Detecting regions with copy number variation

- Choosing a method for copy number detection
- Detecting amplifications and deletions with Genomic Segmentation
- Visualizing regions of interest
- Analyzing shared regions of copy number variation
- Visualizing shared regions of copy number variation

Starting with copy number estimates for each marker (either taken directly from the vendor's input file or calculated previously), the next step is to create a list of regions where adjacent markers share the same copy number.

Choosing a method for copy number detection

There are two algorithms available for copy number region detection: Genomic Segmentation and Hidden Markov Model (HMM). Both algorithms look for trends across multiple adjacent markers. The genomic segmentation algorithm identifies breakpoints - changes in copy number between two neighboring regions. The HMM algorithm looks for discrete changes of whole number copy number states (e.g., 0, 1, 2... with no upper limit) and will find regions with those numbers of copies. Therefore, the HMM model performs better in cases of homogeneous samples such as clinical syndromes with underlying copy number aberrations. Genomic segmentation is preferable for heterogeneous samples such as cancer because tumor biopsies often contain "contaminating" healthy tissue and a tumor can have cells with different genomic aberrations.

Detecting amplifications and deletions with Genomic Segmentation

The number of copies of each marker created in the previous step will be used to detect the genomic regions with copy number variation, i.e., to identify amplifications and deletions across the genome.

- Select the IC_IntensitiesSNP6pairedcopynumber spreadsheet in the Analysis tab
- Select Detect Amplifications and Deletions from the Copy Number Analysis section of the workflow (Figure 1)



Figure 11. Invoking Detect Amplifications and Deletions

The Detect Amplifications and Deletions dialog will give you the option to choose Genomic Segmentation or HMM Region Detection (Figure 2).



Figure 12. Select a method for detecting amplifications and deletions

Select Genomic Segmentation

Select OK

The Genomic Copy Number Segmentation dialog gives options for setting segmentation parameters and the configuring the region report (Figure 3).

🧐 Genomic Copy Num	ber Segmentation Spreadsheet 2	×
-Segmentation Paramete	rs	
Minimum genomic markers	50	•
P-value threshold	0.001	
Signal to noise	0.3	•
Region Report		
Diploid copy number from	1.7 to 2.3	
Result file	segmentation.txt	Browse
	Report unchanged regions.	
Report SNP and CN	/ counts	0
	ОК	Cancel
\odot		

Figure 13. Configuring the Genomic Copy Number Segmentation dialog

- Set *Minimum genomic markers* to **50**
- Leave the rest of the parameters set to default values as shown (Figure 3)
- Select OK

The *Genomic Segmentation* task is divided into two steps. In the first step, each region is compared to an adjacent region to determine whether both have the same average copy number and whether a breakpoint can be inserted. This is determined by first using a two-sided t-test to compare the average intensities of adjacent regions and then checking whether the corresponding cut-off p-value is below the specified *P-value threshold*. The genomic size of a region is defined by the number of genomic markers in the region, *Minimum genomic markers*, while the magnitude of the significant difference between two regions is controlled by *Signal to noise*, which can be thought of as the difference in copy numbers between the regions. If the t-test is significant, the copy number of the region differs significantly from its nearest neighbors. However, a second step is needed to detemine whether the difference is due to compare the mean copy number in the region with the expected diploid copy number. For a detailed explanation of the genomic segmentation procedure, please consult our guide, Optimizing Copy Number Segmentation.

The resulting spreadsheet, segmentation, shows one row per genomic region per sample (Figure 4). The columns provide the following information:

1-4. Genomic location of the region

- 5. Sample ID
- 6. Description of the copy number change

7. The length of the region (in base pairs)

- 8. The number of markers in the region
- 9. Markers density in the region (region length in base pairs divided by the number of markers)

