Starting with a list of genomic regions

Importing a region list

A region list must contain the chromosome, start location, and stop locations as the first three columns. The chromosome number in the region list must be compatible with the genomic annotation for the species if you plan to use any feature (like motif detection) that requires reference sequence information.

- Import the region list as described above for text files with the following options
  - Select **Other** for data type
  - Set chromosome as a *text* field
  - Set location start and stop as either *integer* or *text* fields
- Right-click on the imported spreadsheet in the spreadsheet tree
- Select **Properties**
- Select **List of genomic regions** from the *Configure Spreadsheet* dialog to add *region* to the properties (Figure 1)

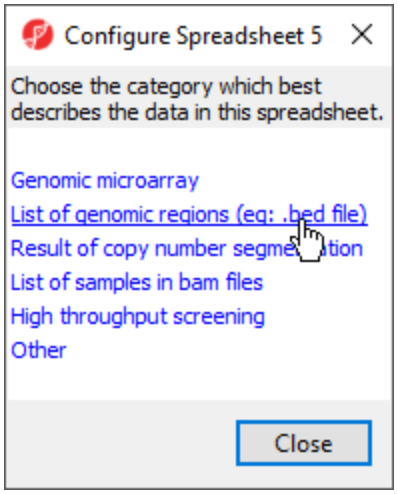


Figure 10. Adding region to the properties of a spreadsheet

The spreadsheet properties will now include *region*. Alternatively, *region* can be added as a spreadsheet property from the *Configure Genomic Properties* dialog by selecting **Advanced..** , choosing **region** from the drop-down menu, selecting **Add**, and selecting **OK**.

If you would like to do any operation that requires looking up the reference genomic sequence information for the regions based on genomic location, you will need to specify the species for this region list.

- Right-click on the imported spreadsheet in the spreadsheet tree
- Select **Properties**
- Select **species** from the *Add Property* drop-down menu and click **Add**
- Specify the *Species Name* and *Genome Build* from the drop-down menus
- Select **OK**

Motif detection

Starting with a region list, you may detect either known or de novo motifs using the ChIP-Seq workflow if your spreadsheet has been associated with a species and a reference genome.

- Select **ChIP-Seq** from the *Workflows* drop-down menu
- Select **Motif detection** from the *Peak Analysis* section of the workflow

Both *Discover de novo motifs* and *Search for known motifs* can be performed. Motif detection sequence information of the genome, you can specify either . 2bit file or .fa file which can be used to create .2bit file

Determining the average values for a region list

If you have a region list or a .BED file and you have a microarray experiment with data, you can summarize the microarray data by the genomic coordinates contained in the region list. For example, the region list contains a list of CpG islands, the experiment contains methylation percentage values for probes (values), and you would like to summarize the methylation values of all probes in each CpG island.

- Import the region list (or .BED file)

Be sure that you have added the *region* property. The list of region coordinates (chromosome, start, stop) from the region list will be mapped against the reference genome specified for the microarray data so specifying *Species and Genome Build* for your region list is unnecessary.

- Open the microarray data spreadsheet, this spreadsheet should have annotation file associated to, and there are genomic location information in the annotation file.

Samples should be on rows and data on columns in the microarray data spreadsheet.

- Select the region list spreadsheet
- Right-click any column header in the region list spreadsheet
- Select **Insert Average** from the pop-up menu (Figure 2)

