

1 Subjects & Methods

1.1 Samples

For assay assessment, we sequenced AML cell lines ($n = 3$), samples from healthy volunteers ($n = 2$, from peripheral blood (PB)) and AML patient samples ($n = 57$ samples, from bone marrow aspirates (BM)). In total, and including all technical replicates, we sequenced and analyzed 90 samples.

The human AML cell lines HL-60 (*FLT3*-ITD negative control¹), MOLM-14 (*FLT3*-ITD positive control, 21 bp ITD¹) and PL-21 (*FLT3*-ITD positive control, 126 bp ITD¹) were obtained from the German cell line repository DSMZ (Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). The 57 AML samples were derived from 28 AML patients who were all included in the AMLSG BiO Registry study (ClinicalTrials.gov identifier: NCT01252485) and gave their informed consent for both biobanking and genomic analysis of leukemia samples according to the Declaration of Helsinki.

For initial assay assessment, we sequenced 3 *FLT3*-ITD negative controls (healthy volunteers, $n = 2$; AML cell line HL-60), 2 *FLT3*-ITD positive controls (AML cell lines MOLM-14 and PL-21) and diagnosis samples of 28 AML patients who previously tested *FLT3*-ITD positive by PCR- and capillary electrophoresis-based fragment analysis (FA). To further evaluate our assay in the context of MRD monitoring of *FLT3*-ITD positive AML, we sequenced 29 follow-up samples of 10 of these AML patients, five of whom achieved continuous complete remission and five of whom relapsed from the disease. For all patients, at least one sample from complete remission, following induction therapy, and at least one additional follow-up sample, either from continued complete remission or the time of relapse,

were sequenced. All of these 10 patients were also enrolled in the AMLSG 16-10 trial (ClinicalTrials.gov identifier: NCT01477606). Within this trial, patients were treated with intensive induction chemotherapy in combination with midostaurin followed by allogeneic stem cell transplantation (HSCT) as first priority for consolidation or an age-adapted high-dose cytarabine (HDAC) based chemotherapy regimen in combination with midostaurin. A one-year maintenance therapy with midostaurin was intended for all patients, starting after chemotherapy-based consolidation or 30 days after HSCT.

A detailed sample overview is provided in Table S2, including the attained sequencing coverage and *FLT3*-ITD statistics.

1.2 FLT3-ITD detection by NGS and getITD

Next-generation sequencing: Patient samples had been enriched for mononuclear cells by Ficoll gradient centrifugation and purified cells were biobanked at -80°C . Of these samples, genomic DNA was isolated using the AllPrep Mini Kit (Qiagen, Hilden, Germany) according to the manufacturers instructions. *FLT3* exons 14–15 were amplified by PCR, using 1 μL (50 ng) of genomic sample DNA, 20 μL of 2x KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, Massachusetts, USA) and 1.2 μL of 10 μM forward and reverse primer (forward primer: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGA + GCAATTTAGGTATGAAAGCCAGCTAC, reverse primer: CCAACGGCAGTTTTTACGACTTTC + AGACAGAGAATATGTGTAGAGGCTCGGGTGCTCTG, each consisting of *FLT3* locus-specific sequence and the required sequencing adapter). The PCR comprised an initial denaturation step (95 $^{\circ}\text{C}$ 3 min), 30 amplification cycles

(denaturation 98 °C 20 s, annealing 65 °C 30 s, elongation 72 °C 1 min) and a final elongation step (72 °C 5 min). PCR products were prepared for sequencing as detailed in the Illumina 16S Metagenomics Sequencing Protocol (16S Metagenomic Sequencing Library Preparation, Illumina, San Diego, California, USA) using either single or double unique indices for sample multiplexing. The resulting libraries were sequenced to high-coverage on the Illumina MiSeq using 250 bp paired-end reads (500-cycles MiSeq Reagent Kit V2, Illumina) and 7% phiX control spike-in (Kit V3, Illumina; coverage: 0.04–4.2 million, mean 1.6 million paired-end reads).

