

5-10-2017

# The Clinical Impact of Copy Number Variants in Inherited Bone Marrow Failure Syndromes

Nicolas Waespe

*The Hospital for Sick Children*

Santhosh Dhanraj

*The Hospital for Sick Children*

Manju Wahala

*The Hospital for Sick Children*

Elena Tsangaris

*The Hospital for Sick Children*

Tom Enbar

*The Hospital for Sick Children*

*See next page for additional authors*

Follow this and additional works at: <https://ir.lib.uwo.ca/paedpub>

 Part of the [Pediatrics Commons](#)

---

## Citation of this paper:

Waespe, Nicolas; Dhanraj, Santhosh; Wahala, Manju; Tsangaris, Elena; Enbar, Tom; Zlateska, Bozana; Li, Hongbing; Klaassen, Robert J; Fernandez, Conrad V; Cuvelier, Geoff D E; Wu, John K; Pastore, Yves D; Silva, Mariana; Lipton, Jeffrey H; Brossard, Joséé; Michon, Bruno; Abish, Sharon; Steele, MacGregor; Sinha, Roona; Belletrutti, Mark J; Breakey, Vicky R; Jardine, Lawrence; Goodyear, Lisa; Kofler, Liat; Cada, Michaela; Sung, Lillian; Shago, Mary; Scherer, Stephen W; and Dror, Yigal, "The Clinical Impact of Copy Number Variants in Inherited Bone Marrow Failure Syndromes" (2017). *Paediatrics Publications*. 168.

<https://ir.lib.uwo.ca/paedpub/168>

---

**Authors**

Nicolas Waespe, Santhosh Dhanraj, Manju Wahala, Elena Tsangaris, Tom Enbar, Bozana Zlateska, Hongbing Li, Robert J Klaassen, Conrad V Fernandez, Geoff D E Cuvelier, John K Wu, Yves D Pastore, Mariana Silva, Jeffrey H Lipton, José Brossard, Bruno Michon, Sharon Abish, MacGregor Steele, Roona Sinha, Mark J Belletrutti, Vicky R Breakey, Lawrence Jardine, Lisa Goodyear, Liat Kofler, Michaela Cada, Lillian Sung, Mary Shago, Stephen W Scherer, and Yigal Dror

## ARTICLE OPEN

# The clinical impact of copy number variants in inherited bone marrow failure syndromes

Nicolas Waespe<sup>1,2</sup>, Santhosh Dhanraj<sup>1,3</sup>, Manju Wahala<sup>1</sup>, Elena Tsangaris<sup>1</sup>, Tom Enbar<sup>1</sup>, Bozana Zlateska<sup>1,3</sup>, Hongbing Li<sup>1</sup>, Robert J. Klaassen<sup>4</sup>, Conrad V. Fernandez<sup>5</sup>, Geoff D. E. Cuvelier<sup>6</sup>, John K. Wu<sup>7</sup>, Yves D. Pastore<sup>8</sup>, Mariana Silva<sup>9</sup>, Jeffrey H. Lipton<sup>10</sup>, José Brossard<sup>11</sup>, Bruno Michon<sup>12</sup>, Sharon Abish<sup>13</sup>, MacGregor Steele<sup>14</sup>, Roona Sinha<sup>15</sup>, Mark J. Belletrutti<sup>16</sup>, Vicky R. Breakey<sup>17</sup>, Lawrence Jardine<sup>18</sup>, Lisa Goodyear<sup>19</sup>, Liat Kofler<sup>1</sup>, Michaela Cada<sup>3</sup>, Lillian Sung<sup>20</sup>, Mary Shago<sup>21</sup>, Stephen W. Scherer<sup>1</sup> and Yigal Dror<sup>1,2,3</sup>

Inherited bone marrow failure syndromes comprise a genetically heterogeneous group of diseases with hematopoietic failure and a wide array of physical malformations. Copy number variants were reported in some inherited bone marrow failure syndromes. It is unclear what impact copy number variants play in patients evaluated for a suspected diagnosis of inherited bone marrow failure syndromes. Clinical and genetic data of 323 patients from the Canadian Inherited Marrow Failure Registry from 2001 to 2014, who had a documented genetic work-up, were analyzed. Cases with pathogenic copy number variants (at least 1 kilobasepairs) were compared to cases with other mutations. Genotype-phenotype correlations were performed to assess the impact of copy number variants. Pathogenic nucleotide-level mutations were found in 157 of 303 tested patients (51.8%). Genome-wide copy number variant analysis by single-nucleotide polymorphism arrays or comparative genomic hybridization arrays revealed pathogenic copy number variants in 11 of 67 patients tested (16.4%). In four of these patients, identification of copy number variant was crucial for establishing the correct diagnosis as their clinical presentation was ambiguous. Eight additional patients were identified to harbor pathogenic copy number variants by other methods. Of the 19 patients with pathogenic copy number variants, four had compound-heterozygosity of a copy number variant with a nucleotide-level mutation. Pathogenic copy number variants were associated with more extensive non-hematological organ system involvement ( $p = 0.0006$ ), developmental delay ( $p = 0.006$ ) and short stature ( $p = 0.04$ ) compared to nucleotide-level mutations. In conclusion, a significant proportion of patients with inherited bone marrow failure syndromes harbor pathogenic copy number variants which were associated with a more extensive non-hematological phenotype in this cohort. Patients with a phenotype suggestive of inherited bone marrow failure syndromes but without identification of pathogenic nucleotide-level mutations should undergo specific testing for copy number variants.

npj Genomic Medicine (2017)2:18; doi:10.1038/s41525-017-0019-2

## INTRODUCTION

Inherited bone marrow failure syndromes (IBMFs) are complex genetic disorders characterized by hematopoietic failure, which results in single or multi-lineage cytopenias, and an increased prevalence of physical malformations.<sup>1–3</sup> A wide variety of specific syndromes have been described so far with more than 80 different genes associated to IBMFs. These include genes involved in ribosome biogenesis that are associated with Diamond–Blackfan anemia<sup>4</sup> and Shwachman–Diamond syndrome,<sup>5</sup> genes involved in telomere maintenance, associated with dyskeratosis congenita,<sup>6</sup> and genes involved in DNA repair associated with Fanconi anemia.<sup>7</sup> Pathogenic mutations in these

pathways perturb cell homeostasis and lead ultimately to failure of the bone marrow to produce a sufficient amount of blood cells. Based on the inheritance patterns of IBMFs in multiplex families and the segregation of mutated alleles in known IBMF genes of phenotypically affected family members, the disorders are considered monogenic in the vast majority of patients.

