A Multi-Platform Approach for the Detection of Mutations in 54 genes with Relevance to Myeloid Neoplasms in a Clinical Laboratory Setting

Charles Ma PhD, Venkata Thodima PhD, Asha Reddy MS, Kshitija Desai, Weiyi Chen PhD, Sitharthan Kamalakaran PhD, and Jane Houldsworth PhD
Cancer Genetics, Inc.

Introduction

Somatic tumor mutations within specific genes offer significant diagnostic/prognostic values for many myeloid malignancies and detection of them have been incorporated into clinical guidelines. Most of the reported mutations are SNVs or small indels (<25bp) and can be reliably detected by NGS. However, larger indels important in these diseases (FLT3-ITD, and CALR and KIT deletions) as well as high GC content gene (CEBPA) are challenging by NGS. In order to implement comprehensive mutation testing in a clinical setting for the myeloid neoplasms (AML, MDS, and MPN), we have developed a multiplatform approach combining NGS, Sanger sequencing, and PCR fragment analysis.

FOCUS::Myeloid

- Designed by Illumina for MDS, MPN, MDS/MPN, and AML
- Genes with both diagnostic and prognostic significance

Gene Name	Target (exons)	Gene Name	Target (exons)
ABL1	4 to 6	*STAG2*	full
ASXL1	12	*TET2*	3 to 11
BCOR	full	*TP53*	2 to 11
CALR	9	*U2AF1*	2, 6
CBL	8 to 9	*ZRSR2*	full
CEBPA	full	*CSF3R*	14-17
CUX1	full	*ATRX*	8-10, 17-31
DNMT3A	full	*BCORL1*	full
ETV6	full	*BRAF*	15
EZH2	full	*CBLB*	9 to 10
GATA2	2 to 6	*CBLC*	9 t0 10
IDH1	4	*CDKN2A*	full
IDH2	4	*FBXW7*	9 to 11
JAK2	12 , 14	*FLT3*	14, 15, 20
KIT	2, 8-11, 13, 17	*GATA1*	2
KMT2A (MLL)	5 to 8	*GNAS*	8 to 9
KRAS	2 to 3	*HRAS*	2 to 3
MPL	10	*IKZF1*	full
NPM1	12	*JAK3*	13
NRAS	2 to 3	*KDM6A*	full
PDGFRA	12, 14, 18	*MYD88*	3 to 5
PTPN11	3, 13	*NOTCH1*	26-28, 34
RUNX1	full	*PHF6*	full
SETBP1	4 partial	*PTEN*	5, 7
SF3B1	13-16	*RAD21*	full
SMC1A	2,11,16,17	*SMC3*	10,13, 19, 23,25,28
SRSF2	1	*WT1*	7,9

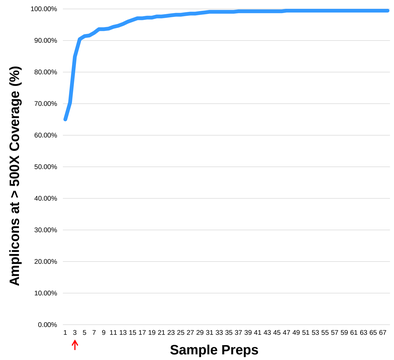
Mutation analysis recommended in NCCN Guideline for MDS 2015

Materials and Methods

Extracted DNAs from a combination of clinical samples (AML, MDS, MPN, and 7 normals, majority with prior known variants by Sangers sequencing or quantitative PCR) and cell line cultures were submitted to targeted, multiplex-PCR based sequencing for 54 genes (TruSight Myeloid, Illumina, Inc.). Diluted Jurkat DNA and NA12878 DNA served as positive and negative controls respectively. Somatic Variant Caller within MiSeq Reporter was used to detect variants, with annotation by Variant Studio (Illumina). Verification of filtered variants was done by Sangers sequencing (>20% AVF) or repeat NGS (<20% AVF). Sangers sequencing was performed for CEBPA (due to high GC content), KIT (exon 11), and CALR (exon 9). Fragment analysis was used to detect FLT3-ITD.

Sample Performance Profile

Figure 1



65 out 68 independent library preparations achieved greater than 90% of amplicons with a coverage of at least 500x. Based on this, samples with less than 90% of amplicons with 500x are not considered further for analysis and recommended for re-assay.

Amplicon Performance Profile

To identify poor-performing amplicons out of the total of 573 amplicons, the % sample preparation and reactions with greater than 500x coverage for each of the 573 amplicons were plotted as shown in Figure 2. Figure 3 is the same plot but shows in greater detail those amplicons for which less than 90% of samples have less than 500x coverage.

