Analysis of a Tiling Regulation Study in Partek® Genomics Suite® 6.6

The example data set used in this tutorial consists of 6 technical replicates from the same human cell line, 3 are SP1 treated, and 3 are non-specific antibody (Ab) negative controls. This experiment was performed using Affymetrix GeneChip® Human Promoter Tiling Arrays.

This tutorial will illustrate how to:

- Import and normalize data from Human Promoter Tiling CEL files
- Normalize the treated samples to the control samples
- Statistically and visually identify regions of binding on the genome
- Generate list of genes in the binding regions

Note: It is recommended that you go through **Chapter 6 The Pattern Visualization System®** in the *Partek On-line Help* before going through this tutorial if you have not done so. In addition, this tutorial covers specific information; for general information covering a variety of subjects, see the *Partek On-line Help*.

Importing Affymetrix® CEL Files

The data files for this experiment can be downloaded by selecting **Help > On-line Tutorials** from the Partek main menu. For this example, put the unzipped *Data* files (.CEL and other files) that are underneath the Tiling heading in the folder *C:\Partek Training Data\Tiling*.

- Select **Tiling** from the *Workflows* drop down (Figure 1)

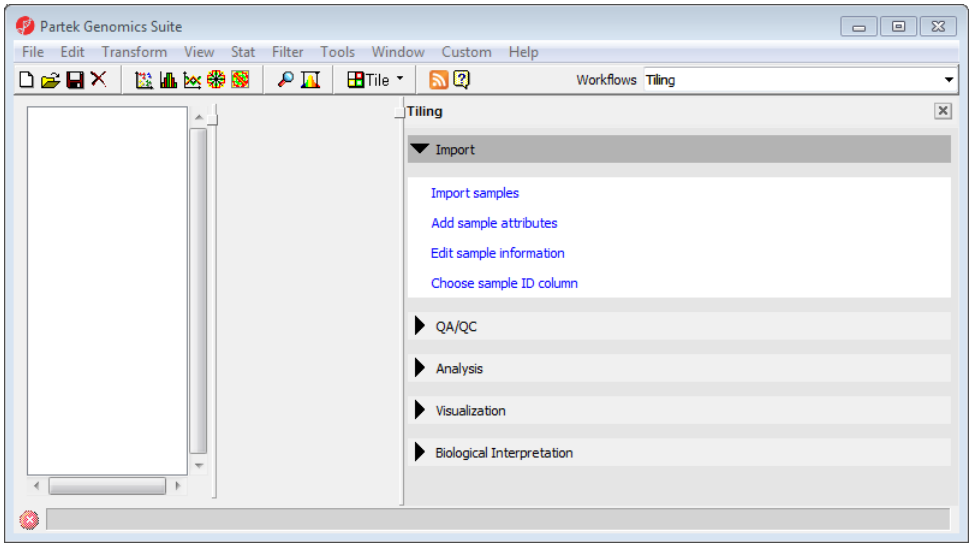


Figure 1: Choosing the pre-defined tiling workflow

- Select **Import Samples** (Figure 1)
- The *Load Tiling Data* dialog will appear, select **Import from Affymetrix CEL Files,** and select **OK**
- Select the **Browse** button to select the *C:\Partek Training Data\Tiling* folder. By default, all the files with a .cel extension are selected (Figure 2)

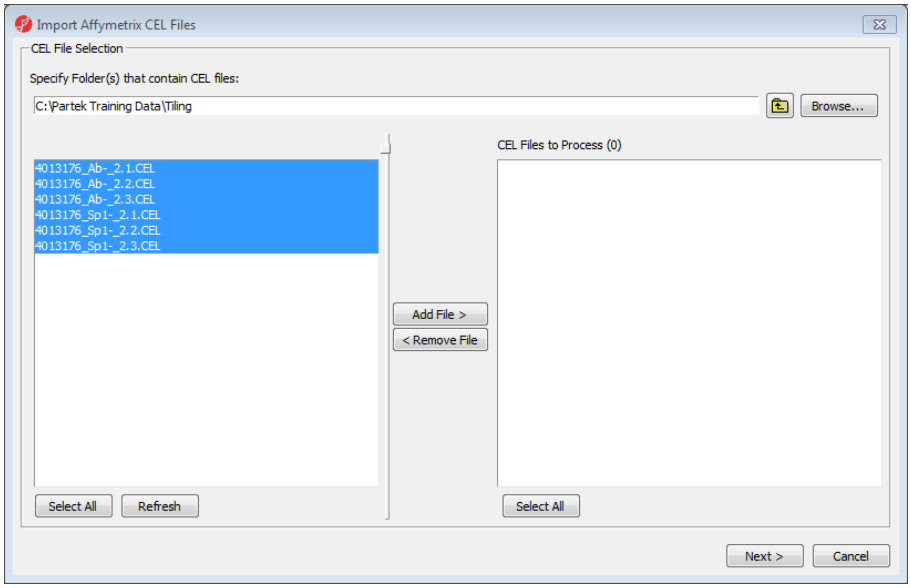


Figure 2: Selecting the CEL files

- Click the **Add File** > button to move all the .cel files to the right panel. 6 files will be processed
- Click **Next >**, the dialog in Figure 3 will appear

