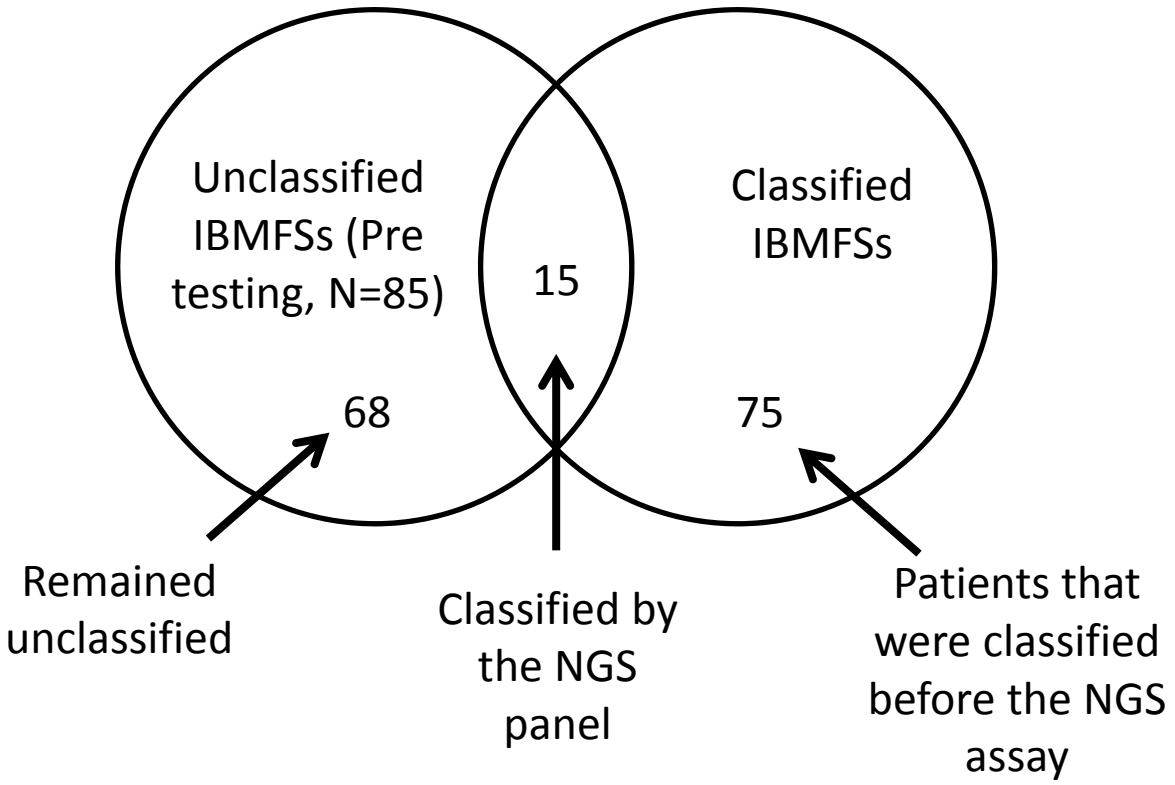


57. Patient 14 in Supplemental Table 4.
TERT c.2371 G>A



58. Patient 16 in Supplemental Table 4.
MASTL c.811+2 T>G

Figure 59. A Venn diagram showing how many unclassified patients were added to the classified pool after being tested by the NGS gene panel assay (15), and how many patients whose diagnosis remained unclassified after being tested by the assay (68).



Supplementary Tables

Improving diagnostic precision, care and syndrome definitions using comprehensive next generation sequencing for the inherited bone marrow failure syndromes

Ibrahim Ghemlas, Hongbing Li, Bozana Zlateska, Robert Klaassen, Conrad V Fernandez, Rochelle A Yanofsky, John Wu, Yves Pastore, Mariana Silva, Jeff H Lipton, Josse Brossard, Michon Bruno, Sharon Abish, MaCregor Steele, Roona Sinha, Mark Belltrutti, Vicky Breaky, Lawrence Jardine, Lisa Goodyear, Lillian Sung, Santhosh Dhanraj, Emma Reble, Amanda Wagner, Joseph Beyene, Peter Ray, Stephen Meyn, Michaela Cada, Yigal Dror

Supplementary Table 1: A list of 72 inherited bone marrow failure syndrome genes that were included in next generation sequencing panel assay and their coverage. We designed a custom NGS IBMFS Gene Panel Assay that allows the discovery of mutations in a comprehensive panel of 72 known IBMFS genes. The assay is based on a hybridization oligonucleotide pool, which covers coding regions, 50bp flanking intronic regions that include splicing sites, 3'-untranslated regions that include potential translation regulatory elements, and upstream promoter regions. The oligonucleotide size was set at 150 mers for read length with 3x tiling and a maximum of 10bp overlap between oligonucleotides. The design was submitted to the Agilent HaloPlex Design Wizard program (http://www.halogenomics.com/haloplex/custom-reagent-kits).