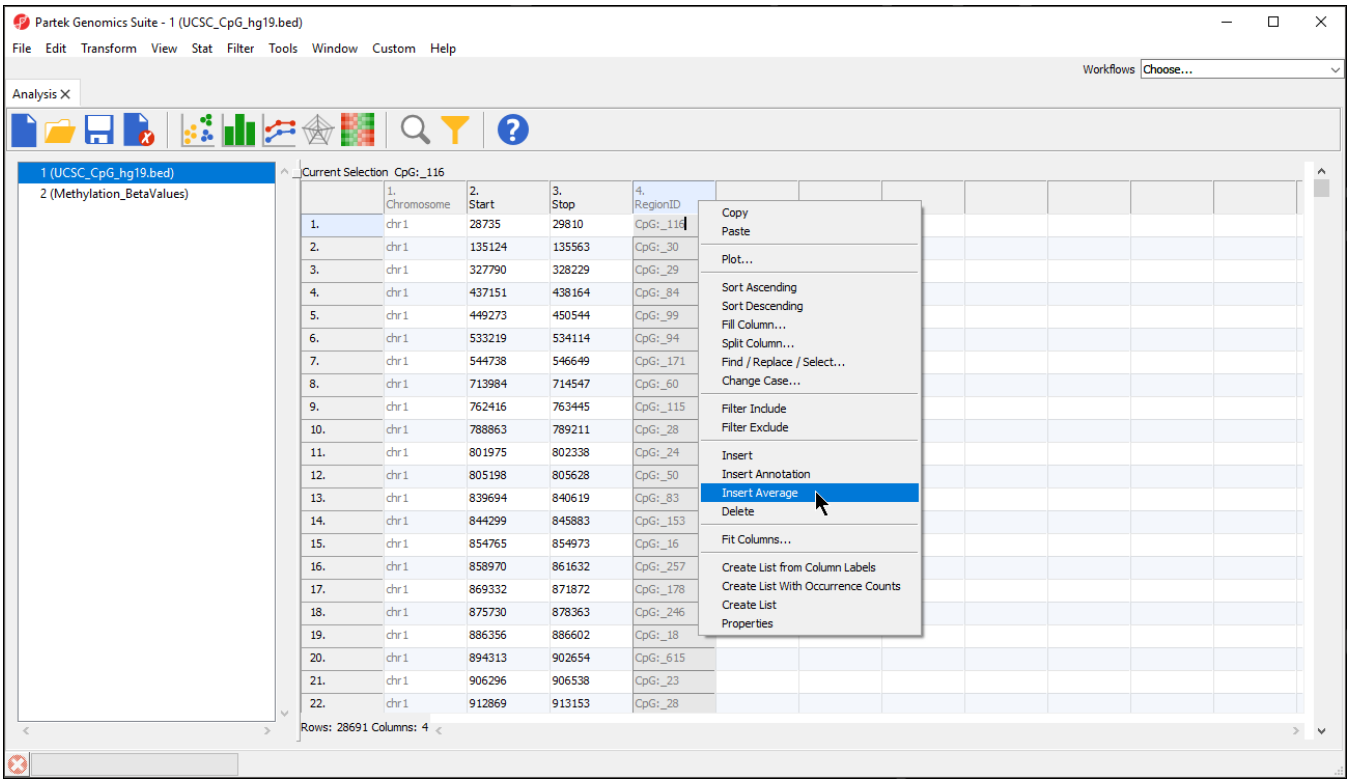


Figure 11. Adding the average values for a region list

- Select the microarray data spreadsheet containing the values you want to average for each region from the *Get average from spreadsheet* drop-down menu

There are three options for averaging the data (Figure 3). *Mean of samples significant in region* is used when the region list has SampleIDs from the microarray data set associated with each region. In this case, only the microarray data set samples specified for each region would be included in the mean calculation. *Mean of all samples* will add columns for the mean value of all probes for all samples and the number of probes for all samples in each region. *Mean value for all samples separately* will add two columns for each sample with the mean value of all probes for that sample and the number of probes for that sample in each region.

- We have selected **Mean value for all samples**
- Select **OK** (Figure 3)

