MGZ
Medical Genetics Center

Prof. Elke Holinski-Feder, MD
Angela Abicht, MD
Clinical Geneticists

All forms are available on our website:
www.mgz-muenchen.com

ilac-MRA DAkkS Deutsche Akkreditierungsstelle D-ML-13242-01-00

VALIDATION OF COPY NUMBER VARIATION ANALYSIS WITH NGS

Overview

The ability to detect copy number variants (CNVs) with a high degree of sensitivity and specificity is the basis for comprehensive gene analysis. Many laboratories perform multiplex ligation-dependent probe amplification (MLPA) and microarray-based technologies to identify changes in the number of copies, but these methods are either not scalable for the parallel analysis of a large number of genes or the detection limit is not sufficient for the identification of small events (i.e. fewer exons). We have developed a bioinformatics method to identify CNVs from targeted next-generation sequencing (NGS) data sets and confirmed the accuracy of the results in a validation study. By developing an approach that maximizes the use of our established workflows and capabilities, we are able to offer parallel analysis of single nucleotide and CNVs from the same NGS dataset while maintaining our commitment to best practices genetic testing.

Method

CNVs were determined by NGS in all coding exons and their flanking intronic regions using DNA-targeted enrichment and Illumina® sequencing technology. The NGS data were analyzed using a custom analysis copy number caller consisting of four different published tools as well as a proprietary additional prediction method. All software tools were executed in parallel via an in-house script that provided a reference set as input for the CNV tools. The last CNV calls contained all individual calls that were recognized by at least two methods and completely overlapped in the exon annotation.

Validation Data

This approach was validated with individuals known to have specific pathogenic deletions/duplications in 99 distinct genes associated with a predisposition to hereditary cancer or rare diseases. For validation of the CNVs called by the CNV pipeline, two orthogonal methods, MLPA or SNP array analysis (Illumina® Infinium CytoSNP-850K Assay), were used. The results of MLPA and SNP array analyses were considered to be "true-set" of variants. We analyzed a total of 3753 individuals, 122 single samples with pathogenic deletions or copy number gains, comprising a single exon up to the complete gene and 3631 individuals with no pathogenic CNVs. Accuracy, reproducibility, analytical sensitivity and specificity are described in Table 1.

		True variant (MLPA or SNP array)		
		deletion/duplication	no CNVs	**Precision (PPV) / (NPV)**
Predicted variant (NGS)	deletion/duplication	115	31	78,8 % [71,4 %]
	no CNVs	7	3600	99,8 % [99,6 %]
	Sensitivity (TPR) / Specificity (TNR)	94,3 % [88,6 %]	99,1 % [98,8 %]	N = 3753

Table 1: Contingency table.

The lower bound for the confidence interval at 95 % is given in parenthesis. The confidence interval was calculated using Wilson's method and is only an approximation. PPV: positive predictive value, NPV: negative predictive value, TPR: true positive rate, TNR: true negative rate.

MGZ – Medical Genetics Center | Bayerstrasse 3-5 | D-80335 Munich | Germany
phone +49 (0)89 / 30 90 886 - 0 | fax +49 (0)89 / 30 90 886 - 66 | email inquiry@mgz-muenchen.com
© 2019

Results and Discussion

We compared the accuracy of copy number detection with a custom caller applied to targeted NGS data sets to CNV events detected by MLPA or SNP array for 3753 individuals. The sensitivity and specificity applied to these specific NGS data sets was determined to be 94.3 % and 99.1 %, respectively. A total of 115 out of 122 true positive (TP) CNVs were correctly identified (70 deletions and 45 copy number gains). Among the true positive calls, 14.8 % were individual exon variants demonstrating the tool's ability to detect smaller events (Figure 1). Almost all of the false positive (FP) and false negative (FN) calls comprised a single exon in regions with uneven coverage or high homologous sequences (i.e. pseudogenic regions). We found that the inter- and intra-sample variability in the depth of the sequencing coverage and the composition of the analytical domains are important determinants for successful CNV monitoring by our tool.

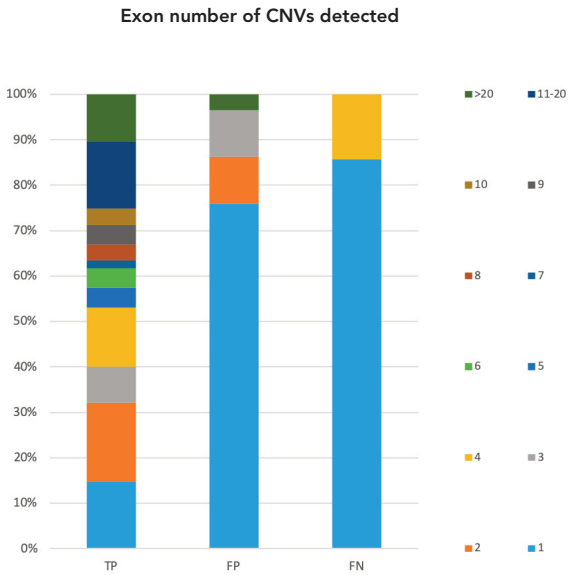

Figure 1:

Comparison of the CNV sizes of TP, FP and FN calls detected by the custom CNV tool on the validation set. Among the true positive calls were 14.8 % single exon events, 74.8 % from 2 to 20 exons, and 10.4 % more than 20 exons.

Overall, the addition of CNV detection to the current single nucleotide variant (SNV) analysis from the same NGS individual data set increased the diagnostic yield by 4 %. Our results support the integration of CNV detection into the routine services of NGS diagnostics for Mendelian diseases. The implementation of this strategy increases diagnostic yield and improves clinical care.

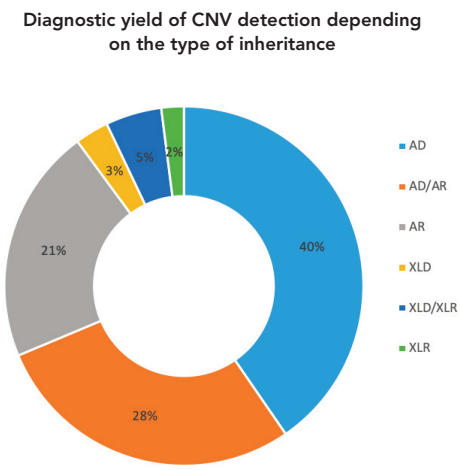

Figure 2:

The detection of CNVs improved the diagnostic yield by 4 % and concerned all types of hereditary phenotypes. Of particular interest were cases with autosomal recessive mode of inheritance in which the combination of a single nucleotide variant combined with a copy number variation clarified the phenotype of the patient. AD: autosomal dominant, AR: autosomal recessive, XLD: X-linked dominant, XLR: X-linked recessive.

References

(1) Backenroth et al. Nucleic Acids Res. 2014 Aug 1; 42(12): e97.
(2) Packer et al.; Bioinformatics. 2016 Jan 1;32(1):133-5
(3) Plagnol et al.; Bioinformatics. 2012 Nov 1;28(21):2747-54
(4) Jiang et al.; Nucleic Acids Res. 2015 Mar 31;43(6):e39

MITTELSTAND GLOBAL HEALTH MADE IN GERMANY

MGZ – Medical Genetics Center | Bayerstrasse 3-5 | D-80335 Munich | Germany
phone +49 (0)89 / 30 90 886 - 0 | fax +49 (0)89 / 30 90 886 - 66 | email inquiry@mgz-muenchen.com

© 2019