Ned Tijdschr Klin Chem Labgeneesk 2016; 41: 244-252

Oorspronkelijke artikelen

Validation and implementation of a custom 21-gene panel next-generation sequencing assay for myeloid neoplasms

H.M. LOOVERS[1#], E.N. de BOER[2#], A. SIMPELAAR[1], K.M. ABBOTT[2], E. VELLENGA[3], E. van den BERG[2] and A.B. MULDER[1]

Rapid and reliable mutational analysis of myeloid neoplasms is increasingly important for diagnostic, prognostic and therapeutic reasons. In this article we describe the development and validation of a custom next-generation sequencing (NGS) assay that reliably tests across a broad range of myeloid neoplasms, including AML, MDS, and myeloproliferative neoplasms. The lllumina TruSeq Custom Amplicon panel was designed to detect variants in 21 genes. The validation protocol included sequencing of cell lines (n=3) and patient samples (n=36) on an Illumina MiSeq platform. A read depth ≥100x was observed for >97% of targeted bases. After filtering for artifacts, a specificity of 100% was obtained. A detection limit of ≤5% was observed for variants present in cell lines. On average two reportable variants were present in samples from patients with a myeloid neoplasm. In conclusion, the custom NGS assay provides an adequate routine assay for genetic analysis of variants present in myeloid neoplasms. Practical considerations on choice of targeted genes, type of assay and method of data analysis are provided in this report.

Recent advances in genomic sequencing methods resulted in an exponential increase in information on acquired mutations in myeloid neoplasms (reviewed in e.g. (1-6)). Detection of such genetic aberrations has several therapeutic consequences. Characterization of disease will be more specific, allowing medication to be directed specifically towards mutated proteins or linked pathways, individual prognosis will be more accurate and sensitivity of monitoring can be increased. Because of these improvements in therapy clinicians are eager to incorporate extensive mutational analysis into routine practice.

Large cohorts of AML and MDS have been studied to identify variants in myeloid neoplasms in a research setting (reviewed in e.g. (1-6)). In contrast, limited literature is available concerning the actual implementation of NGS-based assays into routine clinical work-up. Fundamentally different to research, a clinical application directly affects individual patient treatment. The clinical considerations of broad mutational screening in routine settings are addressed in several publications (e.g. (7-10)). Furthermore, general guidelines have been published on the implementation of next-generation sequencing (NGS) into clinical practice (e.g. (11-13)).

For clinical application of an NGS-based assay a number of fundamental requirements should be fulfilled. Of course, the assay should provide reliably information on clinically relevant variants. The turn-around time should be sufficiently short to allow for result-based adjustment of therapy. Consequently, the validation process has to indicate an reliable variant calling without the necessity for confirmation using a second method. Furthermore, sequence read depth should be high enough to allow for detection of mutations in a background of wild-type cells. To our knowledge, only three reports describe the validation of a NGS-based assay for routine work-up of myeloid neoplasms. A 54-gene targeted panel was validated for diagnosis and disease monitoring in AML/MDS (14). A 194-gene and a 48-gene panel were validated with samples from a broad range of malignancies, including AML and myeloproliferative neoplasms, (15, 16). These reports describe the test validation in detail, and address to different extents the considerations regarding data analysis, coverage variability and variant reporting.

The 21-gene panel assay described in this report reliably tests across a broad range of myeloid neoplasms, including AML, MDS, and myeloproliferative neoplasms. Choice of targeted genes was based on frequencies of gene mutations reported in literature, including genes with a mutation frequency >5% in AML and/or MDS patients. Presented are results of assay validation, including

Department of Laboratory Medicine, Department of Hematology[1]; Department of Genetics[2]; Department of Hematology[3], University Medical Center Groningen, University of Groningen, The Netherlands

# authors contributed equally to this manuscript

Correspondence: André B. Mulder, University Medical Centre Groningen, P.O. box 30.001, 9700 RB Groningen, The Netherlands
E-mail: a.b.mulder@umcg.nl

validation of platform, test and pipeline. Considerations regarding choice of data filters are discussed.

Patients and methods

Patient and cell line samples

Peripheral blood (n=13) or bone marrow samples (n=23) were obtained from 36 patients as part of routine diagnostics; no adjustments were made with respect to amount and type of sample. Patient samples were collected in EDTA. Sample and patient characteristics are shown in Table I. Used cell lines are OCI-AML3, MOLM13 and KASUMI-1 (Leibniz – Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). Mutations in these cell lines are DNMT3A R822C (OCI-AML3), NPM1 W288fs (OCI-AML3), *CBL delGGTACGGATCTAAA* (at border exon 8 and intron 8; MOLM13), FLT3 V592VDFREYEF (MOLM13), *TP53* R248Q (KASUMI-1) and KIT N822K (KASUMI-1) (Leibniz –Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). Published allele frequencies are 50% DNMT3A R822C (18), 50% NPM1 W288fs (18), 50% *CBL delGGTACGGATCTAAA* (19), ca. 85% KIT N822K (20), 100% *TP53* R248Q (21) and ca. 66% FLT3 V592VDFREYEF (22). Dilution series of cell line samples were used in a ratio of 98:2:0, 10:10:80, and 2:98:0 of OCI-AML3:MOLM13:KASUMI-1. A selection of patient samples tested positive during routine work-up with mutation-specific assays for CALR Type I and Type II, JAK2 V617F, KIT D816V, MPL W515A and NPM1 W288fs were included in the validation process.