A large proportion of IBMFs are important cancer predisposition syndromes, with a particularly increased risk of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML).<sup>1, 2</sup> In particular, Fanconi anemia,<sup>8</sup> Shwachman–Diamond syndrome,<sup>9</sup> severe congenital neutropenia,<sup>10</sup> and familial thrombocytopenia with predisposition to MDS/AML<sup>11</sup> were associated with the

<sup>1</sup>Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada; <sup>2</sup>Marrow Failure and Myelodysplasia Program, Division of Hematology/Oncology, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; <sup>3</sup>Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada; <sup>4</sup>Department of Pediatrics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; <sup>5</sup>Pediatric Hematology/Oncology, IWK Health Centre, Halifax, NS, Canada; <sup>6</sup>Pediatric Hematology/Oncology, University of Manitoba, CancerCare Manitoba, Winnipeg, MB, Canada; <sup>7</sup>Division of Hematology/Oncology, UBC & B.C. Children's Hospital, Vancouver, BC, Canada; <sup>8</sup>CHU Sainte-Justine, Montreal, QC, Canada; <sup>9</sup>Kingston General Hospital, Kingston, ON, Canada; <sup>10</sup>Allogeneic Blood and Marrow Transplant Program, Princess Margaret Cancer Centre, University of Toronto, Toronto, ON, Canada; <sup>11</sup>Centre Hospitalier Universitaire, Sherbrooke, QC, Canada; <sup>12</sup>Centre Hospitalier Universitaire, Québec, QC, Canada; <sup>13</sup>Pediatric Hematology/Oncology, Montreal Children's Hospital, Montreal, QC, Canada; <sup>14</sup>Alberta Children's Hospital, Calgary, AB, Canada; <sup>15</sup>Royal University Hospital, Saskatoon, SK, Canada; <sup>16</sup>Department of Pediatrics, University of Alberta, Edmonton, AB, Canada; <sup>17</sup>Department of Pediatrics, McMaster University, Hamilton, ON, Canada; <sup>18</sup>Children's Hospital, London Health Sciences Centre, London, ON, Canada; <sup>19</sup>Pediatric Hematology/Oncology, Janeway Child Health Centre, St. John's, NF, Canada; <sup>20</sup>Population Health Sciences, Research Institute, Division of Hematology/Oncology, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada and <sup>21</sup>Cytogenetics Laboratory, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada

Correspondence: Yigal Dror (yigal.dror@sickkids.ca)

Nicolas Waespe and Santhosh Dhanraj contributed equally to this work.

Received: 5 October 2016 Revised: 15 March 2017 Accepted: 10 April 2017

Published online: 10 May 2017

development of malignancies. There is an overlap of classical IBMFSs and secondary MDS due to MDS-predisposing germline mutations (e.g., *RUNX1*, *ETV6*, and *GATA2*-related disorders), as patients with these mutations can present with MDS without any additional features of IBMFSs.<sup>12</sup> The risk of development of cancers differs greatly between the various IBMFSs, and identification of the underlying etiology of marrow failure is imperative to assess the need and type of cancer screening.<sup>2</sup>

Identifying the causal genotype in cases with IBMFSs is challenging.<sup>13</sup> In part, the challenge derives from a large number of patients who cannot be classified into specific syndromic diagnoses, in spite of having features highly suggestive of an inherited disorder.<sup>14</sup> To improve the efficiency of genetic testing, utilization of comprehensive next generation sequencing panels was shown to be successful. In a previous report from our group using a panel of 72 genes, the genotype was elucidated in 59% of the patients with clinically classified disorders, and 18% of the patients with marrow failure who were not clinically classified into one particular IBMFS.<sup>15</sup>

Copy number variants (CNVs, i.e. insertions and deletions of at least 1 kilobasepairs) have been found as the underlying genetic mutation in a variety of disorders. Specifically, neurocognitive illnesses, such as autism spectrum disorders and schizophrenia,<sup>16</sup> epilepsy,<sup>17</sup> cerebral palsy,<sup>18</sup> and congenital heart disease<sup>19, 20</sup> were found to be associated with this type of genetic alterations. Short stature has also been associated with rare CNVs.<sup>21</sup> Cases of IBMFSs with pathogenic CNVs have been reported previously, most notably in Diamond-Blackfan anemia<sup>22–24</sup> but also in Fanconi anemia.<sup>25, 26</sup> In particular, an association of these types of mutations with neurodevelopmental delay<sup>22</sup> and short stature<sup>23</sup> has been suggested in small series of patients with Diamond-Blackfan anemia.

Identification of CNVs is not efficiently done through Sanger sequencing and targeted next-generation gene panel sequencing.<sup>27</sup> Therefore, CNVs are often missed when a genome-wide analysis, such as a single nucleotide polymorphism (SNP) array or comparative genomic hybridization (CGH) array, is not performed.<sup>28–30</sup> The cost of these tests is relatively modest around USD 300–400 per assay. In contrast, commercially available targeted next-generation sequencing gene panels, used to detect smaller aberrations of the genetic code, harbor costs of around USD 1500–4000 per assay.<sup>15</sup>

The proportion of cases harboring pathogenic CNVs among patients who have, or are suspected to have IBMFSs, and the importance of finding such genetic lesions in this population is unknown. The aims of this study were to determine the utility of genome-wide microarray analysis to detect pathogenic CNVs, to characterize the genetic lesions and phenotypes of patients with CNVs, and to identify clinical features that might be associated with CNVs. Since CNVs directly disrupt all or a large sequence of the respective gene, and sometimes adjacent genomic regions, it was hypothesized that patients with pathogenic CNVs would likely have a more extensive clinical phenotype compared to nucleotide-level mutations (i.e., single nucleotide variants and small insertions/deletions).

## RESULTS

### Characteristics of the patients

Of the 451 patients enrolled in the Canadian Inherited Marrow Failure Registry (CIMFR) from January 2001 to December 2014, genetic test results were available for 323 (71.6%) patients. Genetic testing was done at the discretion of the treating physicians and/or as part of systematic genotyping of enrolled patient into our research laboratory (YD). One hundred eighty of the 323 patients (55.7%) were found to have causal mutations, which included nucleotide-level mutations, CNVs, or both (Supplementary Table 1). Four patients had insufficient data to

determine the type of mutation, and they were excluded from further analysis. Patients were diagnosed with a wide variety of IBMFSs according to established criteria (Supplementary Table 2).

### Nucleotide-level variant analysis

Among the group of 323 patients who had genetic testing, 303 patients (93.8%) underwent investigation for nucleotide-level mutations by either Sanger sequencing, disease-specific next-generation sequencing (NGS) panels, or by a comprehensive research panel of 72, and later 77 IBMFS genes (ibmfNGS Panel).<sup>15</sup> These genetic evaluations were undertaken either at our research laboratory (YD), or accredited clinical laboratories. Causal nucleotide-level mutations were found in 157 patients (51.8%, Supplementary Table 3). These included 37 patients with small indels involving 1–20 bp. The majority of unclassified patients had undergone extensive genetic testing, which was negative.<sup>13–15</sup>

### Genome-wide CNV analysis

Sixty-seven of 323 patients had genome-wide CNV analysis by SNP/CGH arrays (20.7%). Of those, 39 were assessed in a research laboratory (YD) and the rest in a clinical laboratory. The identified median number of CNVs in the patients assessed in the research laboratory was 141 (interquartile range 133–158) per patient. Further filtering identified a median of 65.5 (59–72) CNVs being >1 kb with five or more probes and 5 (3–7) were found to be rare. Finally, zero to four CNVs per patient were found to affect coding regions in all but one patient. The latter was found to have a significantly higher number of rare CNVs, which might be due to ethnical differences.

