4294

Bridging the gap between targeted NGS and FISH gene-level CNV detection capabilities in hematologic malignancies

Christophe N. Magnan • Hyunjun Nam • Shashikant Kulkarni • Segun C. Jung
NeoGenomics Laboratories, Carlsbad, CA, San Diego, CA, Houston, TX, and Aliso Viejo, CA

Background

Copy Number Variations (CNVs) are prominent features of cancer cells. From a clinical standpoint, their accurate detection at a low cost is a priority. With regular increases in the number of markers to be tested, the cost effectiveness and practicality of gold standard techniques like Fluorescence In Situ Hybridization (FISH) are slowly decreasing. Cost-efficient Next Generation Sequencing (NGS) targeted gene panels can be scaled up but accurately detecting CNVs from the resulting data remains challenging. We demonstrate large amounts of data and machine learning can help bridge the gap between the two techniques.

Methods

We collected the sequencing data for 6,277 patients tested using a custom amplicon based NGS assay designed to detect somatic alterations in 297 hematological cancer relevant genes such that at least one concurrent FISH test was also performed. FISH results were used to infer the gain, loss, or normality information for both the gene directly targeted by the FISH probe (reported as direct strategy in the various tables) or by using inference rules such as the observed loss of centromere 7 results in the loss of all targeted genes on chromosome 7 (reported as indirect strategy in the tables). The annotated genes were then used to curate a training set by extracting 20 features per gene from the alignment results. 10 of these features were collected from existing CNV detection methods (PureCN [1], CNVkit [2]) while 10 others are custom normalizations of the gene coverage designed to correct the high coverage variability that comes with amplicon assays. A random forest classifier was trained using this dataset. The selected model was evaluated on a distinct set of 2,738 patients sequenced using the same NGS assay and for which at least one concurrent FISH test was available.

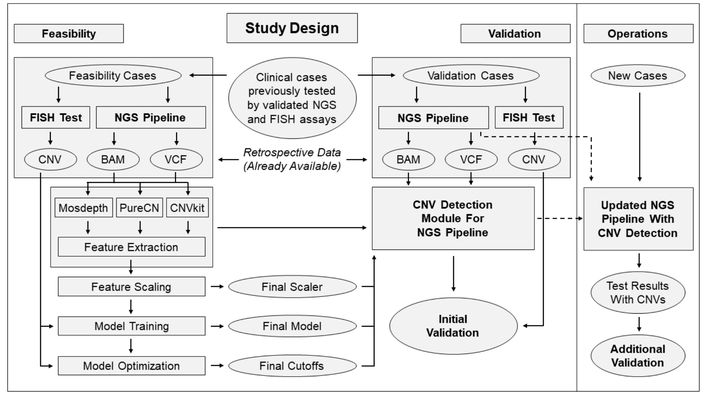
Results

Evaluation results are provided in the various tables on the side for both the 8 genes for which the FISH probe used to infer the gene gain, loss, or normality information directly spanned the gene region and for all 62 genes that could be evaluated using either a direct or an indirect marker. The predicted CNVs are almost a perfect match with the FISH results with a limit of detection at 20% abnormal cells. In most cases, the model reduces discordant calls by over 50% compared to using existing CNV detection software only. The last two tables allow to compare the NGS performances with the actual position of the FISH probe(s) used to label the genes and show that the concordance between NGS and FISH is maximal when compared at the exact same location. These results suggest the lower measured accuracy on some genes located far from the FISH probes may simply be a side-effect of using indirect markers to label the genes.

Conclusion

We show the CNV detection capabilities of a targeted NGS assay can closely match the gold-standard FISH technique by analytically correcting the biases introduced by the targeting procedures. The model presented here is used to detect CNVs in ALL patients after a successful formal validation in our laboratory.