Gene	Regions	Coverage (%)	High Coverage (>= 90%)	Low Coverage (< 90%)
ABCB7	15	100	15	0
AK2	11	100	11	0
ALAS2	12	100	12	0
ANKRD26	37	99.7	37	0
BTHS	5	100	5	0
CDAN1	25	99.29	25	0
CTC1	14	100	14	0
CXCR4	3	100	3	0
DKC1	14	100	14	0
ELANE	5	100	5	0
FANCA	35	98.61	34	1
FANCB/FAAP95	10	100	10	0
FANCC	21	100	21	0
FANCD1/BRCA2	27	99.31	27	0
FANCD2	45	97.15	40	5
FANCE	10	100	10	0
FANCF	1	100	1	0
FANCG	11	100	11	0
FANCI	37	99.98	37	0
FANCJ/BRIP1	24	98.09	23	1
FANCL	14	100	14	0
FANCM	25	99.57	24	1
FANCO/RAD51C	11	100	11	0
FANCP/PALB2	14	99.69	14	0
FECH	11	98.25	11	1
G6PC3	37	100	37	0
GATA1	6	100	6	0

GATA2	10	100	10	0
GFI1	10	98.98	10	0
GLRX5	3	100	3	0
GP1BA	2	100	2	0
HAX1	3	100	3	0
HOXA11	12	100	12	0
KLF1	3	100	3	0
LIG4	4	99.94	4	0
MASTL	12	99.96	12	0
MPL	12	98.17	12	0
MYH9	46	99.97	46	0
NBEAL2	34	99.93	34	0
NHP2	3	100	3	0
NOP10	2	98.81	2	0
PALB2	15	100	15	0
PUS1	5	100	5	0
RBM8A	4	81.41	2	2
RMRP	1	100	1	0
RPL11	6	100	6	0
RPL27	4	100	4	0
RPL35A	7	99.24	7	0
RPL5	7	99.44	7	0
RPS10	7	100	7	0
RPS19	6	100	6	0
RPS24	8	100	8	0
RPS27	2	100	2	0
RPS26	3	99.34	3	0
RPS29	4	94.86	4	0
RPS7	2	96.87	2	0
RTEL1	33	100	33	0
RUNX1	26	99.34	26	0
SBDS	5	90.61	2	3
SEC23B	20	100	20	0
SLC19A2	6	100	6	0
SLC25A38	8	100	8	0
SLC37A4	2	100	2	0
SMARCAL1	21	95.97	20	1
SRP72	17	99.16	17	0
TERC	1	100	1	0
TERT	16	100	16	0

TINF2	3	94.86	2	1
***USB1/* C16orf57**	8	93.33	7	1
WAS	11	99.57	11	0
WRAP53	7	100	7	0
XRCC2	4	95.62	4	0

[*] *FANCQ/ERCC4* was identified as an IBMFS gene after January 2013, and was added to the panel when the second batch of patients was tested.

The gene list is modified and updated from Dror Y. Genetic Basis of Inherited Bone Marrow Failure Syndromes. InTech Open Access Publisher 2011: pp 357-392.