Sample processing, sequencing chemistry and platform

Patient samples were processed the same day (weekdays) or within 3 days (weekends). DNA was extracted using a salt extraction method (23). Quality of DNA was assessed using OD260/280nm and gel electrophoresis. Library preparation was performed according to the protocol described in the TruSeq Custom Amplicon Library Preparation Guide (Illumina Inc, San Diego, CA, USA). During processing, DNA yield and quality where assessed using Tapestation 2200 (Agilent, Santa Clare, CA, USA). Sequencing reactions were performed using V3 chemistry (Paired end 250 bp, Illumina Inc, San Diego, CA, USA). The sequencer was operated according to the manufacturers protocol (MiSeq, Illumina Inc, San Diego, CA, USA).

Panel design

Based on frequencies of gene mutations reported in literature (frequency >5% in AML or MDS patients), 21 genes were selected as targets for sequencing: ASXL1, CALR, CBL, CEBPA, DNMT3A, EZH2, FLT3, IDH1, IDH2, JAK2, KRAS, KIT, MPL, NPM1, NRAS, RUNX1, SRSF2, SF3B1, TET2, *TP53* and WT1. For genes in which mutational hotspots are reported, only corresponding exons were selected (Table II). For genes with mutations located across a large part of the gene, all coding exons were selected. The target region was restricted to coding exons +/- 30bp of intron sequence. Probes were designed using the web-based application DesignStudio v1.7.0.108 (Illumina Inc, San Diego, CA, USA), resulting in 91 targets covered by 232 amplicons. Amplicons were designed to cover up to 250 bp, suitable for MiSeq V3 technology. The web-based application designed probes to theoretically cover 100% of the selected regions, with exception of *TP53* exon 2, exon 3 and exon 11. Despite several adjustments to the design, predicted coverage of *TP53* exon 2 and exon 3 remained zero. The reported frequencies of mutations in these exons is low (e.g.(24)), though exclusion of these exons in previous studies (e.g. (25)(26)) likely contributed to the low frequencies. Based on the predicted lack of coverage and possible lack of clinical relevance we excluded *TP53* exons 2 and 3 from the final design.

Data analysis and filter steps

Sequence alignment and a first filtering step was performed using NextGENe version 2.3.4.2 (SoftGenetics Pennsylvania, US). Format conversion was done using the settings MedianScore threshold ≥30, Max uncalled bases ≤3 Called Base Number of each Read ≥25 Trim or Reject Read when ≥3 Base(s) with score ≤25 and Paired Reads Date selected. Alignment, Sample Trim, Mutation Filter and filter Type setting were as follows. Alignment: Allowable Mismatched Bases 1, Allowable Ambiguous Alignments 50, Seeds 30 bases, Move Step 5 Allowable Alignments 100 Overall Matching Base Percentage ≥85. Sample Trim: Hide Unmatched Ends. Mutation Filter: Mutation Percentage for SNPs Indels and HomopolymerIndels (except for Homozygous) ≤1.5 Percentage SNP Allele Count ≤3 and Total Coverage count ≤5 Balance Ratios and frequency for Indels 0.1 and 80% and for HomopolymerIndels 0.8 and 80%. Filter Type: Load Paired Reads selected with library Size Range from 150 to 350 Bases. This filter returned all mutations at an allele frequency ≥1.5% and with a quality score Q>Q30 (Figure 1). Subsequent data analysis on the VCF file was performed using Cartagenia Bench Lab NGS (Cartagenia N.V., Leuven, Belgium). The variant triangle had filters on minimal read depth, population frequency, gene structure, variant allele frequency and position. The minimal read depth was set on ≥20x. Population frequency testing was done with the 1000 Genome Phase 1 (27), ESP6500 (28), dbSNP (29) and GoNL-v4 (30) databases, the allele frequency had to be lower than 2 percent (5% for the GoNL) and the allele count was set on 400, only validated not suspected SNP's were excluded. After this filtering only exon variants (+/- 2 bp of intron sequence to cover splice sites) were listed in the variant list. We chose to not to list variants in the remaining intron regions as consequences for prognosis and therapy are difficult to interpret at this moment. The filtered data were processed manually by at least two laboratory analysts specialized in molecular diagnostics for hemato-oncology. Non-synonymous variants with allele frequencies ≥5% and coverage ≥100x were considered reportable, unless