In total, library preparation takes 6 h, with roughly 3 h of hands-on time. Sequencing takes an additional 39 h, so that samples are fully processed within 48 h. For our validation experiments, we have aimed for a rather high coverage and pooled 8 samples per MiSeq run, generating on average around 2 million paired-end reads per sample. With this setup, assay costs amount to about 150 Euro per sample. However, more than 90% of these costs are spent on sequencing and only a minor fraction covers the sequencing library preparation. Costs could thus be reduced to about 100 Euro per sample if 16 instead of 8 samples were pooled, which would still generate a sufficient coverage of around 1 million reads per sample.

Sample indexing: We tested our assay with both single and double unique indices for sample barcoding and multiplexing. Comparing 14 technical replicates, derived from 14/28 *FLT3*-ITD positive AML patients analyzed for assay assessment, we found significant index mis-assignment, so-called index hopping, when using single indices only². Index hopping causes reads to be mis-assigned to the wrong sample in a pooled library when *FLT3*-ITD mutated reads are mis-assigned, they are then detected as false positive *contaminated* variants in these samples. For exam-

ple, in our 14 replicates sequenced with single unique indices, 22/61 *FLT3*-ITDs could be traced to index hopping. We therefore sequenced all of our AML patient samples thereafter using double unique indices, found zero ITDs attributable to index hopping, and thus recommend others also use double indexing whenever low-allelic variants are of interest. Details of the ITDs detected with either index design are provided in Table S3.

getITD: We developed the bioinformatics program *getITD* for the analysis of our assays sequencing data and the identification of respective ITDs. In brief, *getITD* aligns high-quality sequencing reads to the wild type (WT) amplicon, identifies insertions relative to this reference and then determines whether these qualify as ITDs. Insertions and ITDs are both extensively annotated and reported to the user.

By default, high-quality sequencing reads are those with an average base quality score (BQS) of at least 30, equivalent to a sequencing error probability below 0.1%, whose sequence is not unique in a given sample. Each of these reads is aligned to the WT sequence of the target amplicon using the Needleman-Wunsch alignment algorithm³, with alignment scores optimized for the detection of long insertions (match: 5, mismatch: -15, gap opening: -36, gap extension: -0.5). Only alignments with gap-free primer sequences and an alignment score of at least 50% of the maximum possible score are processed further.

This alignment score, which quantifies the similarity between the reference sequence and each of the samples reads, is calculated for each read as a running sum that increases whenever bases match between the read and the reference and decreases when they do not. Very low scores are indicative of PCR artefacts and

respective reads are thus filtered out.

From passing reads, insertions relative to the reference are extracted which are (i) in-frame, (ii) at least 6 bp / 2 amino acids long and (iii) free of ambiguous bases (N). Inserts at the very 5' or 3' end of a read, which we call trailing inserts, are not required to be in-frame - since they are not fully covered by the read, their exact length is unknown. Detected insertions, both trailing and non-trailing, are considered ITDs if they are adjacent to a second repeat of their own sequence (Figure 1B). These repeats, the respective ITDs WT tandems, are identified by realigning the insert to the WT reference and again require an alignment score of at least 50% of the maximum possible score. From all ITDs identified in this way, a set of high-confidence calls is obtained by filtering for those with a VAF of at least 0.006% and two or more distinct supporting reads.

While this is the primary output of interest, all of the insertions and ITDs identified as part of the analysis are annotated and reported. Specifically, each insertion and ITD is annotated with insertion length, site, sequence and frequency. ITD frequencies are provided as VAFs and, calculated from these as shown below in equation 1, allelic ratios (ARs). Insertion sites are given as coordinates relative to the amplicon, genome, transcript and protein.

Detailed results of *getITD* are presented in Table S3 for all 210 *FLT3*-ITDs described in the manuscript.

$$AR = \frac{VAF}{100 - VAF} \quad (1)$$

All analysis parameters are user-adjustable and can be tuned to accommodate for example single-end, merged or trimmed reads as well as reads generated from

other sequencing technologies or target regions⁴. Limits on the maximum detectable ITD length are imposed only by the sequencing data used for the analysis, not by *getITD* itself.