Supplementary Table 2: List of genes and mutation in previously genotyped patients which were validated with NGS IBMFS gene panel

Patient	Gene	Nucleotide change	Protein change	Mutation type
1*	*SBDS* (hom)	c.258+2T>C	Splicing	Splicing
2	*RPL5* (het)	c.83delC	p.Thr28Metfs*10	Frameshift
3	*ELANE* (het)	c.597+5 G>A	Splicing	Splicing
4	*FANCA* (het)	c.3788-3790 delTCT	p.F1263SFS*194	Frameshift
5	*TINF2* (het)	c. 845G>A	R282H	Missense
6*	*cMPL* (hom)	c.304C>T	R102C	Missense
7	*RPS19* (het)	c.250_251delAG	p.R84Lfs*69	Frameshift
8*	*DKC1* (com hem)	c.112delA; c.116InsC	I38S ; K39T	Frameshift
9	*TERT* (het)	c.2383-15C>T	Splicing	Splicing
10	*RPL11* (het)	c.158-1G>C	Splicing	Splicing

*Two different mutations that were present in one allele (combined mutations) or two identical mutations that were found in both alleles (homozygous mutations) were counted twice. Thus, the total number of mutations (*i.e.* mutant alleles) in these 10 patients is 13.
Com, two mutations combined on the same allele; hem, hemizygous; het, heterozygous; hom, homozygous; NA, not applicable

Supplemental Table 3: List of identified damaging mutations in patients with classified IBMFSs without known genes

Number	Clinical Diagnosis	Gene	Mode	Nucleotide change	Protein change	Mutation type	Previous reporting
1	DBA	*RPS26* (het)	AD	c. 243delC	p.Ser81Argfs*3	Indel/Frameshift	Novel[2]
2	DBA	*RPS26* (het)	AD	c.1A>G	p.Met1?	Start code lost	Reported
3	DBA	*RPS26* (het)	AD	c.4-32_21 delGTTTTCCTAACA	Splicing change	Splicing	Novel[2]
4	DBA	*RPS19* (het)	AD	c.10_13 delAGTT	p.Val4Leufs*2	Indel/Frameshift	Reported
5	DBA	*RPS19* (het)	AD	c. 3G>T	p.Met1?	Start code lost	Reported
6	DBA	*RPL11* (het)	AD	c.60_61delCT	p.Cys21Serfs*33	Indel/Frameshift	Reported
7	DBA	*RPL11* (het)	AD	c.60_61delCT	p.Cys21Serfs*33	Indel/Frameshift	Reported
8	DBA	*RPS24* (het)	AD	c.1A>G	p.Met1?	Start code lost	Reported
9	DBA	*RPS24* (het)	AD	c.4-2A>G	Splicing change	Splicing	Novel[2]
10	DBA	*RPL35A* (het)	AD	c.78_80delTCT	p. Leu28del	Indel/Inframe	Reported
11	DBA	*RPL11* (het)	AD	c.372C>G	p.Ile124Met	Missense	Novel[2]
12	DBA	*RPS29* (het)	AD	c.63-3 C>A	Splicing change	Splicing	Novel[2]
13	DBA	*RPS19* (het)	AD	c.185G>A	p.Arg62Gln	Missense	Reported
14	DBA	*RPS19* (het)	AD	c.16delG	p.Val6*	Nonsense	Novel[2]
15	DBA	*RPS7* (het)	AD	c.398T>C	p.Leu133Ser	Missense	Novel[2]

16	DBA	*SBDS* (combined)	AR	c.127G>T c.258+2T>C	p.Val43Leu Splicing change	Missense Splicing	Novel[2] Reported
17	FA	*FANCE* (hom)	AR	c. 1111C>T	p.Arg371Trp	Missense	Reported
18	FA	*FANCA* (hom)	AR	c.1115_1118delTTGG	p.Val372Alafs*42	Indel/Frameshift	Reported
19	FA	*FANCA* (hom)	AR	c.1645C>T	p.Gln549*	Nonsense	Reported
20	FA	*FANCA* (combined)	AR	c.2851 C>T c.1470G>A	p.Arg951Trp p.Gln490Gln	Missense Splicing	Reported Reported
21	FA	*FANCA* (hom)	AR	c.2830_2831Ins GAAATTCAACCTGAAGCTG	p.Asp944Glyfs*5	Indel/Frameshift	Reported
22	FA	*BRIP1* (hom)	AR	c.2392C>T	p.Arg798*	Nonsense	Reported
23	FA	*BRIP1* (hom)	AR	c.2392C>T	p.Arg798*	Nonsense	Reported
24	FA	*ERCC4* (hom)	AR	c.2065C>A	p.Arg689Ser	Missense	Reported
25	FA	*TINF2* (het)	AD	c.734C>A	p.Ser245Tyr	Missense	Reported
26	DC	*RTEL1* (combined)	AR	c.49C>T c.3442delC	p.Pro17Ser p.Gln1148Argfs*96	Missense Indel/Frameshift	Novel[2] Novel[2]
27	DC	*TERC* (combined)	AD	n.37A>G n.216_229del GGCGGGTCGCCTGC	NA NA	ncRNA ncRNA	Reported Reported