Table 1. Variants detected in patient samples

Patient sample	Diagnosis	Material	Variant	Variant allele frequency	Coverage
P1	AML	BM	IDH1 R132H#	0.26	9706
			WT1 S381*#	0.05	10457
P2	AML	BM	NRAS G12C#	0.50	18312
			SRSF2 P95_R102del#	0.25	1224
			RUNX1 N112Kfs#	0.49	7149
P3	AML	BM	DNMT3A R882H#	0.52	4800
			TET2 M1333Wfs#	0.47	6400
			JAK2 V617F#	0.06	5727
P4	AML	BM	DNMT3A R882H	0.46	101
			IDH1 R132C	0.32	24625
			TET2 V1718L	0.26	35096
P5	AML	BM	none	-	-
P6	AML	PB	IDH2 R140Q	0.51	2128
			NPM1 W288fs#	0.39	12231
P7	AML	BM	none	-	-
P8	AML	PB	DNMT3A R729W	0.47	4584
			NPM1 W288fs#	0.39	12714
P9	AML	BM	none	-	-
P10	AML	BM	IDH2 R140Q	0.51	6206
			DNMT3A V375Wfs	0.47	2780
			NPM1 W288fs#	0.39	4054
P11	AML	BM	TET2 E782Dfs	0.34	4405
			TET2 R1404*	0.10	9217
P12	AML	BM	ASXL1 G645fs	0.30	4395
			RUNX1 D171N	0.36	4773
P13	AML	BM	DNMT3A R882H	0.49	2600
			NPM1 W288fs	0.41	7912
			IDH2 R140W	0.41	7830
			FLT3 Y599_E608dup (ITD)#	0.14	11328
P14	AML	BM	KIT I935I	0.52	3965
			SRSF2 P95H	0.50	173
			TET2 A1505T	0.43	10607
P15	MDS	BM	SRSF2 P96H	0.43	109
			RUNX1 P398L	0.50	922
P16	MDS	BM	NRAS G13D	0.13	23071
			TET2 S1290L	0.33	25126
			EZH2 N673S	0.48	14911
			ASXL1 Y591*	0.08	16498
P17	MDS	BM	SF3B1 K700E	0.33	14320
			TET2 L34F	0.50	7875
			TET2 Y867H	0.30	24038
			TET2 P1723S	0.37	53511
			ASXL1 A1312V	0.58	3740
P18	MDS	PB	TET2 D1384N	0.36	30393
P19	MDS/MPN	BM	TET2 E807fs	0.46	26947
			TET2 Q1170*	0.50	21891
			CBL F418S	0.44	29558
P20	AML/MDS/CMML	BM	TET2 H1881L	0.49	11071
			EZH2 Q653E	0.96	34445
			RUNX1 G172A	0.44	11672
P21	CMML	PB	TET2 T229fs	0.43	19181
			TET2 Q548*	0.45	7903
			SRSF2 P95L	0.55	93
			ASXL1 H630_T639del	0.34	6454
			RUNX1 G367fs	0.07	941
P22	PV	BM	TET2 R544*	0.38	24147
P23	PV	BM	JAK2 V617F#	0.19	8896
P24	ET	PB	DNMT3A R749C	0.40	2492
			CALR K385fs (5bp ins)#	0.37	699
P25	ET	PB	MPL W515A#	0.65	833
			ASXL1 E1102D	0.48	17291
P26	ET	PB	JAK2 V617F#	0.12	11427
P27	ET	PB	CALR E364fs (52bp del)#	0.11	874
P28	MF	PB	CALR Q365fs (46bp del)#	0.15	729
			ASXL1 R693*	0.30	16235
P29	MF in CR after Allo-SCT	PB	none	-	-
P30	Mastocytosis	PB	KIT D816V#	0.09	35152
P31	Pancytopenia eci	PB	none	-	-
P32	Trombopenia eci	PB	none	-	-
P33	Anemia and trombopenia eci	BM	none	-	-
P34	Cyclic neutropenia eci	BM	none	-	-
P35	ALL	BM	none	-	-
P36	ALL	BM	NRAS G12D	0.65	37387

confirmed by alternative method.; MPN: myeloproliferative neoplasm; CMML: chronic myelomonocytic leukemia; PV: polycythemia vera; MF: myelofibrosis; ET: essential thrombocytemia; CR: complete remission.

Table 2. Panel characteristics[#]

Genes	Exons targeted	Number of bases targeted	Mean read depth	Percentage of bases above cut-off value*	
				≥100x	≥1000x
ASXL1	12	2912	26658	100	100
CALR	9	205	3527	100	92
CBL	8,9	340	30907	100	100
CEBPA	1	1081	1771	41	30
DNMT3A	8,9,13-15,18-23	1299	11457	100	99
EZH2	2-20	2242	20354	100	100
FLT3	14,15,20	373	18760	100	100
IDH1	4	296	23082	100	100
IDH2	4	165	15694	100	100
JAK2	12,14	220	9958	100	100
KIT	1-21	2916	25216	100	98
KRAS	2,3	293	14766	100	100
MPL	10	101	122	67	0
NPM1	12	43	13084	100	100
NRAS	2,3	292	15738	100	84
RUNX1	2-9	1489	8332	98	85
SF3B1	12,14-16	748	22953	100	100
SRSF2	1	366	1355	98	55
TET2	2-11	6069	29257	100	100
TP53	4-11	1086	10934	100	94
WT1	7,9	247	11017	100	100

data from one representative run; * calculated as mean of percentage per sample

high suspicion of an assay artifact was present (see discussion for reasons for cut-off values). Artifacts were suspected when variants were present at similar low allele frequencies (<10%) in ≥2 independent samples within one run. In case a sample showed two variants within 10bp distance, the data was analyzed by hand to assure an accurate variant calling.

Platform and pipeline validation

Validation was accomplished using cell line samples and patient samples with known single nucleotide mutations and small insertions/deletions. Criteria on platform and pipeline performance were established: (i) the assay should fulfill test validation criteria (see test validation) (ii) the assay should be suitable for both peripheral and bone marrow samples (iii) library preparation should be easily incorporated in a routine molecular biology laboratory (iv) a turn-around time ≤2 weeks for at least 7 patients simultaneously should be feasible (v) total costs should be <800 euro per sample and (vi) the average yield should be ≥1 variant per sample for patients with a confirmed myeloid neoplasm. Validation of the data filtering steps with Cartagenia was accomplished by manual check of NextGENe data using Excel software (Microsoft Corporation, Redmond, USA). The entire pipeline was validated using well-characterized cell lines and patient samples containing variants according to alternative assays.