Study Design



Complete Validation Results

Gene	FISH Test	FISH Probes	Strategy	Direction	Positive Cases			Negative Cases			All Cases		
					Total	Conc.	Sensitivity	Total	Conc.	Specificity	Total	Conc.	Accuracy
APC	Del(5q)	EGR1 & RPS14	Indirect	loss	164	155	**94.51%**	1,546	1,525	**98.64%**	1,710	1,680	**98.25%**
ARID2	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	569	**99.48%**	698	693	**99.28%**
ATM	ATM Deletion	ATM	Direct	loss	72	72	**100.00%**	641	629	**98.13%**	713	701	**98.32%**
ATP2A2	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	567	**99.13%**	698	691	**99.00%**
BRAF	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	106	**100.00%**	1,602	1,581	**98.69%**	1,708	1,687	**98.77%**
	Monosomy 7	CEN 7	Indirect	loss	56	56	**100.00%**	1,659	1,593	**96.02%**	1,715	1,649	**96.15%**
CARD11	Monosomy 7	CEN 7	Indirect	loss	56	52	**92.86%**	1,659	1,636	**98.61%**	1,715	1,688	**98.43%**
CBFB	Rearrangement	CBFB	Direct	loss	23	22	**95.65%**	517	506	**97.87%**	540	528	**97.78%**
CCND2	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	565	**98.78%**	698	689	**98.71%**
CDC25C	Del(5q)	EGR1 & RPS14	Indirect	loss	164	161	**98.17%**	1,546	1,533	**99.16%**	1,710	1,694	**99.06%**
CDK2	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	564	**98.60%**	698	688	**98.57%**
CDK4	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	567	**99.13%**	698	691	**99.00%**
CDK6	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	100	**94.34%**	1,602	1,583	**98.81%**	1,708	1,683	**98.54%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,599	**96.38%**	1,715	1,654	**96.44%**
CDKN1B	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	564	**98.60%**	698	688	**98.57%**
CSF1R	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,538	**99.48%**	1,710	1,700	**99.42%**
CUX1	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	103	**97.17%**	1,602	1,571	**98.06%**	1,708	1,674	**98.01%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,583	**95.42%**	1,715	1,638	**95.51%**
DDX41	Del(5q)	EGR1 & RPS14	Indirect	loss	164	143	**87.20%**	1,546	1,528	**98.84%**	1,710	1,671	**97.72%**
EBF1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	151	**92.07%**	1,546	1,477	**95.54%**	1,710	1,628	**95.20%**
EGFR	Monosomy 7	CEN 7	Indirect	loss	56	54	**96.43%**	1,659	1,621	**97.71%**	1,715	1,675	**97.67%**
EGR1	Del(5q)	EGR1	Direct	loss	171	169	**98.83%**	1,541	1,538	**99.81%**	1,712	1,707	**99.71%**
	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,538	**99.48%**	1,710	1,700	**99.42%**
ERBB3	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	563	**98.43%**	698	687	**98.42%**
ETNK1	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	564	**98.60%**	698	688	**98.57%**
ETV6	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	568	**99.30%**	698	692	**99.14%**
EZH2	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	104	**98.11%**	1,602	1,579	**98.56%**	1,708	1,683	**98.54%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,593	**96.02%**	1,715	1,648	**96.09%**
FGFR1	Trisomy 8	CEN 8	Indirect	gain	93	86	**92.47%**	1,590	1,590	**100.00%**	1,683	1,676	**99.58%**
GNA12	Monosomy 7	CEN 7	Indirect	loss	56	51	**91.07%**	1,659	1,600	**96.44%**	1,715	1,651	**96.27%**
IKBKB	Trisomy 8	CEN 8	Indirect	gain	93	89	**95.70%**	1,590	1,575	**99.06%**	1,683	1,664	**98.87%**
IKZF1	Monosomy 7	CEN 7	Indirect	loss	56	54	**96.43%**	1,659	1,603	**96.62%**	1,715	1,657	**96.62%**
IRAK4	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	568	**99.30%**	698	692	**99.14%**
KMT2A	Rearrangement	KMT2A	Direct	gain	27	25	**92.59%**	474	472	**99.58%**	501	497	**99.20%**
KMT2C	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	105	**99.06%**	1,602	1,580	**98.63%**	1,708	1,685	**98.65%**
	Monosomy 7	CEN 7	Indirect	loss	56	56	**100.00%**	1,659	1,593	**96.02%**	1,715	1,649	**96.15%**
KMT2D	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	568	**99.30%**	698	692	**99.14%**
KRAS	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	569	**99.48%**	698	693	**99.28%**
