

Using NextGENe[®] Software (v2.4.2) Batch CNV Tool as an NGS replacement for MLPA[®] and Similar Analysis Techniques

July 2016

John McGuigan, Jacie Wu, Michael Wiegand, Jonathan Liu

Introduction

Detection of copy number variants is often performed using multiplex ligation-dependent probe amplification (MLPA). While it is sensitive, this method can be slow and labor intensive. High Throughput Sequencing (HTS) using targeted panels offers a different approach that is simpler, cheaper, and faster. Using a simple coverage normalization method with compatible data, it is possible to generate CNV calls with HTS and outperform the use of MLPA and Sanger sequencing¹.

Procedure

1. Align data to the whole human genome using NextGENe, with one project per sample.
2. Add all multiplexed samples from a single run to the Batch CNV Tool in the viewer “Comparisons” menu (Figure 1).
A BED file defining the targeted regions must be loaded to perform the analysis. A second file with gene names (matching the BED file region names) may also be loaded to limit report results.

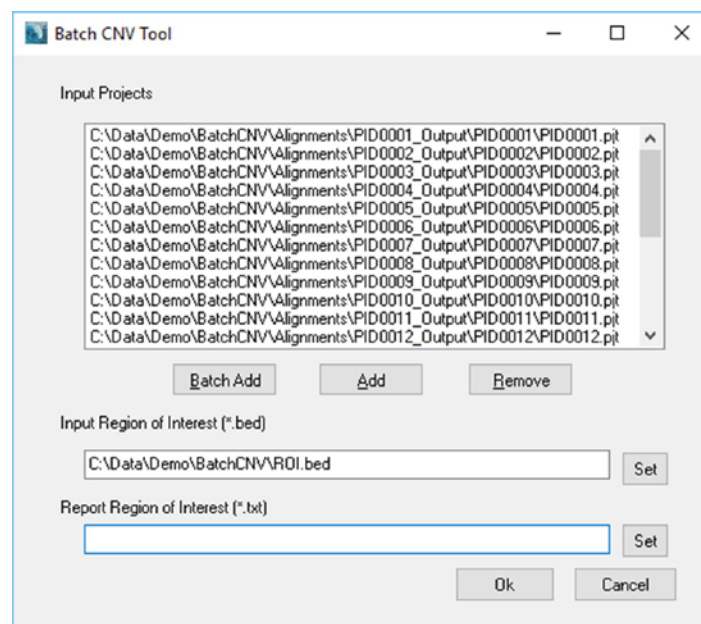


Figure 1: Loading projects in the Batch CNV Tool

Discussion & Processing

The choice of library preparation method is very important for this analysis. Any method that produces inconsistent coverage from sample to sample will result in poor sensitivity and specificity. Samples prepared using SeqCap EZ Choice (Roche NimbleGen, Inc) have been shown to work well with the tool.

The tool will calculate the pileup coverage at each position in the targeted panel, normalize it by a constant value in each sample, and report the normalized values (figure 2). The normalization procedure is:

1. Get the total coverage in each sample (adding the coverage at each position in the ROI bed file).
2. Divide these values by the average across all samples to get the normalization factor for each sample.
3. Divide the coverage at each position by the normalization factor for that sample.
4. When reporting a value for a given region, report the average of all bases within that region.

Index	Chr	Chr Start	Chr End	Description	S16.pit	S21.pit	S3.pit	S19.pit	S14.pit	S23.pit	S1.pit
1	chr1	12049206	12049420	MFN2_NM_014874.3_cds_1	1.007	1.016	1.014	0.930	1.046	0.953	0.984
2	chr1	12052592	12052767	MFN2_NM_014874.3_cds_2	0.938	1.030	0.997	1.057	1.100	0.913	1.037
3	chr1	12056193	12056395	MFN2_NM_014874.3_cds_3	0.922	1.002	0.960	1.047	0.873	1.108	0.941
4	chr1	12057334	12057498	MFN2_NM_014874.3_cds_4	1.021	0.990	0.970	0.979	0.888	0.968	0.940
5	chr1	12058807	12058955	MFN2_NM_014874.3_cds_5	0.967	1.054	0.984	1.012	0.957	0.936	0.916
6	chr1	12059025	12059172	MFN2_NM_014874.3_cds_6	0.937	0.977	1.030	1.032	0.963	0.955	0.902
7	chr1	12061438	12061631	MFN2_NM_014874.3_cds_7	0.989	0.965	1.049	0.979	0.919	1.123	0.966

Figure 2: Table of Results

Any regions with an average normalized value outside the set thresholds (0.8 and 1.2 by default) will be called (figures 3 and 4). The region names are taken from the 4th column of the BED file used to define the regions. Sub-region results (for example, within an exon) can be visualized (figure 5) in order to see potential sub-exon copy number variations.