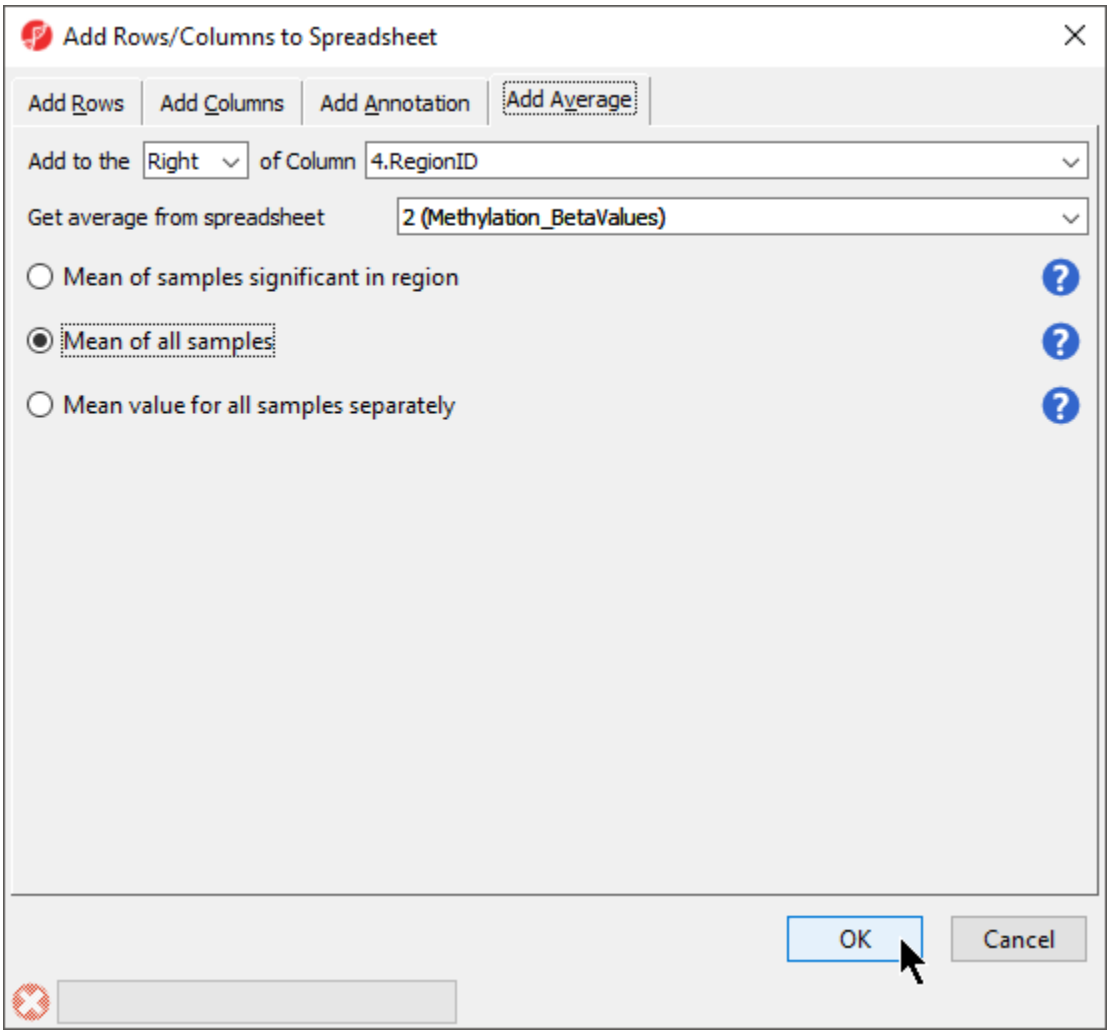


Figure 12. Selecting options for adding average values for regions

Columns will be added to the regions list spreadsheet. Here, we have added two columns with the average -value for all samples in each CpG island and the number of probes in each CpG island (Figure 4).