LUC7L2	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	104	**98.11%**	1,602	1,579	**98.56%**	1,708	1,683	**98.54%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,594	**96.08%**	1,715	1,649	**96.15%**
MAP3K1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	146	**89.02%**	1,546	1,487	**96.18%**	1,710	1,633	**95.50%**
MDM2	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	563	**98.43%**	698	687	**98.42%**
MET	Del(7q)	D7S2460	Direct	loss	113	113	**100.00%**	1,593	1,582	**99.31%**	1,706	1,695	**99.36%**
	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	106	**100.00%**	1,602	1,584	**98.88%**	1,708	1,690	**98.95%**
	Monosomy 7	CEN 7	Indirect	loss	56	56	**100.00%**	1,659	1,596	**96.20%**	1,715	1,652	**96.33%**
MYC	Trisomy 8	CEN 8	Indirect	gain	93	89	**95.70%**	1,590	1,559	**98.05%**	1,683	1,648	**97.92%**
NBN	Trisomy 8	CEN 8	Indirect	gain	93	89	**95.70%**	1,590	1,551	**97.55%**	1,683	1,640	**97.45%**
NF1	Monosomy 17	NF1	Direct	loss	15	15	**100.00%**	908	908	**100.00%**	923	923	**100.00%**
NHP2	Del(5q)	EGR1 & RPS14	Indirect	loss	164	153	**93.29%**	1,546	1,496	**96.77%**	1,710	1,649	**96.43%**
NPM1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	122	**74.39%**	1,546	1,508	**97.54%**	1,710	1,630	**95.32%**
NSD1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	157	**95.73%**	1,546	1,499	**96.96%**	1,710	1,656	**96.84%**
PDGFRB	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,537	**99.42%**	1,710	1,699	**99.36%**
PIK3R1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	149	**90.85%**	1,546	1,468	**94.95%**	1,710	1,617	**94.56%**
PMS2	Monosomy 7	CEN 7	Indirect	loss	56	51	**91.07%**	1,659	1,576	**95.00%**	1,715	1,627	**94.87%**
POT1	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	102	**96.23%**	1,602	1,582	**98.75%**	1,708	1,684	**98.59%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,599	**96.38%**	1,715	1,654	**96.44%**
PRPF40B	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	564	**98.60%**	698	688	**98.57%**
PTPN11	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	567	**99.13%**	698	691	**99.00%**
RAC1	Monosomy 7	CEN 7	Indirect	loss	56	51	**91.07%**	1,659	1,543	**93.01%**	1,715	1,594	**92.94%**
RAD21	Trisomy 8	CEN 8	Indirect	gain	93	88	**94.62%**	1,590	1,581	**99.43%**	1,683	1,669	**99.17%**
RHEB	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	98	**92.45%**	1,602	1,573	**98.19%**	1,708	1,671	**97.83%**
	Monosomy 7	CEN 7	Indirect	loss	56	54	**96.43%**	1,659	1,590	**95.84%**	1,715	1,644	**95.86%**
RPS26	Trisomy 12	CEN 12	Indirect	gain	126	125	**99.21%**	572	563	**98.43%**	698	688	**98.57%**
SAMD9	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	101	**95.28%**	1,602	1,570	**98.00%**	1,708	1,671	**97.83%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,583	**95.42%**	1,715	1,638	**95.51%**
SAMD9L	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	103	**97.17%**	1,602	1,571	**98.06%**	1,708	1,674	**98.01%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,583	**95.42%**	1,715	1,638	**95.51%**
SBDS	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	71	**66.98%**	1,602	1,588	**99.13%**	1,708	1,659	**97.13%**
	Monosomy 7	CEN 7	Indirect	loss	56	53	**94.64%**	1,659	1,628	**98.13%**	1,715	1,681	**98.02%**
SH2B3	Trisomy 12	CEN 12	Indirect	gain	126	124	**98.41%**	572	566	**98.95%**	698	690	**98.85%**
SMO	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	103	**97.17%**	1,602	1,576	**98.38%**	1,708	1,679	**98.30%**
	Monosomy 7	CEN 7	Indirect	loss	56	55	**98.21%**	1,659	1,591	**95.90%**	1,715	1,646	**95.98%**
STAT6	Trisomy 12	CEN 12	Indirect	gain	126	123	**97.62%**	572	566	**98.95%**	698	689	**98.71%**
TERT	Monosomy 5	hTERT	Direct	loss	10	10	**100.00%**	1,723	1,709	**99.19%**	1,733	1,719	**99.19%**
TP53	TP53 Deletion	TP53	Direct	loss	76	73	**96.05%**	1,567	1,556	**99.30%**	1,643	1,629	**99.15%**
UBR5	Trisomy 8	CEN 8	Indirect	gain	93	88	**94.62%**	1,590	1,580	**99.37%**	1,683	1,668	**99.11%**
ZFHX4	Trisomy 8	CEN 8	Indirect	gain	93	89	**95.70%**	1,590	1,576	**99.12%**	1,683	1,665	**98.93%**