Of the 67 patients tested with genome-wide CNV analysis, 11 were found to harbor pathogenic CNVs, resulting in a detection rate of 16.4%. Seven patients were identified with a monoallelic CNV (four patients with Diamond-Blackfan anemia and one with Potocki-Lupski syndrome, supernumerary ring chromosome 1 syndrome, and *RUNX1*-associated familial thrombocytopenia/AML, each), one patient (with Fanconi anemia) harbored two different compound-heterozygous CNVs, and three patients were compound-heterozygous for a CNV and a nucleotide-level mutation (one patient with dyskeratosis congenita and two patients with Thrombocytopenia absent radius syndrome, respectively). In four of 11 cases who were genotyped by SNP/CGH array, genotyping was critical for establishing the diagnosis, as their clinical phenotype was not explicit. The diagnoses in these cases included *PARN*-associated dyskeratosis congenita (published previously),<sup>31</sup> *RPL11*- and *RPL35a*-associated Diamond-Blackfan anemia, and Potocki-Lupski syndrome (Table 1).<sup>32</sup>

In addition to the 11 patients identified with genome-wide CNV analysis, patients were also found to have pathogenic CNVs by metaphase cytogenetics, multiplex ligation-dependent probe amplification assay, fluorescence in-situ hybridization, DNA-quantitative real-time PCR analysis, and Southern blotting (Supplementary Table 3). Altogether, 19 patients were found to harbor pathogenic CNVs in our study. The CNVs ranged from 5 kb to 145.1 Mb in size (Table 1). The most common disease associated with pathogenic CNVs in this population was Diamond-Blackfan anemia, found in four of 35 genotyped patients (11.4%); followed by thrombocytopenia absent radius syndrome ( $n=3$ , 100%), trisomy 8 syndrome ( $n=3$ , 100%), and Fanconi anemia found in two of 28 patients (7.1%). Four patients had compound heterozygosity for a CNV and nucleotide-level mutations (Supplementary Table 1).

### Differences in clinical phenotypes between patients with CNVs and nucleotide-level mutations

The clinical characteristics of patients who harbor pathogenic CNVs or nucleotide-level mutations are shown in Table 2.

**Table 1.** Characteristics of the identified copy number alterations

Patient number	Syndrome	Gene/genomic region (inheritance pattern)	HumanGenome19 coordinates or cytogenetic location	Approximate size of the CNV	Detection method
1#	Diamond–Blackfan anemia	<i>RPL35a</i> (AD)	del(chr3:186,550,246–197,837,050)x1	11.3 Mb	Array 1 and FISH
2#	Diamond–Blackfan anemia	<i>RPS17</i> (AD)	del(chr15:83,196,738–84,812,671)x1	1.6 Mb	Array 2
3#	Diamond–Blackfan anemia	<i>RPL5</i> (AD)	del(chr1:92,932,837–94,018,138)x1	1.1 Mb	Array 3
4#	Diamond–Blackfan anemia	<i>RPL11</i> (AD)	del(chr1:23,449,116–24,147,166)x1	698 kb	Array 4
5	TAR syndrome	Chr1q21.1 <i>RBM8A</i> (AR)	del(chr1:145,390,101–145,792,052)x1 c.–21G>A (chr1:145,507,646G>A)	402 kb SNV	Array 5
6	TAR syndrome	Chr1q21.1 <i>RBM8A</i> (AR)	del(chr1:145,372,550–145,792,064)x1 c.–21G>A (chr1:145,507,646G>A)	420 kb SNV	Array 5
7	TAR syndrome	Chr1q21.1 <i>RBM8A</i> (AR)	del(chr1:144,100,000–144,290,000)x1* NA	190 kb SNV	Quantitative PCR
8	Trisomy 8 syndrome	Chr8 (aneuploidy)	dup(chr8)	145.1 Mb	Cytogenetics and FISH
9	Trisomy 8 syndrome	Chr8 (aneuploidy)	dup(chr8)	145.1 Mb	Cytogenetics and FISH
10	Trisomy 8 syndrome	Chr8 (aneuploidy)	dup(chr8)	145.1 Mb	Cytogenetics and FISH
11#	Fanconi anemia	<i>FANCA</i> (AR)	FANCA deletion exon 31* FANCA deletion exons 4–29*	5 kb 49 kb	MLPA MLPA
12#	Fanconi anemia	<i>FANCA</i> (AR)	del(chr16:89,799,574–89,847,471)x1 del(chr16:89,824,684–89,869,755)x1	47.9 kb 45 kb	Array 2 Array 2
13#	Dyskeratosis congenita	<i>PARN</i> (AR)	del(chr16:14,658,272–14,679,880)x1 c.1045C>T (chr16:14,676,047G>A)	21.6 kb SNV	Array 5
14#	Familial Thrombocytopenia/AML	<i>RUNX1</i> (AD)	del(chr21:34,965,815–36,781,907)x1	1.8 Mb‡	Array 6 and FISH
15	Pearson syndrome	mitDNA (mitochondrial)	del(mitDNA)*	6 kb	Southern blot
16	SRC1 syndrome	Chr1p13.1	+r(chr1:116,673,235–152,748,194)	36.1 Mb	Array 7 and cytogenetics
17	Wolf–Hirschhorn syndrome	Chr4p	der(4)t(4;8)(p16.3;p23.1)*	2.5 Mb(del), 9 Mb(dup)	Cytogenetics and FISH
18	Jacobson syndrome	11q23	del(11)(q23.3)x1*	17 Mb	Cytogenetics
19	Potocki–Lupski syndrome	17p11.2	dup(chr17:16,778,108–18,252,450)x1	1.5 Mb	Array 8 and FISH

#, patients included in the subset analysis; \*, for methods without identification of the exact breakpoint, the genomic locations were estimated as well as the total size of the respective CNV; ‡, complex intrachromosomal rearrangement of chromosome 21 including a deletion involving the *RUNX1* gene  
AD autosomal dominant, AML acute myeloid leukemia, AR autosomal recessive, Array 1 Agilent 105 K Human Genome Oligonucleotide array, Array 2 Agilent 180 K Human Genome Oligonucleotide array, Array 3 GeneDx 180 K microarray v4, Array 4 Agilent Oligo Array – EmArray cyto 6000 custom design, Array 5 Affymetrix SNP Array 6.0, Array 6 CytoSure Syndrome Plus V2, Array 7 Roche NimbleGen 135 K oligonucleotide array, Array 8 Signature Genomic SignatureChipWG Whole genome BAC array, FISH fluorescence in-situ hybridization, MLPA Multiplex Ligation-dependent Probe Amplification, NA not available, SNV single nucleotide variant, SRC1 supernumerary ring chromosome 1, TAR thrombocytopenia absent radius

Demographic data were similar among the groups. The age at presentation and age at diagnosis were assessed where available, which was the case in 17 and 19 patients harboring CNVs, respectively, and 143 and 146 patients without CNVs, respectively. There was no statistical difference between cases with CNVs and cases with nucleotide-level mutations (Fig. 1).

