Peak Callers (MACS2)

The Peak calling task is used to detect enriched genomic regions on reads generated from nucleic acid enrichment experiments such as ChIP-seq, DNase-seq, and MeDIP-seq. experiments. Partek Flow provides the widely used method of MACS2-model-based analysis¹ (http://liulab.dfci.harvard.edu/MACS/) to find peaks. It can be performed with or without control sample.

MACS2 is used to demonstrate the task setup in this manual. However, Flow also provides the MACS3 task which has the same interface. If you would like to make the switch, please talk to the Flow Admin or tech support team.

MACS2 dialog

Selecting MACS2 from the context sensitive menu will bring up the MACS2 task dialog. The interface will appear differently depending on the input aligned data node and whether there are sample attributes available in the Data tab.

If the selected aligned data node was imported, the reference assembly used during data alignment needs to be specified. Choose the **Assembly** from the drop-down list within the MACS2 dialogue (Figure 1). If the selected aligned data node was generated by Partek Flow, this option will not appear.

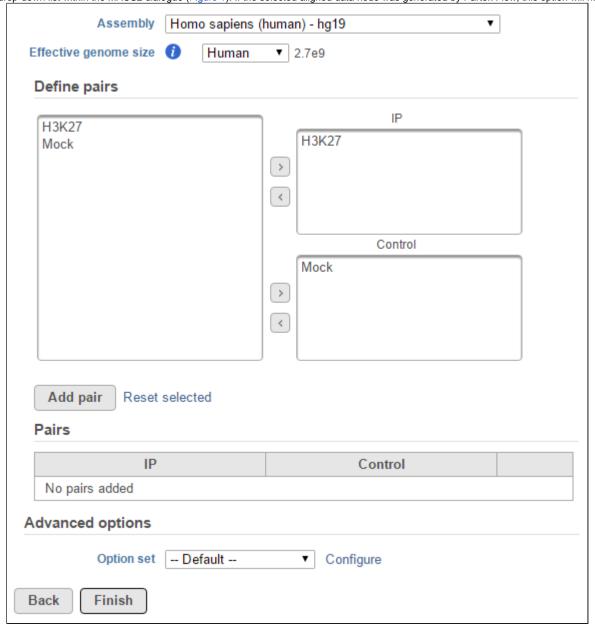


Figure 1. MACS2 dialog: manually add ChIP vs control pairs for peak detection

The **Effective genome size** must be configured prior to running the peak caller. It refers to the size of the genomic regions that are actually mappable. This size is smaller than the actual size of an organism's whole genome because of the presence of repetitive features. They are typically about 70%-90% of the whole genome. The MACS2 authors¹ have recommended presets available for four different species. Select from the drop-down menu the preset that best describes the genome you are working with. They are as follows:

- Human (*Homo sapiens*) 2.7 x 10⁹
- Mouse (*Mus musculus*) 1.87 x 10⁹
- C. elegans 9 x10⁷
- Fruitfly (*Drosophila melongaster*) 1.2 x 10⁸

If none of these presets match your genome of interest, select *Other...* Then enter the effective genome size (Figure 2). The values are in base pairs (bps). Consult the MACS documentation for guidance on selecting the best effective genome size for your experiment.

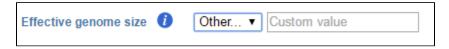


Figure 2. Specify other species effective genome size by manually type in the value

For data where no sample attributes are specified, the peak detection pairs need to be manually defined. In the example in Figure 1, there only two samples. Under the **Define pairs** section, the left panel lists all the sample names uploaded to the project (*H3K27* and *Mock*). Add one pair at a time by dragging the corresponding samples to either the *IP* panel on the top-right or the *Control* panel on the bottom-right. If no control samples are present in the experiment, leave the *Control* panel blank. If more than one ChIP or Control samples are added, the samples will be combined (or pooled) during the analysis. After defining a pair, click the **Add pair** button.

For data where the sample attributes are defined (Figure 3), you will have an additional option to add pairing or grouping based on sample attributes. Figure 3 shows an example dataset with 4 samples, 2 time points, and there is one IP sample and one Input sample in each time point.

	Sample name	Attributes	
		ChIP	Time
1	T1_input	Input	2h
2	T1_IP	IP	2h
3	T2_input	Input	6h
4	T2_IP	IP	6h
Show data files Download			wnload

Figure 3. Experiment example data illustrate sample with two attributes: IP and Time

When running the MACS2 task, sample attributes will be used to define the multiple pairs (Figure 4). There is an IP-Input pair for each time point, so the Pa ir attribute is the *Time attribute*. The Control attribute is the attribute that differentiates between the *Input* and *IP* groups, and in this example, it is the *ChIII* attribute. Finally, the Control term is labeled as *Input* in the example.

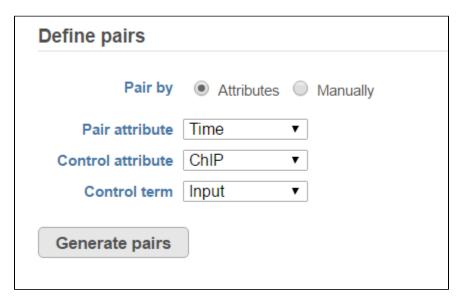


Figure 4. Specify IP vs control pairs based on sample attributes

Click Generate pairs and the two pairs will be automatically added to the Pairs table (Figure 5).

Pairs				
IP	Control	Clear all		
T2_IP	T2_input	×		
T1_IP	T1_input	×		
11_IP	I 1_input	*		

Figure 5. Two IP vs input sample pairs are added in the Pairs table

If multiple pairs are added in the Pairs table, the peak detection is performed on each pair independently.

Peaks report

In the task report, each pair will generate a list of peaks displayed in a table (Figure 6). Use the drop down menu next to **Peaks detected for...** to select the pair.

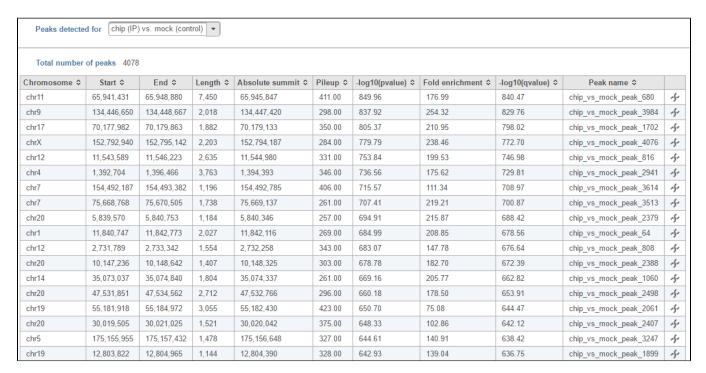


Figure 6. Peaks report on each IP vs control pair

In the report table, each row is a region of a peak and includes the following information:

- Absolute summit: base pair location of peak summit
- Pileup: pileup height at peak summit
- -log10(pvalue): negative log10 pvalue for the peak summit
- Fold enrichment: fold enrichment for the peak summit against random Poisson distribution with local lambda
- -log10(qvalue): negative log10 qvalue at peak summit
- a peak name generated by the MACS2 algorithm

Click the browse to peak button () to invoke chromosome view and zoom into that location.

Click the **Download** button at the lower-right corner to download the peaks in a text file.

References

1. Zhang Y, Liu T, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

Additional Assistance

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

