Tumor Biol. (2016) 37:13391–13401
DOI 10.1007/s13277-016-5142-7

ORIGINAL ARTICLE

Parallel targeted next generation sequencing of childhood and adult acute myeloid leukemia patients reveals uniform genomic profile of the disease

Irena Marjanovic[1] · Jelena Kostic[1] · Bojana Stanic[1] · Nadja Pejanovic[1] · Bojana Lucic[1] · Teodora Karan-Djurasevic[1] · Dragana Janic[2,3] · Lidija Dokmanovic[2,3] · Srdja Jankovic[2] · Nada Suvajdzic Vukovic[3,4] · Dragica Tomin[3,4] · Ognjen Perisic[5] · Goran Rakocevic[5] · Milos Popovic[5] · Sonja Pavlovic[1] · Natasa Tosic[1]

Received: 29 January 2016 / Accepted: 11 July 2016 / Published online: 27 July 2016

Abstract The age-specific differences in the genetic mechanisms of myeloid leukemogenesis have been observed and studied previously. However, NGS technology has provided a possibility to obtain a large amount of mutation data. We analyzed DNA samples from 20 childhood (cAML) and 20 adult AML (aAML) patients, using NGS targeted sequencing. The average coverage of high-quality sequences was 2981 × per amplicon. A total of 412 (207 cAML, 205 aAML) variants in the coding regions were detected; out of which, only 122 (62 cAML and 60 aAML) were potentially protein-changing. Our results confirmed that AML contains small number of genetic alterations (median 3 mutations/patient in both groups). The prevalence of the most frequent single gene AML associated mutations differed in cAML and aAML patient cohorts: *IDH1* (0 % cAML, 5 % aAML), *IDH2* (0 % cAML, 10 % aAML), *NPM1* (10 % cAML, 35 % aAML). Additionally, potentially protein-changing variants were found in tyrosine kinase genes or genes encoding tyrosine kinase associated proteins (*JAK3*, *ABL1*, *GNAQ*, and *EGFR*) in cAML, while among aAML, the prevalence is directed towards variants in the methylation and histone modifying genes (*IDH1*, *IDH2*, and *SMARCB1*). Besides uniform genomic profile of AML, specific genetic characteristic was exclusively detected in cAML and aAML.

Keywords Childhood AML · Adult AML · Next generation sequencing · SNVs

Introduction

Acute myeloid leukemia (AML) is a genetic disease of somatic cells that mostly affects middle aged and elderly population [1]. Development of AML represents unlimited proliferation and impaired differentiation of early myeloid cells leading to accumulation of immature blast cells in the bone marrow and peripheral blood, thus resulting in hematopoietic failure. The incidence of AML increases with age. In children, AML is present in 15–20 % of cases, while in adults, the representation of this type of leukemia is up to 80 % of all acute leukemia [2]. Also, the distribution of cytogenetic aberrations and molecular characteristics are different in children and adults. The incidence of abnormal karyotypes is, in general, lower in adult than in pediatric de novo AML (55 vs. 76 %), and the incidence of specific recurrent aberration varies between these two groups. While the incidence of t(15;17) and inv(16) is similar in adults and children, t(8;21) is two times more frequent in pediatric AML. On the other hand, some rare recurrent aberrations were found only in childhood AML, but not among adult patients [3].

Even greater differences were observed in terms of the molecular characteristics of children's and adult AML.

Electronic supplementary material The online version of this article (doi:10.1007/s13277-016-5142-7) contains supplementary material, which is available to authorized users.

✉ Natasa Tosic
nmtosic@sezampro.rs; natasa.tosic@imgge.bg.ac.rs

[1] Laboratory for Molecular Biomedicine, Institute for Molecular Genetica and Genetic Engineering University of Belgrade, Vojvode Stepe 444a, Belgrade 11010, Serbia

[2] University Children's Hospital, Belgrade, Serbia

[3] School of Medicine University of Belgrade, Belgrade, Serbia

[4] Clinic of Hematology, Clinical Center of Serbia, Belgrade, Serbia

[5] Seven Bridges Genomics, Belgrade, Serbia

Mutations in exon 12 of *NPM1* gene are the most frequent mutations found in AML, occurring in 30 % of adult AML and more than 50 % in AML-NK patients [4]. *NPM1* mutations are distinctly less common in childhood AML ranging from 0 to 6.5 %. Internal tandem duplication of *FLT3* gene (*FLT3/ITD*) is present in approximately 25 % of adult AML while the frequency among children is much lower, approximately 5–15 % [5, 6]. This suggests that there are significant differences in the molecular mechanisms in leukemogenesis in children compared to adults.

Development of various approaches in next generation sequencing (NGS) has provided more comprehensive insight into the origin and evolution of carcinogenesis including AML pathogenesis. NGS technology has been applied in several studies enabling better understanding of mutational profiling that underline AML pathology. The first sequencing of primary AML genome was reported by Ley TJ and coworkers in 2008 using whole genome sequencing approach [7]. Subsequently, the sequence of additional AML genomes have been reported [8–10], among others a study from Mardis et al., in which mutations in isocitrate dehydrogenase 1 (*IDH1*) gene was reported for the first time in AML-NK patients [9]. Further analysis of a large number of AML-NK patients, treated with different clinical protocols, has confirmed that *IDH* mutations (*IDH1* and *IDH2*) occur in approximately 20 %, placing them among the most common molecular aberrations in AML-NK [11]. It is interesting that the frequency of these mutations are also age-specific, with the highest frequency in adult AML-NK patients and lowest in pediatric patients, from 0 to 9.8 % in AML-NK patients [12, 13].