Code availability & details: *getITD* is implemented in python3, runs on Linux, Mac and Windows and is freely available for download at <https://github.com/tjblaette/getitd>. It may be run directly from the command-line or via a wrapper program that offers a simple graphical interface to supply input files and parameters. Analysis of a single sample with one million paired-end reads takes, depending on the computational resources available, 20 min with and up to 60 min without parallelization across multiple cores. Analysis of the data described in this manuscript was performed with default parameters and without any manual curation of results.

Fragment analysis: All patient samples were independently screened for *FLT3*-ITDs at diagnosis and relapse using the established diagnostic FA, a PCR-based amplification of *FLT3* exons 14 and 15 followed by capillary electrophoresis as described previously⁵. ITD insertion sites were determined by conventional Sanger sequencing⁵. To compare ITDs detected by FA and by NGS, which report mutation frequencies as ARs and VAFs respectively, the AR determined using the GeneScan platform was converted to the corresponding VAF as follows:

$$VAF = \frac{AR}{AR + 1} * 100 \quad (2)$$

Supplementary references

1. Quentmeier H, Reinhardt J, Zaborski M, and Drexler HG. FLT3 mutations in acute myeloid leukemia cell lines. *Leukemia* 2003; **17**: 120.
2. Kircher M, Sawyer S, and Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 2011; **40**: e3–e3.
3. Needleman SB and Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970; **48**: 443–453.
4. Rücker FG, Du L, Blätte TJ, Benner A, Krzykalla J, Gathmann I, et al. Prognostic Impact of Insertion Site in Acute Myeloid Leukemia (AML) with FLT3 Internal Tandem Duplication: Results from the Ratify Study (Alliance 10603). *Blood* 2018: 435 (abstract).
5. Kayser S, Schlenk RF, Londono MC, Breitenbuecher F, Wittke K, Du J, et al. Insertion of FLT3 internal tandem duplication in the tyrosine kinase domain-1 is associated with resistance to chemotherapy and inferior outcome. *Blood* 2009; **114**: 2386–2392.

MOLM-14 : HL-60	1:100	1:1000	1:10000
Expected VAF (%)	0.67	0.067	0.0067
VAF by NGS (%)	0.88043	0.08545	0.00483
VAF by NGS (%)	0.72653	0.07486	0.01509
VAF by NGS (%)	0.72068	0.05967	0.00073
VAF by NGS (%)			0.00075
Mean	0.77588	0.07333	0.00535
Standard deviation	0.09059	0.01296	0.00677
Coefficient of variation (CV)	0.11676	0.17677	1.26592

Table S1: Replicates of the serial dilution: For the three dilutions with expected *FLT3*-ITD frequencies 0.67%, 0.067% and 0.0067%, 3–4 independent technical replicates were analyzed using our NGS-based assay. Reported here are the exact variant allele frequencies (VAFs) estimated in each replicate, also plotted in Figure 1A of the manuscript, and the standard deviation, mean and coefficient of variation (CV, standard deviation / mean) for each of the three dilutions. Note that independent serial dilutions were pipetted for all replicates.