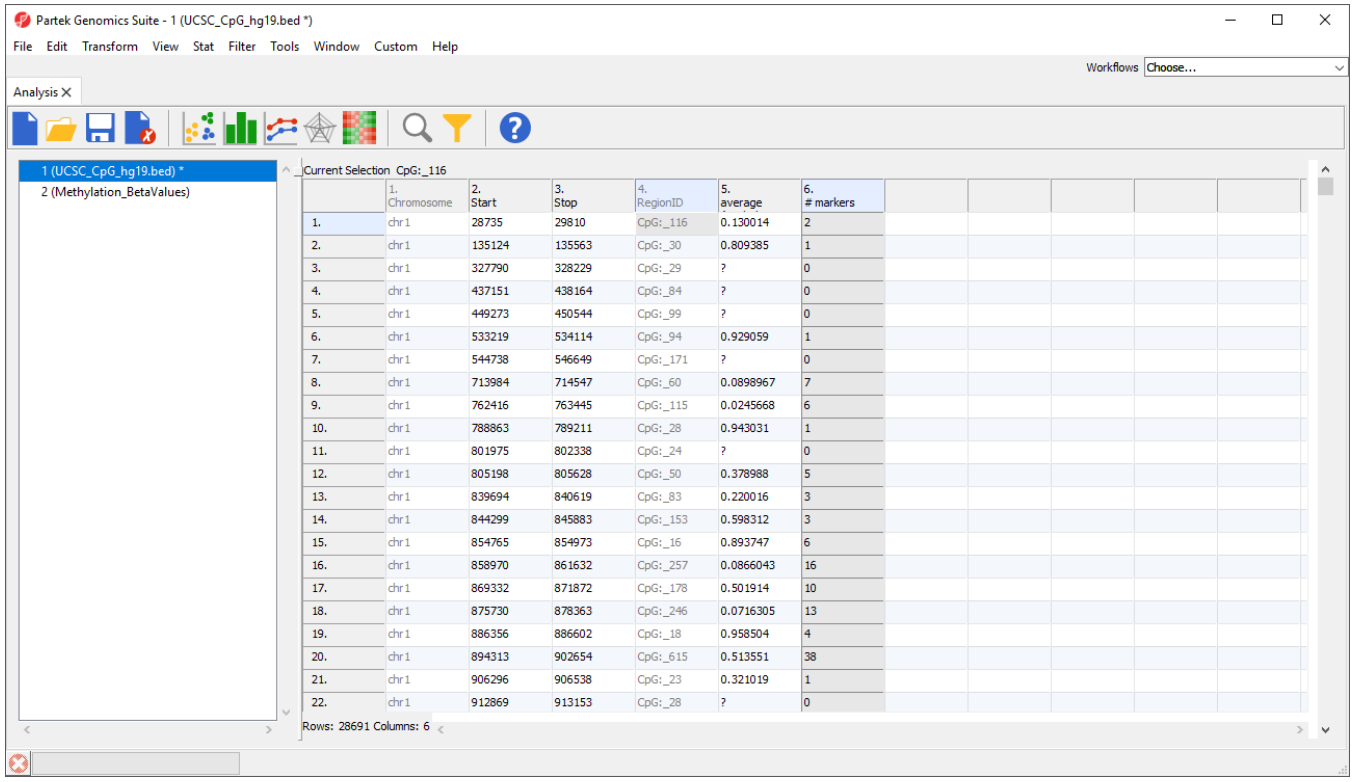


Figure 13. Added average beta values and number of probes per CpG island

Find region overlaps

If you have two or more region lists with coordinates on the same reference genome, you can compare them to identify overlapping regions.

- Open all region list spreadsheets that you want to compare
- Select Tools from the main toolbar
- Select **Find Region Overlaps** (Figure 5)

