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



JOHANNES KEPLER UNIVERSITY LINZ

Gundula Povysil¹, A. Tzika², J. Vogt², V. Haunschmid¹, L. Messiaen³, K. Wimmer², G. Klambauer¹, S. Hochreiter¹ 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.

Data													
	Training Cohort				Test Cor	nort	Total						
CN	13				110		123						
Multi Exon Deletions				5			11		16				
Multi Exon Duplications				2			1			3			
Single Exor	י Dele	etions					13			13			
Single Exon	Dupli	cations					2			2			
Whole Gen	e Dele	etions		4			1		5				
Subt	total		24				138		162				
			Cases excluded from calculations										
Deletions	S < 1	ROI		1			2		3				
Duplication	ns < 1	ROI					2			2			
De novo Alu Insertions				3				3					
To	tal			25 145						170			
Results: Test Set													
	nanal		Evomo	Donth			VicCon	NovtCE		SocNov	tontimal		
ТD	paner	86						NexiGENE		Sequex			
		7843		8156 7		7787 7200		8080		700	7 8173		
		043		17	17 770		11	11		1901			
FN		0		1	l 1		1						
		330		0	0 3		386 873		77				
Total		8259		8259 8		8259 8259		8259		8259	8259		
rotai		0200				0200 0200		0200		0200	0200		
Sensitivity		1.0000 0).9884 0.9		9882 0.9882		1.0000		0.9884	1.0000		
Specificity		1.0000 0		0.9979 0.		9999 0.9985		0.9979		1.0000	0 1.0000		
No Call Rate		0.0400	0.	.0000 0.04			467 0.1057 0.00			0.0322	2 0.0000		
numbers corre	spon	d to numh	ers of	ROIs									
				R	esults	S							
pan		panelcn.N	/IOPS	ExomeDepth		CoNVaDING		VisCap Ne		ktGENe	SeqNext		
Sensitivity		+++		++		++		++ ++		+	++		
Specificity		+++		+		++		+ +			+++		
No Call Rate		+		+++		+		- ++			+		
Low Quality Filter		+++		+		+++		+ +			+++		
CNVs < 1 ROI		++		++		+++		++ +			-		
Whole Gene CNVs		+++		+++		+++					+		
Incidental Findings		+++		-		-		- +			+		
Runtime		+++		++		-	-		+ +		+		
GUI		+++		-		-		-	- ++		+++		
Non-Commercial		+++		+++		+++		+++ -			-		

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 qu 	ality"
--	--------

• samples with low median RC across all ROIs excluded from controls, warning for test

Read Counting	panelcn.MOPS Pipeline						
Quality Control	for n test samples						
Control Sample Selection	for each test sample separately						
Normalization							
CN Detection	for each ROI separately						

Graphical User Interface

CNV Detective powered by panelcn.mops

8	Samples St	Samples Show Control Base Create Control Base Results Plot Edit BED File App settings										
d Control Base	Show CN2								Export as CSV			
	Show 25 🔹 ent	Show 25 • entries Search:										
S	Sample 💧	Chr 🗄	Gene 🖕	Exon	¢	Start 🔶	End 🔶	RC 🔶	medRC 🖕	RC.norm 🔶	medRC.norm 🔶	lowQual CN 🔶
d Test Samples	CNV_IBK1.bam	17	NF1	NF1.E1.chr17.2942229	7.29422418	29422297	29422418	779	828	739	765	CN2
f used samples	CNV_IBK1.bam	17	NF1	NF1.E2.chr17.2948297	0.29483175	29482970	29483175	562	1097	533	1007	CN1
le	CNV_IBK1.bam						29486142	676			1150	CN1
nknown	CNV_IBK1.bam	17	NF1	NF1.E4.chr17.2949017	3.29490425	29490173	29490425	666	1211	632	1148	CN1
	CNV_IBK1.bam	17	NF1	NF1.E5.chr17.2949687	8.29497046	29496878	29497046	550	1069	522	987	CN1
eters	CNV_IBK1.bam	17	NF1	NF1.E6.chr17.2950840	9.29508538	29508409	29508538	755	1276	717	1212	CN1
	CNV_IBK1.bam										763	CN1
ames.txt 💌	CNV_IBK1.bam	17	NF1	NF1.E8.chr17.2950949	5.29509714	29509495	29509714	699	1410	663	1288	CN1
ilable Selected	CNV_IBK1.bam	17	NF1	NF1.E9.chr17.2952740	9.29527644	29527409	29527644	628	1176	596	1068	CN1
▲ NF1 ▲	CNV_IBK1.bam	17	NF1	NF1.E10.chr17.295280	24.29528208	29528024	29528208	817	1572	775	1436	CN1
æ	CNV_IBK1.bam	17	NF1	NF1.E11.chr17.295283	98.29528534	29528398	29528534	696	1212	661	1171	CN1
	CNV_IBK1.bam	17	NF1	NF1.E12.chr17.295332	27.29533420	29533227	29533420	228	464	216	429	CN1
G	CNV_IBK1.bam	17	NF1	NF1.E13.chr17.295414	38.29541634	29541438	29541634	728	1401	691	1320	CN1
2 🗸	CNV_IBK1.bam						29546167	313			573	CN1
median RC / exon	CNV_IBK1.bam	17	NF1	NF1.E15.chr17.295488	37.29548978	29548837	29548978	660	1165	626	1097	CN1
	CNV_IBK1.bam							322			601	CN1
20 30 40 50 60 70 80 90 100	CNV_IBK1.bam	17	NF1	NF1.E17.chr17.295520	82.29552299	29552082	29552299	1188	1340	1128	1244	CN2
	CNV_IBK1.bam	17	NF1	NF1.E18.chr17.295534	22.29553733	29553422	29553733	1653	1788	1569	1554	CN2
Run	CNV_IBK1.bam	17	NF1	NF1.E19.chr17.295542	05.29554340	29554205	29554340	1319	1385	1252	1252	CN2
Close window	CNV_IBK1.bam	17	NF1	NF1.E20.chr17.295545	10.29554655	29554510	29554655	510	561	484	512	CN2
	CNV/ IBK1 ham	17	NF1	NF1 F21 chr17 295560	12 29556514	29556012	29556514	1327	1386	1260	1267	CN2
								1				

CNV_IBK8.bam

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[®] \rightarrow 300 cycles with paired end reads



NF1

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

• standalone app

1500

based on R shiny

• simple installer for Windows

results exportable as .csv

quality control for samples and ROIs

• reports only CNs for genes of interest

read count plots for genes of interest

• option for building up control base

• 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.

