High-throughput mutational screening adds clinically important information in myelodysplastic syndromes and secondary or therapy-related acute myeloid leukemia

Mohsen Karimi,[1] Christer Nilsson,[1] Marios Dimitriou,[1] Monika Jansson,[1] Hans Matsson,[2] Per Unneberg,[3] Sören Lehmann,[1] Juha Kere,[2,4] and Eva Hellström-Lindberg[1]

[1]Center for Hematology and Regenerative Medicine, Department of Medicine Huddinge, Karolinska Institutet, Sweden; [2]Department of Biosciences and Nutrition and Center for Innovative (CIMED), Karolinska Institutet, Sweden; [3]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm, Sweden; [4]Molecular Neurology Research Program, University of Helsinki and Folkhälsan Institute of Genetics, Helsinki, Finland

Correspondence: Eva.Hellstrom-Lindberg@ki.se
doi:10.3324/haematol.2014.118034

Supplementary material

Supplementary methodological description

Targeted sequencing

BM mononuclear cells were isolated using Lymphoprep (Axis-Shield, Oslo, Norway) density gradient centrifugation according to standard protocol and genomic DNA were extracted using GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich AB, Sweden) or DNAeasy Kit (Qiagen, Qiagen AB, Sweden). Halogenomics target amplification technology (Halogenomics AB, Uppsala, Sweden) was applied to amplify all exons of 22 selected recurrently mutated genes in MDS and AML (design details are presented in Online Supplementary Table S2). To analyze samples in a high-throughput and cost effective way we used a sample pooling approach in which 10 samples were pooled and targeted regions were amplified according to manufacturer´s protocol. A total of 22 pools, 20 with patient samples and 2 with control DNA, were defined by 6 bp barcoding and samples were sequenced in 2 Illumina Hiseq2000 sequencing system using the 100 bp paired end protocol at Science for Life Laboratory facilities (Stockholm, Sweden).

Bioinformatic analysis

Sequencing data were filtered based on the barcodes of each pool and aligned to the human genome reference hg19 with BWA version 0.5.9-r16,[1] and sequence alignment/map (sam) files were converted to their binary format (bam) using SAMtools version 0.1.18.[2] Mapped reads were sorted and duplicates marked with Picard version 1.41 (http://picard.sourceforge.net), generating analysis-ready bam files. SNP and indel detection in pooled sequences was performed with CRISP version 5.[3] Only variants covered by a minimum number of 4 reads

were considered for variant calling, and reads with more than 4 mismatches were disregarded. Finally, variants were annotated with ANNOVAR version 2011-11-28.[4]

Non-synonymous variants in protein coding regions found in our pooled sample analysis (234 in total) were selected and analyzed by Sequenom technology. In addition, we analyzed hotspot mutations in 3 splicing factor genes, *SF3B1*, *SRSF2* and *U2AF1* (Online Supplementary Table S2).

Sequenom analysis

Sequenom® is a clinically accredited genotyping system for analyzing point mutations and SNPs in high-throughput manner.[5] To screen for mutations we analyzed each individual patient for all of the variants found by Halogenomic sequencing. Sequenom analysis was performed according to standard procedure[5] at Mutation Analysis Facility core facility (Karolinska Institutet, Stockholm, Sweden).

Statistical analysis:

Categorical data were compared using Fisher's exact test. Survival was estimated using Kaplan-Meier method and compared using the log-rank test. The Cox proportional hazards model was used for uni- and multivariable analyses of survival. Variables included in the model for MDS were the mutations found to be significant in the univariable analyses (Online Supplementary Table S6) together with age, sex and IPSS-R classification. In the corresponding multivariable model for AML, IPSS-R was replaced by the cytogenetic risk group.[6] Two sided p-values with a significance level of 0.05 were used in all analyses. All analyses were performed using the R statistical computing software.[7]

Refrences.

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95

2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.

3. Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics. 2010 Jun 15;26(12):i318-24.

4. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010 Sep;38(16):e164.

5. Millis MP. Medium-throughput SNP genotyping using mass spectrometry: multiplex SNP genotyping using the iPLEX(R) Gold assay. Methods in molecular biology. 2011;700:61-76.

6. Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010 Jul 22;116(3):354-65.

7. R Core Team. R: A language and environment for statistical computing. 3.0.3 ed. Vienna, Austria: R Foundation for Statistical Computing, 2014.