In this study, we applied targeted next generation sequencing and MiSeq System for analyzing somatic mutations in groups of adult (aAML) and childhood (cAML) AML patients, in order to facilitate recognition and better understanding of the genetic profile of the disease.

Material and methods

Patients

Bone marrow samples from the 20 adult and 20 childhood AML patients at diagnosis were collected. Adult AML patients came from the Clinic of Hematology, Clinical Center of Serbia, and childhood patients came from the Department of Hematology, University Children Hospital in Belgrade. The study was approved by the Ethics Committee of the Clinical Center of Serbia. Mononuclear cells were separated by Ficoll density gradient centrifugation and cryopreserved until mutational analyses. Some clinical characteristics of the patients are listed in Supplementary Tables 1 and 2.

TruSeq Amplicon – Cancer Panel library preparation and sequencing

TruSeq Amplicon – Cancer Panel, TSACP (Illumina Inc., San Diego, CA, USA) targets mutational hotspots in 48 cancer-related genes. TSACP consists of 212 amplicons captured by pairs of oligonucleotides designed to hybridize flanking targeted regions of interest. Genomic DNA from mononuclear cells of 20 cAML and 20 aAML patients and three healthy controls were extracted using the Qiagen Blood Mini Kit (Valencia, CA, USA). The library preparation was performed using 250 ng of genomic DNA, according to the manufacturer's protocol. The DNA was hybridized with upstream and downstream oligonucleotides (oligos) specific for target regions of interest, followed by the removal of unbound oligos and the extension and ligation step. PCR amplification was carried out using specific primers that add index sequences for sample multiplexing (i5 and i7) as well as common adapters for cluster generation (P5 and P7). The PCR products were purified using AMPure XP beads. The equal volumes of normalized libraries were pooled and prepared for subsequent cluster generation and sequencing on the MiSeq system (Illumina Inc., San Diego, CA, USA). Paired-end sequencing was performed using the MiSeq Reagent Kit v3 (600-cycle), and the sequencing quality was demonstrated by the percentage of bases having the Q30 score (1 error in 1000 bases) of 97.2 %.

Bioinformatics analysis

FASTQ files produced upon library sequencing were processed in four stages: basic quality control and trimming, alignment and preprocessing, additional quality control, variant calling and filtration. The processing pipeline was assembled by Seven Bridges Genomics (SBG). It was composed of both freely available open source bioinformatics tools as well as tools developed in-house by SBG. The first processing step was composed of the basic quality control performed with FastQC [14] and the trimming of low quality bases (base quality <20) from read ends, which was performed with FastqMcf [15]. The alignment to the GRCh37 reference genome, which produced BAM file(s), was done with BWA-MEM [16–18]. The indel realignment over the reads overlapping target regions was performed with the RealignerTargetCreator and IndelRealigner tools from GATK [19, 20]. The additional quality control was done using custom scripts developed by SBG. The scripts counted reads in each amplicon and identified amplicons with systematically low read-coverage across all samples. The variant calling and filtration was carried out with the GATK UnifiedGenotyper and VariantFiltration tools [19, 20]. UnifiedGenotyper produced a VCF file containing single nucleotide variants (SNVs) and indel variants in relation to the GRCh37 reference genome by applying a Bayesian approach. As the final step, the Variant Filtration tool

Table 1 Mutations identified in childhood acute myeloid leukemia using NGS, Sanger sequencing, and PCR

Sample no.	Mutation detected by MySeq	Mutation status	dbSNP	COSMIC	Mutation detected by PCR/ Sanger	Coverage (x)
1	None					
2	KDR, c.1416A>T, p.Q472H	Homozygous	rs1870377	COSM149673		10,894
	GNAQ, c.842A>G, p.E281G	Heterozygous				82
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		62
3	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		252
	JAK3, c.2164G>A, p.V722I	Heterozygous	rs3213409	COSM34213		236
4	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		66
	KIT, c.2447A>T, p.D816V	Heterozygous	rs121913507	COSM1314		3356
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		2892
	FLT3, c.2522A>T, p.N841I	Heterozygous				705
5	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		60
	NRAS, c.182A>T, p.Q61L	Heterozygous	rs11554290	COSM583		979
6	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		11,062
	ATM, c.2572T>C, p.F858L	Heterozygous	rs1800056	COSM21826		2249
	FLT3, c.2028C>A, p.N676K	Heterozygous		COSM303886		2479
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		184
7	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		198
	NRAS, c.182A>T, p.Q61L	Heterozygous	rs11554290	COSM583		1486
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		3806
8	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		413
	KDR, c.1416A>T, p.Q472H	Homozygous	rs1870377	COSM149673		3411
	NRAS, c.35G>A, p.G12D	Heterozygous	rs121913237	COSM564		3807
	NRAS, c.34G>A, p.G12S	Heterozygous	rs121913250	COSM563		3431
	KIT, c.1621A>C, p.M541L	Heterozygous	rs3822214	COSM28026		2871
9	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		4443
10	EGFR, c.2368. A>T, p.T790S	Heterozygous				36
	RET, c.2770T>C, p.F924L	Heterozygous				60
	TP53, c.442G>A, p.D148N	Heterozygous		COSM44043		78
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		65
11	NPM1, c.863_864insCGGA P.W288fs$^{\#}$12	Heterozygous		COSM28066		2182
	NPM1, c.871G>T,p.R291M	Heterozygous				2182
	NRAS, c.34G>A, p.G12S	Heterozygous	rs121913250	COSM563		557
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		12,063
12	NRAS, c.34G>A, p.G12S	Heterozygous	rs121913250	COSM563		387
	KDR, c.1416A>T, p.Q472H	Homozygous	rs1870377	COSM149673		14,782
	SMO, c.984C>T, p.A235V	Heterozygous	rs142599757	COSM1226876		216
	ABL1, c.806C>T, p.A269V	Heterozygous				125
	KRAS, c.35G>A, p.G12D	Heterozygous	rs121913529	COSM521		3552
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		386
	JAK3, c.2164G>A, p.V722I	Heterozygous	rs3213409	COSM3413		579
13	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		10,653
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		218
					FLT3/ITD	
14	EGFR, c.2437T>G, p.Y813D	Heterozygous				30
	MET, c.3029C>T, p.T1010I	Heterozygous	rs56391007	COSM707		8695
	SMO, c.618G>T, p.W206L	Heterozygous				42