To assess the potential effect of pathogenic CNV mutations on hematological phenotype across the different types of IBMFSs, the age at onset of severe cytopenias was assessed where this information was available. No differences were found between patients with CNVs ( $n = 17$ ) and those without CNVs ( $n = 146$ ) (Fig. 2a). Similarly, the overall survival was not different between the two groups (Fig. 2b).

The correlation of CNV mutations with the extent of involvement of non-hematological organ systems was analyzed. Agreement in assignment of non-hematological organ system

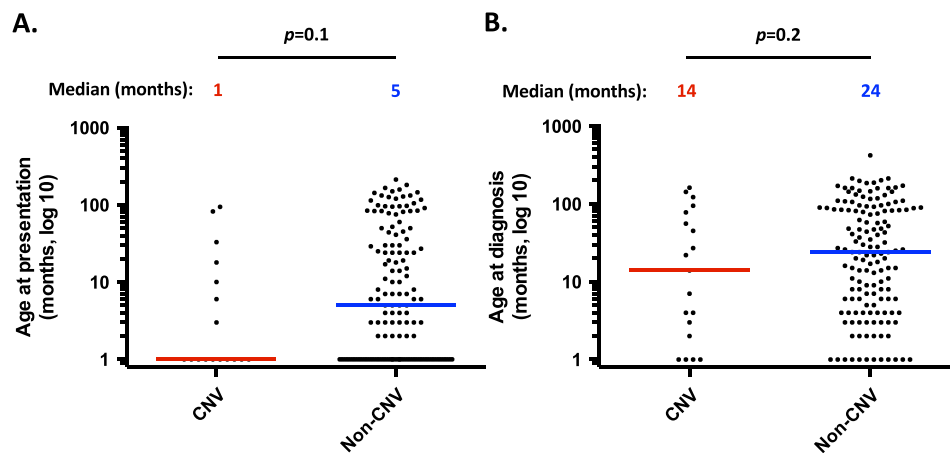
involvement between two evaluators was excellent ( $\kappa = 0.87$ ; 95%-CI 0.85–0.89). Patients with pathogenic CNVs were found to have symptoms in significantly more non-hematological organ systems compared to patients with nucleotide-level mutations ( $p = 0.0006$ ; Fig. 3). Patients with CNVs were also found to have a significantly higher proportion of developmental delay ( $p = 0.006$ ) and short stature ( $p = 0.04$ ) compared to patients with nucleotide-level mutations (Fig. 4).

Since the above analyses included several syndromes with small numbers of patients, a subgroup analysis of only IBMFSs with at least five cases that included patients with CNVs, as well as patients with nucleotide-level mutations, was performed. Patients with Diamond–Blackfan anemia, Fanconi anemia, dyskeratosis congenita, and familial thrombocytopenia/AML and MDS predisposition syndromes due to *RUNX1*, *ANKRD26* gene mutations, and *GATA2*-associated MDS/AML met these criteria. Patients harboring

**Table 2.** Phenotypic characteristics of patients included in this study, sorted by cases with copy number variants (CNVs) and other mutations

	Patients with CNVs ( <i>n</i> = 19)	Patients with other mutations ( <i>n</i> = 157)	<i>p</i> -value
Sex (females, %)	7 (36.8%)	76 (48.4%)	<i>p</i> = 0.5
Median age at presentation (IQR)	1 month (1–14)	5 months (1–41)	<i>p</i> = 0.1
Median age at diagnosis (IQR)	14 months (2–78)	24 months (5–88.25)	<i>p</i> = 0.2
Median age at last follow up (IQR)	12.6 years (6.5–16.7)	12.2 years (7–18)	<i>p</i> = 0.6
Severe anemia ( <i>n</i> , %)	9 (50%)	64 (46.4%)	<i>p</i> = 0.8
NA ( <i>n</i> )	1	19	
Severe neutropenia ( <i>n</i> , %)	5 (35.7%)	56 (41.8%)	<i>p</i> = 0.8
NA ( <i>n</i> )	5	23	
Severe thrombocytopenia ( <i>n</i> , %)	5 (31.3%)	18 (13.3%)	<i>p</i> = 0.07
NA ( <i>n</i> )	3	22	
Severe aplastic anemia ( <i>n</i> , %)	4 (21.1%)	33 (21%)	<i>p</i> = 1.0
NA ( <i>n</i> )	0	0	
MDS/ AML ( <i>n</i> , %)	5 (26.3%)	26 (16.7%)	<i>p</i> = 0.3
NA ( <i>n</i> )	0	1	
Median number of non-hematological systems involved (IQR)	6 (3.75–8)	3 (1–5)	<i>p</i> = 0.0006
Developmental delay ( <i>n</i> , %)	11 (61.1%)	40 (27.4%)	<i>p</i> = 0.006
NA ( <i>n</i> )	1	11	
Short stature ( <i>n</i> , %)	12 (66.7%)	59 (40.1%)	<i>p</i> = 0.04
NA ( <i>n</i> )	1	10	

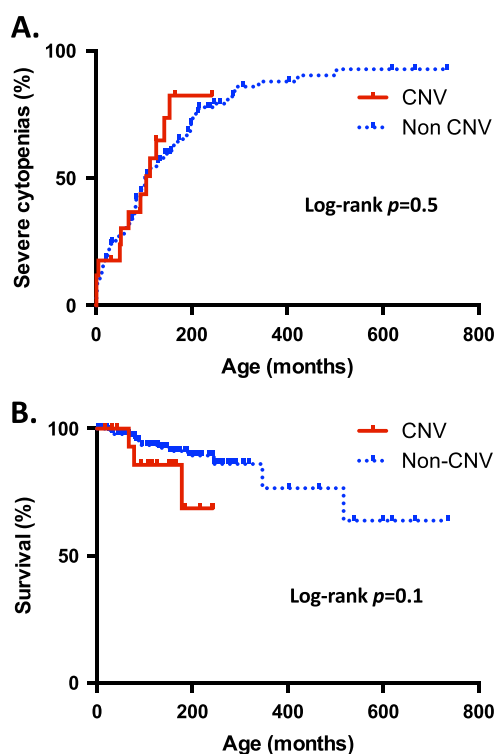
CNV copy number variant, IQR inter-quartile range, NA data not available

