panelcn.MOPS reaches clinical standards as a CNV detection tool for targeted panel sequencing data

Verena Haunschmid[1], Gundula Povysil[1], Julia Vogt[2], Katharina Wimmer[2], Günter Klambauer[1], Sepp Hochreiter[1]
[1] Institute of Bioinformatics, Johannes Kepler University Linz, Austria, [2] Division of Human Genetics, Medical University Innsbruck, Austria

Abstract

Targeted panel sequencing is becoming increasingly important as a cost-effective strategy to identify disease-causing variants in clinical and research applications. While various copy number variation (CNV) detection methods exist for whole-genome and whole-exome sequencing data, highly accurate methods for panel sequencing data that are suitable for clinical purposes are still missing. The challenges with this kind of data are 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 even just parts thereof as well as longer CNVs that affect multiple exons or even an entire gene.
We present panelcn.MOPS for copy number detection which extends our previously developed method cn.MOPS to targeted panel sequencing data. The method is well suited for this type of data since it can estimate technical and biological characteristics influencing the read counts of each targeted region by a mixture of Poissons model. The design of the count windows, the read counting procedure, the parameters of the model and the segmentation algorithm have been optimized for targeted panel sequencing. cn.MOPS supplies integer copy numbers together with probabilities which inform users about the reliability of the copy number estimates.
We have tested panelcn.MOPS on simulated and real sequencing data. On 240 simulated data sets, that resembled the characteristics of targeted panel sequencing data, panelcn.MOPS has reached an average accuracy of 99.99%. The real sequencing data was enriched with the TruSight cancer panel that targets 94 cancer predisposition genes including *NF1/2*, *BRCA1/2* and *APC*. panelcn.MOPS detected 100% of CNVs known from previous MLPA analyses without any false positives. The size of the CNVs ranged from an 80bp deletion starting in the intron and affecting only part of one exon over duplications of several exons to deletions of 350kb affecting the entire gene.
These results show that CNVs in targeted panel sequencing data can accurately be predicted with panelcn.MOPS. Consequently additional biotechnologies to detect CNVs, such as MLPA, can be omitted in order to reduce time and costs.

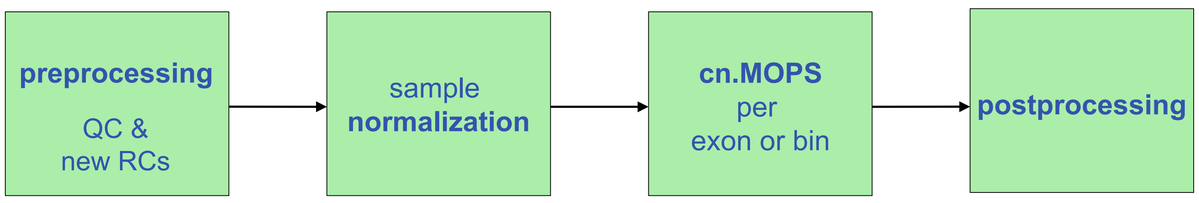
cn.MOPS

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 (intellectually disability, ASD, ...)

panelcn.MOPS

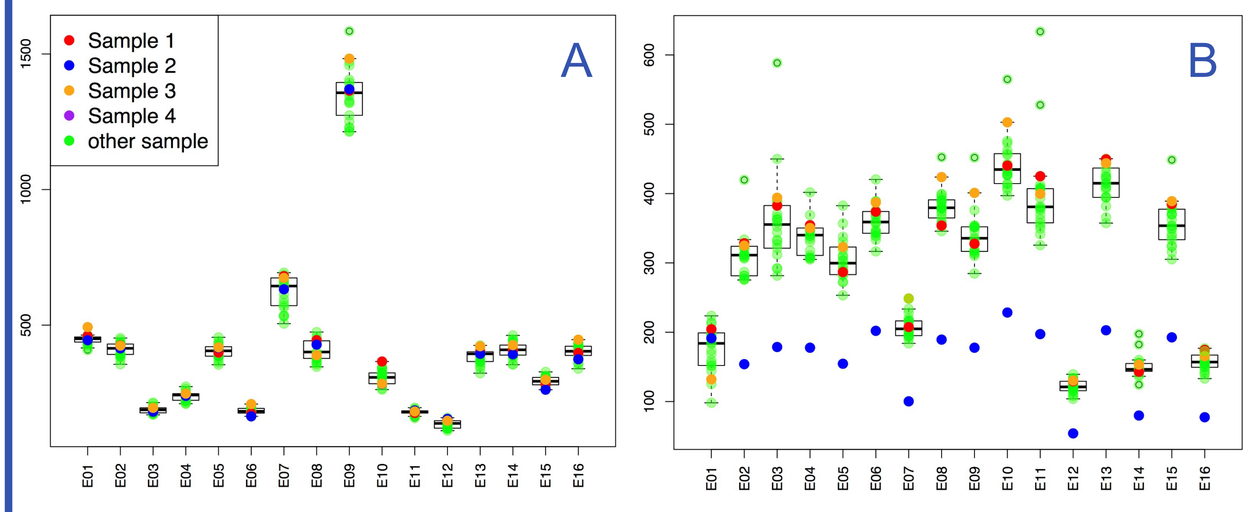
- **extension of cn.MOPS** for targeted panel sequencing data
- **special preprocessing:** quality control & read counting
- **binning option:** count reads in small overlapping bins (30-50bp) to increase resolution
- **optimized parameters**
- **special postprocessing** instead of standard segmentation
- **filter** for displaying CNVs only for genes of interest