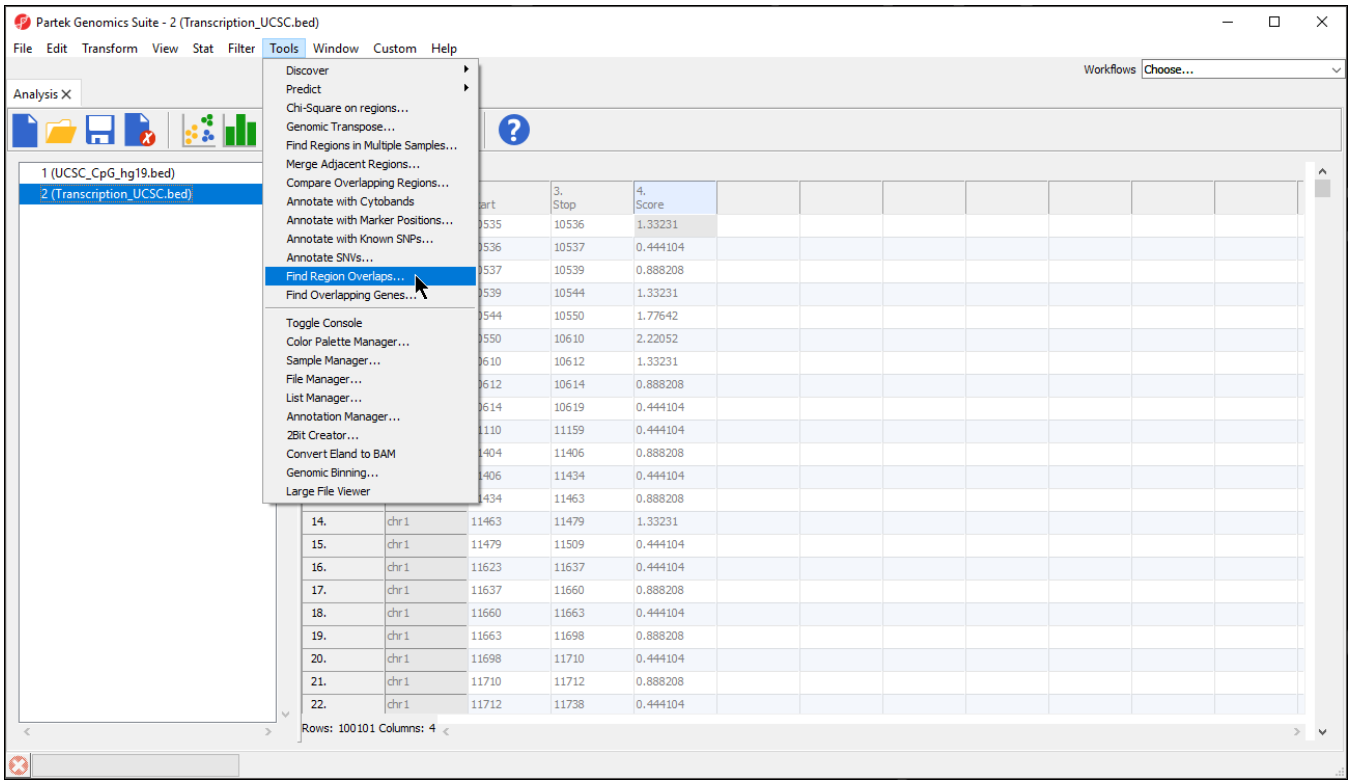


Figure 14. Selecting Find Region Overlaps

The *Find Region Overlaps* tool has two modes of operation. The first, *Report all regions,* creates a new spreadsheet with any regions that did not intersect and all regions of intersection between any of the input lists. For each intersection, the start and stop coordinates of the intersection and the percent overlap between the intersected region with each of the regions in the input lists are reported. The second, *Only report regions present in all lists* creates a new spreadsheet with the intersected regions found in all the lists.

- Select your preferred mode; we have selected **Only report regions present in all lists**
- Select **Add New Spreadsheet** to add any spreadsheets you want to compare; we are comparing two region list spreadsheets (Figure 6)
- Select **OK**

