Copy Number Variation (CNV) Detection using Targeted Sequencing Data with NextGENe® Software v2.3.1

November 2012

John McGuigan, Megan Manion, Kevin LeVan, Ning Wan, Ni Shouyong, and CS Jonathan Liu

Introduction

NextGENe v2.3.1 has a new Copy Number Variation (CNV) analysis tool designed to make variant calls on a case-control basis. It uses a proprietary normalization method, and works well with targeted sequencing data such as Ion AmpliSeq™ panels or the HaloPlex™ Target Enrichment System from Agilent Technologies. Ideally the two samples used in a comparison will be as close as possible in experimental conditions. No special processing is needed to use the tool- any aligned NextGENe projects can be loaded.

Procedure

The CNV tool can be opened from the Viewer "Tools" menu. One "sample" and one "control" alignment project is loaded, and the regions are defined (Figure 1). For targeted sequencing it is best to load a BED file that specifies the location of each amplicon.

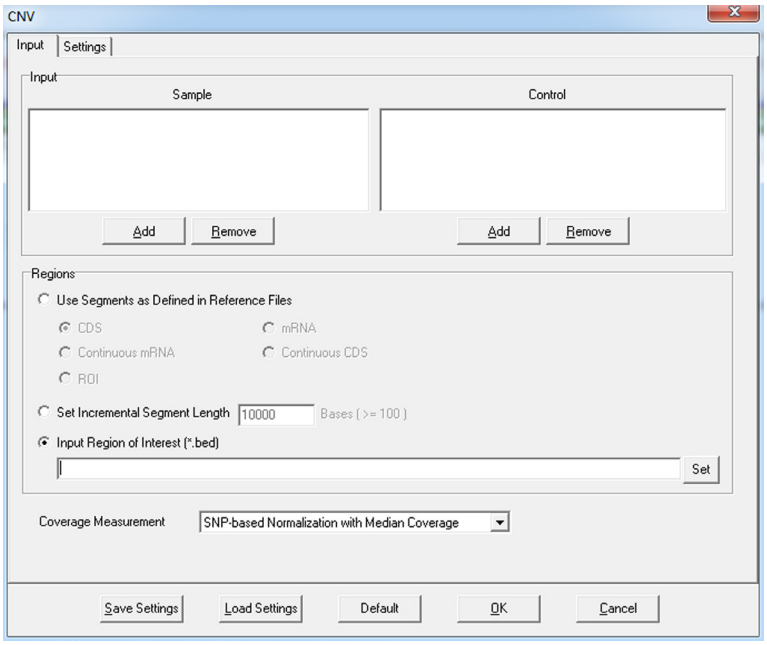

Figure 1

Filters can be set on the next page for log2 ratio, Z score, and minimum coverage in one project (Figure 2). If a region does not meet the minimum coverage requirement, the region will be hidden in the report and the calculations for that region will not be performed but there is an option to show these regions (with "NA" reported for log2 ratio and Z-score) in order to keep the report consistent between comparisons.

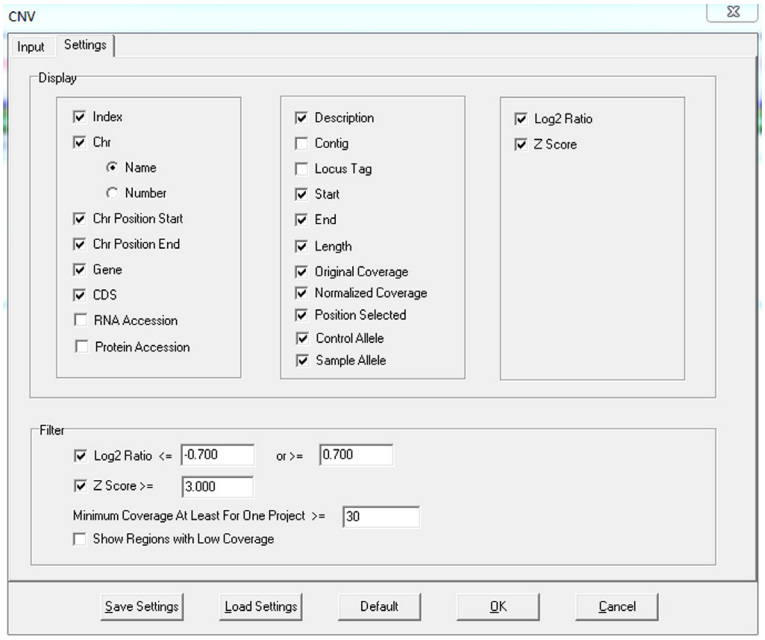

Figure 2

Results

Figure 3 shows the detection of a known deletion in the KCNH2 gene using HaloPlex Cardiac Panel data. The log2 ratio (-1.03) was very close to the expected value of -1.0.