Table 1 (continued)

Sample no.	Mutation detected by MySeq	Mutation status	dbSNP	COSMIC	Mutation detected by PCR/ Sanger	Coverage (x)
	PTPN11, c.206A>T, p.E76V	Heterozygous	rs121918465	COSM13025		7569
	TP53, c.308A>T, p.Y103F	Heterozygous				54
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		75
15	KRAS, c.38G>A, p.G13D	Heterozygous		COSM532		978
	PTPN11, c.179G>T, p.G60V	Heterozygous		COSM13028		1929
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		88
16	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		343
					FLT3/ITD	
17	KDR, c.1416A>T, p.Q472H	Homozygous	rs1870377	COSM149673		14,429
	MET, c.1124A>G, p.N375S	Heterozygous	rs33917957	COSM5020653		6643
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		36
18	FLT3, c.2503G>T, p.D835Y	Heterozygous		COSM783		299
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		42
19	MET, c.1124A>G, p.N375S	Heterozygous	rs33917957	COSM5020653		12,193
	KRAS c.37G>T, p.G13C	Heterozygous	rs121913535	COSM527		1175
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		142
20	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		10,697
	ATM, c.7357C>T, p.R2453C	Heterozygous		COSM1351001		4009
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		326

was used to filter out low quality variants from the VCF file. Next, for variant annotation, we used Ensembl Variant Effect Predictor (VEP) [21]. Finally, a report was generated that summarizes per sample the results for all amplicons, including the sequence depth and, if present, the called mutation (both on DNA and protein level) and dbSNP identifier. The Integrated Genomics Viewer [22] was used for visual evaluation of the data.

Polymerase chain reaction and Sanger sequencing

The detection of *FLT3/ITD* PCR performed as previously described [23]. For the detection of *NPM1*, *IDH1*, and *IDH2* mutations, the samples were directly sequenced with Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on 3130 Genetic Analyzer (Applied Biosystems) [23].

Results

To detect mutational profile involved in the pathogenesis of childhood and adult AML, we analyzed approximately 14×10^8 bp sequence from 20 cAML and 20 aAML patients by targeted NGS using TSACP that covers 212 hotspot regions from 48 genes. The average coverage of high-quality sequences was 2981× per amplicon. Nine genes were discarded due to insufficient coverage; therefore, a total of 188 amplicons from 39 genes was used for subsequent analysis. Since matched germline material was unavailable, variants were identified in relation to the reference genome by applying a Bayesian approach and compared to public genetic variation databases and in-house databases. A number of different variants detected in both coding and non-coding targeted regions was 114, out of which 75 (55 cAML, 42 aAML) variants were in the coding regions and 39 (31 cAML, 34 aAML) outside of the targeted exons. Out of them, we identified four different indels (3 cAML, 3 aAML) in the coding regions and 13 different indels (9 cAML, 13 aAML) in the non-coding regions, whereas 71 different SNVs (52 cAML, 39 aAML) were identified in the coding and 26 different SNVs (22 cAML, 21 aAML) in the non-coding regions (Fig. 1).

In our cohort of patients, we identified a total of 412 (207 cAML, 205 aAML) variants in the coding regions (median per patient: 10, range: 5–16; median per cAML: 10, range: 5–14; median per aAML: 10, range: 7–16) and 527 (260 cAML, 267 aAML) variants in the non-coding regions (median per patient: 14, range: 8–17; median per cAML: 14, range: 8–16; median per aAML: 14, range: 9–17). Only potentially protein-

Table 2 Mutations identified in adult acute myeloid leukemia using NGS, Sanger sequencing, and PCR