28	SCN	*ELANE* (het)	AD	c.466T>G	p.Trp156Gly	Missense	Novel[2]
29	SCN	*ELANE* (het)	AD	c. 452G>C	p.Cys151Ser	Missense	Novel[2]
30	SCN	*ELANE* (het)	AD	c.176 T>C	p.Leu59Pro	Missense	Novel[2]
31	CN	*ELANE* (het)	AD	c.377C>T	p.Ser126Leu	Missense	Reported
32	CN	*ELANE* (het)	AD	c.182C>T	p.Ala61Val	Missense	Reported
33	CN	*ELANE* (het)	AD	c.574_581dupGGCCGGCA	p.Val197Argfs*18	Indel/Frameshift	Novel[2]
34	SCN	*HAX1* (hom)	AR	c. 131InsA	p.Trp44*	Nonsense	Reported
35	CDA	*CDAN1* (combined)	AR	c.2015C>T c.2081C>T	p.Pro672Leu p.Pro694Leu	Missense Missense	Reported Novel[2]
36	CDA	*CDAN1* (combined)	AR	c.2015C>T c.2081C>T	p.Pro672Leu p.Pro694Leu	Missense Missense	Reported Novel[2]
37	SDS	*SBDS* (combined)	AR	c.120delG c.258+2T>C	p.Ser41Alafs*17 Splicing change	Indel/Frameshift Splicing	Novel[2] Reported
38	FT	*MYH9* (het)	AD	c. 4562A>G	p.His1521Arg	Missense	Novel[2]
39	FT	*ANKRD26* (het)	AD	c.4976dupA	p.Ile1659Tyrfs*3	Indel/Frameshift	Novel[2]
40	FT	*TERT* (het)	AD	c.2383-15T>C	Splicing change	Splicing	Novel[2]
41	FT	*TERT* (het)	AD	c.2383-15T>C	Splicing change	Splicing	Novel[2]

42	CSA	*SLC25A38* (hom)	AR	c.560G>A	p.Arg187Gln	Missense	Novel[2]
43	TAR	*RBM8A*[1]	AR	c. -21G>A/ Large deletion[1]	Reduced translation Large deletion[1]	5'- UTR Large deletion[1]	Reported Reported
44	TAR	*RBM8A*[1]	AR	c. -21G>A/ Large deletion[1]	Reduced translation Large deletion[1]	5'- UTR Large deletion[1]	Reported Reported

AD, autosomal dominant; AR, autosomal recessive; CDA, congenital dyserythropoietic anemia; CSA, congenital sideroblastic anemia; CN, cyclic neutropenia; DBA, Diamond-Blackfan anemia; DC, Dyskeratosis congenita; FA, Fanconi anemia; FT, familial thrombocytopenia; ncRNA, non-coding RNA; SCN, severe congenital neutropenia; SDS, Shwachman–Diamond syndrome; TAR, Thrombocytopenia with Absent Radii

[1]Analysis of the NGS data by the SureCall CNV algorithm identified a large deletion on the allele without the mutation (Supplementary Fig 42, 43). This confirmed a compound heterozygosity state, which is the commonest genotype combination in this disease.

[2]See information in Supplementary Table 5 about damage prediction of this mutation. The criteria for calling a variant a novel/most likely damaging mutation are provided in Supplementary Table 5 and in the Methods Section (in the paragraph "Variant analysis and filtering strategy").

Supplementary Table 4: List of patients with unclassified IBMFSs who were Genotyped and diagnosed based on this study.