Test validation

Criteria for test specifications were established: (i) >95% of targeted bases should be covered ≥100x, (ii) variants should be detected if present at frequencies ≥5% and coverage ≥100x (iii) single nucleotide variants and small insertions/deletions (<10 bp) should be detected with >99% specificity at an allele frequency of ≥5%; i.e. all low-quality reads and likely artifacts should be dismissed in the pipeline, (iv) variant calling should be reproducible; i.e. variants should be detected in repetitive assays at similar allele frequencies. Specificity, accuracy and sensitivity were validated using three cell line samples (MOLM13, KASUMI-1, and OCI-AML3) and three patient samples (P1-P3). To determine limit of detection, DNA samples of well-defined cell lines were mixed in 3 different ratios of DNA concentration before enrichment and labeling. Used ratios were 98:2:0, 10:10:80, and 2:98:0 of OCI-AML3:MOLM13:KASUMI-1. To confirm sensitivity of the assay, a selection of patient samples tested positive with mutation-specific assays for CALR Type I and Type II, JAK2 V617F, KIT D816V, MPL W515A and NPM1 W288fs during routine work-up were included in the validation process (see below). Furthermore, all samples that tested positive with the 21-gene panel assay for these variants were subjected to the mutation-specific assays. To assess inter-run reproducibility, a dilution sample of 10:10:80 of OCI-AML3:MOLM13: KASUMI-1 was analyzed in two separate runs with different amplicon reagents and sequencing reagents lot numbers.

Mutation specific assays

Quantitative PCR was used for detecting the JAK2 V617F mutation (LightCycler 480II, Hoffmann-La Roche, Basel, Switzerland) and KIT D816V mutation (ABI7500, ThermoScientific, Waltham, Massachusetts, United States). Fragment analysis was used for

detecting *NPM1* exon 12 and *CALR* exon 9 mutations (ABI3130, ThermoScientific, Waltham, Massachusetts, United States). High Resolution Melting was used for detecting *MPL* exon 10 mutations (LightCycler 480II, Hoffmann-La Roche, Basel, Switzerland). The *CALR* and *MPL* exon 10 mutations were confirmed by Sanger sequencing (ABI3130, ThermoScientific, Waltham, Massachusetts, United States). Assay conditions and primer or probe sequences are available on request.

Reporting, data systems and storage
In close collaboration between a Hematologist, Clinical Chemist with hemato-oncology as specialty, and Clinical Geneticist with hemato-oncology as specialty a report format was developed. All variants fulfilling the filter requirements are reported to the clinic, including the corresponding allele frequencies. Data is reported to the clinic after authorization by a group of specialists, consisting of at least one clinical chemist with hemato-oncology as specialty and one clinical geneticist with hemato-oncology as specialty. A more comprehensive report, including all variants at ≥2.6 allele frequency and including extensive coverage data is generated for easy access in case of future clinical questions.

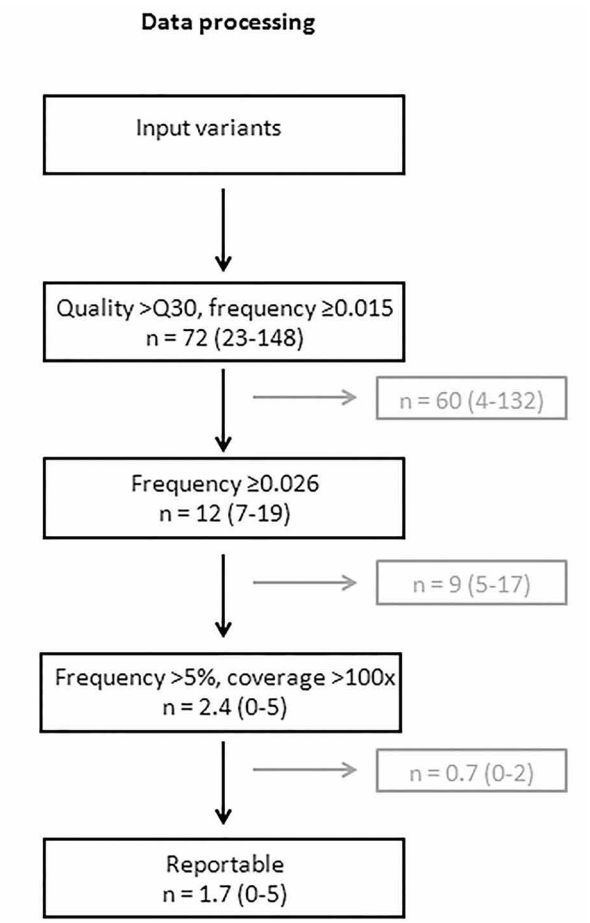

Figure 1. Flow chart of filter steps during data processing (n = average number of variants per patient)

Results

Target read depth and coverage
After alignment and filtering, a mean read depth ranging from 122x to 30907x was observed for the targeted regions (Table II). Validation criteria of >95% of bases covered ≥100x were routinely obtained (Table II, Figure 2 and data not shown). For *CEBPA, MPL* and *RUNX1* part of the targeted exons showed a read depth <100x in more than one sample (Figure 2). Coverage of *MPL* was sample dependent, with either 0% or 100% of bases covered ≥100x. For *CEBPA*, all patient samples showed low coverage for a defined part of the gene. *CEBPA* contains a high percentage of regions with increased GC content (>70% GC), likely reducing the capture or library amplification efficiencies (31)(32). Due to this consistent lack in adequate read depth for a large part of the gene, and the clinical importance of detecting bi-allelic mutations, data on *CEBPA* was deemed unreportable to the clinic. All other targeted regions showed clinical useful coverage data. Inter-assay variation, including variation between runs with different amplicon reagents and sequencing reagents lot numbers, was within acceptable range. Data on read depth were similar for DNA isolated form bone marrow aspirates or peripheral blood samples (data not shown).