Dilution	Patient ID	Cohort_sensitivity	Cohort_specificity	Cohort_accuracy	Cohort_reproducibility	Cohort_mrd-rl	Cohort_mrd-rl	Material	Time point	Number of ITDs (FA)	AR of ITDs (FA)	VAF of ITDs (FA) (calculated)	ITD1 length & site (FA / Sanger)	ITD2 length & site (FA / Sanger)	Number of ITDs (getITD)	Total ITD VAF (getITD)	Number of FA-found ITDs (getITD)	Total VAF of FA-found ITDs (getITD)	Sample indexing	Coverage (read pairs)
-	1			yes	yes			BM	dx	1	0.192	16.10738255	54 (AS 598599)		2	13.2897	12.958	double	121444	
-	1							BM	dx	1	0.192	16.10738255	54 (AS 598599)		6	13.1415	12.68	single	502823	
-	2			yes				BM	dx	1	0.617	38.15708101	72 (AS 606607)		2	33.7071	33.695	double	1139145	
-	3			yes	yes			BM	dx	1	0.789	44.10285075	21 (AS 606601)		1	43.489	43.489	double	1142537	
-	3							BM	dx	1	0.789	44.10285075	21 (AS 606601)		3	44.4894	44.4894	single	283201	
-	4			yes	yes			BM	dx	1	17.169	94.49611976	39 (AS 593)		1	94.653	94.653	double	143104	
-	4							BM	dx	1	17.169	94.49611976	39 (AS 593)		6	89.1975	89.111	single	642035	
-	5			yes	yes			BM	dx	1	0.861	46.26544868	42 (AS 602)		1	44.728	44.728	double	2092251	
-	5			yes	yes			BM	dx	1	0.861	46.26544868	42 (AS 602)		3	47.007	47.007	single	321840	
-	6			yes	yes	yes		BM	dx	1	0.552	35.56701031	27 (AS 599)		5	37.2114	37.2114	double	2110204	
-	6							BM	dx	1	0.552	35.56701031	27 (AS 599)		5	38.6549	37.755	single	271565	
-	6				yes			BM	cr_cy2_c1						0	0	0	double	2688487	
-	6				yes			BM	cr_fu_m11						0	0	0	double	1987005	
-	7			yes				BM	dx	1	0.752	42.92274443	33 (AS 596597)		5	43.7485	43.702	double	1602251	
-	8			yes	yes	yes		BM	dx	1	0.686	40.68801898	36 (AS 608609)		2	44.7312	44.718	double	1358468	
-	8				yes			BM	dx	1	0.686	40.68801898	36 (AS 608609)		1	46.784	46.784	single	285460	
-	8				yes			BM	cr_cy2_c1						0	0	0	double	2415099	
-	8				yes			BM	cr_fu_m3						0	0	0	double	2956217	
-	8				yes			BM	cr_fu_m12						0	0	0	double	1895959	
-	9			yes				BM	dx	1	0.103	9.338168631	48 (AS 610)		1	3.4012	3.4012	double	779831	
-	10			yes				BM	dx	2	0.347	25.76095026	24 (AS 598)	33 (AS 594)	12	25.5057	24.677	double	2324362	
-	11			yes	yes			BM	dx	1	0.452	31.12947658	45 (AS 590591)		1	24.702	24.702	double	182304	
-	11				yes			BM	dx	1	0.452	31.12947658	45 (AS 590591)		3	27.636	27.6	single	1344506	
-	12			yes	yes	yes		BM	dx	1	0.768	43.43891403	45 (AS 598)		2	42.2052	41.874	double	2076011	
-	12				yes			BM	dx	1	0.768	43.43891403	45 (AS 598)		2	42.3822	42.105	single	543655	
-	12				yes			BM	cr_cy2_c1						0	0	0	double	1886671	
-	12				yes			BM	cr_fu_m12						0	0	0	double	1887112	
-	13			yes	yes	yes		BM	dx	2	0.276	21.63009404	60 (AS 613)	39 (AS 595596)	2	22.2246	22.2246	double	1423217	
-	13				yes			BM	dx	2	0.276	21.63009404	60 (AS 613)	39 (AS 595596)	3	22.2663	22.2529	single	383536	
-	13				yes			BM	cr_cy2_c1						0	0	0	double	1708522	
-	13				yes			BM	cr_eot_c4						0	0	0	double	2962904	
-	13				yes			BM	cr_fu_m3						0	0	0	double	2013235	
-	13				yes			BM	cr_fu_m6						0	0	0	double	2220894	
-	13				yes			BM	cr_fu_m9						0	0	0	double	1426831	
-	13				yes			BM	rl_fu_m11	0	0	0			2	0.21863	0	double	1888212	
-	14			yes				BM	dx	2	0.765	43.3427762	18 (AS 603)	30 (AS 593)	2	42.581	42.581	double	2474806	
-	15			yes	yes			BM	dx	1	0.525	34.42622951	24 (AS 601602)		1	31.827	31.827	double	38989	
-	15				yes			BM	dx	1	0.525	34.42622951	24 (AS 601602)		9	31.0303	30.846	single	549621	
-	16			yes				BM	dx	1	0.863	46.32313473	18 (AS 598)		5	42.8743	41.651	double	2601573	
-	17			yes				BM	dx	2	0.412	29.17847025	54 (AS 613)	90 (AS 606)	11	29.2862	28.646	double	1203484	
-	18			yes	yes	yes		BM	dx	1	5.419	84.42124942	27 (AS 608)		2	78.789	78.789	double	1160877	
