Lack of splice factor and cohesin complex mutations in pediatric myelodysplastic syndrome

Julia C. Obenauer,[1,2] François G. Kavelaars,[1] Mathijs A. Sanders,[1] Remco M. Hoogenboezem,[1] Andrica C. H. de Vries,[2] Paulina M. H. van Strien,[1] Valerie de Haas,[3] Franco Locatelli,[4] Henrik Hasle,[5] Peter J. M. Valk,[1] Ivo P. Touw[1] and Marry M. van den Heuvel-Eibrink[2,6,*]*

**IPT and MMH-E contributed equally to this work.*

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands; [2]Department of Pediatric Hemato-Oncology, Erasmus University Medical Center / Sophia Children's Hospital, Rotterdam, the Netherlands; [3]Dutch Childhood Oncology Group (DCOG/SKION), The Hague, the Netherlands; [4]Department of Pediatric Hematology-Oncology, IRCCS Bambino Gesù Children's Hospital, University of Pavia, Rome, Italy; [5]Department of Pediatrics, Aarhus University Hospital Skejby, Aarhus, Denmark and [6]Princess Màxima Center for Pediatric Oncology (PMC), Utrecht, the Netherlands

Correspondence: m.m.vandenheuvel-eibrink@prinsesmaximacentrum.nl or i.touw@erasmusmc.nl
doi:10.3324/haematol.2016.151753

Supplementary Methods

Patient material

Samples taken at time of diagnosis from 38 pediatric patients with MDS were studied; defined as being under the age of 19 years. 24 were Dutch (Dutch Childhood Oncology Group, The Hague, the Netherlands), 7 German (University Hospital Freiburg, Freiburg, Germany), 3 Italian (IRCCS Ospedale Bambino Gesù, Rome, Italy), 2 Czech (University Hospital Motol, Prague, Czech Republic) and 2 Danish (Aarhus University Hospital Skejby, Aarhus, Denmark). Informed consent was obtained from each participant, or their legal guardians, as part of sample collection by the European working group (EWOG). Institutional review board approval was obtained in all participating centers. The cohort comprised 24 primary and 14 secondary MDS cases. Secondary MDS occurred after aplastic anemia (N=2), acute lymphoid leukemia (N=3), chemotherapy (N=1), Burkitt non-Hodgkin lymphoma (N=1), Diamond-Blackfan Anemia (N=1), Ewing sarcoma (N=2), Fanconi anemia (N=1) and acute myeloid leukemia (N=2). The primary disease was unknown in one case. Material was obtained from bone marrow aspirates, and mononuclear cells were obtained using Ficoll gradient separation. Further patient details are provided in Supplementary table 1.

DNA sequencing

Next generation DNA sequencing was performed for 37 patients on the MiSeq platform (Illumina, San Diego, CA). The TruSight Myeloid Sequencing Panel (Illumina) was used to generate libraries and prepared according to the TruSight DNA Amplicon Sequencing Panel Guide Part#15054779Rev.B (Illumina). Patient 5 was sequenced by whole exome sequencing (WES). WES sample preparation was performed as described previously by Beekman *et al*[1].