**Fig. 1** Analysis of clinical parameters comparing cases with CNVs vs. other mutations. **a** Age at presentation, color bars representing corresponding median; **b** Age at diagnosis, color bars representing corresponding median

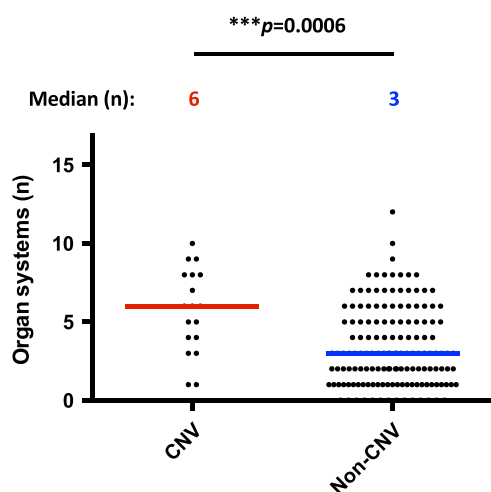
whole or large segmental chromosome alterations that are typically identified through metaphase cytogenetics and patients with thrombocytopenia-absent radius syndrome who are typically identified by combined sequencing and targeted copy number analysis of the *RBM8A* gene were thus excluded. Patients included in this sub-analysis were all identified by genome-wide CNV analysis. This subgroup analysis revealed no differences between patients with CNVs (*n* = 8) and nucleotide-level mutations (*n* = 81) with regard to age at presentation, age at diagnosis, onset of severe cytopenias, and overall survival (Supplementary Figs. 1 and 2). However, similar to the whole group analysis, a significantly higher number of organ system involvement was found among cases

with CNVs compared to other types of mutations (*p* = 0.007). Importantly, a significantly higher incidence of developmental delay (*p* = 0.04), and short stature (*p* = 0.02) was also found in the subgroup analysis.

Finally, to further adjust for potential bias caused by unequal distribution of syndromes with high or low number of organ system involvement between the two genetic groups, the extent of non-hematological organ system involvement was dichotomized. A trend was seen towards more cases showing “extensive non-hematological organ system involvement” for the specific syndromes with CNVs (75%) compared to those without (39%) but the difference in incidence did not reach statistical significance (*p* = 0.07).



**Fig. 2** Analysis of onset of hematological complications and survival comparing cases with CNVs vs. other mutations. **a** Age at onset of severe cytopenias; **b** Overall survival



**Fig. 3** Analysis of non-hematological organ system involvement comparing cases with CNVs vs. other mutations. Color bars representing corresponding median

## DISCUSSION

Patients who present with bone marrow failure often cannot be directly assigned to a specific syndromic and genetic diagnosis. Our group and others showed that utilization of a comprehensive nucleotide-level testing approach, that targets a large number of known IBMFS genes in one next-generation sequencing panel, have resulted in a high identification rate of causal mutations with sometimes unexpected genetic findings and amendment of clinical diagnoses.<sup>15, 33</sup> The present study shows that genome-wide CNV analysis can reveal causal mutations in a substantial number of patients with IBMFSs. Importantly, there was a

difference in non-hematological phenotypes in patients affected with pathogenic CNVs compared to other types of mutations.

It is well known that patients with nucleotide-level mutations in IBMFS genes can express hematological and non-hematological symptoms as a consequence of a single gene mutation. It is unknown what role CNVs play in IBMFSs as a group. The CIMFR database offers the opportunity to assess a large cohort of patients with IBMFSs in Canada. The registry data reflects the unselected spectrum of diseases that get attention for workup of a suspected IBMFS. The vast majority of patients was tested for nucleotide-level mutations which revealed the causal genetic germline mutation in about half of the patients. The current analysis included 67 patients tested with a genome-wide approach to identify CNVs. Of those, 16.4% harbored a pathogenic CNV, which helped establish the genetic diagnosis in this patient cohort.

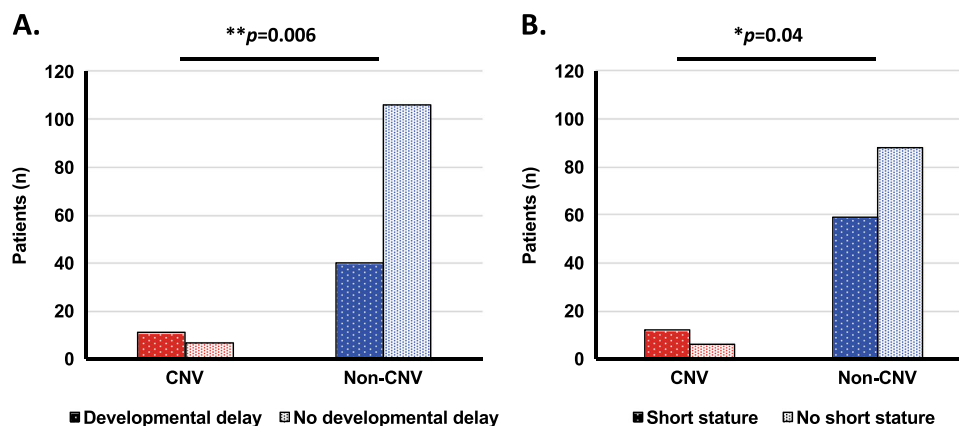
Depending on the source of DNA, somatic mutations can contaminate germline DNA sequences. The present study reduced the risk of falsely identifying mutations as being germline by the choice of sampled cells (see Methods section). Nevertheless, in a small number of patients, peripheral blood myeloid cells were not removed from the analysis and a possibility of somatic CNVs cannot be completely excluded.

Previous studies assessed the extent of non-hematological symptom manifestations by simply counting the number of congenital defects.<sup>34</sup> Since patients with IBMFSs may have multiple defects in only one organ system (e.g., café-au-lait spots, hyperpigmentation, and leukoplakia on the tongue and mucosal surfaces), this approach does not provide information about the potential biological links between a genetic lesion and the extent of involvement of different organ systems. To overcome this limitation, the symptoms were grouped by affected organ system and the number of organ systems summed up. It was previously unknown whether the presence of pathogenic CNVs increases the risk of having more extensive organ system involvement. In this study, patients with CNVs were found to be affected in significantly more non-hematological organ systems than those with only nucleotide-level mutations. This was true for the whole cohort including patients with structural chromosomal changes. Interestingly, this association remained true for a subgroup of IBMFSs where at least five patients were present with at least one patient harboring a CNV excluding structural chromosomal changes, rare IBMFSs, and IBMFSs where no CNVs were present.

Short stature and developmental delay are common features among patients with IBMFSs. These complications are primarily skeletal and neurological abnormalities, respectively. However, both might be affected by defects in various other organ systems (e.g., gastrointestinal, endocrine, cardiac etc.). We speculated that larger mutations (i.e., CNVs) might affect several areas of the genome, influencing several organ systems and therefore have an effect on development and stature as markers for multisystem involvement.