-	18				yes			BM	dx	1	5.419	84.42124942	27 (AS 608)		1	77.659	77.659	single	237902	
-	18				yes			BM	cr_cy2_c1						0	0	0	double	2331331	
-	18				yes			BM	cr_eot_c4						0	0	0	double	2178063	
-	19			yes	yes	yes		BM	dx	2	0.195	16.31799163	36 (AS 596)	93 (AS 611612)	16	13.6331	8.0268	double	1268768	
-	19				yes			BM	cr_cy2_i2						4	0.65465	0	double	1460411	
-	19				yes			BM	cr_eot_bx						1	0.03156	0	double	1241329	
-	19				yes			BM	rl_fu_m4	1	0.299	23.01770593	93 (AS 611612)		8	17.9518	17.484	double	1071694	
-	20			yes				BM	dx	1	0.51	33.77483444	39 (AS 593)		3	28.6368	28.364	double	1454110	
-	21			yes	yes	yes		BM	dx	1	0.632	38.7254902	87 (AS 609610)		4	38.2656	38.089	double	2480855	
-	21				yes			BM	dx	1	0.632	38.7254902	87 (AS 609610)		3	34.866	34.724	single	2387119	
-	21				yes			BM	cr_cy2_i2						1	0.93816	0	double	2709486	
-	21				yes			BM	cr_eot_bx						0	0	0	double	2857166	
-	21				yes			BM	cr_fu_m3						1	0.060832	0	double	1131351	
-	21				yes			BM	dx	1	0.646	39.24665857	87 (AS 609610)		1	35.688	35.688	double	745374	
-	22			yes				BM	dx	1	0.892	47.14587738	39 (AS 611612)		3	46.1544	46.136	double	1095411	
-	23			yes		yes		BM	dx	1	0.76	43.18181818	54 (AS 600)		2	41.2423	41.156	double	1633097	
-	23				yes			BM	cr_cy2_c1						0	0	0	double	1942717	
-	23				yes			BM	cr_eot_c4						0	0	0	double	1394662	
-	24			yes				BM	dx	1	0.788	44.07158837	66 (AS 610)		1	39.744	39.744	double	2525191	
-	25			yes	yes	yes		BM	dx	2	0.398	28.46924177	42 (AS 590)	60 (AS 598)	7	29.083	27.903	double	736243	
-	25				yes			BM	dx	2	0.398	28.46924177	42 (AS 590)	60 (AS 598)	8	29.5654	28.573	single	789474	
-	25				yes			BM	cr_cy2_c1						6	2.22909	0	double	2705254	
-	25				yes			BM	cr_eot_c4						1	0.63748	0	double	1862886	
-	25				yes			BM	rl_fu_m1	1	0.744	42.66055046	60 (AS 598)		1	37.1815	37.103	double	2254869	
-	26			yes				BM	dx	1	0.178	15.11035654	36 (AS 608609)		3	15.5536	15.505	double	2254869	
-	27			yes				BM	dx	1	0.861	46.26544868	96 (AS 613)		2	48.3012	48.29	double	1160877	
-	28			yes	yes	yes		BM	dx	1	0.386	27.84992785	198 (AS 614615)		6	10.326	9.9464	double	1022779	
-	28				yes			BM	dx	1	0.386	27.84992785	198 (AS 614615)		8	20.9549	20.593	single	346073	
-	28				yes			BM	cr_cy2_c1						0	0	0	double	1483295	
-	28				yes			BM	rl_fu_m4	0	0	0			1	0.055593	0	double	2069770	
-	HL-60			yes				cell_line	-	0	0	0			0	0	0	double	4164257	
-	healthy_volunteer1			yes				PB	-	0	0	0			0	0	0	double	1137581	
-	healthy_volunteer2			yes				PB	-	0	0	0			0	0	0	double	2441834	
10-0	MOLM-14	yes						cell_line	-	1	KNOWN 2_FA.2.451	KNOWN 66.7_FA.71.02289192	21 (AS 599)		1	64.979	64.979	double	1797368	
10-1	MOLM-14	yes						cell_line	-	1	KNOWN 0.0718	KNOWN 6.7	21 (AS 599)		1	8.6377	8.6377	double	2698207	
10-2	MOLM-14	yes						cell_line	-	1	KNOWN 0.00675	KNOWN 0.67	21 (AS 599)		1	0.72068	0.72068	double	1688834	
10-2	MOLM-14	yes						cell_line	-	1	KNOWN 0.00675	KNOWN 0.67	21 (AS 599)		1	0.72653	0.72653	double	1629789	
10-2	MOLM-14	yes						cell_line	-	1	KNOWN 0.00675	KNOWN 0.67	21 (AS 599)		1	0.89043	0.89043	double	2661959	
10-3	MOLM-14	yes						cell_line	-	1	KNOWN 0.000670	KNOWN 0.067	21 (AS 599)		1	0.059666	0.059666	double	1139435	
10-3	MOLM-14	yes						cell_line	-	1	KNOWN 0.000670	KNOWN 0.067	21 (AS 599)		1	0.074857	0.074857	double	1261903	
10-3	MOLM-14	yes						cell_line	-	1	KNOWN 0.000670	KNOWN 0.067	21 (AS 599)		1	0.085454	0.085454	double	2850841	
10-4	MOLM-14	yes						cell_line	-	1	KNOWN 0.000670	KNOWN 0.0067	21 (AS 599)		1	0.0046347	0.0046347	double	2408074	
10-4	MOLM-14	yes						cell_line	-	1	KNOWN 0.000670	KNOWN 0.0067	21 (AS 599)		1	0.015097	0.015097	double	1609750	
10-4	MOLM-14	yes						cell_line	-	1	KNOWN 0.000670	KNOWN 0.0067	21 (AS 599)		1	0.000730				