Bioinformatics

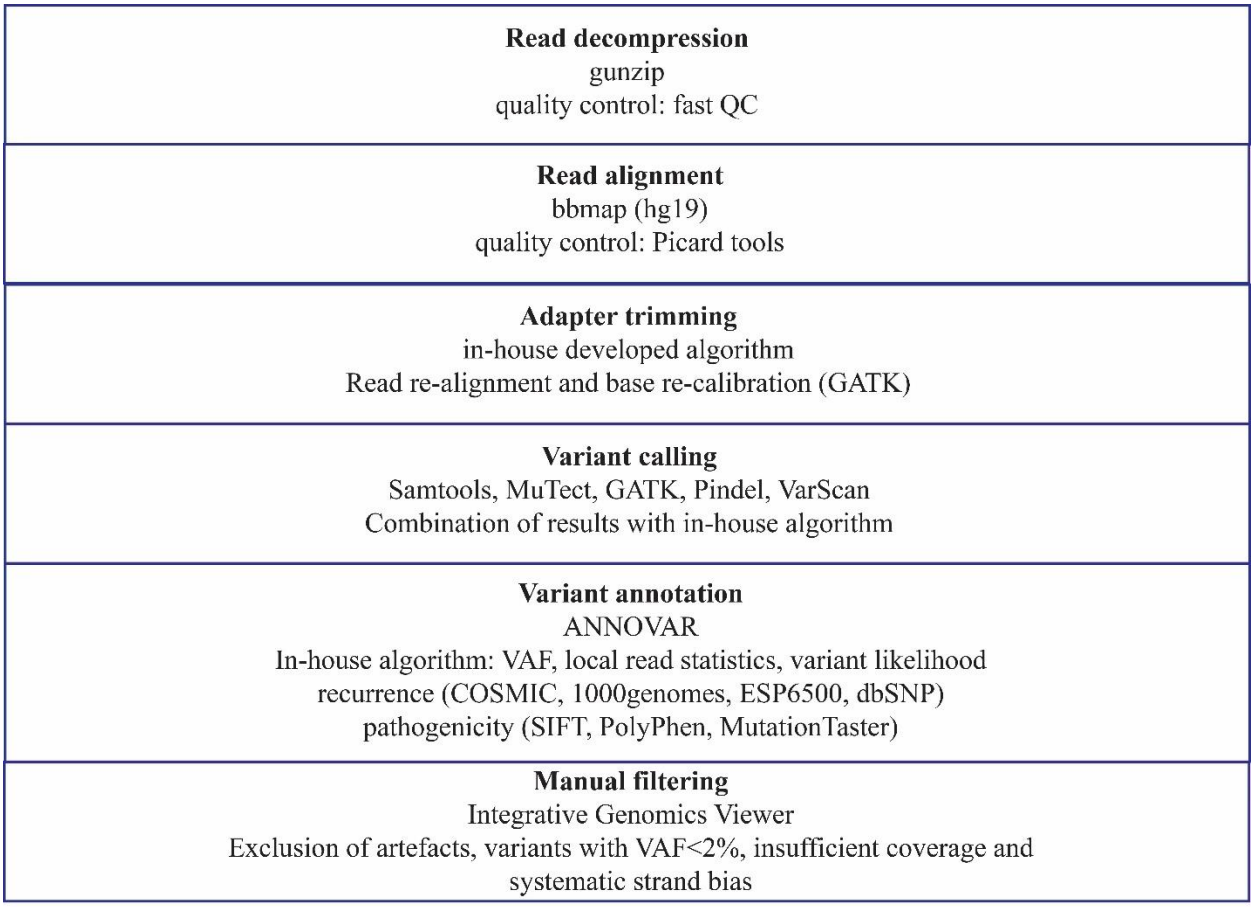
MiSeq data analysis was performed using a pipeline developed at the Erasmus Medical Center (M.A.S., R.M.H.): reads were decompressed using gunzip, and resulting fastq files controlled for quality using fastQC (Babraham Institute, Cambridgeshire, UK). Reads were subsequently aligned using bbmap, using genome build hg19. The resulting sam and bam files were sorted and controlled for quality using Picard tools (Broad Institute, Cambridge, MA). Adapter reads were trimmed using an in-house developed algorithm, and read re-alignment and base re-calibration performed using GATK (Broad Institute). Finally, variants were called with multiple programs: Samtools (Genome Research Ltd, UK), MuTect and GATK unified genotyper (Broad Institute), Pindel and VarScan (McDonnell Genome Institute, St. Louis, MO); the resulting lists were combined using an in-house developed program and variants annotated using ANNOVAR (University of Southern California) and an in-house developed program, as described previously[2]. In brief, the algorithm determined the variant allele frequency (VAF), local read statistics and mutation likelihood for each variant. In addition, recurrence of the variant is given from the catalogue of somatic mutations in cancer (COSMIC)[3] and from population-based sequencing studies (1000 genomes, ESP6500 and ExAC)[4, 5]. Variants were further annotated for entries in dbSNP137, 138 and 144 databases, and for the predicted pathogenicity using PolyPhen[6], SIFT[7-10] and MutationTaster[11]. Variant lists were further minimally scrutinized using the Integrative Genomics Viewer (Broad Institute) for artifact exclusion, and the remaining variants filtered using a VAF threshold at 2% as well as for coverage at the mutation location and systematic bias in the strand. Mutations were selected when at least 10 high-quality reads were present at the variant site, the variant was not present in all population-based sequencing data and the variant function was found to be deleterious by at least 2 independent prediction