Sample no.	Mutation detected by MySeq	Mutation status	dbSNP	COSMIC	Mutation detected by PCR/Sanger	Coverage (x)
1	NRAS,c.38G>A,p.G13D	Heterozygous	rs121434596	COSM573		1159
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		22,343
	PTPN11,c.1508G>A,p.G503E	Heterozygous		COSM13021		1140
	HNF1A, c.863_864insC p.P289fs#28	Heterozygous		COSM1476243		2306
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		166
	FLT3, c.2503G>T, p.D835Y	Heterozygous		COSM783	FLT3, c.2503G>T, p.D835Y	358
					IDH2,c.419G>A, p.R140Q	
2	SMO, c.704C>T, p.A235V	Heterozygous	rs142599757	COSM1226876		170
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		184
3	KIT, c.1621A>C, p.M541L	Heterozygous	rs3822214	COSM28026		2585
	RET, c.2975A>G, p.T929A	Heterozygous				151
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		252
	GNAS, c.631A>T, p.K211stop	Heterozygous				113
4	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		12,563
	NPM1, c.859_860insTCTG p.W288fs#12	Heterozygous	rs758959453	COSM158604	NPM1, c.859_860insTCTG p.W288fs#12	2009
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		58
					FLT3/ITD	
5	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		11,343
	FLT3, c.2522A>C, p.N841T	Heterozygous				652
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		198
	NPM1, c.859_860insTCTG p.W288fs#12	Heterozygous	rs758959453	COSM158604	NPM1, c.859_860insTCTG p.W288fs#12	1836
6	NPM1, c.859_860insTCTG p.W288fs#12	Heterozygous	rs758959453	COSM158604		2671
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		247
7	APC, c.3920T>A, p.I1307K	Heterozygous	rs1801155			6274
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		89
8	IDH1, c.394C>A, p.R132S	Heterozygous	rs121913499	COSM28748		1575
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		15,785
	NPM1, c.859_860insTCTG p.W288fs#12	Heterozygous	rs758959453	COSM158604		1592
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		109
9	ATM, c.3925G>A, p.A1309T	Heterozygous	rs149711770	COSM22507		1988
	FLT3, c.2027A>G, p.N676S	Heterozygous				906
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		102
10	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		12,686
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		98
11	NPM1, c.859_860insTCTG p.W288fs#12	Heterozygous	rs758959453	COSM158604		1782
					FLT3/ITD	
12	KIT, c.2446G>T, p.D816Y	Heterozygous	rs28933969	COSM1310		3329
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		142
	SMARCB1, c.622A>T, p.M208L	Heterozygous				160
13	KRAS, c.437C>T, p.A146V	Heterozygous		COSM19900		1230
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		59
14	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		9183
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		70
15	KIT, c.1621A>C, p.M541L	Heterozygous	rs3822214	COSM28026		6057
	KDR, c.1416A>T, p.Q472H	Heterozygous				8539

Table 2 (continued)

Sample no.	Mutation detected by MySeq	Mutation status	dbSNP	COSMIC	Mutation detected by PCR/Sanger	Coverage (x)
	HNF1A, c.863_864insC p.P289fs$^{\#}$28	Heterozygous		COSM1476243		100
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		68
	RET, c.2372A>T, p.Y791F	Heterozygous	rs77724903	COSM1159820		40
16	KIT, c.2446G>T, p.D816Y	Heterozygous	rs121913507	COSM1314		4710
	KDR, c.1416A>T, p.Q472H	Homozygous	rs1870377	COSM149673		5500
	NPM1, c.859_860insTCTG p.W288fs$^{\#}$12	Heterozygous	rs758959453	COSM158604		1978
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		186
17	NRAS,c.38G>A,p.G13D	Heterozygous	rs121434596	COSM564		439
	NRAS, c.37G>T, p.G13C	Heterozygous	rs121434595	COSM570		1160
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		14,553
	NPM1, c.859_860insTCTG p.W288fs$^{\#}$12	Heterozygous	rs758959453	COSM158604		2562
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		102
18	MET, c.1124A>G, p.N375S	Heterozygous	rs33917957	COSM5020653		4403
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		126
					IDH2,c.515G>A, p.R172K	
19	NRAS, c.35G>A, p.G12D	Heterozygous	rs121913237	COSM564		539
	KDR, c.1416A>T, p.Q472H	Heterozygous	rs1870377	COSM149673		17,024
	MET, c.1124A>G, p.N375S	Homozygous	rs33917957			8726
	TP53, c.215C>G, p.P72R	Homozygous	rs1042522	COSM250061		186
20	NPM1, c.859_860insTCTG p.W288fs$^{\#}$12	Heterozygous	rs758959453	COSM158604		1991
	TP53, c.215C>G, p.P72R	Heterozygous	rs1042522	COSM250061		72

changing variants were considered for further analysis. A total of 122 variants, 62 cAML, and 60 aAML (median per patient: 3, range: 0–7; median per cAML: 3, range: 0–7; median per aAML: 3, range: 1–5) were potentially protein-changing, including nonsense, frameshift, and missense (NFM) mutations. Six patients had five or more NFM mutations, including three cAML (#8, #12, #14) and three aAML (#1, #15, #17) (Fig. 2). It is important to emphasize that, since we did not perform analyses on parallel "normal" tissue control samples of the patients, we cannot clearly define detected variants as somatic mutations. Nevertheless, TruSeq Amplicon Cancer Panel is designed to "cover" mutational hotspot in cancer genes and to detect somatic mutations. TSACP comprises oncogenes and tumor suppressor genes involved in cell proliferation, apoptosis, genome stability, and chromatin regulation. Our analysis reveals that 26 different genes had at least one mutation in the coding regions (23 in cAML, 23 in aAML), whereas 21 different genes had at least one NFM mutation (16 in cAML, 17 in aAML). Out of these, we identified four cAML-specific genes (*JAK3*, *ABL1*, *GNAQ*, and *EGFR*) and five genes containing NFM mutations only in aAML patients (*IDH1*, *APC*, *HNF1A*, *GNAS*, and *SMARCB1*). More than 10 NFM mutations were detected in targeted sequences of two genes, *KDR* and *TP53* (Fig. 3).

In six cAML patients and in four aAML patients, we detected 14 novel NFM mutations (9 in cAML and 5 in aAML patients) in 10 genes (Tables 1 and 2). The largest number of new mutations were detected in the gene *FLT3* with three, followed by *RET* and *EGFR* gene with two mutations. All of the mutations were substitution—missense type, and many of them are novel substitution in the codons which represent mutational "hot spots." For example, we detected mutation in *FLT3* gene c.2522 A>T; p.N841I, but substitutions in codon 841 have already been reported (N841K, N841H, N841S).