Filter steps
Of all mutations detected, 99.8% of the variants had an allele frequency <2.6%. The majority of these mutations were presumed to originate from assay artifacts. The number of detected mutations increased dramatically with decreasing cut-off values for allele frequency. The average numbers of mutations with allele frequencies ≥1.5% was 72 mutations per patient sample (range 23-148; Figure 1). After data processing and filtering, samples concerning a confirmed myeloid neoplasm (n=30) showed an average of two reportable variants (Figure 1, Table I). In total, 13 variants at 8 different positions (i.e., ASXL1 G645, CBL Y455, DNMT3A E294, NRAS Q61, RUNX1 L415, *RUNX1* *454fs, SRSF2 E53 and SRSF2 S119) were regarded likely artifacts based on presence in ≥2 independent samples within one test run at similarly low allele frequencies (<10%). Likelihood of being artifacts was supported by all of these variants not being detected in repeat sample measurements (data not shown). Samples from five patients not diagnosed with a myeloid neoplasm were included in the validation process. No variants were detected in these samples, with the exception of a NRAS mutation in an ALL patient sample. No discrepancies were observed using either Cartagenia (Cartagenia N.V., Leuven, Belgium) or Excel (Microsoft Corporation, Redmond, USA) for data analysis.

Sensitivity, specificity and accuracy
A limit of detection of ≥5% allele frequency was set, based on the wish to detect heterozygous mutations in samples with a bone marrow blast count as low as 10% (i.e. MDS RAEB II). Limits of detection were determined by a dilution series of three well-defined

cell lines, using a cut-off value of total base coverage ≥100x. A limit of detection ≤5% was observed for the variants DNMT3A R822C and NPM1 W288fs (Table III). The gene variant *CBL delGGTACGGATCTAAA* was not readily detected at low allele frequencies, likely due to the position of the deletion (at the border of the amplicon). The allele frequency of FLT3-ITD was underestimated, possibly due to the relatively large insertion (30bp). Specificity and accuracy were determined using three well-defined cell lines and three exome-sequenced patient samples (P1-P3 in Table I; Table III). Using the data filter steps, all detected variants in the three cell lines were variants previously described in literature (18)(19)(20)(21)(22). All mutations detected in patient samples P1-P3 were confirmed by exome sequencing. The data show that all variants were enumerated correctly, showing specificity and accuracy up to 100%. Small insertions were detected, as shown for FLT-ITD (30bp insertion) and CALR Type II (5bp insertion) (Table I). Relative large deletions were detected as well, as shown for CALR Type I-like mutations (52bp and 46bp deletions). Sensitivity and specificity were more extensively addressed for the variants CALR Type I and Type II, JAK2 V617F, KIT V816F, MPL W515A and NPM1 W288fs present in patient samples. No false-positives were detected with the 21-gene panel assay (Table IV). False- negatives were detected for JAK2 V617F and KIT D816V. JAK2 V617F allele frequencies in the two false-negative samples were <6% (frequencies are relative to HEL cell line; measured with in-house mutation-specific assay). Similarly, the false-negative KIT D816V mutation was present at low frequency as well (Ct value of 37 cycles using the mutation-specific assay). As the NGS assay was not designed to detect low-frequency variants, they were not considered true false-negatives.

Reproducibility and quality control

Verification of inter-assay reproducibility showed detection of the same variants at similar allele frequencies. Based on these results, a QC sample was defined. This sample contains MOLM13, OCI-AML3, and KASUMI-1 in a ratio of 10:10:80. This QC sample will be used to assess quality of new batches of reagents. Acceptance criterion is detection of DNMT3A R822C, NPM1 W288fs, FLT3 V592VDFREYEF, *TP53* R248Q and KIT N822K at similar allele frequencies and no detection of additional variants except for *CBL delGGTACGGATCTAAA* when using the described filter steps.

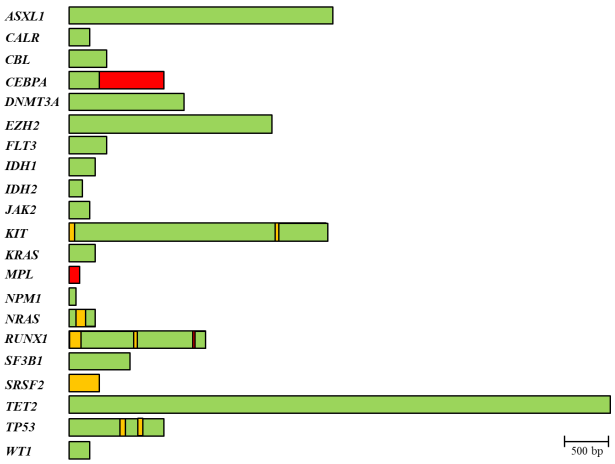

Figure 2. Coverage of targeted bases. Bases with read depth <100x in >1 sample are shown in red, bases with read depth <1000x in >1 sample are shown in orange. The other regions are shown in green. Multiple targeted exons within one gene are shown in sequence, with omission of non-targeted exons.

Platform and pipeline validation

Both bone marrow aspirates and peripheral blood samples produced acceptable test results, with the minimal required amount of DNA being 250 ng. No adjustments to laboratory facilities were required; only small investments in laboratory equipment were required (see TruSEQ Custom Amplicon Library Preparation Guide (Illumina, San Diego, USA)). As 12 samples were included within one run, a turn-around times of 12 patient samples per 1 week was feasible at a cost of <800 euro per sample. In addition, all claims for test performance were met, yielding >1 mutation per sample (see above). In conclusion, all criteria set for platform and pipeline performance were met in the validation process.