programs. All data for each variant are provided in Supplementary Table 2. A graphical summary of the described data analysis pipeline is shown in Supplementary Figure 1. WES data analysis was performed as described previously[1].

Legend Supplementary Figure 1

MiSeq data analysis workflow. Schematic representation of the major data analysis steps (from top to bottom), as outlined in the Supplementary methods section and in Gröschel *et al*[2].

Supplementary Figure 1



Supplementary Table 1: Clinical characteristics of the total cohort

Number of patients (%)	38 (100)
Gender	
Male (%)	23 (61)
Female (%)	13 (33)
MDS type	
Primary (%)	24 (65)
Secondary (%)	14 (34)
MDS subtype	
RC (%)	7 (19)
Advanced (%)	29 (76)
Other (%)	2 (5)
Median age at diagnosis (years)	10.2
Range	0.8 – 18.8
Karyotype	
Normal	9 (24)
Monosomy 7	8 (21)
Trisomy 8	1 (2)
Complex / other	17 (45)
unknown	3 (8)
Outcome	
Alive (%)	21 (55)
Deceased (%)	14 (37)
Median blast % bone marrow	10
Range	0-29
Median blast % peripheral blood	1
Range	0 – 18
Median hemoglobin (mmol)	5.5
Range	2.5 – 9.8
Median white blood cell count ($x10^9$)	3.6
Range	1.1 – 47.0
Mean platelet count	77
Range	3 - 215
Median MCV	99
Range	71 - 114
Treatment	
SCT (%)	32 (84)
Other (%)	3 (8)
Unknown (%)	3 (8)

Supplementary Table 2

patient number	Gene name	Chromosome	position	reference sequence	HQ Total read coverage	HQ Mutated read coverage	variant allele frequency	strand bias	nucleotide change	amino acid change	mutation type	pathway	COSMIC	ESP6500	1000genomes	ExAC	PolyPhen score	PolyPhen prediction	SIFT score	SIFT prediction	Mutation Taster score	Mutation Taster prediction
1	TP53	chr17	7576626	NM_001126116	1445	666	46	0.51	c.A629C	p.X210S	stoploss	other		0	0	0					0.758	D
	BCOR	chrX	39933244	NM_001123383	1428	31	2	0.50	c.C1355T	p.A452V	nsSNV	chromatin modification		0	0	0	0.994	D	0	D	1	D
2																						
3																						
4																						
5	STAG2	chrX	123184039	NM_001042749	159	31	20	0.50	c.G902insG	p.A300fs	frameshift insertion	cohesin complex		0	0	0					1	D
	GATA2	chr3	128205776	NM_001145662	57	10	18	0.50	c.T96insAGTT	p.N32_Y33insfsX	stopgain	transcription		0	0	0					1	D
6	ASXL1	chr20	31022441	NM_015338	2674	580	21	0.51	c.1927dupG	p.G642fs	frameshift insertion	chromatin modification	COSM1411076	0	0	0					1	D
	RUNX1	chr21	36231783	NM_001001890	3953	446	11	0.52	c.C520T	p.R174X	stopgain	transcription	COSM24771	0	0	0					1	D
7																						
8																						
9																						
10	WT1	chr11	32417907	NM_000378	939	212	23	0.52	c.1093_1094insTCGG	p.A365fs	frameshift insertion	transcription	COSM21392, COSM1166613	0	0	0					1	D
11	GATA2	chr3	128204640	NM_001145662	1078	570	53	0.57	c.801delC	p.P267fs	frameshift deletion	transcription		0	0	0						
	IKZF1	chr7	50444401	NM_001220765	1244	811	65	0.52	c.C331T	p.R111X	stopgain	transcription	COSM303891	0	0	0					1	D
	ASXL1	chr20	31022441	NM_015338	1606	484	30	0.50	c.1927dupG	p.G642fs	frameshift insertion	chromatin modification	COSM1411076	0	0	0					1	D
12																						
13	CSF3R	chr1	36932108	NM_000760	533	39	7	0.46	c.T2361G	p.Y787X	stopgain	signaling		0	0	0					1	D