Additionally, we used PCR and Sanger sequencing in order to analyze common mutated genes in AML disease (*FLT3/ITD* and *IDH2*). Our study demonstrated that the prevalence of the most AML associated mutations, including *NPM1*, *IDH1*, and *IDH2* gene, differed in cAML and aAML patient cohorts: *IDH1* (0 % cAML, 5 % aAML), *IDH2* (0 % cAML, 10 % aAML), *NPM1* (10 % cAML, 35 % aAML). The list of different mutations detected by MiSeq using TruSeq

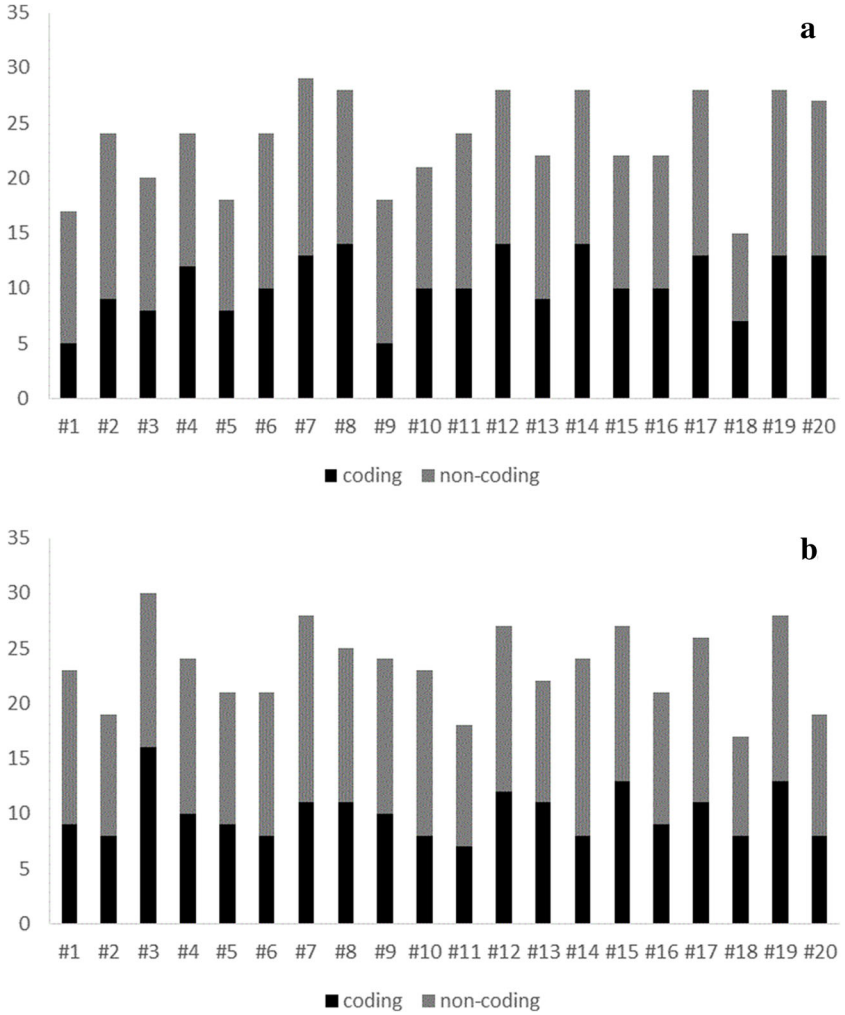

Fig. 1 Total number of variants in coding and non-coding regions identified by targeted NGS in cAML (**a**) and aAML (**b**)

Amplicon Cancer panel, PCR, and Sanger sequencing has been summarized in Tables 1 and 2.

Discussion

Acute myeloid leukemia represents a hematological malignancy with an excessive production of immature blood cells from myeloid lineage [1]. Recent advances in NGS methodology have provided deeper insight into mutational landscape of this complex disease, including both childhood and adult AML. Despite of this, contribution of mutational profile to AML is not completely defined and the knowledge related to epigenetic status underlying AML pathogenesis still remains poorly understood [24]. At the same time, new findings give better understanding of the AML disease and lead to discovery of novel prognostic markers and potential therapeutic strategies.

In this study, we have analyzed 40 AML patients (20 cAML and 20 aAML), using TSACP cancer panel for detection of somatic variants. Using this platform enabled us to gain insight into the mutational pattern of childhood and adult AML and to analyze the role of genes previously described primarily in solid tumors. Also, the use of targeted re-sequencing has enabled the high accuracy in the detection of the SNVs, which is reflected in the average coverage of 2981× per amplicon. As in the study by Luthra R et al. conducted on the same TSACP panel, our research also recognized the necessity for high accuracy mutation detection [25]. High coverage is required for detection of somatic mutations in the samples with large number of subclones, characteristic for malignancies, including AML. Using deep targeted sequencing, we detected 122 variants in the coding regions which were potentially protein-changing (62 cAML, 60 aAML).

Methodological approach used in our study is based primarily on the detection of potentially protein-changing mutations. In this way, our attention is focused on mutations which have significant role in the pathogenesis of leukemia, the so-called driver mutations that disrupt essential cellular processes

