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 analysis[1] (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.

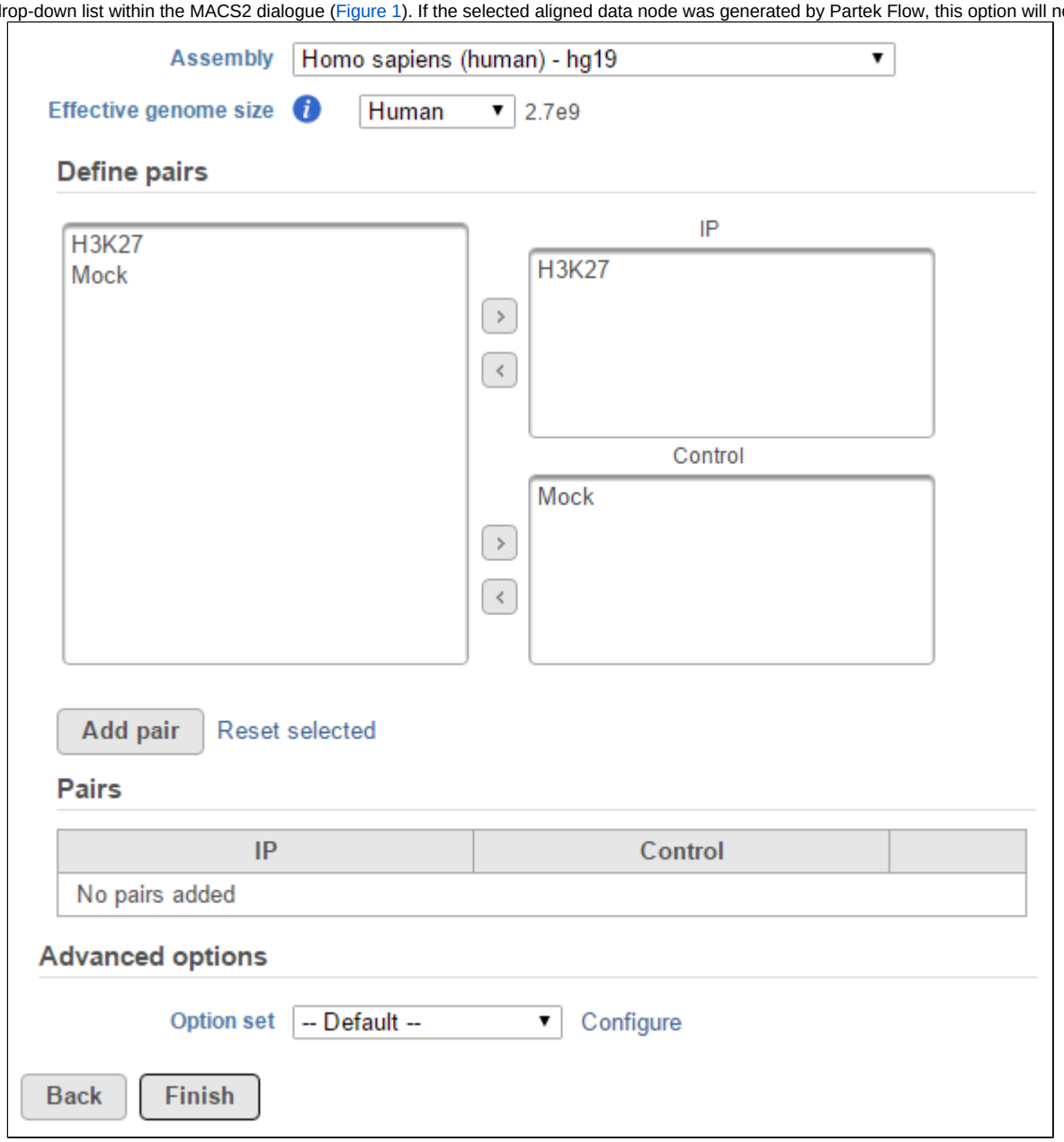

Figure 1. MACS2 dialog: manually add ChIP vs control pairs for peak detection

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

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 authors[1] 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 10^9
- Mouse (*Mus musculus*) – 1.87 x 10^9
- C. elegans – 9 x10^7
- Fruitfly (*Drosophila melongaster*) – 1.2 x 10^8

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.

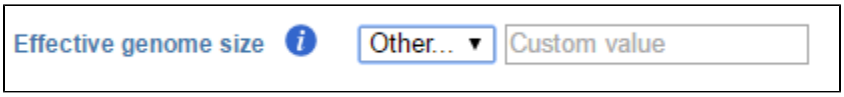

Figure 2. Specify other species effective genome size by manually type in the value

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

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 **Pair 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 *ChIP* attribute. Finally, the **Control term** is labeled as *Input* in the example.