10. Geometric mean of the copy number of all the markers in the region

11. Minimum p-value of the one-sided t-tests of the difference of the copy number in column 10 vs. the diploid range

														Workflows Copy Number	
elysis X Scatter Plot X Chromosome V	iew X												-0	Copy Number	>
		1 III III III	\bigcirc	0										/ Import	
			~											Import samples	
1 (IC_Intensities_SNP6)	Curre	nt Selection chr:	1										、 ·	Add Sample Attributes	
2 (IC_Intensities_SNP6_pairedcopynum		1. chromosome	2. start	3. end	4. cytoband	5. Sample ID	6. Conv Number	7. length(bps)	8. # markers	9. marker density	10. mean	11. p-value(region)	11	View Sample Information	
segmentation (segmentation.txt)	1.	chr1	61735	761550	1p36.33	IC_151T_FF.C	E Unchanged	699815	50	13996.3	2.25919	0.637753			
	2.	dr1	761550	12192261	1p36.33 -	IC_151T_FF.C	E Unchanged	11430711	6216	1838.92	1.8545	1		Choose Sample ID Column	
	3.	dr1	12192261	12315320	1p36.22	IC_151T_FF.C	E Deletion	123059	84	1464.99	1.50716	0.000128171		Create Copy Number (from Allele Intensities Only)	
	4.	dr1	12315320	25588234	1p36.22 -	IC_151T_FF.C	E Unchanged	13272914	8062	1646.35	1.87173	1		∕ QA/QC	
	5.	dr1	25588234	25699232	1p36.11	IC_151T_FF.C	E Deletion	110998	50	2219.96	1.27462	0.000856841		PCA Scatter Plot	
	6.	dr1	25699232	26328074	1p36.11	IC_151T_FF.C	E Unchanged	628842	288	2183.48	1.81903	0.999996		Sample Histogram	
	7.	dr1	26328074	33393918	1p36.11 -	IC_151T_FF.C	E Deletion	7065844	3535	1998.82	1.41851	0		Chromosome View	
	8.	dr1	33393918	33606055	1p35.1	IC_151T_FF.C	E Deletion	212137	107	1982.59	1.1932	4.13048e-24		/ Come Number Analysis	
	9.	chr 1	33606055	35546698	1p35.1 - 1p34.	3IC_151T_FF.C	E Deletion	1940643	1339	1449.32	1.49033	0	F		
	10.	dr 1	35546698	57198030	1p34.3 - 1p32.	2 IC_151T_FF.C	E Amplification	21651332	12623	1715.23	2.39176	0		Detect Amplifications and Deletions	
	11.	dr1	57198030	72760190	1p32.2 - 1p31.	1IC_151T_FF.C	E Unchanged	15562160	11233	1385.4	2.14522	1		Analyze Detected Segments	
	12.	chr 1	72760190	72818834	1p31.1	IC_151T_FF.C	E Amplification	58644	57	1028.84	2.6741	3.90572e-05		View Detected Regions	
	13.	dr1	72818834	115322035	1p31.1 - 1p13.	2IC_151T_FF.C	E Unchanged	42503201	27612	1539.3	2.10198	1		Create Region List	
	14.	chr1	115322035	145107483	1p13.2 - 1q21.	1 IC_151T_FF.C	E Amplification	29785448	3883	7670.73	2.55142	0		Find Overlapping Genes	
	15.	dr1	145107483	229146879	1q21.1 -	IC_151T_FF.C	E Unchanged	84039396	54664	1537.38	2.27472	1			
	16.	dr1	229146879	249224389	1q42.13 - 1q44	IC_151T_FF.C	E Unchanged	20077510	14732	1362.85	1.94588	1		Overlap with Known SNPs	
	17.	dr2	12784	29628028	2p25.3 - 2p23.	2IC_151T_FF.C	E Amplification	29615244	20895	1417.34	2.44363	0		Test for Known Abnormalities	
	18.	dr2	29628028	30725801	2p23.2 - 2p23.	1 IC_151T_FF.C	E Amplification	1097773	852	1288.47	3.00951	0		Visualization	
	19.	dr2	30725801	37843472	2p23.1 - 2p22.	2IC_151T_FF.C	E Amplification	7117671	5071	1403.6	2.49528	0		Biological Interpretation	
	20.	chr2	37843472	37911962	2p22.2	IC_151T_FF.C	E Unchanged	68490	69	992.609	2.06658	0.999748		Genomic Integration	
	21.	dhr2	37911962	43356725	2p22.2 - 2p21	IC_151T_FF.C	E Amplification	5444763	4091	1330.91	2.47305	0			
	22.	chr2	43356725	43466379	2p21	IC_151T_FF.C	E Unchanged	109654	67	1636.63	2.00051	0.999984			
	23.	chr2	43466379	44192732	2p21	IC_151T_FF.C	E Amplification	726353	479	1516.39	2.55679	7.221e-23			
	24.	chr2	44192732	44246131	2p21	IC_151T_FF.C	E Unchanged	53399	50	1067.98	2.05307	0.864468			
	25.	chr2	44246131	65497154	2p21 - 2p14	IC_151T_FF.C	E Amplification	21251023	14349	1481.01	2.46436	0			
>	Rows:	: 3170 Columns:	11 <									> \	/ 1	Related: LOH Allele Specific Comy Number	