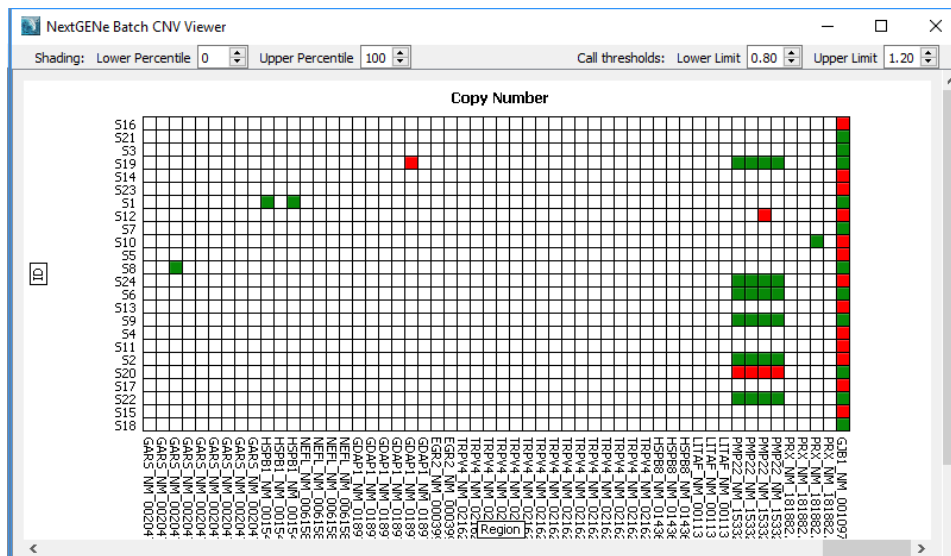


Figure 3: Global plot with samples in rows and regions in columns. Red squares indicate a deletion call while green squares indicate a duplication call.

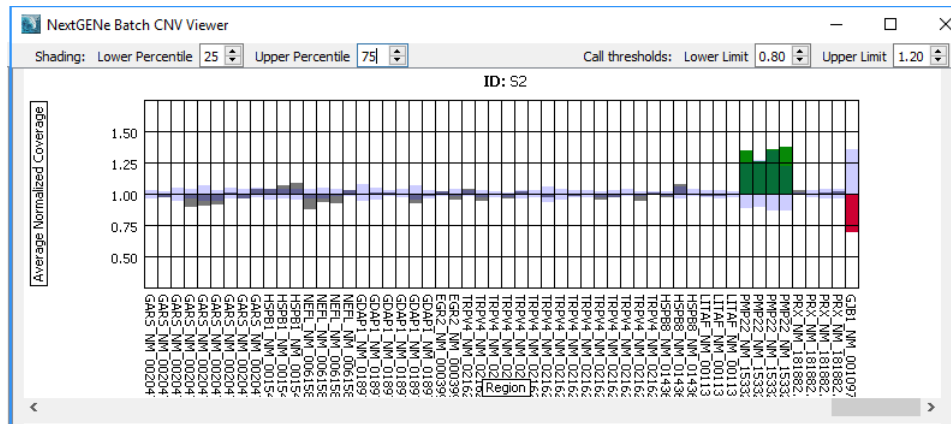


Figure 4: Results across several regions for a given sample. 4 exons in the PMP22 gene were called as duplications and an exon from a gene on chrX was called as a deletion.

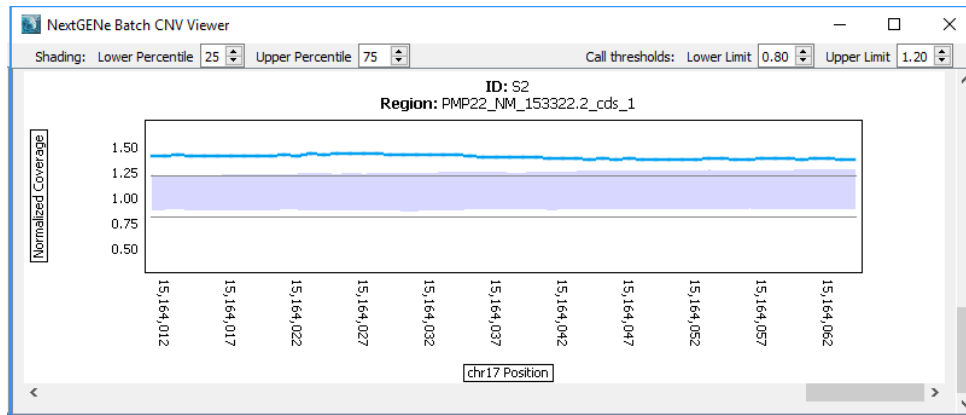


Figure 5: Results within an exon (that was called as a duplication) in a single sample. The normalized coverage (light blue) is consistently above the cutoff for a duplication call at 1.20 (green).

References & Acknowledgements

We would like to thank Dr. Bekim Sadikovic and colleagues (London Health Science Center, London Canada) for their suggestions, ideas, and feedback during development of this analysis tool and data used to demonstrate the analysis.

1. Schenkel, Laila C., et al. "Clinical Next-Generation Sequencing Pipeline Outperforms a Combined Approach Using Sanger Sequencing and Multiplex Ligation-Dependent Probe Amplification in Targeted Gene Panel Analysis." *The Journal of Molecular Diagnostics* (2016).

Trademarks are property of their respective owners. For research use only.