Supplementary Table S1. Clinical characteristics of Patients

MDS (n = 100)		n	(%)
Age			
	Median	72	
	Range	32-88	
Gender			
	Female	49	(49%)
WHO classification			
	RS*	23	(23%)
	5q-	19	(19%)
	RCMD	2	(2%)
	RAEB I	17	(17%)
	RAEB II	24	(24%)
	CMML	15	(15%)
IPSS-R risk classification			
	Very low	6	(6%)
	Low	33	(33%)
	Intermediate	19	(19%)
	High	21	(21%)
	Very high	8	(8%)
	Not classified	13	(13%)

AML (n = 92)			
Age			
	Median	71	
	Range	27-88	
Gender			
	Female	41	(45%)
AML subtypes			
	Following MDS	28	(30%)
	MDS-like cytogenetics**	38	(41%)
	Following MPN***	11	(12%)
	Therapy-related	15	(16%)
Cytogenetic risk****			
	Low	5	(5%)
	Intermediate	35	(38%)
	High	52	(57%)

* 11 RCMD-RS, 10 RARS, 2 RARS-T

** AML with myelodysplasia-related changes according to WHO 2008

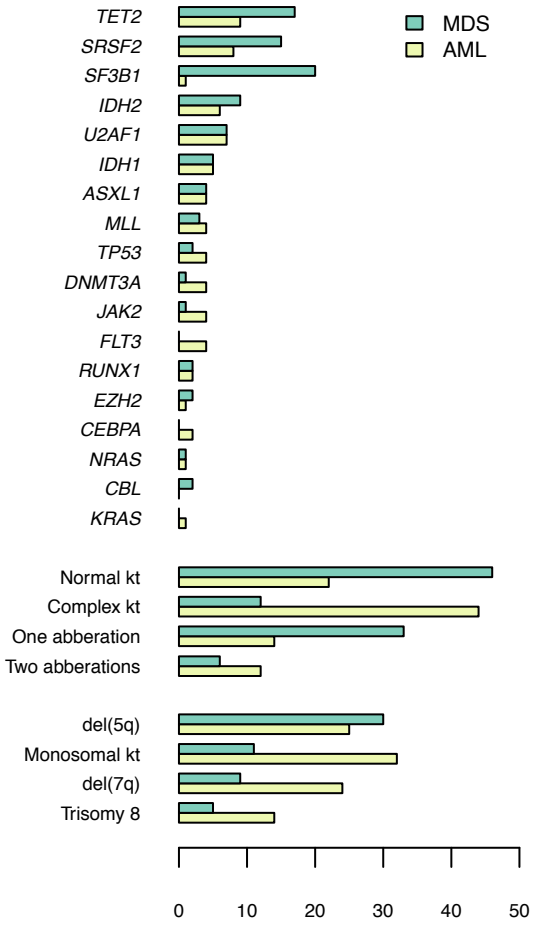
*** 5 polycythemia vera, 2 essential thrombocythemia, 1 PCV/ET, 1 myelofibrosis, 2 unspecified MPN

**** According to the MRC criteria[14]

Supplementary Table S2. Hazard ratios for death among MDS patients with mutation present vs mutation absent

		Univariable		Adjusted for gender, age and IPSS-R			
	N	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
SF3B1	20	0.39 (0.18 - 0.85)	0.018 *	0.34 (0.14 - 0.82)	0.016 *	0.30 (0.13 - 0.73)	0.008 *
SRSF2	15	2.26 (1.14 - 4.45)	0.019 *	1.65 (0.82 - 3.32)	0.161	2.25 (1.00 - 5.06)	0.050 *
IDH2	9	3.85 (1.71 - 8.64)	0.001 *	2.22 (0.92 - 5.37)	0.077	3.31 (1.28 - 8.56)	0.013 *
U2AF1	7	3 (1.26 - 7.12)	0.013 *	2.78 (1.07 - 7.25)	0.036 *	4.78 (1.71 - 13.38)	0.003 *
RUNX1	2	4.75 (1.12 - 20.18)	0.035 *	3.08 (0.69 - 13.85)	0.142	2.63 (0.52 - 13.3)	0.242

Supplementary Figure S3 Frequency of mutations and cytogenetic aberrations in MDS (n=100) and AML (n=92). Abbreviations: KT = karyotype.



Supplementary Figure S4. Comparison of overall survival in years between mutated vs wildtype splice factor genes in MDS and AML.