panelcn.MOPS: CNV detection in targeted panel sequencing data for diagnostic use

Gundula Povysil[1], A. Tzika[2], J. Vogt[2], V. Haunschmid[1], L. Messiaen[3], K. Wimmer[2], G. Klambauer[1], S. Hochreiter[1]
[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

Abstract

While various copy number variation (CNV) detection methods exist for whole-genome and whole-exome sequencing data, highly accurate methods for targeted panel sequencing data that are suitable for a diagnostic setting are still missing. The challenges with analyzing this kind of data include the small size and number of target regions as well as their uneven coverage. For clinical applications a method should furthermore be able to detect both short CNVs affecting only single exons or just parts thereof as well as longer CNVs that affect multiple exons or even an entire gene. Another important issue is the risk of incidental findings.

Our new method panelcn.MOPS for copy number detection extends cn.MOPS to targeted panel sequencing data. We optimized the design of the count windows, the read counting procedure, the parameters of the model and the segmentation algorithm for targeted panel sequencing. Additionally, several quality control criteria both for samples and targeted exons have been implemented to increase the confidence in called CNVs. In contrast to other CNV detection methods all targeted regions are exploited for the detection of CNVs, but only results for user-selected genes are reported to avoid the risk of incidental findings.

We have tested panelcn.MOPS on simulated and real sequencing data. The real sequencing data was enriched with the TruSight cancer panel that targets 94 cancer predisposition genes including NF1/2, BRCA1/2 and APC. The performance of panelcn.MOPS was compared on a data set of 170 samples against several CNV detection tools including NextGENe, SeqNext, ExomeDepth, CoNVaDING, and VisCap. The size of the CNVs ranges from a 20bp deletion affecting only part of an exon over duplications of several exons to a 350kb deletion of an entire gene. In contrast to the other methods, panelcn.MOPS not only achieved a sensitivity of 100%, but also the highest specificity. Furthermore, we do not only provide panelcn.MOPS as an R package, but also as a standalone program with a practical graphical user interface (GUI). Therefore, panelcn.MOPS can be conveniently used by users without any programming experience.

Our results show that panelcn.MOPS accurately predicts CNVs in targeted panel sequencing data. Consequently complementary biotechnologies to detect CNVs, such as MLPA, can be omitted in order to reduce time and costs.

cn.MOPS

Mixture Of PoissonS for discovering Copy Number variations:

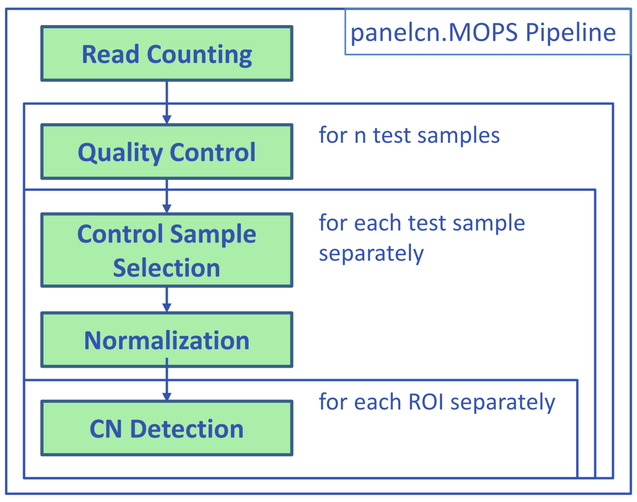
- low FDR by local modeling across samples
- model decomposes read count variation into:
  - noise variation (Poisson)
  - copy number variation (mixture components)

Best performance on

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

panelcn.MOPS

- **extension of cn.MOPS** for targeted panel sequencing data
- **adapted read counting**
- **quality control** for **samples** and **ROIs**
- **selection of control samples**
- **improved normalization**
- **increased sensitivity**
- **no segmentation**
- **filter** for displaying CNVs only for genes of interest
- boxplots of normalized RCs



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

TruSight® Cancer Panel

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

Data

	Training Cohort	Test Cohort	Total
CN2	13	110	123
Multi Exon Deletions	5	11	16
Multi Exon Duplications	2	1	3
Single Exon Deletions		13	13
Single Exon Duplications		2	2
Whole Gene Deletions	4	1	5
Subtotal	24	138	162
	Cases 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	86	85	84	84	85	85	86
TN	7843	8156	7787	7290	8080	7907	8173
FP	0	17	1	11	17	0	0
FN	0	1	1	1	0	1	0
No Call	330	0	386	873	77	266	0
Total	8259	8259	8259	8259	8259	8259	8259
Sensitivity	1.0000	0.9884	0.9882	0.9882	1.0000	0.9884	1.0000
Specificity	1.0000	0.9979	0.9999	0.9985	0.9979	1.0000	1.0000
No Call Rate	0.0400	0.0000	0.0467	0.1057	0.0093	0.0322	0.0000

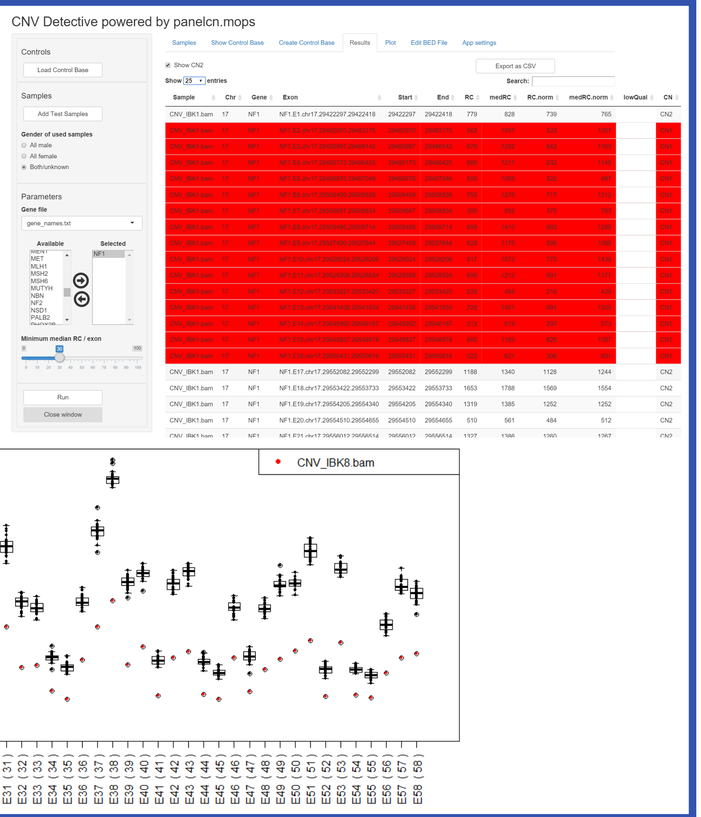
numbers correspond to numbers of ROIs

Results

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

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



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
- GUI available for better usability

Klambauer G et al. (2012) cn.MOPS: mixture of Poissons for discovering copy number variations in next generation sequencing data with a low false discovery rate. *Nucleic Acids Res* 40(9):e69.