panelcn.MOPS: CNV detection in targeted NGS panel data for clinical diagnostics

Gundula Povysil[1], A. Tzika[2], J. Vogt[2], V. Haunschmid[1], L. Messiaen[3], J. Zschocke[2], G. Klambauer[1], S. Hochreiter[1], K. Wimmer[2]
[1] Institute of Bioinformatics, Johannes Kepler University Linz, Austria, [2] Division of Human Genetics, Medical University Innsbruck, Austria,
[3] Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA

Summary

Targeted next-generation-sequencing (NGS) panels have largely replaced Sanger sequencing in clinical diagnostics. Enrichment-based targeted NGS panels allow for the detection of copy-number-variations (CNVs) in addition to single-nucleotide-variants and small insertions/deletions. However, existing computational CNV detection methods have shortcomings regarding accuracy, quality control, incidental findings, and user-friendliness.

To solve these problems we developed panelcn.MOPS, a novel pipeline for detecting CNVs in targeted NGS panel data. Using NGS panel data from 170 samples, we compared panelcn.MOPS with 5 state-of-the-art methods.

We present the first thorough comparison of CNV detection methods for targeted NGS panel data. Most methods achieved comparably high sensitivity and/or specificity, but panelcn.MOPS led the field with a sensitivity and specificity of 100%. panelcn.MOPS reliably detected CNVs ranging in size from 20 nucleotides (only part of a region-of-interest - ROI), to whole genes, which may comprise all ROIs investigated in a given sample. The latter is enabled by analyzing reads from all ROIs of the panel, but presenting results exclusively for user-selected genes, thus, avoiding incidental findings. Additionally, panelcn.MOPS offers quality control criteria not only for samples but also for individual ROIs within a sample which increases the confidence in called CNVs.

panelcn.MOPS is freely available both as an R package and as standalone software with an intuitive graphical user interface (GUI). It can therefore readily be used by clinical geneticists without any programming experience or integrated into existing analysis pipelines. Taken together, panelcn.MOPS combines high sensitivity and specificity with user-friendliness rendering it highly suitable for routine clinical diagnostics.

Characteristics of cn.MOPS

Mixture Of PoissonS for discovering Copy Number variations:

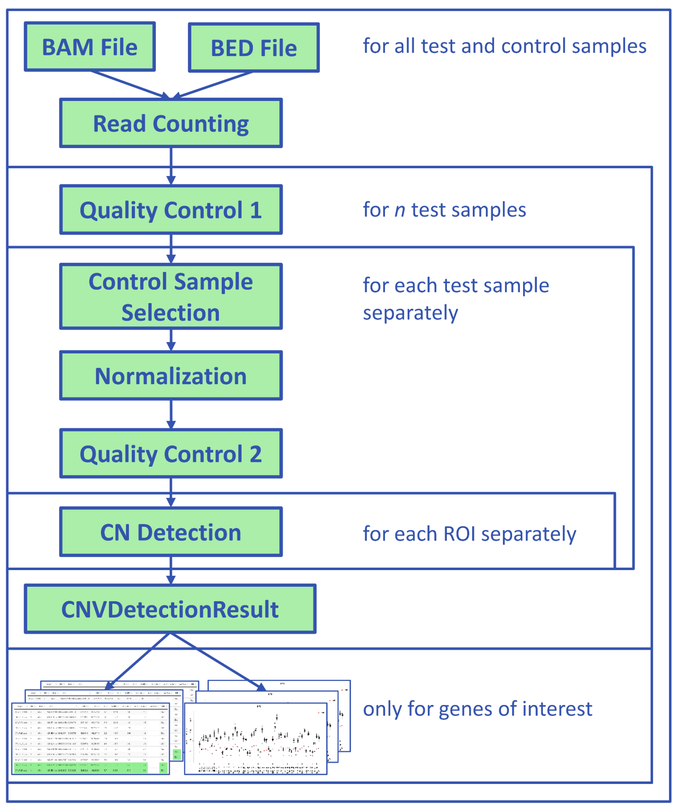
- low FDR by local modeling across samples
- model decomposes read count variation into:
  - noise variation (Poisson)
  - copy number variation (mixture components)

Best performance on

- Whole-Genome Sequencing data (1000 Genomes) and
- Whole-Exome Sequencing data (intellectual disability, ASD, ...)

panelcn.MOPS Overview

- **extension of cn.MOPS**
  for targeted NGS panel data
- **adapted read counting**
- **2 quality controls** for **samples**
- **2 quality controls** for **ROIs**
- **selection** of **best control samples**
- **improved normalization**
- **increased sensitivity**
- **no segmentation**
- **binning option** for large ROIs
- **filter** for displaying copy numbers (CNs) only for genes of interest
- **boxplots** of normalized read counts (RCs) for visual inspection



panelcn.MOPS Details

Read counting: all reads that overlap respective window (ROI) counted

Quality control:

- **ROIs** with **low median RC** across all samples excluded
- **ROIs** with **high variation of** RCs across all samples marked as "low quality"
- **samples** with **low median RC** across all ROIs excluded from controls, warning for test samples
- **samples** with **high variation in ratios** between normalized RCs of sample compared to median across all samples excluded from controls**,** warning for test samples

Control sample selection: control samples with high correlation of RCs to the RCs of the test sample, ROIs of gene(s) of interest for specific test sample excluded for calculating correlation

Competing Methods

ExomeDepth (Plagnol et al. 2012)
CoNVaDING (Johansson et al. 2016)
VisCap (Pugh et al. 2016)
NextGENe (Softgenetics): commercial tool with GUI
SeqNext (JSI medical systems): commercial tool with GUI