Number	Clinical phenotype	Gene	Nucleotide change	Protein change	Mutation type	Previous reporting	Diagnosis based on this study
1	Neutropenia, MDS, mother with neutropenia. Negative for *ELA2* mutations.	*GATA2* (het)	c.1009C>T	p.Arg337*	Nonsense	Reported	Familial MDS
2	Neutropenia, solitary kidney, maternal grandfather and granduncle with neutropenia and AML. Negative screens for FA, DC and mutations in common neutropenia genes.	*WAS* (hemi)	c.881T>C	p.Ile294Thr	Missense	Reported	SCN
3	Neutropenia, atrial septal defect, negative for mutations in common neutropenia genes.	*G6PC3* (hom)	c.911dupC	p.Gln305Serfs*82	Indel/Frameshift	Novel[1]	SCN
4	Chronic pancytopenia from the age of 2 years; hypocellular bone marrow, brother with neutropenia and failure to thrive; short telomeres	*TERT* (het)	c.1234C>T	p.His412Tyr	Missense	Reported	DC
5	Pancytopenia from the age of 2.5 years. Decreased marrow cellularity, and reduced erythropoiesis and megakaryopoiesis. Red blood cell transfusion until the age of 18 years. Spontaneous elevation of chromosome fragility (stimulation with cross linking agents were not done), but repeat testing with mitomycin C and diepoxybutane in	*TERT* (het)	c. 2014C>T	p.Arg672Cys	Missense	Novel[1]	DC

	adulthood was normal. Telomere testing has not been done.						
6	Moderate to severe anemia, moderate neutropenia, mildly hypoplastic thumbs, developmental delay; semilobar holoprosencephaly; hypertonus and contractures; mildly hypocellular marrow, short telomeres, high eADA.	*TERT* (het)	c. 2014C>T	p.Arg672Cys	Missense	Novel[1]	DC
7	Chronic moderate pancytopenia, hypocellular marrow; paternal grandmother with thrombocytopenia; short telomeres	*TERC* (het)	n.83T>C	NA	ncRNA	Reported	DC
8	SAA, developmental delay, diabetes mellitus, pyloric stenosis, cerebral calcifications	*TINF2* (het)	c. 844C>G,	p.Arg282Gly	Missense	Novel[1]	DC
9	Severe neutropenia, anemia, thrombocytopenia; marrow active without dysplasia; reduced T & B lymphocytes, low IgG	*CXCR4* (het)	c.1000C>T	p.Arg334*	Missense	Reported	WHIM syndrome
10	Hydrops fetalis, severe anemia that persisted, early moderate thrombocytopenia, intermittent neutropenia; marrow at the age 2 months showed increased celluarity, normal erythropoiesis, reduced granulopoiesis, reduced megakaryopoiesis,	*RPL5* (het)	c.174_175delAG	p.Arg58Argfs*53	Indel/Frameshift	Novel[1]	DBA
11	Congenital thrombocytopenia, persistent neutropenia, large platelets.	*MYH9* (het)	c.3493C>T	p.Arg1165Cys	Missense	Reported	MYH9- related disorder

12	Thrombocytopenia and anemia from early childhood, arthritis and vasculitis (low T and B cells while on prednisone). Active marrow.	*WAS* (hemi)	c.157_162delCTGTAC	p.Leu53_Tyr54del	Indel/Inframe	Novel[1]	WAS
13	SAA; short telomeres; no response to IST	*RTEL1* (comb)	c.1373C>T c.1416G>C	p.Thr458Met p.Lys472Asn	Missense	Novel[1] Novel[1]	DC
14	SAA; no response to IST	*TERT* (het)	c.2371G>A	p.Val791Ile	Missense	Reported	DC
15	SAA; no response to IST	*MASTL* (het)	c.811+2 T>G	Splicing change	Splicing	Novel[1]	*MASTL* associate disorder

AD, autosomal dominant; AR, autosomal recessive; DBA, Diamond Blackfan anemia; DC, Dyskeratosis congenita; eADA, adenosine deaminase; IgG, Immunoglobulin G ; IST, immunosuppressive therapy; MDS, myelodysplastic syndrome; ncRNA, non-coding RNA; SAA, severe aplastic anemia; SCN, severe congenital neutropenia; TL, telomere length ; WAS, Wiskott-Aldrich syndrome ; WHIM, Warts-Hypogamaglobulinemia-Infection-Myelokathexis.

[1]See information in Supplementary Table 5 about damage prediction of this mutation. The criteria for calling a variant a novel/most likely damaging mutation are provided in Supplementary Table 5 and in the Methods Section (in the paragraph “Variant analysis and filtering strategy”).