patient number	Gene name	Chromosome	position	reference sequence	HQ Total read coverage	HQ Mutated read coverage	variant allele frequency	strand bias	nucleotide change	amino acid change	mutation type	pathway	COSMIC	ESP6500	1000genomes	ExAC	PolyPhen score	PolyPhen prediction	SIFT score	SIFT prediction	Mutation Taster score	Mutation Taster prediction
14	CUX1	chr7	101882719	NM_001202543	1534	741	48	0.51	c.C3775T	p.R1259W	nsSNV	transcription	COSM1083478	0	0	0	1	D	0	D	1	D
	PTPN11	chr12	112888210	NM_002834	1921	195	10	0.50	c.G226C	p.E76Q	nsSNV	signaling	COSM13016	0	0	0	0.967	D	0	D	1	D
	PHF6	chrX	133549124	NM_001015877	577	525	91	0.50	c.C808T	p.Q270X	stopgain	transcription	COSM144566	0	0	0					1	D
15																						
16																						
17	DNMT3A	chr2	25470596	NM_153759	109	6	6	0.50	c.G311A	p.G104E	nsSNV	DNA methylation		0	0	0	0.999	D	0	D	1	D
18																						
19																						
20	PTPN11	chr12	112888199	NM_002834	3566	1304	37	0.50	c.C215T	p.A72V	nsSNV	signaling	COSM13015	0	0	0	0.819	P	0	D	1	D
21	BCOR	chrX	39933565	NM_001123383	1012	1006	99	0.49	c.T1034A	p.L345H	nsSNV	chromatin modification		0	0	0	0.999	D	0	D	1	D
22																						
23	TP53	chr17	7578263	NM_001126116	641	599	93	0.52	c.C190T	p.R64X	stopgain	other	COSM99665, COSM1640847, COSM99666, COSM99668, COSM10705, COSM99667	0	0	0					1	D
24																						
25	BCOR	chrX	39930359	NM_001123383	76	16	21	0.50	c.C3051A	p.H1017Q	nsSNV	chromatin modification		0	0	0	0.992	D	0.01	D	1	D
	ATRX	chrX	76778870	NM_138270	169	10	6	0.53	c.C6595A	p.L2199I	nsSNV	chromatin modification		0	0	0	0.996	D	0.02	D	1	D
26																						
27	RUNX1	chr21	36252880	NM_001001890	590	315	53	0.50	c.T401C	p.L134P	nsSNV	transcription	COSM444417	0	0	0	1	D	0	D	1	D
28	CBL	chr11	119149328	NM_005188	688	16	2	0.50	c.G1336A	p.G446R	nsSNV	signaling		0	0	0	0.944	D	0.01	D	1	D
29																						