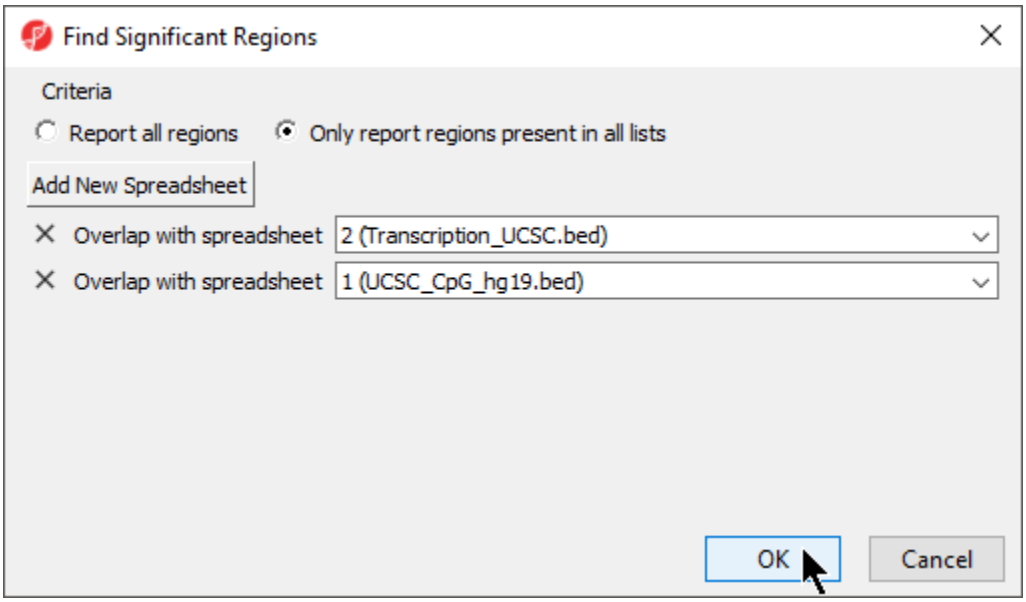
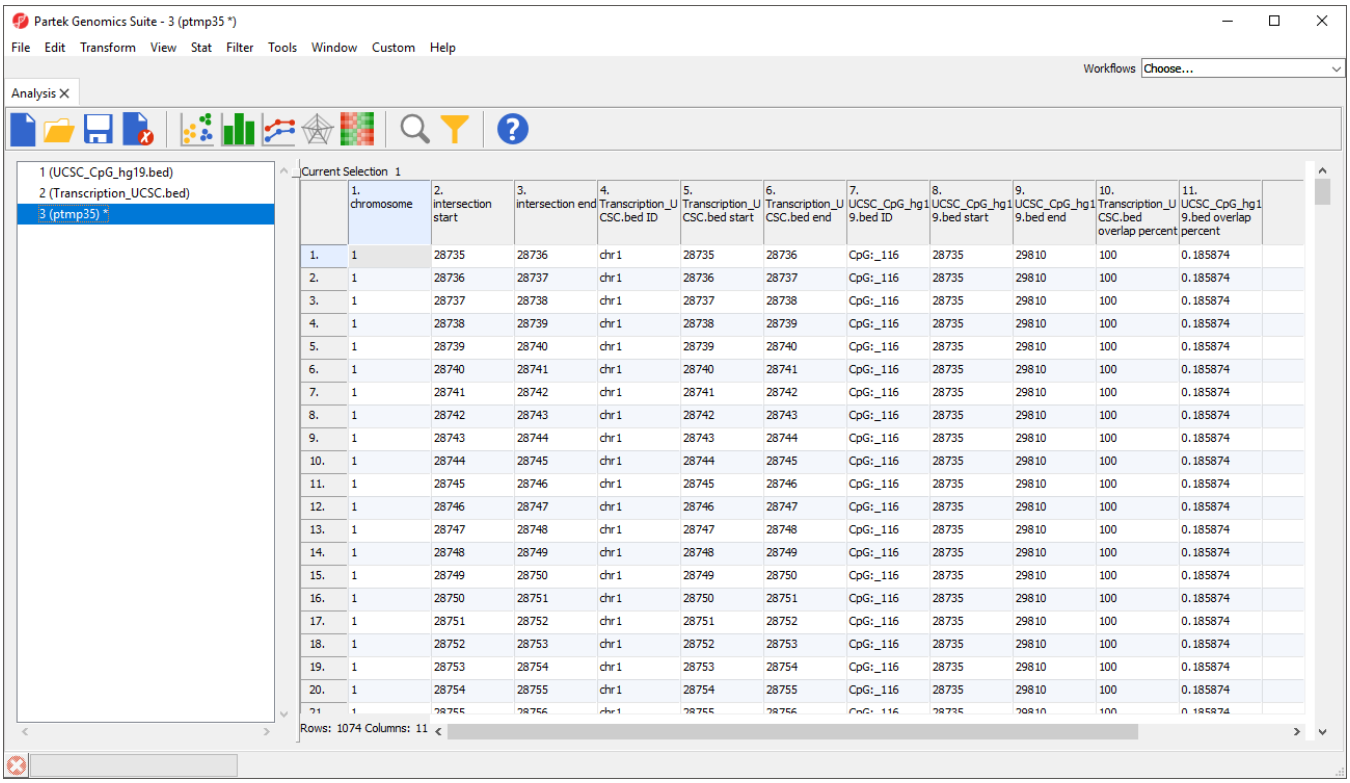