Furthermore, there is previously published data showing that CNVs might be associated with a higher incidence of developmental delay and short stature in patients with Diamond-Blackfan anemia but no such association was seen in other IBMFSs.<sup>22, 23</sup> Developmental delay and short stature were therefore analyzed separately. Both were found to be more prevalent in patients with CNVs in the whole cohort but also in the subgroup analysis of patients with at least 5 patients and CNVs present. Due to the sample size, it was difficult to assess the impact of CNVs on specific IBMFSs. Nevertheless, it is noteworthy that among patients with Diamond-Blackfan anemia, three of four patients with CNVs had growth retardation and three of four had developmental delay (vs. 35.5% and 51.6% of patients without CNVs, respectively).

Assessment of the age at presentation, age at diagnosis and age at onset of severe cytopenias was performed. Most patients



**Fig. 4** Analysis of specific organ system involvement comparing cases with CNVs vs. other mutations. **a** Developmental delay; **b** Short stature

included in this study had this information available. In contrast to the non hematological phenotype in the same patient group, there was no difference in the age at presentation, age at diagnosis and the age at onset of severe cytopenias between patients harboring CNVs and those who do not have this type of mutation. This might be due to the hematological phenotype being influenced exclusively by the gene associated with the IBMFS and not adjacent genomic areas. Since IBMFSs are rare, the number of patients assessed might also have influenced the lack of difference. Future studies with larger patient groups might elucidate if there is a genuine difference in the mechanisms of the hematological and non hematological phenotype in patients affected with CNVs.

No attempt has previously been undertaken to assess the impact of CNVs on the diagnostic investigative work-up of patients who have signs of IBMFSs. Furthermore, for the entire group of IBMFSs, there has not been a comparative assessment of CNVs yet. To our knowledge, this is the largest study in which a comprehensive analysis of various IBMFSs with known causal genotypes has been conducted.

Our study has several limitations. First, although 19 patients with CNVs is a relatively large number when rare disorders such as IBMFSs are analyzed, larger studies are necessary to validate these results. Specifically, the number of patients per IBMFS was too small for intrasyndromic variation analysis. Another limitation is that patients harboring compound-heterozygous and homozygous mutations have been analyzed together in autosomal recessive diseases. In such cases, the genotype severity was graded according to the presumably most damaging lesion, i.e., CNVs. It would be of interest to analyze these groups separately to assess whether there are differences in cases with heterozygous, homozygous and compound-heterozygous CNV mutations. In addition, some mutations (e.g., nonsense or indel with frameshift) can have a major impact on the pathophysiology of the disease by either completely abrogating the translated protein or changing its function. The impact of a variant on the function of the translated protein was not assessed in this analysis. Similarly, some missense mutations have major deleterious effects on protein function and these mutations can cause a severe clinical phenotype. One such example is the Hoyeraal-Hreidarsson syndrome associated with *TINF2* mutations.<sup>35</sup> Further, IBMFS genes may vary in the degree of essentiality and affect the phenotype in different ways. Last, it is noteworthy that our approach to quantify organ system involvement could not assess the impact of encountered defects on morbidity and quality of life. Future methods that focus on these questions should be developed.

The mechanisms for the encountered differences in CNVs compared to other mutations could not be elucidated with this study. CNVs could affect the phenotype of IBMFS patients by

either severely damaging the IBMFS protein itself or by affecting additional genes surrounding the IBMFS gene locus. Future research needs to evaluate what mechanism is underlying this difference in phenotypes.

In summary, most patients with IBMFSs have nucleotide-level mutations. However, an important proportion of patients harbor pathogenic CNVs that are not efficiently detected by current nucleotide-level testing methods. Therefore, genome-wide CNV analysis should be considered in IBMFS cases where nucleotide-level sequencing does not reveal the causal mutation or in cases with autosomal recessive disorders where only one pathogenic allele was found by nucleotide-level analysis. According to our analysis this is true not only for Diamond–Blackfan anemia and Fanconi anemia patients as suggested previously,<sup>22–26</sup> but in general for patients with suspected IBMFS. The costs associated with this additional analysis are relatively modest. In urgent cases, simultaneous investigation to reveal both, nucleotide-level mutations and CNVs, should be considered. In this IBMFS cohort, pathogenic CNVs were associated with more extensive non-hematological organ system involvement, and a higher prevalence of developmental delay and short stature. Patients with these clinical features should be assessed with CNV analysis early in the evaluation of their disease.

## METHODS

### The Canadian inherited marrow failure registry

The Canadian Inherited Marrow Failure Registry (CIMFR) is a multicenter prospective study, which was approved by the Research Ethics Boards of all participating institutions. Individuals who fulfilled diagnostic criteria for an IBMFS<sup>13</sup> were recruited by pediatric hematologists and oncologists at participating centers who care for >95% of the eligible pediatric IBMFS population in Canada. Patients were enrolled in the study registry after informed written consent was obtained from the patient or their guardian.

Eligibility criteria included evidence of chronic bone marrow failure in addition to either a family history of an IBMFS, physical malformations, presentation earlier than 1 year of age, or positive genetic testing for an associated IBMFS gene. When possible, each case was assigned a specific syndromic diagnosis by the participating center. Diagnoses were reviewed centrally, and if necessary were adjusted based on published diagnostic criteria of specific IBMFSs<sup>1, 3, 36, 37</sup> after verification with the respective center. Cases that fulfilled the eligibility criteria, but did not meet the clinical, laboratory and genetic diagnostic criteria for any known IBMFS subtype were defined as unclassified IBMFSs.<sup>14</sup> Patients were excluded if they were diagnosed with (i) acquired aplastic anemia (i.e., without the above criteria for inherited bone marrow failure), (ii) de novo leukemia or leukemia with an unclassified syndrome without preceding bone marrow failure, or (iii) de novo myelodysplastic syndrome (MDS).

The current study included patients who had been enrolled in the CIMFR from its inception in January 2001 until December 2014. Data available for these patients until June 2015 were analyzed. Patient information at study entry and at yearly follow-up was collected and included demographics,



clinical symptoms, family history, physical malformations, laboratory tests, diagnoses, genetic tests, imaging studies, treatment, and outcomes.

**Next generation sequencing panel assay and Sanger sequencing**  
Genomic DNA from either bone marrow fibroblasts, expanded peripheral blood T-cells, or whole blood cells was extracted. Comprehensive testing for 72 and 77 IBMFS genes respectively was done by an NGS IBMFS gene panel using the HaloPlex Capture Kit (Agilent Technologies, Santa Clara, CA) for DNA library preparation, and Illumina HiSeq 2500 platform for sequencing, as described previously.<sup>15</sup> Sanger sequencing was done as previously described.<sup>13</sup> Briefly, targeted genes were analyzed by bidirectional sequencing of individual exons and flanking intronic regions after PCR amplification.

