Copy number alteration detection by targeted next generation sequencing in 14985 real-world samples

Grant Hogg[1], Eric A Severson[2], Qian Zeng[3], Heidi Hoffmann[3], Li Cai[2], Stuart Schwartz[2], Wenjie Chen[3], Sabrina Gardner[2], Kimberly A Holden[1], Lax Iyer[3], Angela Kenyon[3], Deborah Boles[2], Scott Parker[2], Stanley Letovsky[3], Henry Dong[4], Narasimhan Nagan[3], Marcia Eisenberg[5], Anjen Chenn[2], Taylor J Jensen[1,2]
[1]Laboratory Corporation of America® Holdings, San Diego, CA; [2]Laboratory Corporation of America® Holdings, Durham, NC; [3]Laboratory Corporation of America® Holdings, Westborough, MA; [4]Laboratory Corporation of America® Holdings, New York, NY; [5]Laboratory Corporation of America® Holdings, Burlington, NC

1. Introduction

In myeloid neoplasms cytogenetic and molecular analyses are standard of care for initial clinical evaluation. Next generation sequencing (NGS) is routinely used to identify somatic single nucleotide variants and insertions/deletions (<50 bp) (SNV/Indels) in clinically relevant genes to direct treatments, determine prognosis, and confirm diagnoses. Here we investigate whether data from a targeted NGS panel for SNVs/Indels can identify copy number alterations (CNAs).

2. Methods

Real-world data from a targeted NGS panel (IntelliGEN® Myeloid) were retrospectively analyzed to identify CNAs for research purposes. The NGS panel was initially developed, validated, and implemented as a laboratory developed test to detect SNVs/Indels in 50 genes and CNAs in *KMT2A*. A cohort of 14,985 clinical samples tested for myeloid malignancies including acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), and myeloproliferative neoplasms (MPN) were re-analyzed for CNAs in 12 clinically relevant genes (*BCOR, CBL, CDKN2A, CUX1, IKZF1, KMT2A, NF1, RUNX1, TET2, TP53, WT1*, and *ZRSR2*). The historical median of normalized read counts was calculated for each gene analyzed. Read counts for each sample were then normalized to the historical median for each gene. Copy number thresholds were set to 0.85 for deletions and 1.15 for gains. CNAs were filtered for those with robust Z-scores ≤-3 for copy number deletions and ≥ 3 for copy number gains. A CNA p-value was calculated using a two-tailed Wilcoxon rank sum test to determine whether copy numbers for individual gene-specific primers were different to gene-specific primers identified as being in the baseline. Based on performance of *KMT2A* CNV detection, a significance threshold of 0.03 was used to identify CNAs in conjunction with the robust Z-score values. CNAs in 10 of the 12 genes analyzed were orthogonally validated in 66 samples by digital multiplexed ligation dependent amplification (dMLPA) using the SALSA® digitalMLPA Probemix D031-X1-0322 and Coffalyser v220221.1522 (MRC Holland). CNAs for *KMT2A* were orthogonally validated in 463 samples using qPCR and previously reported.

3. Conclusion

These data suggest that a targeted NGS panel for SNVs/indels can also detect somatic CNAs. Gene-specific thresholds could be applied to reduce the number of false positives observed in *KMT2A*. These findings will be used to inform the development of improved NGS panels for somatic mutation and CNA detection. Potential applications of such a panel include monitoring genomic instability in response to treatment.

Tables + Figures

Figure 1. Identification of copy number alterations by NGS

Figure shows robust Z scores for 12 genes calculated by deviation from the historical median. X-axis shows gene name. Y-axis shows robust Z-score. Dotted lines show Z-score thresholds of ±3. Teal colored points identify CNAs where all individual gene-specific primers for each gene had copy numbers greater than baseline (p-value ≤0.03). CNAs were identified when both Z-score and p-value threshold were satisfied.