Discussion

Guidelines for clinical application of NGS recommend a set of validation and decision steps needed for adequate implementation: platform validation, clinical considerations, sample handling, test development/ quality, test selection, turn-around time, bioinformatics, reporting, data systems and storage, and facility management (reviewed in (13)). Platform and pipeline validation should establish that the system consistently detects the type of variant the test is designed to detect. This report describes the process of implementation of a NGS-based assay for myeloid neoplasms, and addresses the decision and validation steps. As concluded in the results section, all criteria set for platform and pipeline performance were met in the validation process for our custom 21-gene panel NGS assay. The greatest challenge proved to be data processing; choices in cut-off criteria directly affecting sensitivity and specificity.

Choices in sequence chemistry and platform

The number of commonly mutated genes is limited in myeloid neoplasms. Most are involved in epigenetic processes (DNA and histone modification), DNA transcription, RNA splicing or signal transduction. Though limited, the number of relevant regions is too large to allow efficient workup by conventional methods. On the other hand, a Whole Genome Sequencing approach would be cost-inefficient for the limited number of genes of interest and the required sequence depth (9, 33). In our opinion, targeted NGS-based assays present the best possibility for adequate variant detection in myeloid neoplasms. The choice for a custom design was based on the unavailability of suitable commercial assays at time of validation. Designing the assay for a broad spectrum of myeloid

Table 3. Frequencies of variants in cell line dilution samples

Variant	calculated allele frequency	detected allele frequency (duplicate)
DNMT3A R882C	0.01	<0.015
	0.05	0.02 (0.06)
	0.49	0.47
NPM1 W288fs	0.01	<0.015
	0.05	0.02 (0.03)
	0.49	0.44
FLT3 V592VDFREYEF	0.01	<0.015
	0.07	0.07 (0.10)
	0.66	0.51
CBL delGGTACGGATCTAAA	0.01	<0.015
	0.05	<0.015 (<0.015)
	0.49	0.34
KIT N822K	0.68	0.76 (0.73)
TP53 R248Q	0.80	0.79 (0.75)

* In brackets are shown repeat measurements with different lot numbers for amplicon panel and sequencing reagents.

neoplasms allowed for adequate turn-around times (≤2 weeks) while at the same time reducing costs (<12 requested AML mutational analyses per 2 weeks). The choice for TruSeq Custom Amplicon method was based on ease of workflow, possibility to work with a relatively small region of interest, good performance and good compatibility with Illumina MiSeq; with relative high and reproducible read depth. Choice of sequencing platform was based on previous laboratory experiences, available hospital facilities, low error rate, and relatively high throughput for a bench top machine.

Choices in data processing

The interplay between detection limit, it's sensitivity and specificity is especially relevant in variant detection of acquired mutations. A low detection limit is required for disease monitoring and variant detection in myeloid neoplasms, but high specificity is required for appropriate treatment-adjusted therapy. As expected using a next-generation technique, a drastic decline in specificity for random variants was observed upon decreasing limits of allele frequency. Reduced specificity can be solved in a research setting by confirmation with an alternative technique and insignificance of a single mutation for outcome parameters within a large cohort. In a routine clinical setting, the first option is generally accepted as time and cost inefficient, and the latter is not applicable. For myeloid neoplasms, allele frequency limits are commonly set at >1% to >5% for random variants (e.g. (14)(15)(34)(35)(36)(37)). Cut-off values for base coverage generally range from total base coverage of 30x to 250x (e.g. (37)(38)(34)(39)). Luthra et al. have shown 100% concordance between a 54-gene panel using MiSeq V2 sequencing chemistry and other sequencing techniques when using cut-off values of allele frequency >5% and base coverage >250x (bidirectional) (14). For the 21-gene panel assay, we opted for cut-off values of allele frequency ≥5% and total base coverage ≥100x (bidirectional). The cut-off value of allele frequency ≥5% was based on the wish to detect heterozygous mutations in samples with a blast count as low as 10% (i.e. MDS RAEB II). The cut-off value of total base coverage ≥100x was chosen to retain confident variant calling (variant base coverage ≥5x) for the maximum number of target regions. With specific care for artifacts, using all data on variants with an allele frequency ≥2.6%, reliable variant calling was indeed obtained at ≥5%.

Table 4. Sensitivity and specificity of specific variants

	TP	TN	FP	FN
CALR Type I + CALR Type II	2+1	2	0	0
JAK2 V617F	3	10	0	2*
KIT D816V	1	1	0	1*
MPL W515A	1	3	0	0
NPM1 W288fs	4	13	0	0

* samples are not considered false-negatives according to assay requirements (see text)

MRD detection

A detection limit of <0.1% for allele frequency would allow assessment of minimal residual disease (MRD) with similar detection limits as currently applied in our lab for immuno-phenotyping. For AML, the detection of MRD can be a challenge, especially for patients without a leukemia-associated phenotype, with cytogenetically normal AML and without traditional molecular markers. Next-generation assays can be designed to detect MRD. Sensitivities as low as 0.2% for the single nucleotide mutation KIT D816V (40) and 0.001% for NPM1 4bp-insertion mutations (41) have been described, without loss of specificity. The read depth at the KIT variant position ranged from 6556 to 10249. In the current validation process the detection of MRD for AML was not included. However, as mean read depth is >6556 for the 21-gene panel, MRD

measurements might be feasible for at least some of the variants detected in the validation process. To confidently report random variants at these low frequencies additional data is warranted. For each individual variant, specificity at low frequencies must be established to prevent false-positive results. Similarly, variants in polycythemia vera, essential thrombocytosis, myelofibrosis and mastocytosis have diagnostic value, even at low frequencies. Replacement of the routinely used mutation-specific assays for e.g. JAK2 V617F and KIT D816V will require further analysis.