#### Genome-wide copy number variant analysis

For genome-wide CNV analysis, Affymetrix SNP6.0 array, Affymetrix Inc., Santa Clara, CA, was performed in our research laboratory (YD). To minimize the possibility of detecting somatic CNVs, genomic DNA was extracted from either bone marrow fibroblasts, expanded peripheral blood T-cells, or whole blood cells. The assay features 1.8 million genic markers, including more than 906,600 SNPs and more than 946,000 probes for detection of CNVs. It was performed according to the manufacturer's instructions as previously reported.<sup>9</sup> A subset of patients were analyzed with other CNV arrays in one of two clinical laboratories: either the clinical Cytogenetics laboratory at the Hospital for Sick Children (Toronto, Ontario, Canada), or at Prevention Genetics (Marshfield, Wisconsin, USA; Table 1 and Supplementary Table 4). In these cases, DNA from either phorbol myristate acetate-stimulated lymphocytes or whole blood was extracted and standard protocols were used for the assays and variant calling.

Three different types of CNV calling algorithms were used, namely Birdsuite,<sup>38</sup> iPattern,<sup>39</sup> and Affymetrix Genotyping Console software. CNVs that were more than 1 kb in size with five or more consecutive array probes were selected. CNVs seen in 2 or 3 algorithms were considered as stringent CNVs and were included in further analysis. CNVs were further filtered based on their occurrence in healthy control datasets that were provided from The Center of Applied Genomics at the Hospital for Sick Children. CNVs that were seen frequently in healthy controls were removed and only rare CNVs were further prioritized based on (1) CNV type (either deletion or gain), (2) region affected (coding or non-coding regions of the gene), (3) CNVs affecting individual genes vs. multiple genes and (4) known function of the gene in hematopoiesis.

#### Metaphase cytogenetics and fluorescence in situ hybridization

Metaphase cytogenetic analysis and fluorescence in situ hybridization (FISH) were performed as previously described.<sup>40</sup> In brief, for metaphase cytogenetics, bone marrow or peripheral blood cells were cultured and stopped in metaphase, stained and assessed by light-microscopy and computer-based imaging techniques. FISH was performed with the locus-specific probe according to clinical suspicion of a specific gene being affected.

#### Grouping of cases according to genetic mutations

A genetic variant was called "causal" if the aberration fulfilled criteria of being "most likely pathogenic" by in silico analysis<sup>15</sup> or previously described as "pathogenic" in public databases with a corresponding clinical phenotype. Patients were grouped according to the causal mutation having one of the following types of mutations: (1) CNVs spanning at least 1 kb,<sup>41</sup> (2) all other types of genetic mutations (i.e., single nucleotide variants, and indels). All cases were grouped by the causal genetic mutation type in dominantly inherited diseases. In diseases with recessive inheritance, cases were assigned as follows: at least one pathogenic CNV qualified for analysis in group 1; all other mutations in group 2.

#### Assessment of hematological and non-hematological phenotype

To analyze the impact of different genetic aberrations on the disease phenotype across all IBMFSs, age at presentation, age at diagnosis, age at onset of severe cytopenias, and overall survival were analyzed. Age at presentation was defined as occurrence of any symptoms attributed to an IBMFS of the hematopoietic or any other organ system. Age at diagnosis was set as the age at which the specific IBMFS was diagnosed either

clinically or by genetic testing. Age at onset of severe cytopenias was the age at which either of the following peripheral blood count values were noted: (i) hemoglobin of <70 g/L and/or need for transfusions or other treatments to increase the hemoglobin (e.g., steroids) for at least 3 months; (ii) platelet count of <20 × 10<sup>9</sup>/L and/or need for platelet transfusions for at least 3 months; (iii) absolute neutrophil count <0.5 × 10<sup>9</sup>/L and/or need for granulocyte colony-stimulating factor treatment for at least 3 months. Overall survival (OS) was defined as the time from birth to last follow up (censored) or death from any cause (event).

As various organ systems are affected in IBMFSs, the extent of non-hematopoietic organ system involvement was evaluated. The absolute number of organ systems affected was assessed by grouping clinical symptoms into 17 different organ systems, and evaluating the number of organ systems affected by at least one symptom (Supplementary Table 5). Acquired symptoms and complications (e.g., acute infections) were excluded. The involvement of each organ system was assigned by two independent investigators (NW, MW). Agreement in assignment was evaluated using the kappa statistic. Discordant findings were reviewed and differences were resolved in a discussion among three investigators (NW, MW, YD).

Cases were considered having "extensive non-hematological organ system involvement" if non-hematological organ system involvement exceeded the median found for the assigned IBMFS in the present study. "Limited non-hematological organ system involvement" was defined as having an equal or lower number of organ systems affected compared to the median for the specific syndrome.

Short stature was defined as body height below the fifth percentile for age. Developmental delay was defined as a delay in the attainment of developmental milestones for young children and/ or inability to attend regular school without special assistance for school-aged children.

#### Statistical analysis

Results were presented using descriptive statistics, including calculation of median and inter-quartile ranges (IQR) for continuous variables, and percentages for categorical variables. Fisher's exact test was used to compare dichotomous variables. Mann-Whitney U test was used to assess for differences in continuous variables. Survival analysis and evaluation of onset of severe cytopenias were performed with the Kaplan Meier estimate and Mantel-Cox log-rank test. Two-tailed  $p < 0.05$  was considered statistically significant. Statistical analysis was conducted through Graphpad Prism 6.0 h (La Jolla, California, USA).

#### Data availability

As many of the IBMFSs included in this study contain only small numbers of individual patients, the risk of identifying single patients is high despite avoidance of explicit identifiers. Therefore, the clinical data are not publicly available due to them containing information that could compromise research participant privacy/consent.

#### ACKNOWLEDGEMENTS

We would like to thank the participating patients and their families who contributed to this study and the funding agencies for their support. This research project was conducted with support from the Canadian Institute of Health Research (funding references 102528 and 286737), from the Nicola's Kids Triathlon Fund and from C17 Research Network, and funded by the Childhood Cancer Canada Foundation. N.W. was supported by a fellowship of the Garron Family Cancer Centre and Sears Canada.

#### AUTHOR CONTRIBUTIONS

N.W. and S.D. performed research, analyzed data, and wrote the paper. M.W., E.T., T.E., B.Z., and L.K. collected and analyzed data. H.L. performed research, and analyzed data. R.J.K., C.V.F., G.D.E.C., J.K.W., Y.D.P., M.Si., J.H.L., J.B., B.M., S.A., M.St., R.S., M.J.B., V. R.B., L.J., L.G., and M.C. are study co-investigators, contributed vital data and wrote the manuscript. L.S., M.Sh., and S.W.S. analyzed data and edited the manuscript. Y.D. designed research, oversaw the project, analyzed data and wrote the paper. All authors contributed substantially to the work, revised it critically, granted final approval of the version to be published, and agreed to be accountable for all aspects of the work.