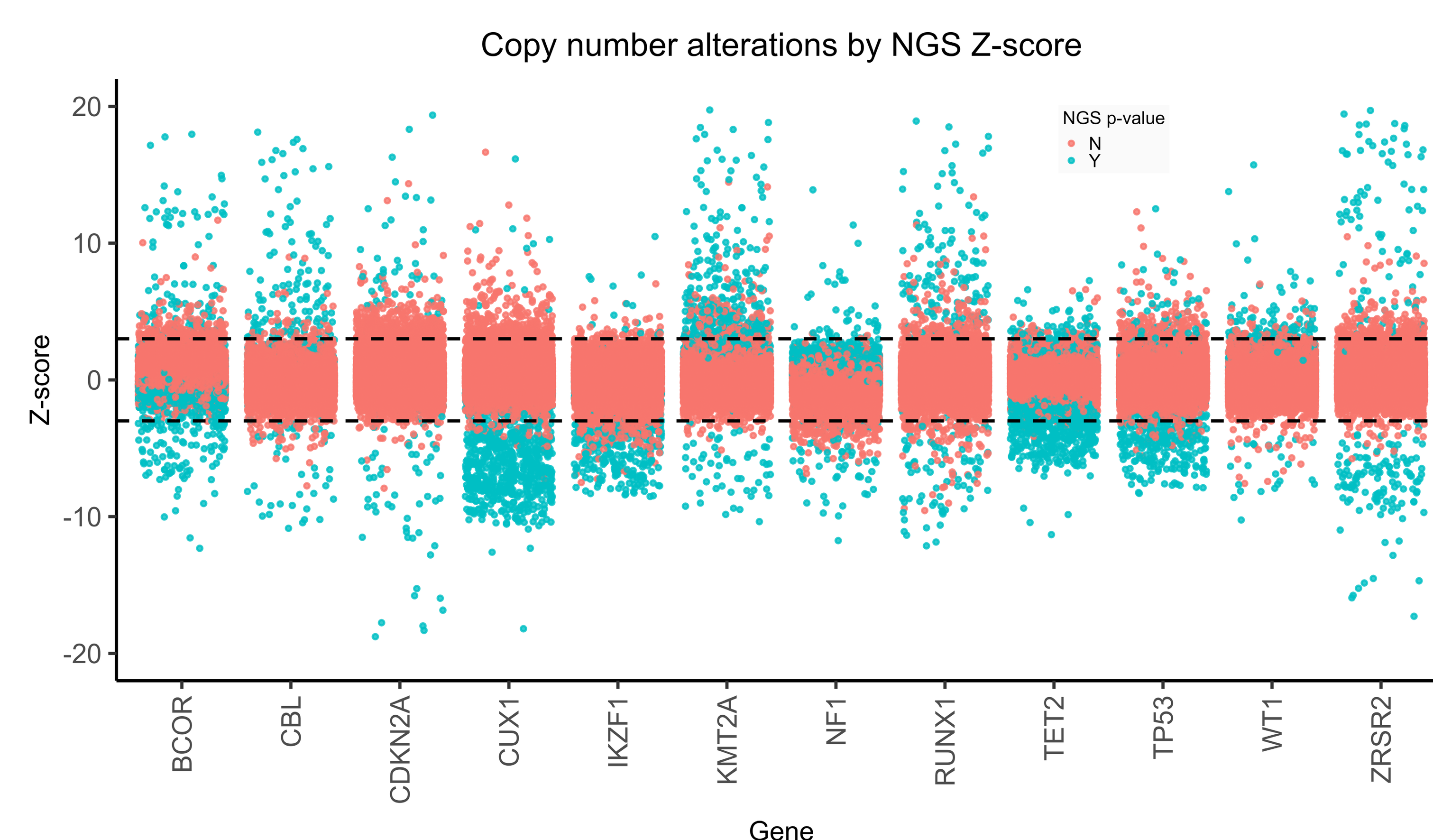


Figure 2. Number of samples with an NGS copy number alteration

A) Number of samples with a copy number alteration (CNA) detected by NGS by test date. Plot shows the percentage of samples with a CNA detected in the 12 genes assessed in 14985 samples. X-axis shows the date of testing by month for the study evaluation period. Y-axis shows the % of samples with a CNA. Dotted blue line is a Loess fit of the percentages observed for each month. This proportion was stable over time, suggesting a consistent detection rate. **B) Number of samples with a CNA detected by NGS by gene.** Plot shows the percentage of samples with a CNA detected in 14985 samples. CNAs in 12 genes were observed in 12.6% of samples (1892/14985). The most common CNA was observed spanning the gene CUX1 (624 samples, 4.2%). *CUX1* is found in the clinically relevant 7q region.