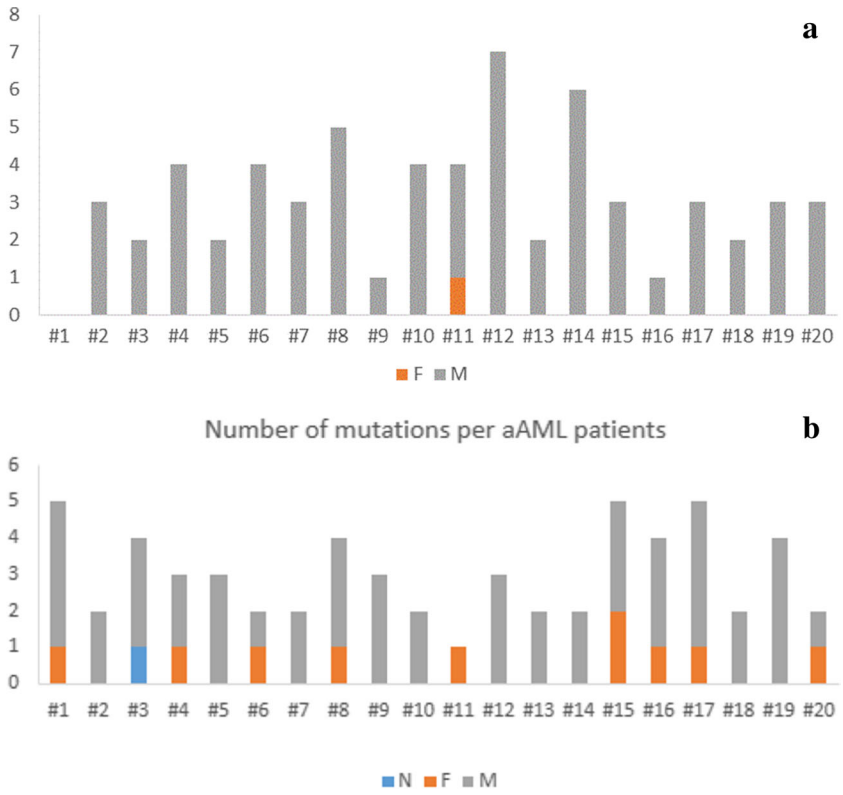

Fig. 2 Number of mutations per patients. Distribution of nonsense (*N*), frameshift (*F*), and missense (*M*) mutations in the coding regions of targeted genes in cAML (**a**) and aAML (**b**)

like proliferation, differentiation, cell cycle, and apoptosis. In other studies, different methodology was used (i.e., whole genome sequencing) enabling them to detect mutations in regions of the genome with regulatory potential. For example, in the study by Mardis et al., the mutations detected in these regions of the genome were found to be present in the dominant clone with moderate frequency (52 SNVs) [9]. Nevertheless, it was concluded that the majority of these mutations represent random, benign changes in the genome of hematopoietic cells. They exist in normal hematopoietic cells even before malignant transformation and are irrelevant for pathogenesis of AML [26].

The whole genome and whole exome sequencing done by the Cancer Genome Atlas Research group done on the 200 AML patients demonstrated an average of only five mutated genes per sample [27]. Comparing with the results obtained by analyses of other tumor tissues, it has been shown that AML has the smallest number of mutated genes. This was confirmed in our study where the average number of coding variants per patient was only 10 in both aAML and cAML, and among cAML in one patient (#1), we could not detect any of the mutations (Table 1). In another study where targeted sequencing analysis was done on the adult AML patients, mean number of mutated genes was only 2.56 [28]. As in our study, Kihara R et al. analyzed very heterogeneous population of patients regarding their cytogenetic finding, with the exception of the fact that our study included AML-M3 patients (both in cAML and aAML group of patients; Supplementary Tables 1 and 2). Indicative was the observation that the majority of mutations are generally detected among AML-NK patients. This all points to a great potential that recurrent cytogenetic aberrations, like translocations and inversions, have for the initiation of AML. The presence of aberrant transcription factor like PML-RARA (due to the presence of t(15;17) in AML-M3 patients) is a dominant driver aberration, sufficient for malignant transformation and onset of the disease.

Analyzing potentially protein-changing mutations (NFM mutations), we found that *TP53* and *KDR* genes contained one or more NFM mutations in over 50 % of AML patients. In a case of *TP53* gene, the shift in codon 72 (P72R), which is present in 17 of 20 in cAML and 19 of 20 in aAML cases, represents the most common *TP53* germline polymorphism, reported to be present in more than 70 % of European population [29]. In two cAML patients, additional *TP53* mutations were found, D148N and Y103F with potentially pathogenic effect. *TP53* was the first identified tumor suppressor gene responsible for maintenance of genome stability, and it is associated with various types of tumor, especially hematologic tumors [30]. In the matter of variants detected in *KDR* gene, all of them were Q472H type, found in 11 cAML and 10 aAML patients. The kinase domain receptor (*KDR*) gene encodes the vascular endothelial growth factor receptor-2 (VEGFR-2). It represents a transmembrane receptor important for angiogenesis and therefore can be important in solid tumor development and metastasis. There is also evidence of increased angiogenesis in acute myeloid leukemia, and