Figure 2

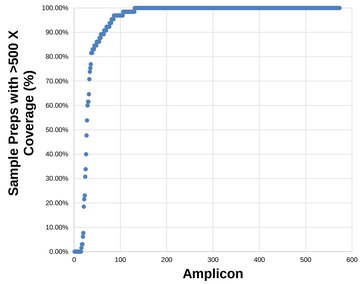
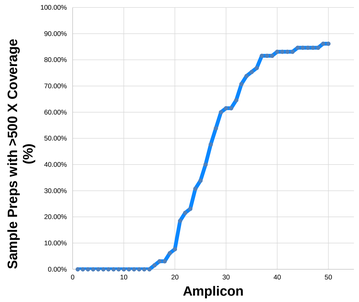


Figure 3



Amplicons that achieved >500X coverage in less than 60% of cases >500x coverage are divided into the following categories: the CEBPA gene which will be sequenced independently due to its significant prognostic value in AML, amplicons exhibit little (<1%) or no evidence of mutation in the target region in the relevant assayed diseases will not be considered further for analysis, and amplicons exhibit mutations in the diseases assayed with at least 100x coverage consistently, these will be analyzed further but with reduced limit of detection (20%).

NGS Accuracy

44 specimens each bearing known variants in genes tested in the NGS assay were used to establish the assay accuracy. The variants included both single nucleotide variants (SNV) and small insertion/deletions (indels) previously identified by Sanger's sequencing or real-time QPCR. Only for the latter methodology was the AVF known, and permitted selection of specimens with AVF down to as low as 2%.

Total number of variants previously known: 47
Reference method: Previous Sanger sequencing or real-time QPCR
Total number of expected variants detected by NGS: 47
Concordance: 100%

NGS Sensitivity

In order to achieve a sensitivity of 5% , dilutions of Jurkat DNA mixed with the Hapmap DNA at around the limit of detection were prepared. There were 12 variants within the target regions assessed in the current NGS panel that were found in the undiluted Jurkat cell line as previously reported in the COSMIC database. Based on the AVF detected for the 12 variants by NGS in a total of 8 runs, the average AVF and SD for each variant were calculated and a cutoff at 3% was established so variants present at 5% frequency can be reliably and consistently detected.

Gene	Variant	Exp AVF	Aver AVF	StDev AVF	Number of Times Variant Not Detected	Average Coverage
					≥ 3% cut-off	
TP53	c.1083delG	6.98	5.90	1.88	0	5,700
CUX1	c.2932G>A	5.29	6.59	2.15	0	10,258
DNMT3A	c.901C>T	5.25	5.17	1.96	1	4,945
ASXL1	c.2841G>T	4.69	5.27	1.41	0	7,723
GATA2	c.1361C>A	4.69	4.09	2.75	2	1,314
IZKF1	c.1227C>T	4.22	3.83	2.08	3	1,044
GATA1	c.44_45insC	5.73	7.22	8.76	1	790
CSF3R	c.2503G>A	4.66	5.72	1.07	0	4,369
TP53	c.586C>T	4.54	5.87	1.36	0	1,487
FBXW7	c.1513C>T	4.07	6.18	4.15	0	3,729
RUNX1	c.445G>A	3.28	4.59	0.89	0	16,245
KMT2A	c.1279C>T	4.57	5.17	0.39	0	9,126
Average		***4.83***	***5.47***	***2.41***		

Similar results were obtained using dilution of two clinical samples in inter and intra runs.

NGS Specificity

Analytical specificity of the NGS Panel was assessed by the repeat sequencing of HapMap DNA in all runs and by sequencing normal samples with no known disease history. No new variants were identified and confirmed across over multiple runs.

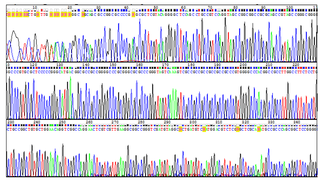
NGS Limitations

- Targeted regions that have significant prognostic value have low or poor coverge i.e. CEBPA
- Large indels are difficult to detect using the standard data analysis workflow (implementing Pindel in the future)

Supplemental Tests

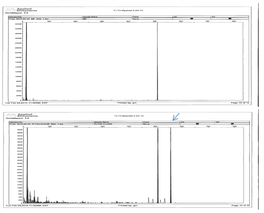
Sanger Sequencing

AML — CEBPA* KIT (exon 11)
MDS — CEBPA*
MPN — CALR (exon 9)



Fragment Analysis

AML — FLT3-ITD
MDS — FLT3-ITD



Conclusions

To facilitate disease identification and/or risk stratification in AML, MDS, and MPN, we have developed a multi-platform approach, permitting robust mutation analysis of 54 relevant genes. This approach is necessary due to intrinsic limitations observed in the NGS sequencing panel. In the future, optimization of amplicon design and data analysis pipeline will enable more robust assaying based on NGS while other platforms can serve as secondary, independent validations of the NGS results.

Reference and Acknowledgement

* Behdad *et al. JMD* 17(1):76-84
We thank Dr. Bryan Betz for suggestions on CEBPA sequencing.