Patient ID	Age at dx (years)	Sex	Type of AML	Time from dx to rl (days)	CR after induction	BM blasts at dx (%)	PB blasts at dx (%)	BM blasts at rl (%)	PB blasts at rl (%)	<i>NPM1</i> -mutated at dx	<i>NPM1</i> transcripts at dx (in BM)	<i>NPM1</i> transcripts at rl (in BM)	Induction cycles (Dauna-ARAC 7+3)	Consolidation cycles (High dose ARAC)	Transplanted in 1 st CR	WBC at dx (G/l)	WBC at rl (G/l)
13	65.15	male	de novo	561.00	CRi	95	98	5	0	yes	104882	22022	1	4	no	166.6	5.2
19	25.03	male	de novo	182.00	CR	69	11	32	na	no	na	na	2	0	yes	10.7	na
21	52.14	male	de novo	325.00	CRi	90	59	43	4	no	na	na	2	0	yes	na	2.3
25	53.49	male	de novo	245.00	CR	80	37	90	25	yes	537470	3682990	1	4	no	na	6.6
28	51.52	male	de novo	289.00	CRi	80	29	na	8	no	na	na	1	3	no	52.8	3.5

Table S4: Clinical data: Provided are clinical details of five *FLT3*-ITD positive AML patients that relapsed during follow-up (Figure 2A).