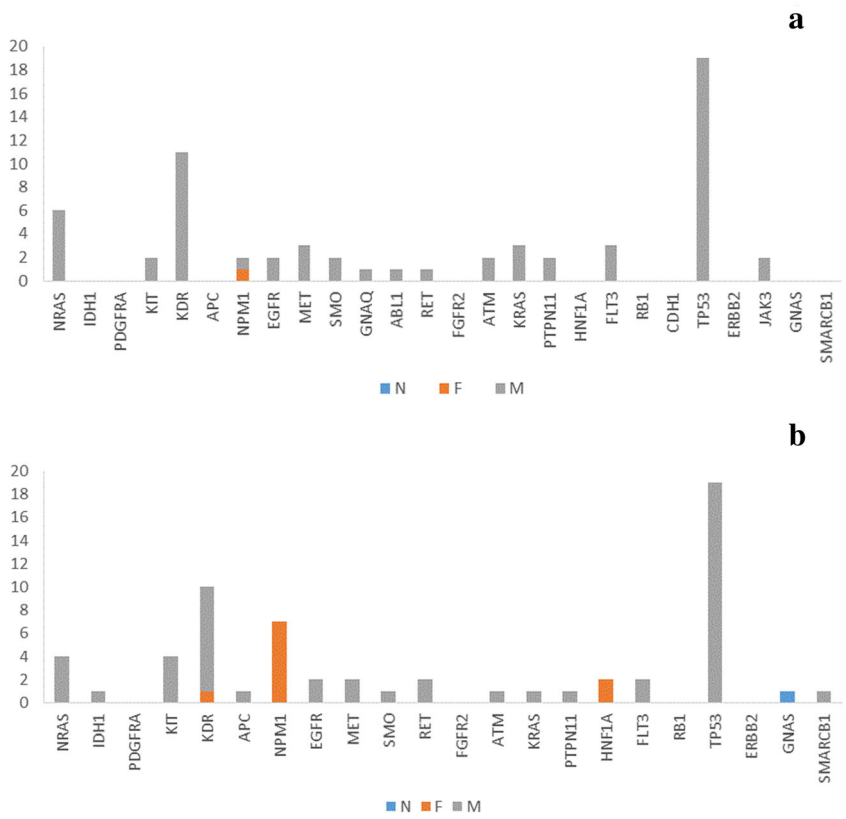

Fig. 3 Number of mutations per targeted genes. Distribution of nonsense (*N*), frameshift (*F*), and missense (*M*) mutations in the coding regions of targeted genes in cAML (**a**) and aAML (**b**)

therefore, it is possible that KDR plays a role in the pathophysiology of AML. It has been found that the presence of Q472H somatic mutation increases microvessel density in patients with lung cancer [31]. The effect of Q472H reflects on modest increase of KDR signaling through increased phosphorylation. In leukemia patients, the presence of Q472H mutation causes the failure of tyrosine kinase inhibitors therapy, like imatinib therapy failure in chronic myeloid leukemia [32].

Variants in the kinase domain resulting in the constitutive activation of tyrosine kinases and their downstream signaling pathways are the prevalent type of mutations in the various types of cancer, as is the case of AML [25, 33]. We have identified a large number of these mutations in both cAML and aAML, such as mutations in *NRAS*, *KRAS*, *RET*, and *FLT3* (Tables 1 and 2). The high prevalence of these mutations in both groups of patients suggests that they are driver mutations, important in the pathogenesis of AML. On the other hand, our study, using amplicon-based sequencing technology, has revealed mutations that exclusively occurred in one group of patients. Namely, NFM mutations in four genes were specific for cAML: *JAK3*, *ABL1*, *GNAQ*, and *EGFR* and five genes contained NFM mutations only in aAML patients: *IDH1*, *APC*, *HNF1A*, *GNAS*, and *SMARCB1*. It is noticeable that all four genes exclusively found in cAML are tyrosine kinase genes or genes encoding tyrosine kinase associated proteins, while among aAML, the prevalence is directed towards mutations in the methylation and histone modifying genes (*IDH1*, *IDH2*, and *SMARCB1*). In one of the early studies of AML genome, Mardis ER et al. reported mutations in *IDH1*, *IDH2*, and *TET2* genes for the first time [9]. After this finding, mutations in other chromatin remodeling genes (e.g., *DNMT3A*, *PBRM1*, *ASXL1*) were reported by others, suggesting that histone modifying proteins are heavily implicated in pathogenesis of AML [25, 33]. The presence of mutations in large number of epigenetic modifying genes does not necessarily imply the existence of the specific expression patterns. Ley TJ et al. in their work in which DNMT3A mutations were for the first time reported in AML tried to conclude whether the presence of the mutations in this epigenetic modifier gene in any way affect the overall methylation pattern of the patients. In paired DNMT3A mutated and wild-type samples, they could not detect alteration of the global methylation patterns [8]. It seems that specific gene expression signature is shared by the AML patients belonging to a special subgroup defined by FAB and cytogenetic findings. Namely, it has been shown that certain subtypes of AML, like AML-M3, have a specific gene expression signature [34].

Regardless of the great heterogeneity of AML, only ten genes appear to have frequency higher than 5 % in AML, such as *FLT3*, *NPM1*, *DNMT3a*, *IDH1*, *IDH2*, *TET2*, *RUNX1*, *p53*,

NRAS, *CEBP*A, and *WT1*. Therefore, they tend to be called leukemia specific genes. In our study, precisely in those genes, the most frequent mutations were detected, like mutations in *FLT3*, *IDH1*, and *IDH2* genes and in *NPM1* gene. Our results contribute to a growing number of evidence about the importance that these genes have in pathogenesis of leukemia.

In this study, conducted on both childhood and adult AML, we confirmed the existence of a very small number of mutations (median 3 protein—changing mutations per patient), suggesting that for the development of AML, fewer genetic alterations are required than for other malignancies. The analysis of larger number of patients and control samples, using the same approach, would enable validation of our results. This study is an initial study that provides the "snapshot" of the genetic structure of AML at the beginning of the disease, and it represents the starting point for further analysis of numerous genetic changes which occur during evolution of the disease, leading to implementation of NGS technology in the follow up of AML.

By applying TSACP cancer panel, we defined the genomic profile of both cAML and aAML. While the previously described mutational pattern for AML was reproduced, some additional novel SNVs were detected in both groups of patients. Given that AML is extremely heterogeneous regarding clinical and genetic characteristics, more informative results for similarities and differences in pathogenesis of these two entities will be obtained in a parallel NGS analysis of the morphological and cytogenetic homogeneous groups of cAML and aAML patients.

Acknowledgments This work was funded by the Ministry of Education, Science and Technological Development, Republic of Serbia (grant no. III 41004) and by the European Commission, EU-FP7-REGPOT-316088, 2013–2016.