Figure 15. Configuring Find Overlapping Regions

A new region list spreadsheet will be created (Figure 7). The new region list is a temporary spreadsheet so be sure to save it if you want to keep it.



	1. chromosome	2. intersection start	3. intersection end	4. Transcription_UCSC.bed ID	5. Transcription_UCSC.bed start	6. Transcription_UCSC.bed end	7. UCSC_CpG_hg19.bed ID	8. UCSC_CpG_hg19.bed start	9. UCSC_CpG_hg19.bed end	10. Transcription_UCSC.bed overlap percent	11. UCSC_CpG_hg19.bed overlap percent
1.	1	28735	28736	chr1	28735	28736	CpG:_116	28735	29810	100	0.185874
2.	1	28736	28737	chr1	28736	28737	CpG:_116	28735	29810	100	0.185874
3.	1	28737	28738	chr1	28737	28738	CpG:_116	28735	29810	100	0.185874
4.	1	28738	28739	chr1	28738	28739	CpG:_116	28735	29810	100	0.185874
5.	1	28739	28740	chr1	28739	28740	CpG:_116	28735	29810	100	0.185874
6.	1	28740	28741	chr1	28740	28741	CpG:_116	28735	29810	100	0.185874
7.	1	28741	28742	chr1	28741	28742	CpG:_116	28735	29810	100	0.185874
8.	1	28742	28743	chr1	28742	28743	CpG:_116	28735	29810	100	0.185874
9.	1	28743	28744	chr1	28743	28744	CpG:_116	28735	29810	100	0.185874
10.	1	28744	28745	chr1	28744	28745	CpG:_116	28735	29810	100	0.185874
11.	1	28745	28746	chr1	28745	28746	CpG:_116	28735	29810	100	0.185874
12.	1	28746	28747	chr1	28746	28747	CpG:_116	28735	29810	100	0.185874
13.	1	28747	28748	chr1	28747	28748	CpG:_116	28735	29810	100	0.185874
14.	1	28748	28749	chr1	28748	28749	CpG:_116	28735	29810	100	0.185874
15.	1	28749	28750	chr1	28749	28750	CpG:_116	28735	29810	100	0.185874
16.	1	28750	28751	chr1	28750	28751	CpG:_116	28735	29810	100	0.185874
17.	1	28751	28752	chr1	28751	28752	CpG:_116	28735	29810	100	0.185874
18.	1	28752	28753	chr1	28752	28753	CpG:_116	28735	29810	100	0.185874
19.	1	28753	28754	chr1	28753	28754	CpG:_116	28735	29810	100	0.185874
20.	1	28754	28755	chr1	28754	28755	CpG:_116	28735	29810	100	0.185874

Figure 16. Spreadsheet with regions present in all lists

Importing a genomic position list for SNV annotation

To be annotated using the *Annotate SNVs* tool, an imported SNV position list must have four columns per locus:

1. Position of the SNP listed as chr.basePosition
2. Sample ID or name
3. The reference base
4. The SNP call (sample genotype base)

- Prepare input list as shown (Figure 8) with four columns describing the position, sample, reference base, and sample genotype base for each SNV

position	sample	reference	genotype
chr1.5609011	sample_1	T	C
chr1.569013	sample_2	C	A
chr1.569902	sample_3	C	T
chr1.9795223	sample_2	C	A
chr1.16154783	sample_1	C	G

Figure 17. An imported SNV list must follow this format to be annotated by the Annotate SNV tool. The first column must be the position and the position must follow the format shown, chr.basePosition

- Save as either a tab-separated or comma separated file
- Import the table as a text file
  - Select **Genomic data** for *What type of data is this file?*
  - Set the position column *Type* to **text**
  - Set the other columns *Type* to **categorical**
- Select **Genomic location instead of marker IDs** from the *Choose the type of genomic data* drop-down menu of the *Configure Genomic Properties* dialog
- Specify the *Species* and *Genome Build*
- Select **OK**

The *Annotate SNVs* tool can now be invoked on this spreadsheet to generate an annotation spreadsheet (Figure 9).