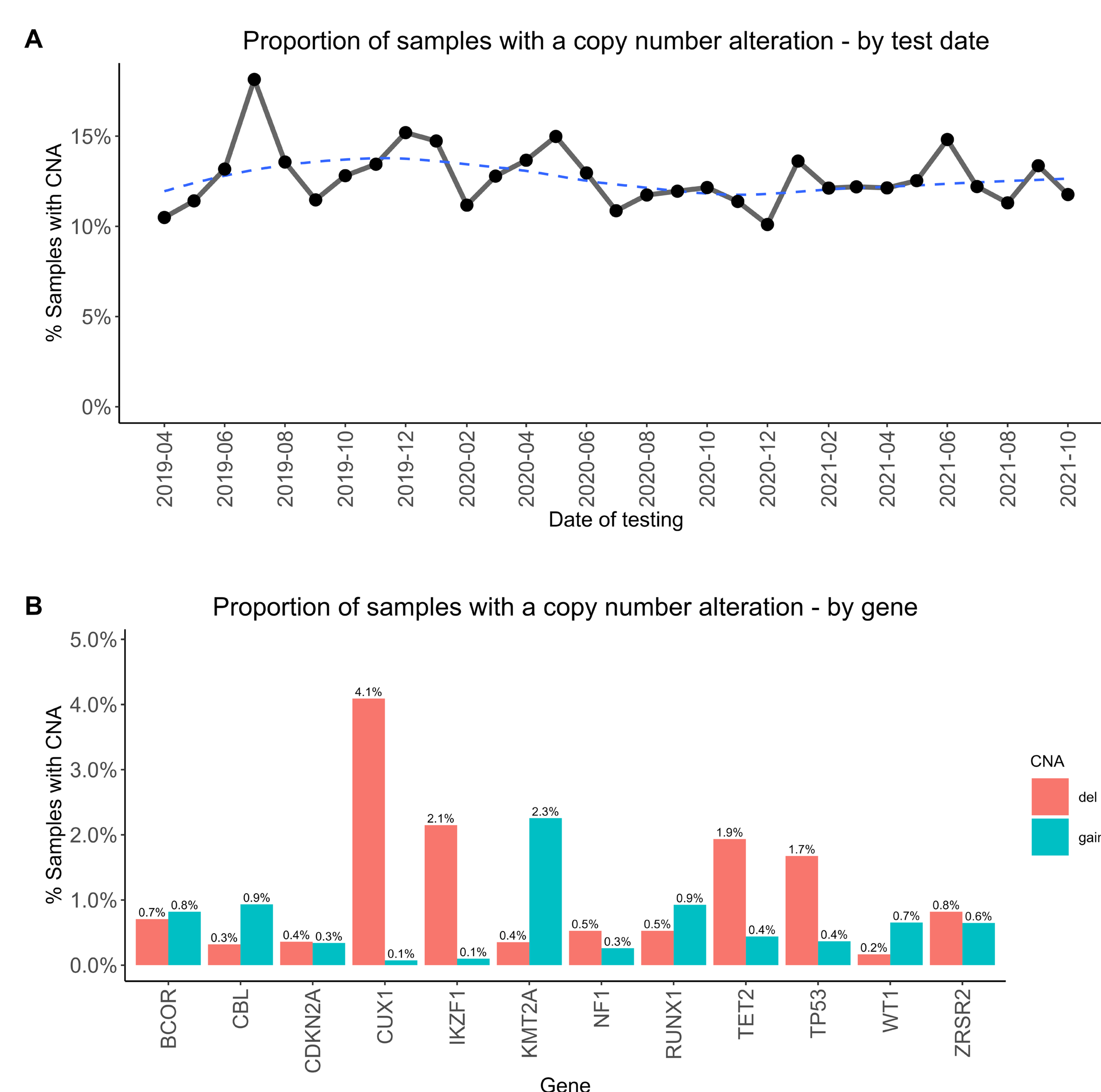
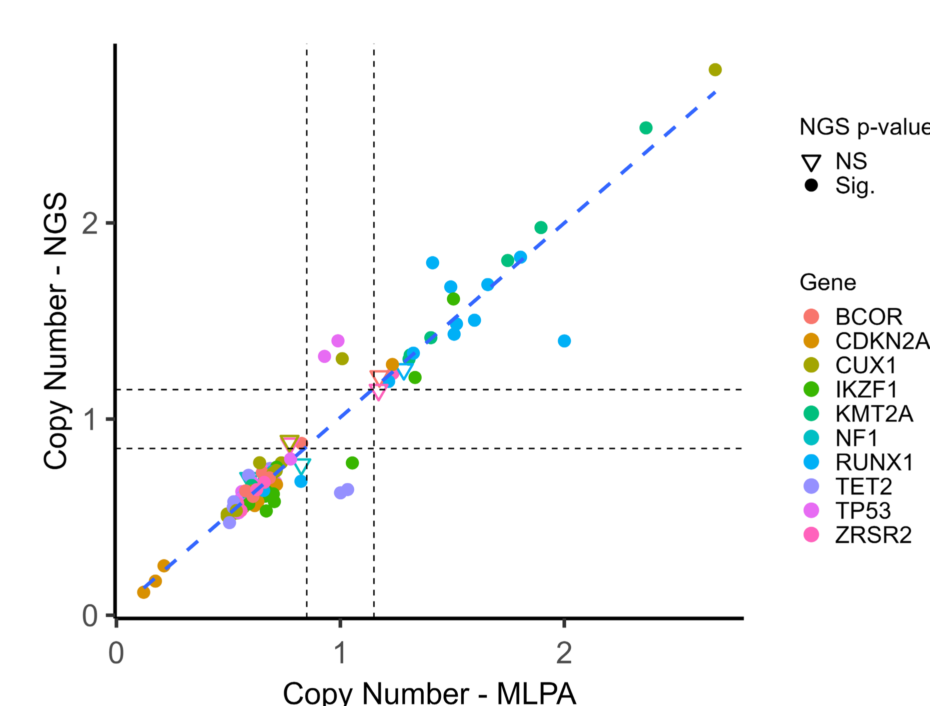