Data Generation

- TruSight® Cancer Panel
- 94 genes associated with cancer predisposition (e.g.: *NF1/2*, *BRCA1/2*, *APC*, *MSH2/6*, *MLH1*, *PMS2)*
- Illumina MiSeq® → 300 cycles with paired end reads

Samples

	Training Set	Test Set	Total
Copy Number 2 (normal)	13	110	123
Multi-Exon Deletions	5	12	16
Multi-Exon Duplications	2	1	3
Single-Exon Deletions		12	13
Single-Exon Duplications		2	2
Whole-Gene Deletions	4	1	5
Subtotal	24	138	162
	Samples excluded from calculations		
Deletions < 1 ROI	1	2	3
Duplications < 1 ROI		2	2
De novo Alu Insertions		3	3
Total	25	145	170

Results: Test Set

	panelcn.MOPS	ExomeDepth	CoNVaDING	VisCap	NextGENe	SeqNext	optimal
TP	91	90	90	91	91	90	91
TN	7889	8203	7705	7279	8141	7954	8222
FP	0	19	1	12	19	0	0
FN	0	1	1	0	0	1	0
No-Call	333	0	516	931	62	268	0
Total	8313	8313	8313	8313	8313	8313	8313
Sensitivity	1.0000	0.9890	0.9890	1.0000	1.0000	0.9890	1.0000
Specificity	1.0000	0.9977	0.9999	0.9984	0.9977	1.0000	1.0000
No-Call-Rate	0.0401	0.0000	0.0621	0.1120	0.0075	0.0322	0.0000

numbers correspond to numbers of ROIs; no-call rate: number of ROIs with low-quality call divided by total number of ROIs

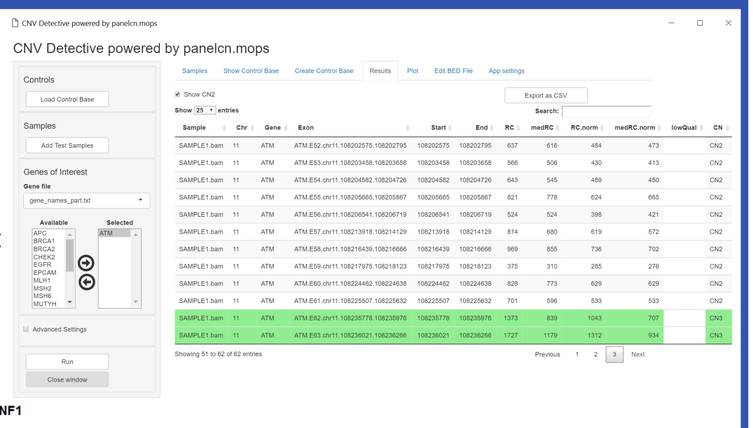
Results

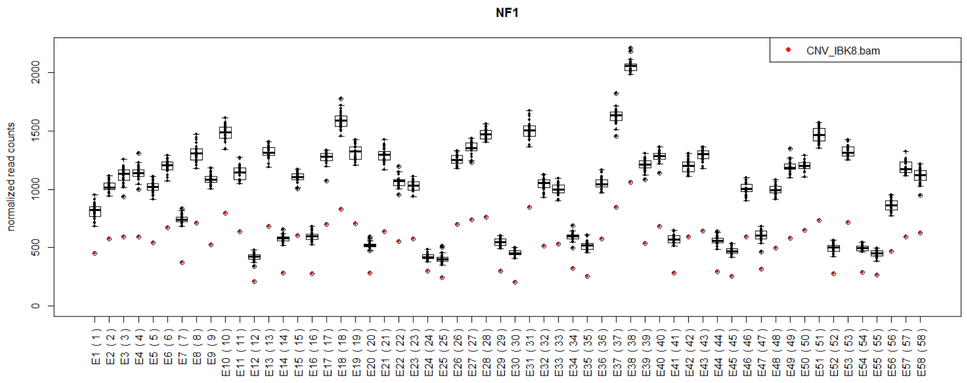
	panelcn.MOPS	ExomeDepth	CoNVaDING	VisCap	NextGENe	SeqNext
Sensitivity	+++	++	++	+++	+++	++
Specificity	+++	+	++	+	+	+++
No-Call-Rate	+	+++	+	-	++	+
Quality Filter	+++	+	+++	+	+	+++
CNVs < 1 ROI	++	++	+++	+++	*	**
Whole-Gene CNVs	+++	+++	++	+	-	-
Incidental Findings	+++	-	-	-	+	+
Runtime	+++	++	-	+	+	+
GUI	+++	-	-	-	++	+++
Non-Commercial	+++	+++	+++	+++	-	-

* except for one detected by variant calling routine of NextGENe
** all detected by variant calling routine of SeqNext

Graphical User Interface

- standalone app
- based on R shiny
- simple installer for Windows
- quality control for samples and ROIs
- option for building up control base
- reports only CNs for genes of interest
- results exportable as .csv
- read count plots for genes of interest
- option for binning (larger) ROIs





R Package

https://github.com/bioinf-jku/panelcn.mops

Conclusion

- panelcn.MOPS for CNV detection in targeted panel sequencing data
- superiority shown for real data on CNVs of a broad range of sizes
- GUI available for better usability

Povysil, G., Tzika, A., Vogt, J., Haunschmid, V., Messiaen, L., Zschocke, J., Klambauer, G., Hochreiter, S., and Wimmer, K. (2017). panelcn.MOPS: Copy number detection in targeted NGS panel data for clinical diagnostics. *Human Mutation*. https://doi.org/10.1002/humu.23237