Supplementary Table 5: Bioinformatic data related to novel/most likely causal mutations. The table shows mutation information, as well as prediction of conservation and damaging effect of the mutations on protein by several different softwares. Variants were considered novel / most likely causal mutations if they fulfilled all the following: 1) they appeared in allelic dosage that was consistent with the known inheritance mode of the disease, 2) the MAF was <0.001, 3) evolutionary conserved amino acid/s are affected, 4) the variant was considered damaging by at least 2 of the following prediction software programs: PolyPhen2, SIFT/SIFT-Indel, Provean, MutationTaster and Human Splicing Finder. In this paper we referred to both, previously published mutations and novel mutations as "causal mutations".

Patient	**Gene**	**Nucleotide change**	**Protein change**	**Mutation type**	**MAF (EVS)**	**Conser-vation[1]**	**MutationTaster[2]**	**Provean[3]**	**Polyphen2[4]**	**SIFT[5]/ SIFT-Indel**	**Human Splicing Finder**
1 (Patient 1 SuppTable 3)	*RPS26*	c. 243delC	p.Ser81Argfs*3	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
2 (Patient 3 SuppTable 3)	*RPS26*	c.4-32_21 delGTTTTCCTA ACA	Splicing change	Splicing	NR	NA	Splice site changes	ID	ID	ID	Splicing branch point broken[9]
3 (Patient 14 SuppTable 3)	*RPS19*	c.16delG	p.Val6*	Indel/ Nonsense	NR	NA	Disease causing	ID	ID	Damaging	NA
4 (Patient 10 SuppTable 4)	*RPL5*	c.174_175delAG	p.Arg58Argfs*53	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
5 (Patient 11 SuppTable 3)	*RPL11*	c.372C>G	p.Ile124Met	Missense	NR	5.94	Disease causing	Deleterious (-2.60)	Probably damaging (0.944)	Intolerant (0)	NA
6 (Patient 9 SuppTable 3)	*RPS24*	c.4-2A>G	Splicing change	Splicing	NR	NA	Disease causing	ID	ID	ID	Site broken
7 (Patient 12 SuppTable 3)	*RPS29*	c.63-3 C>A	Splicing change	Splicing	NR	NA	Disease causing	ID	ID	ID	Site broken

8 (Patient 15 SuppTable 3)	*RPS7*	c.398T>C	p.Leu133Ser	Missense	NR	4.6	Disease causing	Deleterious (-5.462)	Probably damaging (0.99)	Intolerant (0.01)	NA
9 (Patient 40 SuppTable 3)	*TERT*	c.2383-15T>C[6]	Splicing change	Splicing	0.0005	NA	Splice site changes	ID	ID	ID	Site broken
10 (Patient 5 SuppTable 4)	*TERT*	c. 2014C>T[7]	p.Arg672Cys	Missense	0.000082	4.67	Protein might be affected	Deleterious (-3.317)	Probably damaging (0.99)	Intolerant (0.022)	NA
11 (Patient 6 SuppTable 4)	*TERT*	c. 2014C>T[7]	p.Arg672Cys	Missense	0.000082	4.67	Protein might be affected	Deleterious (-3.317)	Probably damaging (0.99)	Intolerant (0.022)	NA
12 (Patient 26 SuppTable 3)	*RTEL1*	c.3442delC	p.Gln1148Argfs *96	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
		c.49C>T	p.Pro17Ser	Missense	NR	4.86	Disease causing	Deleterious (-6.73)	Probably damaging (0.998)	Intolerant (0.02)	NA
13 (Patient 13 SuppTable 4)	*RTEL1*	c.1416G>C	p.Lys472Asn	Missense	NR	4.5	Disease causing	Deleterious (-2.94)	Probably benign (0.094)	Tolerated (0.076)	NA
		c.1373C>T	p.Thr458Met	Missense	0.000539	4.34	Splice site changes; Protein might be affected	Neutral (-1.29)	Probably damaging (0.98)	Intolerant (0.05)	ESE site is broken, and Creates a new ESS site
14 (Patient 8 SuppTable 4)	*TINF2*	c. 844C>G,	p.Arg282Gly	Missense	NR	5.16	Disease causing	Deleterious (-3.60)	Possibly damaging (0.839)	Intolerant (0.000)	NA

