

Copy Number Analysis (CNVkit)

The copy number detection task is used to detect regions of DNA copy number imbalance within the genome for DNA-Seq experiments. Partek Flow provides the CNVkit¹ methodology (<https://cnvkit.readthedocs.io/en/stable/>) to find regions of altered copy number, optimized for targeted resequencing of whole-exome and targeted panels that utilize a hybrid capture approach. The methodology uses both targeted reads and nonspecific off-target reads divided into bins to determine copy number, subsequently normalizing the data to a pooled reference of control samples and correcting for systematic biases.

CNVkit dialog

The CNVkit task can be found under the *Copy number analysis* tab in the context-sensitive menu when any Aligned reads or Filtered reads data node is selected. The dialog consists of four sections: *Define controls*, *Select Reference Assembly*, *Target*, and *Advanced options* (Figure 1).

Define controls

Define controls ☒ By attribute ☐ Manually

Control attribute

Control term

Select Reference sequence

Assembly

Target

Method

Assembly

Target regions

Advanced options

Option set [Configure](#)

[Back](#) [Finish](#)

Figure 1. Example of the CNVkit dialog for copy number detection. The projects contains a sample attribute that specifies control.

Define controls allows for the specification of control samples in the project that will be pooled to create a reference copy number for both on- and off-target genomic bins using bias-corrected read depth from each control sample. In projects that contain matched tumor/normal samples, all normal samples should be included in the control sample pool. Control samples can be group based on a categorical attribute in the data tab or manually selected. If no control samples are available, it is possible to run CNVKit with no controls by leaving the sample pool empty. This will create a "flat" reference for neutral copy number.

Select reference sequence will utilize the species genome build utilized for alignment. If the selected aligned data node was imported, the reference assembly used during data alignment needs to be specified from the drop-down list. The Assembly can be previously associated with Partek® Flow® via [Library File Management](#) or added on the fly.

The Target section allows for the selection of Method, where the type of sequencing experiment can be specified. While CNVkit is optimized for copy number detection in hybrid capture experiments, it can also analyze whole genome and amplicon sequencing data. Assembly should match the reference sequence previously specified. For hybrid capture or amplicon experiments, Target regions should relate to the regions sequenced in this study using a Gene/feature annotation file. For whole genome data, Annotation will specify a Gene/feature annotation file that will be used to annotate regions of copy number imbalance. The Gene/feature annotation can be previously associated with Partek® Flow® via [Library File Management](#) or added on the fly.

Advanced options allows for customization of parameters for CNV detection. Please refer to the CNVkit documentation (<https://cnvkit.readthedocs.io/en/stable/>) for details on the available parameters.

CNVkit report

Selecting the Task report for the Copy number data node will provide a table with test samples in the project (non-control) and the number of copy number regions associated with each sample (Figure 2). Selecting Option columns at the top right of the table will allow for attribute columns to be added to the report.

Optional columns

Sample ↕	# regions ↕
ERR1618872	561
ERR1619104	542
ERR1619286	400
ERR1619381	297
ERR1619394	251
Rows per page 25 (1 of 1)	

Figure 2. Example of the summary report for CNVkit, providing information on all test samples in the project.

Selecting a sample in the table will open the sample-specific CNVkit report (Figure 3). The sample table will provide a row for each detected region of copy number imbalance in the data with Chromosome, Start and End coordinates. The table also includes the following information:

- gene: content in the copy number region as defined by the associated Target regions or Annotation file
- log2: the weighted log2 mean coverage depth across all bins in the segment
- depth: the weighted mean of absolute-scale mean coverage depth for bins in the segment
- probes: the number of bins covered by the segment
- weight: the sum of bin level weights for the segment that denotes reliability based upon bin size and the square of the log2 spread in the pooled reference

	View	↕ chromosome	↕ start	↕ end	↕ gene	↕ log2	↕ depth	↕ probes	↕ weight
1		chr1	65,509	110,224,541	-.1,ref OR4F5,ref...	-0.01	47.88	13,952	9,088.23
2		chr1	110,230,354	110,233,098	ens ENST000004607...	-22.86	0.01	6	3.70
3		chr1	110,235,803	121,611,853	ref GSTM1,ref NM_...	9.52E-3	48.97	1,379	907.28
4		chr1	122,062,889	136,947,095	-	-17.21	4.59E-6	58	24.55
5		chr1	137,849,168	142,509,878	-	-16.60	1.68E-5	16	6.74
6		chr1	142,509,878	152,056,012	ens ENST000004918...	0.02	51.55	1,252	824.38
7		chr1	152,056,512	153,190,751	mRNA KJ900047,mRN...	-0.09	50.76	204	126.94
8		chr1	153,190,751	153,789,100	ref PRR9,ref NM_0...	0.01	50.95	181	119.17
9		chr1	153,789,850	153,957,204	ref GATAD2B,ref N...	0.12	58.86	72	52.69
10		chr1	153,957,209	154,525,416	ref RAB13,ref NM_...	-0.02	49.56	212	141.53
Rows per page 10   (1 of 26)  									

Figure 3. Example of the sample-specific table of regions identified by CNVkit copy number analysis.

Selecting the chromosome icon  in the view column will link to Chromosome view.

References

1. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput Biol*. 2016;12(4):e1004873. doi:10.1371/journal.pcbi.1004873

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:  28 rates