Figure 14. Viewing the segmentation spreadsheet

If desired, you can use Merge Adjacent Regions under Tools in the main toolbar to combine similar regions.

Visualizing regions of interest

Individual regions of interest can be visualized using Chromosome View.

- Right-click a row header in the segmentation spreadsheet
- Select Browse to location from the pop-up menu

Alternatively, you can visualize results at the whole chromosome level.

- Select the segementation spreadsheet
- Select Chromosome View from the QA/QC section of the workflow

The *Genomic Segementation* track displays the segmentation results (Figure 5). Each line in the track represents a sample. Amplified, deleted, and unchanged regions are shown in red, blue, and white, respectively. The *Profile* track now also includes information from the segmentation spreadsheet for the selected sample.



Figure 15. Segmentation results shown as regions of amplification and deletion in each sample

Analyzing shared regions of copy number variation

Amplified and deleted regions in each sample have been detected, we can compare the regions across multiple samples to detect copy number changes that are shared by multiple samples.

• Select Analyze detected segments from the Copy Number Analysis section of the workflow

The Analyze Segments task (Figure 6) can test for associations between copy number variations and sample categories using the ² test. In this tutorial, all pairs share the sample phenotype, so we will not test for associations.

Analyze Segments (2/segmentation)	×
The total number of amplifications and deletions will be reported for each region with the sa samples.	ame aberation state across all
One or more phenotypes may be selected. Each selected phenotype will be tested (indepe for association with the copy number status.	ndent of other phenotypes)
Test association with phenotypes (Optional)	
3. Tumor	•
4. SubjectID	
5. Gender	
🗌 6. Scan Date	
Reporting options	
Add columns with sample ID lists.	0
Result File segment-analysis	Browse
l l l l l l l l l l l l l l l l l l l	OK Canad
	Cancel

Figure 16. Viewing the Analyze segments dialog

- Leave all boxes unchecked
- Select **OK** to run the Analyze Segements task

The task generates a new spreadsheet, summary (segment-analysis) (Figure 7), with one region per row. The columns provide the following information:

1-4. Genomic locations of the regions

5. Total number of samples

6-7. Number of samples with amplifications and the average amplified copy number, respectively

8-9. Number of samples with deletions and the average deleted copy number, respectively

10. Total number of samples with copy number abberations

11-12. Number of samples with no change in copy number and the average copy number in those samples, respectively

- 13. Number of markers in the region
- 14. Length of the region (in base pairs)

15+. Two columns per sample - the average copy number in each sample as well as the copy number change status of the sample sample (e.g., <u>amplified</u>, <u>del</u>eted, <u>unchanged</u>, depending on the copy number and the threshold for unchanged defined in the *Genomic Segementation* dialog)

A "?" indicates that a region with the particular characterisitic does not exist or cannot be computed. For example, if a region is not amplified in any of the samples, the average amplified copy number will be shows as "?". This list may be filtered to contain only regions that meet user-specified criteria as discussed in the next section of the tutorial.