Data interpretation and reporting

Clinical utility of the 21-gene assay will be determined for individual patients by a consultation body consisting of one hematologist and one clinical chemist specialized in hematology. Data will be reported to the clinic after authorization by a clinical chemist specialized in hematology and a clinical geneticist specialized in hemato-oncology. Genes not adequately covered (<95% covered at ≥100x read depth) are specifically addressed in the report. Interpretation of the clinical significance of specific variants will require more research. Furthermore, additional knowledge will hopefully result in diagnostic and therapeutic consequences for a larger number of variants. At the moment, a more general interpretation of the significance of variants is added to the report. For each of the variants detected in a specific patient sample, a general statement is given on the association between that corresponding gene and hematological diseases.

In conclusion, a clinically applicable protocol has been developed for the detection of somatic variants in 20 genes, with a sensitivity of 5% allele frequency. Total base coverage is above the cut-off value of 100x in over 95% of the target regions. Future plans include improvement of coverage distribution; Illumina offers a service to equalize coverage by increasing probe concentrations for regions with low coverage. During validation of the 21-gene panel NGS assay, Illumina marketed a targeted based NGS assay specifically for myeloid neoplasms, TruSight Myeloid Sequencing Panel. Comparison of sensitivity and specificity of our custom 21-gene panel NGS assay with this commercially available assay will be extremely interesting.

Acknowledgements

We acknowledge and appreciate the support from the laboratory technicians Bettie Postema-Veenema and René Mulder of the department of Laboratory Medicine, Department of Hematology, University Medical Centre Groningen.

References

1. Ilyas A, Ahmad S, Faheem M, Naseer M, Kumosani T, Al-Qahtani M, et al. Next Generation Sequencing of Acute Myeloid Leukemia: Influencing Prognosis. BMC Genomics. BioMed Central Ltd; 2015;16(Suppl 1):S5.
2. Zhang L, Padron E, Lancet J. The molecular basis and clinical significance of genetic mutations identified in myelodysplastic syndromes. Leuk Res. 2014;39(1):1–12.
3. Tefferi A. Myeloproliferative Neoplasms: A decade of discoveries and treatment advances. Am J Hematol. 2016;91(1):50-8.
4. Nebbioso A, Benedetti R, Conte M, Iside C, Altucci L. Genetic mutations in epigenetic modifiers as therapeutic targets in acute myeloid leukemia. Expert Opin Ther Targets. 2015;8222(Nov):1–16.
5. Harada H, Harada Y. Recent advances in myelodysplastic syndromes: Molecular pathogenesis and its implications for targeted therapies. Cancer Sci. 2015; 106(4):329-36.
6. Skoda RC, Duek A, Grisouard J. Pathogenesis of myeloproliferative neoplasms. Exp Hematol. 2015;43(8):599–608.
7. Kohlmann A, Bacher U, Schnittger S, Haferlach T. Perspective on how to approach molecular diagnostics in acute myeloid leukemia and myelodysplastic syndromes in the era of next-generation sequencing. Leuk Lymphoma. 2014;55(8):1725–34.
8. Bacher U, Kohlmann A, Haferlach T. Mutational profiling in patients with MDS: Ready for every-day use in the clinic? Best Pract Res Clin Haematol. 2015;28(1):32–42.
9. Duncavage EJ, Tandon B. The utility of next-generation sequencing in diagnosis and monitoring of acute myeloid leukemia and myelodysplastic syndromes. Int J Lab Hematol. 2015;37:115–21.
10. Graubert T, Stone R. AML Genomics for the Clinician. Semin Hematol. 2014;51(4):322–9.
11. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012;30(11):1033–6.
12. Weiss MM, Van der Zwaag B, Jongbloed JDH, Vogel MJ, Brüggenwirth HT, Lekanne Deprez RH, et al. Best Practice Guidelines for the Use of Next-Generation Sequencing Applications in Genome Diagnostics: A National Collaborative Study of Dutch Genome Diagnostic Laboratories. Hum Mutat. 2013;34(10):1313–21.
13. Bennett N, Farah C. Next-Generation Sequencing in Clinical Oncology: Next Steps Towards Clinical Validation. Cancers (Basel). 2014;6(4):2296–312.
14. 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(3):465–73.
15. Pritchard CC, Salipante SJ, Koehler K, Smith C, Scroggins S, Wood B, et al. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. J Mol Diagnostics. 2014;16(1):56–67.
16. Simen BB, Yin L, Goswami CP, Davis KO, Bajaj R, Gong JZ, et al. Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. Arch Pathol Lab Med. 2015;139(4):508–17.
17. Bisling KE, Brewin JN, McGovern AP, Horne G a, Rider T, Stewart HJ, et al. DNMT3A mutations at R882 hotspot are only found in major clones of acute myeloid leukemia. Leuk Lymphoma. 2014;55(3):711-4.
18. Quentmeier H, Martelli MP, Dirks WG, Bolli N, Liso A, Macleod RA, et al. Cell line OCI/AML3 bears exon-12 NPM gene mutation-A and cytoplasmic expression of nucleophosmin. Leukemia. 2005;19(10):1760–7.