## COMPETING INTERESTS

The authors declare no competing financial interests.

## REFERENCES

1. Dror, Y. in *Pediatric Hematology* (eds Arceci, R. J. et al.) 30–63 (Blackwell Publishing Ltd, 2006).
2. Alter, B. P. et al. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *Br. J. Haematol.* **150**, 179–188 (2010).
3. Shimamura, A. & Alter, B. P. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood. Rev.* **24**, 101–122 (2010).
4. Draptchinskaia, N. et al. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat. Genet.* **21**, 169–175 (1999).
5. Boocock, G. R. B. et al. Mutations in SBDS are associated with Shwachman–Diamond syndrome. *Nat. Genet.* **33**, 97–101 (2003).
6. Vulliamy, T. J. & Dokal, I. Dyskeratosis congenita: the diverse clinical presentation of mutations in the telomerase complex. *Biochimie.* **90**, 122–130 (2008).
7. de Winter, J. P. & Joenje, H. The genetic and molecular basis of Fanconi anemia. *Mutat. Res.* **668**, 11–19 (2009).
8. Alter, B. P. Fanconi anemia and the development of leukemia. *Best Pract. Res. Clin. Haematol.* **27**, 214–221 (2014).
9. Dhanraj, S. et al. Molecular characteristics of a pancreatic adenocarcinoma associated with Shwachman–Diamond syndrome. *Pediatr. Blood Cancer* **60**, 754–760 (2013).
10. Donadieu, J. et al. Analysis of risk factors for myelodysplasias, leukemias and death from infection among patients with congenital neutropenia. Experience of the French severe chronic neutropenia study group. *Haematologica* **90**, 45–53 (2005).
11. Wlodarski, M. W. et al. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood* **127**, 1387–1397 (2016).
12. Bannon, S. A. & DiNardo, C. D. Hereditary predispositions to myelodysplastic syndrome. *Int. J. Mol. Sci.* **17**, 838 (2016).
13. Tsangaris, E. et al. Genetic analysis of inherited bone marrow failure syndromes from one prospective, comprehensive and population-based cohort and identification of novel mutations. *J. Med. Genet.* **48**, 618–628 (2011).
14. Teo, J. T. et al. Clinical and genetic analysis of unclassifiable inherited bone marrow failure syndromes. *Pediatrics* **122**, e139–e148 (2008).
15. Ghemlas, I. et al. Improving diagnostic precision, care and syndrome definitions using comprehensive next-generation sequencing for the inherited bone marrow failure syndromes. *J. Med. Genet.* **52**, 575–584 (2015).
16. Cook, E. H. Jr & Scherer, S. W. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455**, 919–923 (2008).
17. Olson, H. et al. Copy number variation plays an important role in clinical epilepsy. *Ann. Neurol.* **75**, 943–958 (2014).
18. Oskoui, M. et al. Clinically relevant copy number variations detected in cerebral palsy. *Nat. Commun.* **6**, 7949 (2015).
19. Mak, C. C. Y. et al. De novo large rare copy-number variations contribute to conotruncal heart disease in Chinese patients. *NPJ Genomic. Med* **1**, 16033 (2016).
20. Costain, G., Silversides, C. K. & Bassett, A. S. The importance of copy number variation in congenital heart disease. *NPJ Genomic. Med* **1**, 16031 (2016).
21. Zahnleiter, D. et al. Rare copy number variants are a common cause of short stature. *PLoS Genet.* **9**, e1003365 (2013).
22. Farrar, J. E. & Dahl, N. Untangling the phenotypic heterogeneity of diamond blackfan anemia. *Semin. Hematol.* **48**, 124–135 (2011).
23. Kuramitsu, M. et al. Extensive gene deletions in Japanese patients with diamond-blackfan anemia. *Blood* **119**, 2376–2384 (2012).
24. Landowski, M. et al. Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond–Blackfan anemia. *Hum. Genet.* **132**, 1265–1274 (2013).
25. Morgan, N. V., Tipping, A. J., Joenje, H. & Mathew, C. G. High frequency of large intragenic deletions in the fanconi anemia Group A gene. *Am. J. Hum. Genet.* **65**, 1330–1341 (1999).
26. Nicchia, E. et al. Identification of point mutations and large intragenic deletions in Fanconi anemia using next-generation sequencing technology. *Mol. Genet. Genomic Med* **3**, 500–512 (2015).
27. Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S. & Salim, A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics.* **28**, 2711–2718 (2012).
28. Carson, A. R., Feuk, L., Mohammed, M. & Scherer, S. W. Strategies for the detection of copy number and other structural variants in the human genome. *Hum. Genomics* **2**, 403 (2006).
29. Keren, B. & Caignec, C. L. Oligonucleotide microarrays in constitutional genetic diagnosis. *Expert. Rev. Mol. Diagn.* **11**, 521–532 (2011).
30. Li, W. & Olivier, M. Current analysis platforms and methods for detecting copy number variation. *Physiol. Genomics.* **45**, 1–16 (2013).
31. Dhanraj, S. et al. Bone marrow failure and developmental delay caused by mutations in poly(A)-specific ribonuclease (PARN). *J. Med. Genet.* **52**, 738–748 (2015).
32. Goh, E. S.-Y. et al. Definition of a critical genetic interval related to kidney abnormalities in the Potocki–Lupski syndrome. *Am. J. Med. Genet. A.* **158A**, 1579–1588 (2012).
33. Zhang, M. Y. et al. Genomic analysis of bone marrow failure and myelodysplastic syndromes reveals phenotypic and diagnostic complexity. *Haematologica.* **100**, 42–48 (2015).
34. Vulliamy, T. J. et al. Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations. *PLoS ONE* **6**, e24383 (2011).
35. Walne, A. J., Vulliamy, T., Beswick, R., Kirwan, M. & Dokal, I. TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. *Blood.* **112**, 3594–3600 (2008).
36. Alter, B. P. Bone marrow failure syndromes in children. *Pediatr. Clin. North. Am.* **49**, 973–988 (2002).
37. Dokal, I. & Vulliamy, T. Inherited bone marrow failure syndromes. *Haematologica* **95**, 1236–1240 (2010).
38. Korn, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
39. Pinto, D. et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat. Biotechnol.* **29**, 512–520 (2011).
40. Dror, Y., Squire, J., Durie, P. & Freedman, M. H. Malignant myeloid transformation with isochromosome 7q in Shwachman–Diamond syndrome. *Leukemia* **12**, 1591–1595 (1998).
41. Connolly, J. J. et al. Copy number variation analysis in the context of electronic medical records and large-scale genomics consortium efforts. *Appl. Genet. Epidemiol.* **5**, 51 (2014).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

Supplementary Information accompanies the paper on the *npj Genomic Medicine* website (doi:10.1038/s41525-017-0019-2).