Targeted NGS Method to Detect Gene Level CNV with Fewer Reads

Stephanie C. Huelga, Zulfiqar Gulzar, Bin Li, Luke Sherlin, Douglas A. Amorese

NuGEN Technologies, 201 Industrial Road, San Carlos, CA 94070 USA

ABSTRACT

To analyze copy number variations (CNVs), next generation sequencing (NGS) approaches have required expensive deep whole genome sequencing and/or complicated analysis procedures. Here we present a simple analytical method capable of extracting CNVs from Single Primer Enrichment Technology (SPET) libraries with good quality DNA as well as DNA derived from formalin-fixed paraffin-embedded (FFPE) tissues. Dynamic range was demonstrated using BAC clones (for genes EGFR and KIT) spiked into 100 ng Promega Male genomic DNA (a normal copy number reference) from 3 copies to 20 copies as verified by qPCR. Libraries were generated using the Ovation® Cancer Panel 2.0 Target Enrichment System, targeting 509 genes whose SNPs and/or CNVs have been associated with specific cancer types. After preprocessing of the sequencing data (approximately 2 million reads), we applied a simple probe-based read counting method to compute copy numbers for all genes as compared to the normal copy number sample, and a t-test to associate a p-value. Our approach confirmed CNVs with statistical significance in only EGFR and KIT for a broad dynamic range. The same approach was applied to 100 ng custom reference samples obtained from HorizonDx with known CNVs for 4 genes with varied levels across the 4 samples. Analysis of the sequencing data identified CNVs with a strong correlation to the HorizonDx validated CNVs (R^2 = 0.97) for each of the 4 genes in all 4 samples. Finally, we performed a retrospective analysis with a custom Ovation Target Enrichment System panel to enrich FFPE-derived patient tumor and normal samples. We were able to confirm all previously detected CNVs, as well as identify additional candidates that the other screen missed. The Ovation Target Enrichment System allows for the accurate and cost effective analysis of CNVs in a simple assay, enabling conservation of patient samples and more efficient use of sequencing resources.

MOLECULAR WORKFLOW

Figure 1

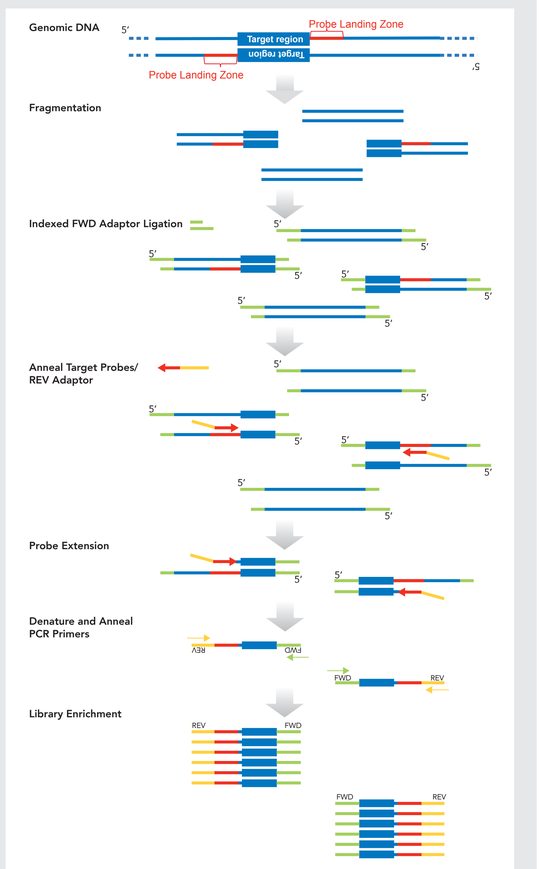

Figure 1: Single Primer Enrichment Technology (SPET) in the Ovation Target Enrichment System.

COMPUTATIONAL WORKFLOW

Figure 2

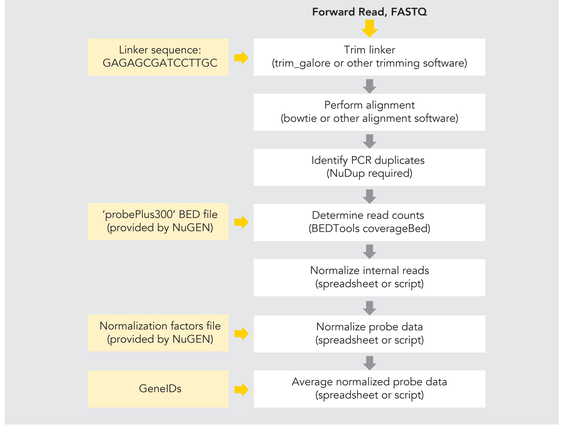

Figure 2: Computational workflow for CNV with Ovation Target Enrichment data.