Figure 3. Concordance of NGS copy number values with dMLPA and qPCR

A Correlation of copy number values by NGS and dMLPA



B NGS concordance with dMLPA

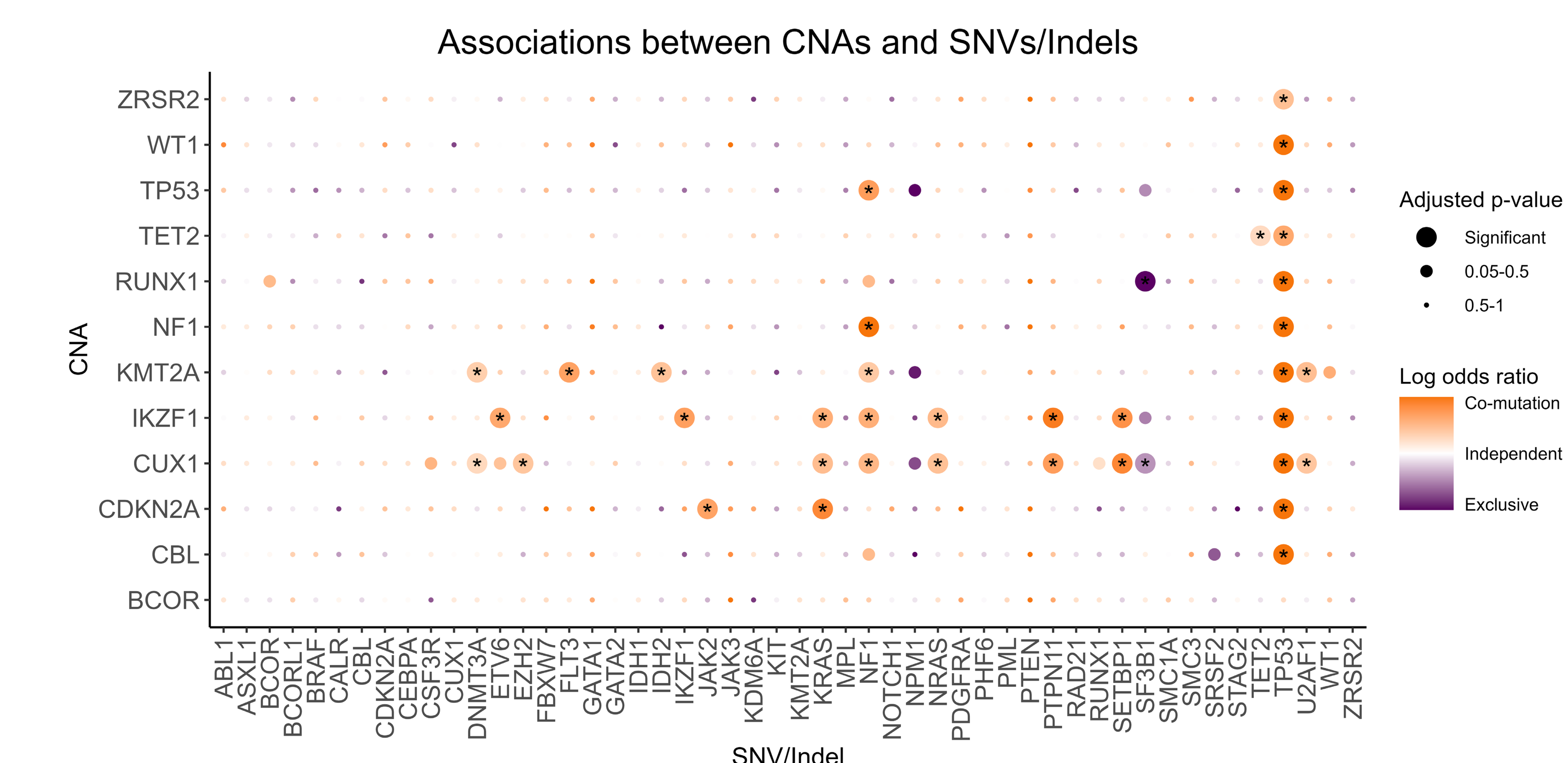
		NGS		
		deletion	gain	normal
dMLPA	deletion	84	0	6
	gain	0	21	3
	normal	3	3	0