19. Kales SC, Ryan PE, Nau MM, Lipkowitz S. Cbl and human myeloid neoplasms: The Cbl oncogene comes of age. Cancer Res. 2010;70(12):4789–94.
20. Beghini A, Magnani I, Ripamonti CB, Larizza L. Amplification of a novel c-Kit activating mutation Asn(822)-Lys in the Kasumi-1 cell line: a t(8;21)-Kit mutant model for acute myeloid leukemia. Hematol J. 2002;3(3):157–63.
21. ATCC: p53 Hotspot Mutation Cell Panels. [cited 2016 May 21].
22. Moore AS1, Faisal A, Gonzalez de Castro D, Bavetsias V, Sun C, Atrash B, et al. Selective FLT3 inhibition of FLT3-ITD+ acute myeloid leukaemia resulting in secondary D835Y mutation: a model for emerging clinical resistance patterns. Leukemia. 2012;26(7):1462-70
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988;16(3):1215.
24. International Agency for Research on Cancer (IARC) TP53 database. [cited 2016 May 24].
25. Grossmann V, Schnittger S, Kohlmann A, Eder C, Roller A, Dicker F, et al. A novel hierarchical prognostic model of AML solely based on molecular mutations. Blood. 2012;120(15):2963–72.
26. Ohgami RS, Ma L, Merker JD, Gotlib JR, Schrijver I, Zehnder JL, Arber DA Next-generation sequencing of acute myeloid leukemia identifies the significance of TP53, U2AF1, ASXL1, and TET2 mutations. Mod Pathol. 2015;28(5):706-14.
27. 1000 Genomes Phase 1. [cited 2016 May 24].
28. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). [cited 2016 May 24].
29. The Database of Short Genetic Variation (dbSNP). [cited 2016 May 24].
30. The Genome of the Netherlands; GoNL v4. [cited 2016 May 24].
31. van Vliet MH, Burgmer P, de Quartel L, Brand JPL, de Best LCM, Viëtor H, et al. Detection of CEBPA double mutants in acute myeloid leukemia using a custom gene expression array. Genet Test Mol Biomarkers. 2013;17(5):395–400.
32. Grossmann V, Schnittger S, Schindela S, Klein HU, Eder C, Dugas M, et al. Strategy for robust detection of insertions, deletions, and point mutations in CEBPA, a GC-rich content gene, using 454 next-generation deep-sequencing technology. J Mol Diagnostics. 2011;13(2):129–36.
33. Meldrum C, Doyle M a, Tothill RW. Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. Clin Biochem Rev. 2011;32(4):177–95.
34. Ok CY, Patel KP, Garcia-Manero G, Routbort MJ, Fu B, Tang G, et al. Mutational profiling of therapy-related myelodysplastic syndromes and acute myeloid leukemia by next generation sequencing, a comparison with de novo diseases. Leuk Res. 2015;39(3):348–54.
35. Kohlmann A, Nadarajah N, Alpermann T, Grossmann V, Schindela S, Dicker F, et al. Monitoring of residual disease by next-generation deep-sequencing of RUNX1 mutations can identify acute myeloid leukemia patients with resistant disease. Leukemia. 2014;28(1):129–37.
36. Fu Y, Schroeder T, Zabelina T, Badbaran A, Bacher U, Kobbe G, et al. Postallogeneic monitoring with molecular markers detected by pretransplant next-generation or Sanger sequencing predicts clinical relapse in patients with myelodysplastic/myeloproliferative neoplasms. Eur J Haematol. 2014;92(3):189–94.
37. Kirschner MMJ, Schemionek M, Schubert C, Chatain N, Sontag S, Isfort S, et al. Dissecting Genomic Aberrations in Myeloproliferative Neoplasms by Multiplex-PCR and Next Generation Sequencing. PLoS One. 2015;10(4):e0123476.
38. Bolli N, Manes N, McKerrell T, Chi J, Park N, Gundem G, et al. Characterization of gene mutations and copy number changes in acute myeloid leukemia using a rapid target enrichment protocol. Haematologica. 2015;100(2):214–22.
39. Kohlmann A, Klein H-U, Weissmann S, Bresolin S, Chaplin T, Cuppens H, et al. The Interlaboratory Robustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. Leukemia. 2011;25(12):1840–8.
40. Kristensen T, Broesby-Olsen S, Vestergaard H, Bindslev-Jensen C, Møller MB. Targeted ultradeep next-generation sequencing as a method for KIT D816V mutation analysis in mastocytosis. Eur J Haematol. 2016;96(4):381-8.
41. Salipante SJ, Fromm JR, Shendure J, Wood BL, Wu D. Detection of minimal residual disease in NPM1-mutated acute myeloid leukemia by next-generation sequencing. Mod Pathol. 2014;27(11):1438–46.

Samenvatting

Loovers HM, de Boer EN, Simpelaar A, Abbott KM, Vellenga E, van den Berg E, Mulder AB. Validatie en implementatie van een custom 21-gene panel next-generation sequencing assay voor myeloid neoplasms. Ned Tijdschr Klin Chem Labgeneesk. 2016; 41:244-252.

Snelle en betrouwbare diagnostiek van verworven mutaties speelt in toenemende mate een rol bij de diagnose, prognose en therapie van myeloïde maligniteiten. In dit artikel beschrijven wij het ontwerp en de validatie van een next-generation sequencing (NGS) assay die betrouwbaar ingezet kan worden voor een breed spectrum van myeloïde maligniteiten, waaronder AML, MDS en myeloproliferatieve aandoeningen. Het Illumina TruSeq Custom Amplicon panel werd ontworpen om mutaties in 21 genen te detecteren. Voor de validatie werden cellijnen (n=3) en monsters van patiënten (n=36) gesequenced met behulp van een Illumina MiSeq. Voor >97% van de basen werd een read depth ≥100x waargenomen. Na filtering werd een specificiteit van 100% gerealiseerd. Daarnaast werd voor mutaties in de gebruikte cellijnen een detectie limiet van ≤5% waargenomen. Gemiddeld werden per patiëntenmonster twee mutaties gedetecteerd. Samengevat, de NGS assay is een adequate methode voor routinematig onderzoek naar mutaties bij de vraagstelling myeloïde maligniteit. Praktische overwegingen met betrekking tot keuzes in geïncludeerde genen, type assay en methode van data analyse worden in dit artikel besproken.