Compliance with ethical standards

Conflicts of interest None

References

1. Pollyea DA, Kohrt HE, Medeiros BC. Acute myeloid leukaemia in the elderly: a review. Br J Haematol. 2011;152:524–42. doi:10.1111/j.1365-2141.2010.08470.x.
2. Weinstein HJ. Acute myeloid leukemia. In: Pui CH, editor. Childhood Leukemias. Cambridge: Cambridge University Press; 1999. p. 322–35.
3. Mrózek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev. 2004;18:115–36.
4. Falini B, Sportoletti P, Martelli MP. Acute myeloid leukemia with mutated NPM1: diagnosis, prognosis and therapeutic perspectives. Curr Opin Oncol. 2009;21:573–81. doi:10.1097/ CCO.0b013e3283313dfa.
5. Krstovski N, Tosic N, Janic D, Dokmanovic L, Kuzmanovic M, Spasovski V, et al. Incidence of FLT3 and nucleophosmin gene mutations in childhood acute myeloid leukemia: Serbian experience and the review of the literature. Med Oncol. 2010;27:640–5. doi:10.1007/s12032-009-9261-5.
6. Rubnitz JE, Inaba H. Childhood acute myeloid leukaemia. Br J Haematol. 2012;159:259–76. doi:10.1111/bjh.12040.
7. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456:66–72. doi:10.1038/nature07485.
8. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010;363:2424–33. doi:10.1056/NEJMoa1005143.
9. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009;361:1058–66. doi:10.1056/NEJMoa0903840.
10. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA. 2011;305:1577–84. doi:10.1001/jama.2011.497.
11. Marcucci G, Maharry K, Wu YZ, Radmacher MD, Mrózek K, Margeson D, et al. IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. J Clin Oncol. 2010;28:2348–55. doi:10.1200/JCO.2009.27.3730.
12. Oki K, Takita J, Hiwatari M, Nishimura R, Sanada M, Okubo J, Adachi M, Sotomatsu M, Kikuchi A, Igarashi T, Hayashi Y, Ogawa S. IDH1 and IDH2 mutations are rare in pediatric myeloid malignancies. Leukemia. 2011;25:382–4.
13. Andersson AK, Miller DW, Lynch JA, Lemoff AS, Cai Z, Pounds SB, Radtke I, Yan B, Schuetz JD, Rubnitz JE, Ribeiro RC. IDH1 and IDH2 mutations in pediatric acute leukemia. Leukemia. 2011;25:1570–7.
14. S, A. www.bioinformatics.babraham.ac.uk/projects/fastqc. 2010; Available from: www.bioinformatics.babraham.ac.uk/projects/fastqc.
15. E, A., http://code.google.com/p/ea-utils. 2011.
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. doi:10.1093/bioinformatics/btp324.
17. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95. doi:10.1093/bioinformatics/btp698.
18. arXiv, 2013.
19. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8. doi:10.1038/ng.806.
20. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. doi:10.1101/gr.107524.110.
21. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010;26:2069–70. doi:10.1093/bioinformatics/btq330.
22. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6. doi:10.1038/nbt.1754.
23. Kuzmanovic M, Tosic N, Colovic N, Karan-Djurasevic T, Spasovski V, Radmilovic M, et al. Prognostic impact of NPM1 mutations in Serbian adult patients with acute myeloid leukemia. Acta Haematol. 2012;128:203–12. doi:10.1159/000339506.
24. Marchesi V. Genetics: the AML mutational landscape. Nat Rev Clin Oncol. 2013;10(6):305. doi:10.1038/nrclinonc.2013.81.
25. Luthra R, Patel KP, Reddy NG, Haghshenas V, Routbort MJ, Harmon MA, et al. Next-generation sequencing-based multigene mutational screening for acute myeloid leukemia using MiSeq:

applicability for diagnostics and disease monitoring. Haematologica. 2014;99:465–73. doi:10.3324/haematol.2013.093765.
26. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006;314:268–74. doi:10.1126/science.1133427.
27. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368:2059–74. doi:10.1056/NEJMoa1301689.
28. Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K, et al. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia. 2014;28:1586–95. doi:10.1038/leu.2014.55.
29. Bodian DL, McCutcheon JN, Kothiyal P, Huddleston KC, Iyer RK, Vockley JG, et al. Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: implications for individual genome sequencing. PLoS One. 2014;9:e94554. doi:10.1371/journal.pone.0094554 .eCollection 2014
30. Oren M, Rotter V. Mutant p53 gain-of-function in cancer. Cold Spring Harb Perspect Biol. 2010;2:a001107. doi:10.1101/cshperspect.a001107.
31. Glubb DM, Cerri E, Giese A, Zhang W, Mirza O, Thompson EE, et al. Novel functional germline variants in the VEGF receptor 2 gene and their effect on gene expression and microvessel density in lung cancer. Clin Cancer Res. 2011;17:5257–67. doi:10.1158/1078-0432.CCR-11-0379.
32. Kim DH, Xu W, Kamel-Reid S, Liu X, Jung CW, Kim S, et al. Clinical relevance of vascular endothelial growth factor (VEGFA) and VEGF receptor (VEGFR2) gene polymorphism on the treatment outcome following imatinib therapy. Ann Oncol. 2010;21:1179–88. doi:10.1093/annonc/mdp452.
33. Ohgami RS, Ma L, Merker JD, Gotlib JR, Schrijver I, Zehnder JL, et al. Next-generation sequencing of acute myeloid leukemia identifies the significance of TP53, U2AF1, ASXL1, and TET2 mutations. Mod Pathol. 2015;28:706–14. doi:10.1038/modpathol.2014.160.
34. Payton JE, Grieselhuber NR, Chang LW, Murakami M, Geiss GK, Link DC, et al. High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. J Clin Invest. 2009;119:1714–26. doi:10.1172/JCI38248.