Validation results for genes with a direct FISH Marker

Gene	FISH - Positive Cases			FISH - Negative Cases			FISH - All Cases		
	Total Cases	Concondant	Sensitivity	Total Cases	Concondant	Specificity	Total Cases	Concondant	Accuracy
ATM	72	72	**100.00%**	641	629	**98.13%**	713	701	**98.32%**
CBFB	23	22	**95.65%**	517	506	**97.87%**	540	528	**97.78%**
EGR1	171	169	**98.83%**	1,541	1,538	**99.81%**	1,712	1,707	**99.71%**
KMT2A	27	25	**92.59%**	474	472	**99.58%**	501	497	**99.20%**
MET	113	113	**100.00%**	1,593	1,582	**99.31%**	1,706	1,695	**99.36%**
NF1	15	15	**100.00%**	908	908	**100.00%**	923	923	**100.00%**
TERT	10	10	**100.00%**	1,723	1,709	**99.19%**	1,733	1,719	**99.19%**
TP53	76	73	**96.05%**	1,567	1,556	**99.30%**	1,643	1,629	**99.15%**

Direct vs Indirect FISH Markers

Gene	FISH Test	FISH Probes	Strategy	Direction	Positive Cases			Negative Cases			All Cases		
					Total	Conc.	Sensitivity	Total	Conc.	Specificity	Total	Conc.	Accuracy
EGR1	Del(5q)	EGR1	Direct	loss	171	169	**98.83%**	1,541	1,538	**99.81%**	1,712	1,707	**99.71%**
	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,538	**99.48%**	1,710	1,700	**99.42%**
MET	Del(7q)	D7S2460	Direct	loss	113	113	**100.00%**	1,593	1,582	**99.31%**	1,706	1,695	**99.36%**
	Del(7q)	D7S2926 & D7S2460	Indirect	loss	106	106	**100.00%**	1,602	1,584	**98.88%**	1,708	1,690	**98.95%**
	Monosomy 7	CEN 7	Indirect	loss	56	56	**100.00%**	1,659	1,596	**96.20%**	1,715	1,652	**96.33%**

NGS performances vs FISH probes position (chromosome 7)

Gene	Cytoband	Strategy	Direction	FISH Probe CEN 7			FISH Probes D7S2926 & D7S2460		
				Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
GNA12	7p22.3	Indirect	loss	**91.07%**	**96.44%**	**96.27%**	N/A	N/A	N/A
CARD11	7p22.2	Indirect	loss	**92.86%**	**98.61%**	**98.43%**	N/A	N/A	N/A
PMS2	7p22.1	Indirect	loss	**91.07%**	**95.00%**	**94.87%**	N/A	N/A	N/A
RAC1	7p22.1	Indirect	loss	**91.07%**	**93.01%**	**92.94%**	N/A	N/A	N/A
IKZF1	7p12.2	Indirect	loss	**96.43%**	**96.62%**	**96.62%**	N/A	N/A	N/A
EGFR	7p11.2	Indirect	loss	**96.43%**	**97.71%**	**97.67%**	N/A	N/A	N/A
FISH PROBE 1 (CEN 7)									
SBDS	7q11.21	Indirect	loss	**94.64%**	**98.13%**	**98.02%**	66.98%	99.13%	97.13%
CDK6	7q21.2	Indirect	loss	98.21%	96.38%	96.44%	**94.34%**	**98.81%**	**98.54%**
SAMD9	7q21.2	Indirect	loss	98.21%	95.42%	95.51%	**95.28%**	**98.00%**	**97.83%**
SAMD9L	7q21.2	Indirect	loss	98.21%	95.42%	95.51%	**97.17%**	**98.06%**	**98.01%**
CUX1	7q22.1	Indirect	loss	98.21%	95.42%	95.51%	**97.17%**	**98.06%**	**98.01%**
FISH PROBE 2 (D7S2926)									
MET	7q31.2	Indirect	loss	100.00%	96.20%	96.33%	**100.00%**	**98.88%**	**98.95%**
FISH PROBE 3 (D7S2460)									
POT1	7q31.33	Indirect	loss	98.21%	96.38%	96.44%	**96.23%**	**98.75%**	**98.59%**
SMO	7q32.1	Indirect	loss	98.21%	95.90%	95.98%	**97.17%**	**98.38%**	**98.30%**
LUC7L2	7q34	Indirect	loss	98.21%	96.08%	96.15%	**98.11%**	**98.56%**	**98.54%**
BRAF	7q34	Indirect	loss	100.00%	96.02%	96.15%	**100.00%**	**98.69%**	**98.77%**
EZH2	7q36.1	Indirect	loss	98.21%	96.02%	96.09%	**98.11%**	**98.56%**	**98.54%**
RHEB	7q36.1	Indirect	loss	96.43%	95.84%	95.86%	**92.45%**	**98.19%**	**97.83%**
KMT2C	7q36.1	Indirect	loss	100.00%	96.02%	96.15%	**99.06%**	**98.63%**	**98.65%**