Sample	E.pjt										
Control	F.pjt										
Index	Description	Chr	Chr Start	Chr End	Gene	CDS	Length	Log2 Ratio	Z Score	Original Coverage (Sample;Control)	Normalized Coverage (Sample;Control)
1	Amplicon206	chr7	91623915	91624109	AKAP9; +	6	195	0.90	3.72	100;55	100;54
2	Amplicon255	chr7	150645512	150645650	KCNH2; -	11	139	-1.03	8.26	197;413	197;402
3	Amplicon358	chr19	35521705	35521783	SCN1B; +	1	79	-2.73	11.00	38;258	38;251

Figure 3

Figure 4 shows Ion Torrent Comprehensive Cancer Panel results for a trisomy sample compared to three non-trisomy samples. The log2 ratio was calculated for each sample against a normal control using NextGENe. A Python script was used to load this log2 ratio data, average replicates into a single value per probe, and graph the values using a moving average of 10 amplicons. Some amplicons appear to be outliers (especially the first few) but most amplicons have log2 ratios near the expected values.

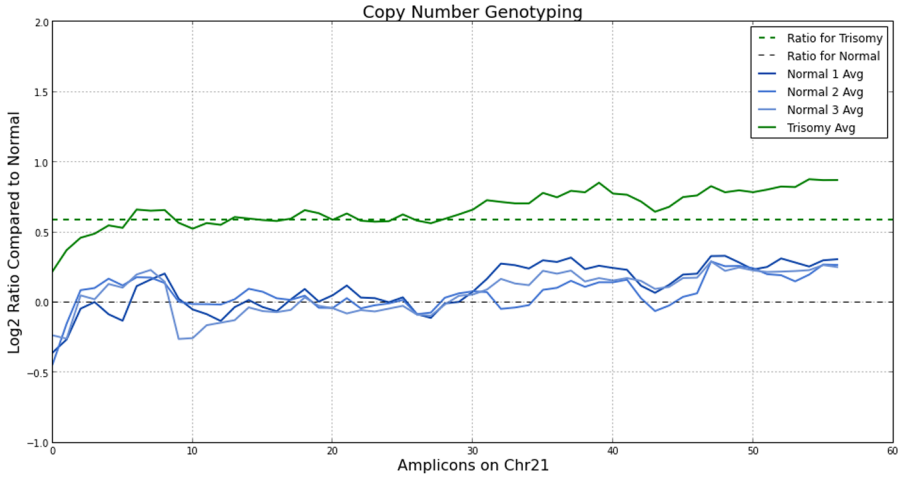

Figure 4

Figure 5 shows the same data for chrX, allowing for detection of duplications and deletions. The control sample was female, so the male sample (red) appears to be a deletion (1/2 the normalized coverage). The XO sample appears to have a deletion in the beginning of chrX.

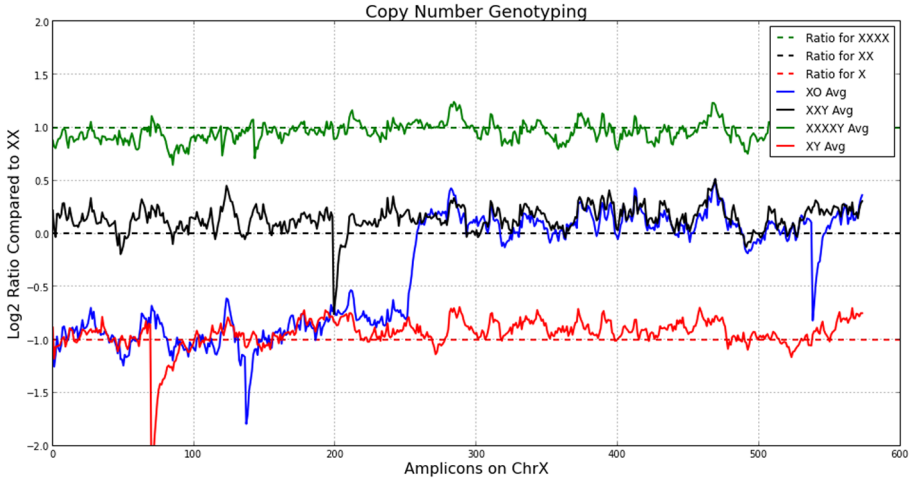

Figure 5

Discussion

The CNV tool first generates a list of heterozygous SNPs in each sample. It uses the median coverage of this list to normalize on a global level. After normalization, a specific position in each region is chosen as representative (based on median coverage) and used to calculate a log2 ratio. A ratio of -1.0 indicates a deletion (the coverage in the sample is 1/2 of the coverage in the control). A ratio of 1.0 indicates a duplication (the coverage in the sample is 2x the coverage in the control). Less drastic duplications are also possible. For example, a certain sequence can increase from 3 copies to 4, resulting in a 33% increase in coverage.

The report can be saved as a tab-delimited text file and graphs can easily be generated outside of NextGENe. A future version of NextGENe will directly incorporate analysis with replicates and graphing of results. The Ion Torrent AmpliSeq results shown here clearly demonstrate the known differences in copy number for chrX and chr21, with a small number of amplicons appearing to be outliers.

Acknowledgements

We would like to thank Agilent Technologies and Berivan Baskin (Clinical Genetics, Uppsala University Hospital; The Centre for Applied Genetics; The Hospital for Sick Children) for supplying the HaloPlex data used in this analysis, and Life Technologies for supplying the Ion Torrent data.