TARGETED LIBRARY METRICS

Table 1

Target	Reads	% Aligned	% On Target	% Uniformity	% Duplicates	Mean Coverage
Ovation Cancer Panel 2.0	3 M	98.8	84.1	82.3	16.0	33.9

Table 1: Average performance metrics for Ovation Cancer Panel 2.0 Target Enrichment System enriched libraries from 100 ng HapMap NA12878 gDNA that were sequenced on a MiSeq. % Duplicates were computed with NuGEN's NuDup software, which considers a random N6 sequence and the read starting position to determine duplicates.

COPY NUMBER VARIATION AT MANY LEVELS

Using a simple read counting approach compared to a normal reference sample, the SPET method accurately identifies CNV calls in the range of 2N to 20N (Figure 3).

Figure 3

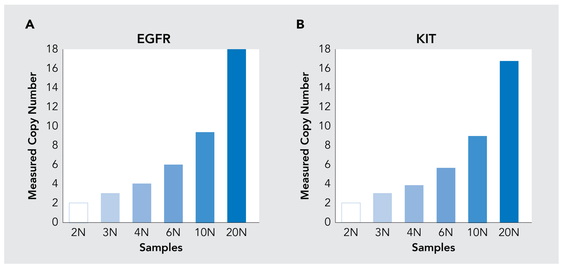

Figure 3: Two different BAC clones were constructed, one expressing EGFR and the other KIT. The BAC clones were spiked-in to Promega Male gDNA, which serves as the normal 2N reference genome. Differing amounts of each BAC were used in separate spike-in samples to achieve 2,3,4,6,10 and 20N levels of EGFR and KIT as measured by qPCR. Each of the BAC + gDNA mixtures were processed with Ovation Cancer Panel 2.0 Target Enrichment System, and libraries sequenced on MiSeq. The sequencing results from the OTE libraries replicate the same gene copy numbers as demonstrated by qPCR. The dynamic range of the assay accurately reports EGFR (A.) and KIT (B.) copy numbers in the range of 2N to 20N.

SINGLE COPY NUMBER LOSSES

The SPET method can detect single copy number losses as shown in a Male vs. Female comparison of chromosome X (Figure 4).

Figure 4

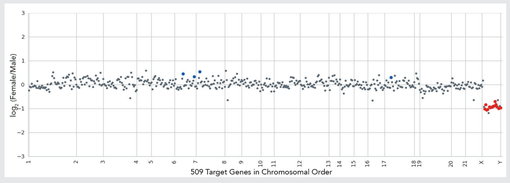

Figure 4: Promega Male genomic DNA and HapMap Female NA12878 genomic DNA were each enriched with the Ovation Cancer Panel 2.0 and sequenced on the MiSeq. The 509 target genes in the Ovation Cancer Panel 2.0 are listed across the x-axis in chromosomal order. The y-axis is the log ratio of the measured copy-number levels in the male compared to the female. Genes that are colored in red show significant copy number losses, and here, those genes are all on the X chromosome as expected when comparing a male to a female sample, displaying that the SPET system can also accurately detect single copy number losses.

HER2 CNV CONCORDANCE WITH ARRAY CGH

Using an alternative internal normalization, a custom Ovation Target Enrichment Panel detects amplification of the HER2 (ERBB2) genomic loci on chromosome 17 in BT-474 Breast Cancer cells concordant with Array CGH aggregate copy number values (from COSMIC, Figure 5).

Figure 5

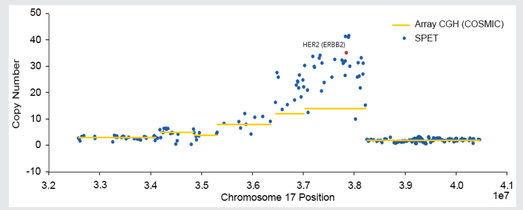

Figure 5: BT-474 breast cancer genomic DNA were each enriched with a custom Ovation Target Enrichment System panel targeting 5 genes on chr1 and 320 genes on chr17. The library was sequenced on a MiSeq. All chr17 genes in the panel are listed across the x-axis in chromosomal order. Each point represents the SPET copy number level represented as a ratio of the gene on chr17 compared to the average of the control genes on chr1. The yellow lines represent the aggregate copy number value as determined by array CGH (obtained from COSMIC database).

CONFIRMED CNV IN REFERENCE SAMPLES

The SPET method accurately confirms the qPCR CNV measurements of 4 genes in 4 custom cancer reference samples provided by Horizon Dx (Figure 6, Table 2).

Figure 6

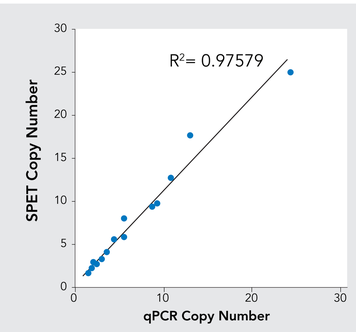

Figure 6: 100 ng of 4 custom cancer reference samples were enriched with the Ovation Cancer Panel 2.0 and sequenced on the Miseq. CNV measurements for the 4 genes with varying copy number levels were measured by qPCR and using the SPET counting approach. Correlation plot of the qPCR levels compared to the SPET levels show an R^2 = 0.9758 (values from Table 2).