NGS performances vs FISH probes position (arm 5q)

Gene	Cytoband	FISH Test	FISH Probes	Strategy	Direction	Positive Cases			Negative Cases			All Cases		
						Total	Conc.	Sensitivity	Total	Conc.	Specificity	Total	Conc.	Accuracy
MAP3K1	5q11.2	Del(5q)	EGR1 & RPS14	Indirect	loss	164	146	**89.02%**	1,546	1,487	**96.18%**	1,710	1,633	**95.50%**
PIK3R1	5q13.1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	149	**90.85%**	1,546	1,468	**94.95%**	1,710	1,617	**94.56%**
APC	5q22.2	Del(5q)	EGR1 & RPS14	Indirect	loss	164	155	**94.51%**	1,546	1,525	**98.64%**	1,710	1,680	**98.25%**
CDC25C	5q31.2	Del(5q)	EGR1 & RPS14	Indirect	loss	164	161	**98.17%**	1,546	1,533	**99.16%**	1,710	1,694	**99.06%**
FISH PROBE 1 (EGR1)														
EGR1	5q31.2	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,538	**99.48%**	1,710	1,700	**99.42%**
CSF1R	5q32	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,538	**99.48%**	1,710	1,700	**99.42%**
PDGFRB	5q32	Del(5q)	EGR1 & RPS14	Indirect	loss	164	162	**98.78%**	1,546	1,537	**99.42%**	1,710	1,699	**99.36%**
FISH PROBE 2 (RPS14)														
EBF1	5q33.3	Del(5q)	EGR1 & RPS14	Indirect	loss	164	151	**92.07%**	1,546	1,477	**95.54%**	1,710	1,628	**95.20%**
NPM1	5q35.1	Del(5q)	EGR1 & RPS14	Indirect	loss	164	122	**74.39%**	1,546	1,508	**97.54%**	1,710	1,630	**95.32%**
NSD1	5q35.3	Del(5q)	EGR1 & RPS14	Indirect	loss	164	157	**95.73%**	1,546	1,499	**96.96%**	1,710	1,656	**96.84%**
DDX41	5q35.3	Del(5q)	EGR1 & RPS14	Indirect	loss	164	143	**87.20%**	1,546	1,528	**98.84%**	1,710	1,671	**97.72%**
NHP2	5q35.3	Del(5q)	EGR1 & RPS14	Indirect	loss	164	153	**93.29%**	1,546	1,496	**96.77%**	1,710	1,649	**96.43%**

References

[1] Riester M, Singh A, Brannon A, Yu K, Campbell C, Chiang D, Morrissey M (2016). PureCN: Copy number calling and SNV classification using targeted short read sequencing. Source Code for Biology and Medicine, 11, 13.

[2] Talevich E, Shain AH, Botton T, Bastian BC (2014). CNVkit: Genome-wide copy number detection and visualization from targeted sequencing. PLOS Computational Biology, 12, 4.