Partek Genomics Suite - 2/segmentation/s File Edit Transform View Stat Filter	summary (segment- Tools Window	analysis) Custom Help										- 0	×
												Workflows Copy Number	~
Analysis × Scatter Plot × Chromosome	View X Chromoso	ome View X									-	Copy Number	\times
🖿 👝 🗔 💽 🗠 🛤	ka 🛃		0									✓ Import	
												Import samples	
1 (IC_Intensities_SNP6)	Current Selection 4									,	^	Add Sample Attributes	
 2 (IC_Intensities_SNP6_pairedcopynu segmentation (segmentation.bxt) 	1. Chromosome	2. Start	3. Stop	4. cytoband	5. Total Number of Samples	6. Total Amplifications	7. Amplification Average Copy	8. Total Deletions	9. Deletion Average Copy	10. Total Aberr		View Sample Information	
summary (segment-analysis)							Number		Number	ations		Choose Sample ID Column	
	1. 4	47231926	48070430	4p12	10	3	2.49926	7	1.55236	10		Create Copy Number (from Allele Intensities Only)	-
	2. 4	48070430	48414701	4p12 - 4p11	10	2	2.45189	8	1.49239	10		✓ QA/QC	
	3. 4	48414701	48814423	4p11	10	3	2.42118	7	1.55236	10		PCA Scatter Plot	
	4. 4	49658417	52689480	4p11 - 4q11	10	3	2.38826	7	1.3896	10		Samala Histogram	-
	5. 4	52689480	52710591	4q11 - 4q12	10	2	2.402	8	1.42555	10		Sample Histogram	
	6. 4	52874472	52887805	4q12	10	3	2.41096	7	1.43175	10		Chromosome View	
	7. 4	52887805	53426984	4q12	10	4	2.38421	6	1.46572	10		✓ Copy Number Analysis	
	8. 4	53426984	55077283	4q12	10	4	2.38421	6	1.54261	10		Detect Amplifications and Deletions	
	9. 4	55141568	55262530	4q12	10	4	2.37853	6	1.54261	10		Analyze Detected Segments	-
	10. 4	103740604	104289463	4q24	10	0	?	10	1.45536	10		View Detected Regions	
	11. 4	182912966	183060840	4q34.3	10	1	2.43982	9	1.41905	10		Ourste Desting Link	
	12. 6	107088052	107306530	6q21	10	/	4.32094	3	1.58/68	10		Create Region List	
	13. 8	39233131	39241139	8p11.22	10	0	2.83273	4	1.09004	10		Find Overlapping Genes	
	14. 8	39386079	39387632	8p11.22	10	0	2.836//	4	1.09004	10		Overlap with Known SNPs	
	15. 8	122963746	124297071	0-24.12	10	7	2.90259	2	1.45702	10		Test for Known Abnormalities	
	17 9	127257071	127410500	0-24.25	10	7	3.00253	2	1.45702	10		> Visualization	
	19 9	127630013	127650015	8024.21	10	7	2.00702	3	1.45702	10		Biological Interpretation	
	10. 0	120557066	129557000	9024.21	10	7	2 26911	2	1.45702	10		Genomic Integration	
	20 8	129580399	134028504	8074 21 -	10	7	3 00977	3	1 45792	10		,g.	
	21. 8	134028504	134092355	8024.22	10	7	2.86928	3	1.45792	10			
	22. 8	134092355	140447722	8024.22 -	10	7	2,99066	3	1,45792	10			
	23. 8	141071354	144280118	8024.3	10	7	3.01896	3	1,45792	10			
	74 8	144790119	146708156	8074 3	10	7	3 0 1896	3	1 52853	10			
< >	Rows: 2887 Columns	: 34 <								> .	~	Related: LOH Allele Specific Conv Number	
	-											Lorr Anere Operate Oupy Number	

Figure 17. Viewing the results of Analyze Detected Segments

Visualizing shared regions of copy number variation

To get an overiew of the common abberations in the group of samples over the entire genome we can use View Detected Regions.

• Select View Detected Regions

The View Detected Regions dialog (Figure 7) allows you to select the spreadsheet with genomic regions and choose between histogram and copy number classification plots.