patient number	Gene name	Chromosome	position	reference sequence	HQ Total read coverage	HQ Mutated read coverage	variant allele frequency	strand bias	nucleotide change	amino acid change	mutation type	pathway	COSMIC	ESP6500	1000genomes	ExAC	PolyPhen score	PolyPhen prediction	SIFT score	SIFT prediction	Mutation Taster score	Mutation Taster prediction
30																						
31																						
32	RUNX1	chr21	36231872	NM_001001890	1710	1074	63	0.53	c.A431T	p.K144I	nsSNV	transcription		0	0	0	0.997	D	0	D	1	D
	EZH2	chr7	148526829	NM_152998	550	185	34	0.50	c.G358A	p.G120R	nsSNV	chromatin modification	COSM1449032, COSM96480	0	0	0	1	D	0	D	1	D
33																						
34																						
35	IDH2	chr15	90631935	NM_002168	1498	209	14	0.51	c.C418T	p.R140W	nsSNV	DNA methylation	COSM41877	0	0	0	1	D	0	D	1	D
36																						
37	CDKN2A	chr9	21974825	NM_000077	93	60	65	0.45	c.T2A	p.M1K	nsSNV	other		0	0	0	0.722	P	0	D	1	D
	TP53	chr17	7579427	NM_001126116	327	11	4	0.52	c.C143A	p.P48Q	nsSNV	other	COSM43544	0	0	0	1	D	0.02	D	0.996	D
38																						

Legend Supplementary Table 1:

Overview of the clinical characteristics. Included are the patient numbers and percentage (of total, in brackets) for the gender, MDS type (primary or secondary MDS), the MDS subtype (RC: refractory cytopenia, advanced: RAEB (refractory anemia with excess blasts) and RAEB-t (RAEB in transformation), and other (including RARS (refractory anemia with ring sideroblasts), MDS-PNH (non-hemolytic MDS with a large (60%) paroxysmal nocturnal hemoglobinuria clone, previously described by van den Heuvel-Eibrink *et al*[12]) and t-MDS (therapy-related MDS)), the median age at diagnosis in years, cytogenetics (complex: 2 or more cytogenetic abnormalities found, other: single cytogenetic abnormality other than monosomy 7 or trisomy 8), outcome, blood count details (median hemoglobin levels, white blood cells (WBC) counts, platelet counts, blast percentages in bone marrow (BM) and peripheral blood (PB), and the mean corpuscular volume (MCV)), treatment (IST: immunosuppressive therapy, SCT: hematopoietic stem cell transplantation, chemo: chemotherapy) and treatment.

Legend supplementary Table 2:

Overview of the mutational status in each patient. Details on patient number, gene name, chromosome position, reference sequence, nucleotide and amino acid changes, mutation type (nsSNV: nonsynonymous single nucleotide variation), pathway attribution, total and mutated high quality (HQ) read counts, VAF, strand bias, frequency in population-based sequencing efforts (ESP6500, 1000 genomes, ExAC), COSMIC entry as well as PolyPhen, SIFT and MutationTaster scores and prediction are provided (D: damaging, P: probably damaging, B:benign).

References:

1. Beekman R, Valkhof MG, Sanders MA, et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood. 2012;119(22):5071-5077.
2. Groschel S, Sanders MA, Hoogenboezem R, et al. Mutational spectrum of myeloid malignancies with inv(3)/t(3;3) reveals a predominant involvement of RAS/RTK signaling pathways. Blood. 2015;125(1):133-139.
3. Forbes SA, Bhamra G, Bamford S, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet. 2008;Chapter 10(Unit 10 11.
4. Exome Variant Server NGESPE, Seattle, WA. [cited 2016 April]; Available from: http://evs.gs.washington.edu/EVS/
5. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75-81.
6. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. 2010;7(4):248-249.
7. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-1081.
8. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11(5):863-874.
9. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res. 2002;12(3):436-446.
10. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-3814.
11. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-362.
12. van den Heuvel-Eibrink MM, Bredius RG, te Winkel ML, et al. Childhood paroxysmal nocturnal haemoglobinuria (PNH), a report of 11 cases in the Netherlands. Br J Haematol. 2005;128(4):571-577.