C NGS concordance with qPCR (*KMT2A* only)

		NGS	
		deletion	gain
qPCR	deletion	44	0
	gain	0	419
	normal	1	161

A) Correlation of copy number values detected by NGS and dMLPA. NGS CNAs for 10 of the 12 genes were orthogonally confirmed in a subset of 66 samples by digital multiplexed ligation-dependent probe amplification (dMLPA). X-axis shows the copy number values measured by dMLPA. Y-axis shows the copy number values measured by NGS. Dotted lines show copy number calling thresholds of 0.85 for deletions and 1.15 for gains. Copy numbers are colored by gene measured. Dashed blue line shows linear fit of NGS and dMLPA copy numbers. Triangle denotes NGS calls that were not called due to CNA p-value ≥0.03. 17 ChrX CNAs from 9 samples that were found to have a discrepant gender between dMLPA and the stated gender on the requisition were excluded from analysis. 5 sub-gene level CNVs in 5 samples identified by dMLPA were also excluded because the NGS algorithm was specifically designed to call whole gene CNAs. A significant correlation was observed between NGS and dMLPA copy numbers (n=120, slope=0.99, Pearson's product-moment correlation adjusted R^2=0.93, $p<2.2 \times 10^{-16}$). **B) NGS concordance with dMLPA.** Concordance is shown as a contingency table. A confirmation rate of 94.6% was observed (105/111 CNAs in 57 samples). **C) NGS concordance with qPCR.** CNAs for *KMT2A* were analyzed in 625 samples by qPCR. Concordance is shown as a contingency table. 463 of 625 (74.1%) *KMT2A* CNAs were concordant between NGS and qPCR.

Figure 4. Pair-wise associations between copy number alteration and small sequence variants in the same sample



X-axis shows the gene with an SNV/Indel, Y-axis shows the corresponding CNA gene. Circular labels are sized by the p-value calculated from the Fisher Exact test with the largest size corresponding to the lowest probability as shown on the upper legend on the right. Associations that are significant after correction for multiple hypotheses are indicated by a * inside the circular label. The effect sizes of the associations are shown by the odds ratio colored according to a gradient from orange (co-mutation) to white (no association) to purple (mutual exclusivity), as shown on the lower legend on the right. *TP53* sequence mutations significantly co-occurred with CNAs in 11 of the 12 CNA genes (odds ratio=2.5 to 42.4, adjusted p<0.002, all pairs) consistent with the known association between *TP53* mutations and complex karyotypes[1]. The largest effect size for co-occurrence was *TP53* CNA with *TP53* sequence mutation (odds ratio=42.4) suggesting that this mutation combination is a common mechanism of *TP53* loss. Two mutually exclusive pairs were identified including SF3B1 sequence mutation with *RUNX1* CNA (odds ratio=0.11, adjusted p=0.002). This is in contrast to the co-mutation between *SF3B1* sequence mutation and *RUNX1* sequence mutation observed in AML[2].

References

1. Rücker, F. G. et al. TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. *Blood* 119, 2114–2121 (2012).
2. K. H. Metzeler et al., Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood*. 128, 686–698 (2016).

Correspondence: Grant Hogg hoggg@labcorp.com
7th AGBT Precision Health Meeting 2022, September 8-10, San Diego
©2022 Laboratory Corporation of America® Holdings. All rights reserved. onc-2165-v1-0822

labcorp