15 (Patient 16 SuppTable 3)	*SBDS*	c.127G>T	p.Val43Leu	Missense	NR	5.19	Disease causing	Neutral (-1.632)	Probably benign (0.005)	Tolerated (0.12)	Site broken
16 (Patient 37 SuppTable 3)	*SBDS*	c.120delG	p.Ser41AlaFs*17	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
17 (Patient 28 SuppTable 3)	*ELANE*	c.466T>G	p.Trp156Gly	Missense	NR	4.42	Disease causing	Deleterious (-11.79)	Probably damaging (0.999)	Intolerant (0.006)	NA
18 (Patient 29 SuppTable 3)	*ELANE*	c. 452G>C	p.Cys151Ser	Missense	NR	4.42	Disease causing	Deleterious (-9.55)	Probably damaging (1.0)	Intolerant (0.000)	NA
19 (Patient 30 SuppTable 3)	*ELANE*	c.176 T>C	p.Leu59Pro	Missense	NR	3.24	Disease causing	Deleterious (-6.51)	Probably damaging (1.0)	Intolerant (0.000)	NA
20 (Patient 33 SuppTable 3)	*ELANE*	c.574_581dupGGCCGGCA	p.Val197Argfs*18	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
21 (Patient 3 SuppTable 4)	*G6PC3*	c.911dupC	p.Gln305Serfs*82	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
22 (Patient 35 SuppTable 3)	*CDAN1*	c.2081C>T[8]	p.Pro694Leu	Missense	NR	5.77	Disease causing	Deleterious (-9.02)	Probably damaging (0.999)	Intolerant (0.003)	NA
23 (Patient 36 SuppTable 3)	*CDAN1*	c.2081C>T[8]	p.Pro694Leu	Missense	NR	5.77	Disease causing	Deleterious (-9.02)	Probably damaging (0.999)	Intolerant (0.003)	NA
24 (Patient 41 SuppTable 3)	*SLC25A38*	c.560G>A	p.Arg187Gln	Missense	0.000077	4.89	Disease causing	Deleterious (-3.78)	Probably damaging (0.999)	Intolerant (0.02)	NA

25 (Patient 15 SuppTable 4)	*MASTL*	c.811+2 T>G	Splicing change	Splicing	NR	5.82	Disease causing	ID	ID	ID	Site broken
26 (Patient 39 SuppTable 3)	*ANKRD 26*	c.4976dupA	p.Ile1659Tyrfs*3	Indel/ Frameshift	NR	NA	Disease causing	ID	ID	Damaging	NA
27 (Patient 38 SuppTable 3)	*MYH9*	c. 4562A>G	p.His1521Arg	Missense	NR	5.25	Disease causing	Deleterious (-6.53)	Probably damaging (1.0)	Intolerant (0.01)	NA
28 (Patient 12 SuppTable 4)	*WAS*	c.157_162delCTGTAC	p.Leu53_Tyr54del	Indel/ Inframe	NR	NA	Protein might be affected	Deleterious (-10.896)	ID	Damaging	NA

[1]The MasterTaster software program evaluates the effect of variants on both, protein function/structure as well as on splicing.
[2]Conservation ranges from -12.3 to +6.17, with +6.17 being the most conserved.
[3]Provean scores: deleterious <-2.50; neutral >-2.50
[4]PolyPhen 2 scores: probably damaging (>0.85–1); possibly damaging (>0.15–0.84); probably benign (< 0.14)
[5]SIFT scores: intolerant (0.00-0.05); potentially intolerant (0.051-0.10); borderline (0.101-0.20), or tolerant (0.201-1.00)
[6,7,8]Patients have recurrent mutations in this study. [6]The same mutation was found in another patient in our registry who had aplastic anemia and very short telomeres and responded to androgen therapy. This mutation was used for validation of the NGS assay (Supplementary Table 2, Patient 9).
[9]The deleted fragment GTTTTCCTAACA contains the only YURAC splicing branch point consensus sequence in intron 1 of *RPS26*, which in this case it is CTAAC. Splicing branch points are typically located 20-50 nucleotide upstream of the acceptor site, similar to the present sequence. Deletion of the branch point abolishes binding of splicing factor 1 and assembly of the spliceosome.
ESE, exonic splicing enhancer; ESS, Exonic splicing silencer; EVS, exome variant server database (http://evs.gs.washington.edu/EVS/). MAF in other databases was analyzed as indicated in Table 1, but not provided herein; ID, indeterminate by the software; MAF, minor allele frequency; NA, not applicable; NR, not reported in this database; SuppTable, Supplementary Table.