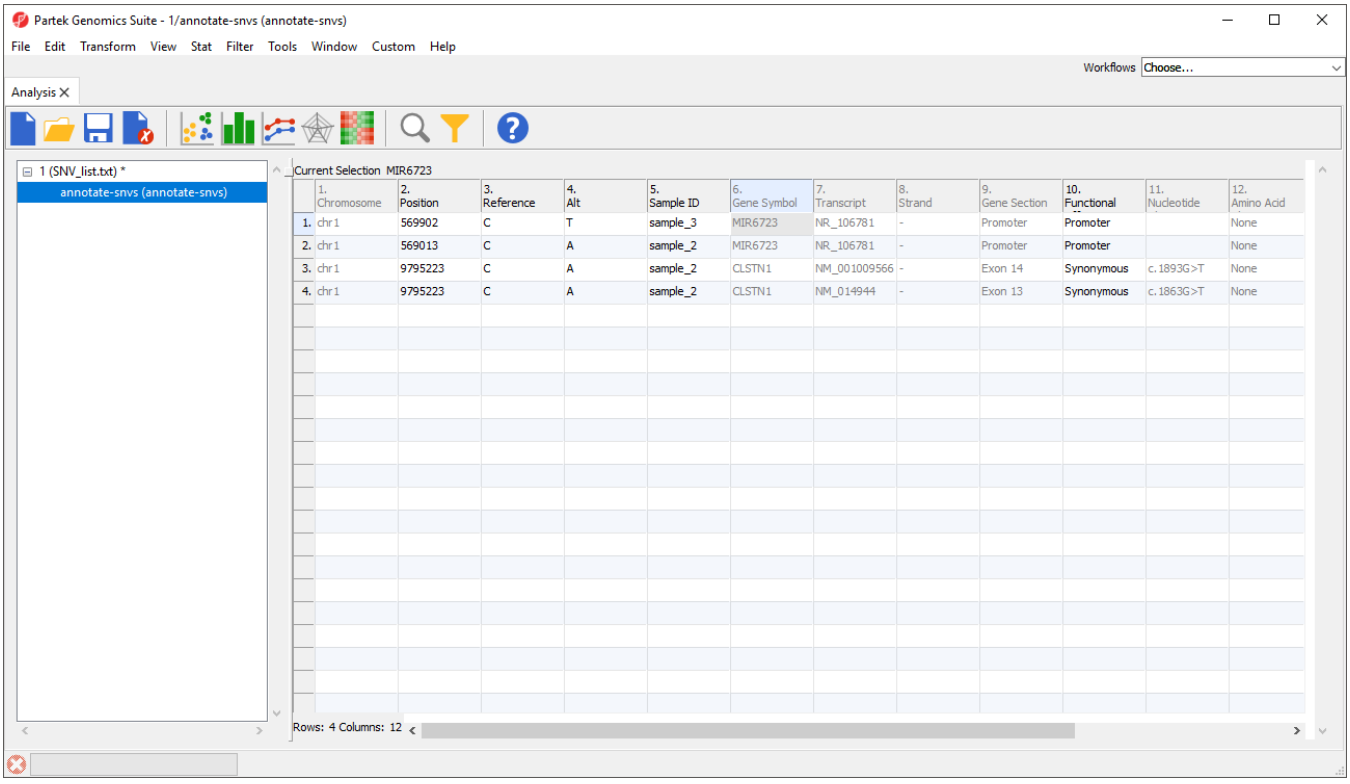


Figure 18. Annotate SNVs creates a new spreadsheet annotating each SNV from the source list

« Tasks available for a gene list Starting with a list of SNPs »

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: 35 rates