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

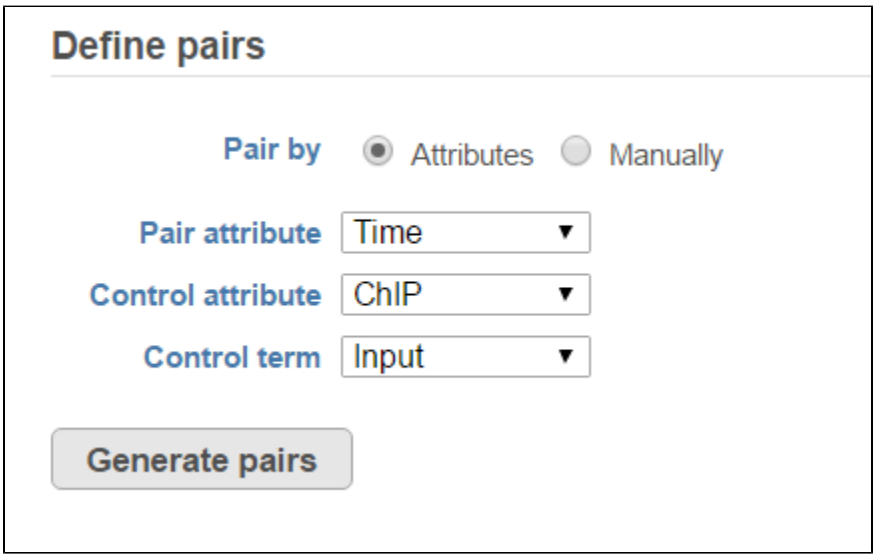

Figure 4. Specify IP vs control pairs based on sample attributes

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

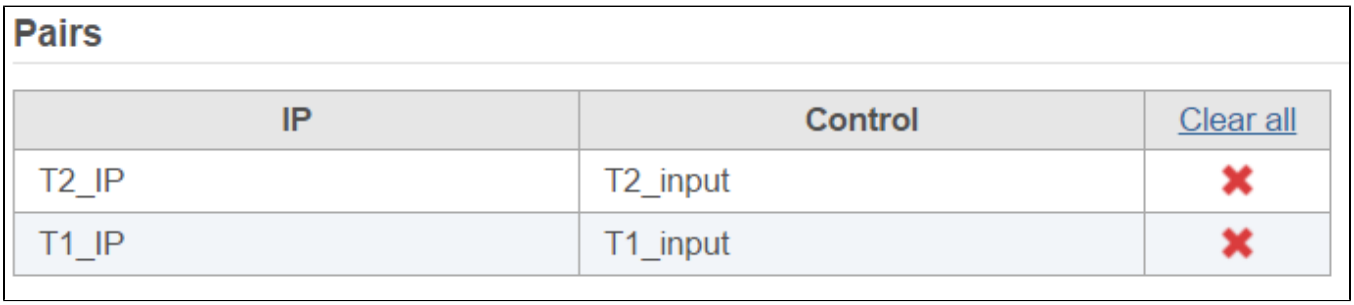

Figure 5. Two IP vs input sample pairs are added in the Pairs table

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.

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

Peaks detected for chip (IP) vs. mock (control)

Total number of peaks 4078

Chromosome	Start	End	Length	Absolute summit	Pileup	-log10(pvalue)	Fold enrichment	-log10(qvalue)	Peak name	
chr11	65,941,431	65,948,880	7,450	65,945,847	411.00	849.96	176.99	840.47	chip_vs_mock_peak_680	
chr9	134,446,650	134,448,667	2,018	134,447,420	298.00	837.92	254.32	829.76	chip_vs_mock_peak_3984	
chr17	70,177,982	70,179,863	1,882	70,179,133	350.00	805.37	210.95	798.02	chip_vs_mock_peak_1702	
chrX	152,792,940	152,795,142	2,203	152,794,187	284.00	779.79	238.46	772.70	chip_vs_mock_peak_4076	
chr12	11,543,589	11,546,223	2,635	11,544,980	331.00	753.84	199.53	746.98	chip_vs_mock_peak_816	
chr4	1,392,704	1,396,466	3,763	1,394,393	346.00	736.56	175.62	729.81	chip_vs_mock_peak_2941	
chr7	154,492,187	154,493,382	1,196	154,492,785	406.00	715.57	111.34	708.97	chip_vs_mock_peak_3614	
chr7	75,668,768	75,670,505	1,738	75,669,137	261.00	707.41	219.21	700.87	chip_vs_mock_peak_3513	
chr20	5,839,570	5,840,753	1,184	5,840,346	257.00	694.91	215.87	688.42	chip_vs_mock_peak_2379	
chr1	11,840,747	11,842,773	2,027	11,842,116	269.00	684.99	208.85	678.56	chip_vs_mock_peak_64	
chr12	2,731,789	2,733,342	1,554	2,732,258	343.00	683.07	147.78	676.64	chip_vs_mock_peak_808	
chr20	10,147,236	10,148,642	1,407	10,148,325	303.00	678.78	182.70	672.39	chip_vs_mock_peak_2388	
chr14	35,073,037	35,074,840	1,804	35,074,337	261.00	669.16	205.77	662.82	chip_vs_mock_peak_1060	
chr20	47,531,851	47,534,562	2,712	47,532,766	296.00	660.18	178.50	653.91	chip_vs_mock_peak_2498	
chr19	55,181,918	55,184,972	3,055	55,182,430	423.00	650.70	75.08	644.47	chip_vs_mock_peak_2061	
chr20	30,019,505	30,021,025	1,521	30,020,042	375.00	648.33	102.86	642.12	chip_vs_mock_peak_2407	
chr5	175,155,955	175,157,432	1,478	175,156,648	327.00	644.61	140.91	638.42	chip_vs_mock_peak_3247	
chr19	12,803,822	12,804,965	1,144	12,804,390	328.00	642.93	139.04	636.75	chip_vs_mock_peak_1899	

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.

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

Your Rating:

Results: 38 rates

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.