Table 2

Sample	Gene	SPET Copy Number	P-Value	qPCR Copy Number
1	ERBB2 (HER2)	24.96	3.84E-32	23.4, gain
	MET	2.64	ns	1.97
	EGFR	3.29	2.88E-09	2.91, gain
	AURKA	9.51	8.93E-09	8.52, gain
2	ERBB2 (HER2)	1.98	ns	1.8
	MET	17.69	7.04E-23	12.8, gain
	EGFR	2.58	ns	2.28
	AURKA	2.85	ns	2.35
3	ERBB2 (HER2)	1.61	ns	1.36
	MET	2.81	ns	2.02
	EGFR	12.74	3.31E-29	10.7, gain
	AURKA	4.05	ns	3.35, gain
4	ERBB2 (HER2)	9.71	6.90E-29	8.88, gain
	MET	8.02	1.17E-18	5.41, gain
	EGFR	5.88	1.68E-22	5.32, gain
	AURKA	5.42	1.81E-06	4.4, gain

Table 2: CNV levels for 4 genes measured across 4 custom reference samples using qPCR and the SPET counting approach. A student's t-test is used to compute the significance of the SPET-computed CNV (ns = not significant). CNVs expected to be significant are confirmed.

CNV VALIDATION IN CLINICAL FFPE SAMPLES

- 32 patient-derived FFPE samples were analyzed with a custom panel of the Ovation Target Enrichment System to detect CNVs and SNPs
- Samples were taken from both tumor and normal tissue
  - Tumor content % not known
  - Unknown if tumor is polyclonal or monoclonal
- Cross-validation was performed with 9 samples to compare 12 copy number changes (Table 3, Figure 7).

Table 3

Gene	SPET Copy Number	P-value	Sample, Orthogonal CNV Validation
CDKN2A/B	0.53	1.01E-15	S8 loss
	1.5	ns	S11 loss
FGFR2	4.8	9.24E-12	S7 gain
IGF1R	3.0	6.27E-10	S11 gain
MAP2K1	2.7	ns	S11 gain
MCL1	13.7	2.75E-15	S4 gain
MYC	25.5	2.16E-14	S7 gain
	6.2	4.29E-08	S11 gain
PRKDC	3.0	5.41E-06	S11 gain
PTEN	2.2	ns	S6 loss
	0.67	1.93E-12	S7 loss
RAF1	7.5	1.99E-11	S8 gain

Table 3: After Ovation Target Enrichment of 32 samples, this table lists the genes that were expected to have copy number changes in some of the 32 samples as determined by an orthogonal method. Using our simple counting approach we also calculated copy numbers with the SPET sequencing data. A student's t-test is used to compute the significance of the SPET-computed CNV (ns = not significant). CNVs expected to be significant are confirmed as well as additional candidate copy number changes across all 32 samples.

Figure 7

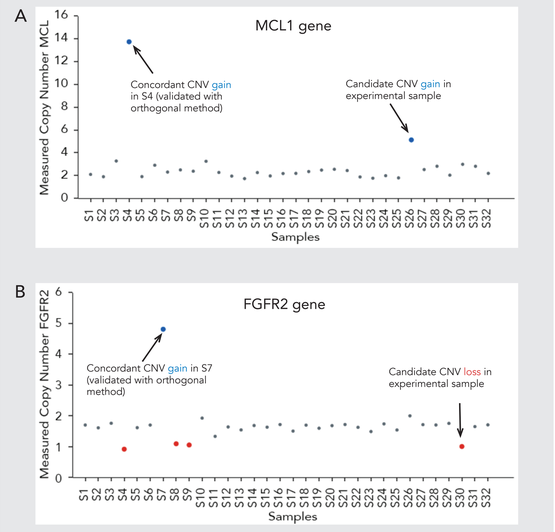

Figure 7: SPET copy number measurements for two genes MCL1 (A) and FGFR2 (B) across all 32 customer samples. CNV measurements validate known aberrations and detect novel events missed in an orthogonal method. Blue dots are samples with a significant gain of the gene and red points are samples with a significant loss of the gene.

CONCLUSIONS

- NGS copy number data can be generated from as little as 10 ng gDNA, or 100 ng FFPE/clinical samples.
- Targeting genes or regions of interest enables accurate CNV measurements obtained with minimal reads.
- Fast, flexible, custom probe design enables enriched libraries capable of monitoring a variety of genomic aberrations (SNPs, indels, CNVs, rearrangements).

NuGEN
imagine more from less®

www.nugen.com