Figure 18. View Detected Regions dialog

- Select summary (segment-analysis) from the drop-down menu
- Select View Histogram
- Select OK

The plot will open in a new tab titled Karyogram View (Figure 8).



Figure 19. Viewing amplification and deletion histograms using Karyogram View

The *Karyogram View* shows each chromosome with red and blue histograms on either side corresponding to amplification and deletion, repsectively. The histogram height reflects the number of samples that share either amplification of deletion a that particular region. For example, the long arms of chromosomes 3 and 7 are amplified in the majority of samples and most samples share a deletion in the long arm of chromosome 4.

Mousing over the chromosome will give cytoband information, mousing over the histogram will give the number of shared regions at each position and the number of samples sharing the type of variation. Both the menu and display may be used to control which chromosomes are displayed; left-click in the menu to toggle a chromosome on/off and right click in the menu or graph to show only that chromosome.

Alternatively, we can use the Copy Number Classification plot to get a more sample-centric view.

- Select View Detected Regions
- Select View Copy Number Classification
- Select OK

The Copy Number Classification also utilizes Karyogram View to provides an overview of all the samples and the copy number of regions on each chromosome (Figure 9).

														Workflows Copy Number	
Analysis \times Scatter Plot \times	Chromosome View X Ch	nromosome V	/iew × Ka	aryogram Vie	w× Kary	ogram View	×							Copy Number	\times
Chromosomes Style		 Amplif 	fication	Dele	etion	Uncha	nged							✓ Import	
Left click to toggle Right click to show only	^													Import samples	
CHR 1														Add Sample Attributes	
CHR 2		_	_	_	_	_	_	_	_	_	_	_	_	View Sample Information	
CHR 3		ß	ŝ	ŝ	ŝ	ŝ	ŝ	ŝ	ŝ	ŝ	ŝ	ŝ	ŝ	Choose Sample ID Column	
CHR 4		51	51	51	511	51	51	511	51	517	511	51	51	Create Copy Number (from Allele Intensities Only)	~
CHR 5		' <u>'</u>	'	' <u>-</u>	' <u>'</u>	' <u>'</u>	' <u>'</u>	' <u>-</u> '	' <u>'</u>	' <u>-</u> '	' <u>'</u>	' <u>'</u>	' <u>'</u>	V QA/QC	
CHR 6		I I I I I I I					EO		Ē	Ē				PCA Scatter Plot	~
CHR 7				Щ	P	P	ЩЩ Ц	P	P	Ē	Ē	P	Ē	Sample Histogram	
CHR 9						et tan								Chromosome View	~
CHR 10			19.0		Y H				AN 11	A 3 5 10				Copy Number Analysis	
CHR 11				VIL.			VII H	V II	Will		ŧ.		96 SC	Detect Amplifications and Deletions	
CHR 12				AF F	- 1			A H 1	A A		A it	YF H	AP IP	Analyze Detected Segments	
CHR 13												Ar li		View Detected Regions	
CHR 14				i b				11						Create Region List	· ·
CHR 15		1	2	3	4	5	6	7	8	9	10	11	12	Find Overlanning Genes	
CHR 16							1							Overlap with Known SNPs	
CHR 17														Test for Known Abasensities	
CHR 19														Viewelization	
CHR 20														Biological Interpretation	
CHR 21														Genomic Integration	
CHR 22		1	7												
CHR X		<u> </u>	A H	Á II DT		0,310	8 III I					A II 1			
CHR Y					A	AL L	- H -	Ø Grade	₹.	8	9		8		
Show all	Ý							ACL	ATH						

Figure 20. Viewing the Copy Number Classification plot

Each sample is drawn as a separate column next to the chromosome. Amplified regions are depicted in red, deleted regions in blue, and regions with no copy number change in white. Sample names are given accross the top of each column. For greater detail, try viewing fewer chromosomes.

« Creating Copy Number from Allele Intensities Creating a list of regions »

Additional Assistance

If you need additional assistance, please visit our support page to submit a help ticket or find phone numbers for regional support.



Your Rating: ☆☆☆☆☆ Results: ★★★★ 34 rates