Preprocessing

Quality control: exclude
- **exons** with median read count (RC) across all samples < 30 (recommended threshold)
- **samples** with median RC across all exons more than 1SD lower than median of all samples

Read counting: adapted to characteristics of targeted panel sequencing data



Read counts for exons 1-16 of the *NF1* gene. Sample 2 (blue) has a deletion from E02 – E16.
Panel A - WGS RC method: deletion can not be detected
Panel B - new RC method: deletion clearly visible

Results: simulated data

Simulated data (resemble the characteristics of targeted panel sequencing data):
240 data sets, different frequencies of CNVs

Deletions:

Method	% CN2	Recall	Precision	Bal. Accuracy	F2-score
panelcn.MOPS	90	1.0000	0.9996	1.0000	0.9999
	95	0.9998	0.9997	0.9998	0.9997
	99	0.9998	0.9997	0.9998	0.9997
	Ø	0.9999	0.9997	0.9999	0.9998
ExomeDepth	90	1.0000	0.9973	0.9997	0.9995
	95	0.9980	1.0000	0.9982	0.9984
	99	0.9915	1.0000	0.9924	0.9932
	Ø	0.9965	0.9991	0.9967	0.9970

Duplications:

Method	% CN2	Recall	Precision	Bal. Accuracy	F2-score
panelcn.MOPS	90	0.9979	0.9919	0.9973	0.9967
	95	0.9979	0.9919	0.9973	0.9967
	99	1.0000	0.9991	0.9999	0.9998
	Ø	0.9986	0.9943	0.9982	0.9977
ExomeDepth	90	0.9973	0.9841	0.9959	0.9944
	95	0.9973	0.9841	0.9959	0.9944
	99	0.7988	1.0000	0.8152	0.8317
	Ø	0.9311	0.9894	0.9357	0.9402

Results: Cancer Predisposition Gene Panel

TruSight cancer panel:
- targets 94 cancer predisposition genes
- e.g.: *NF1/2*, *BRCA1/2*, *APC*, *MSH2/6*, *MLH1* and *PMS2*

36 samples:
- 5 samples excluded due to poor quality
- 16 normal samples
- 13 deletions
- 2 duplications

Size of CNVs:
- Deletions: 80 bp (37 bp overlapping with ROI) – 350 kbp (whole *NF1* gene)
- Duplications: 1 kbp – 80 kbp

Deletions:

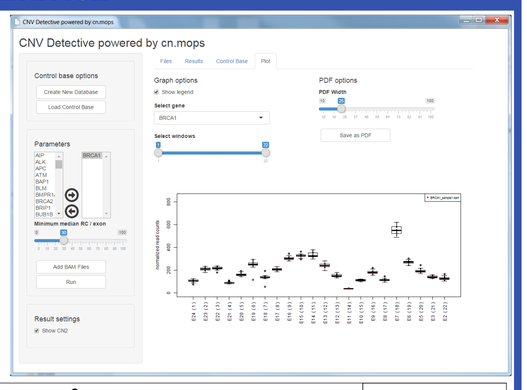
	TP	FP	TN	FN	Recall	Precision	Bal. Acc.	F2-score
panelcn.MOPS	13	0	18	0	1	1	1	1
ExomeDepth	8	0	18	5	0.6154	1	0.8077	0.6667

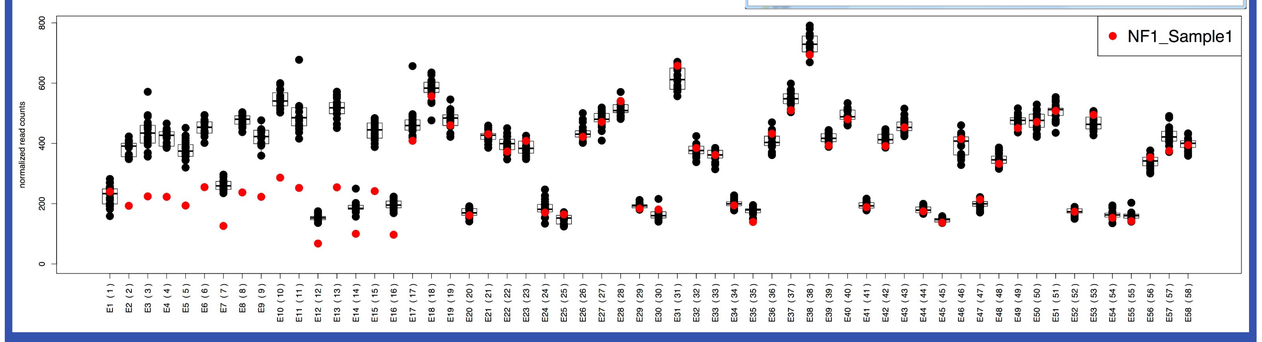
Duplications:

	TP	FP	TN	FN	Recall	Precision	Bal. Acc.	F2-score
panelcn.MOPS	2	0	29	0	1	1	1	1
ExomeDepth	2	0	29	0	1	1	1	1

Graphical user interface

- standalone app for better user-experience
- based on R shiny
- simple installer for Windows
- includes quality control
- option for binning
- option for building up control database
- reports only CNVs for genes of interest
- results exportable as .csv
- read count plots for genes of interest





Conclusion

- panelcn.MOPS for CNV detection in targeted panel sequencing data
- superiority shown on simulated and 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.