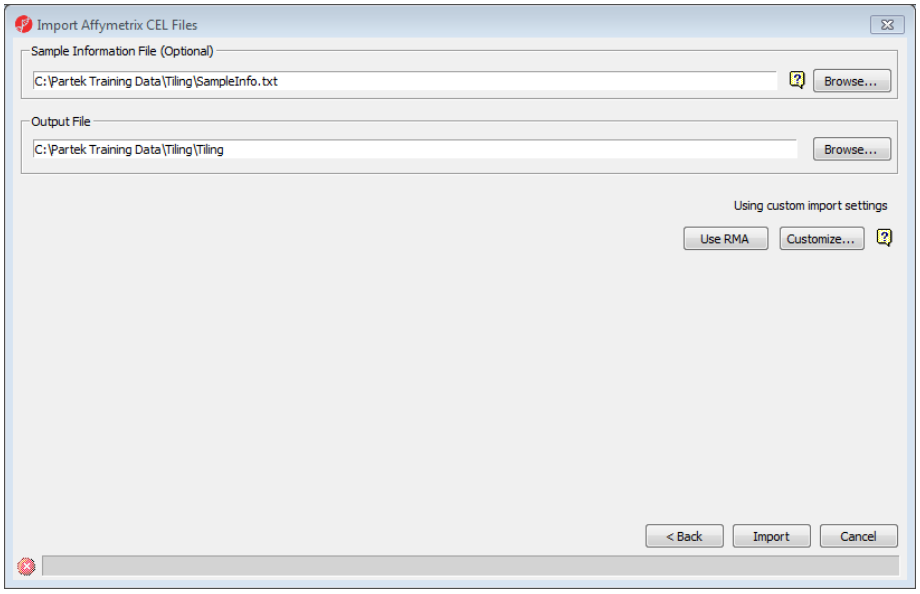


Figure 3: Specifying the sample information file and output file

- In Figure 3, *C:\Partek Training Data\Tiling\SampleInfo.txt.fmt* should be automatically selected as the default *Sample Information File*. If it is not, click the first **Browse** button and choose the **SampleInfo.txt.fmt** file that comes with the tutorial data
- Name the *Output File* as **Tiling**. It is stored at the same location as the CEL files

The default method of normalization is predefined by Partek. It includes **Adjust for probe sequence, RMA Background Correction, Quantile Normalization, and Log (base 2) transformation**. You can view and change the algorithm settings by selecting the **Customize…** button.

- Select **Import** (Figure 3)

You might be asked to **Specify the Library File Root Folder** (Figure 4). You can select to use your current Affymetrix library folder, if you have already installed software from Affymetrix, such as Expression Console™, Genotyping Console™ etc. Here *C:\Microarray Library* is used as the default library folder.

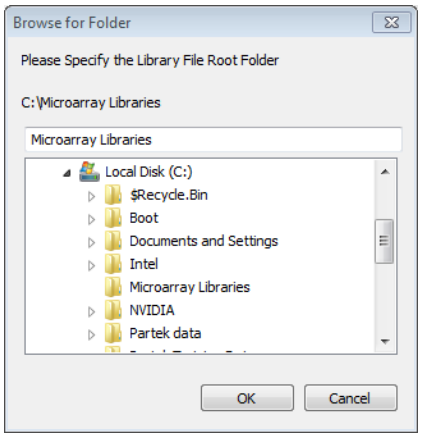


Figure 4: Specifying the default library root folder

Partek will download the needed library files if they could not be found in the default library folder and sub-folders (Figure 5). The file downloading is automated, which does not need the user's involvement unless you have firewall issue.

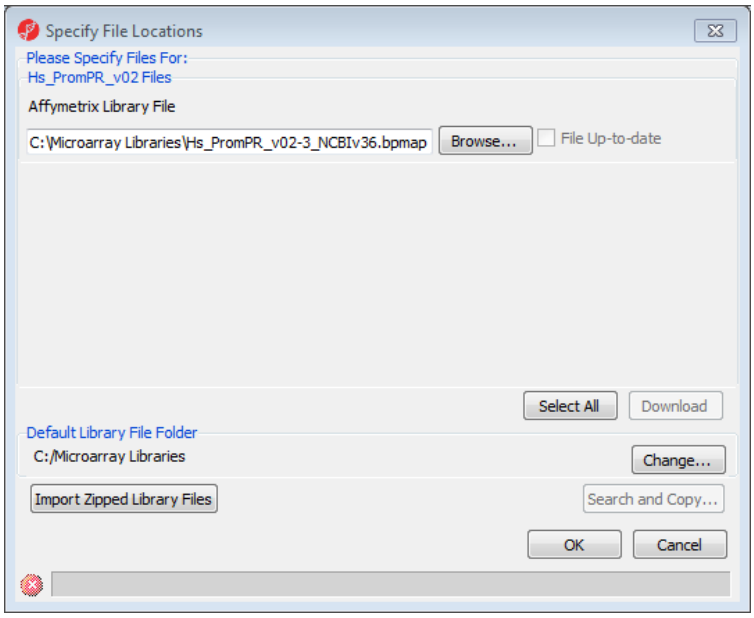


Figure 5: Configuring the Automated Library File Downloading

After all needed files are downloaded, the import will continue. Figure 6 shows the data have been imported into the Analytical Spreadsheet®. Six rows represent the 6 samples, ~4 million probes are on the columns, and the values of the probes are intensities in log base 2. Double clicking on an image allows for quality control inspection.

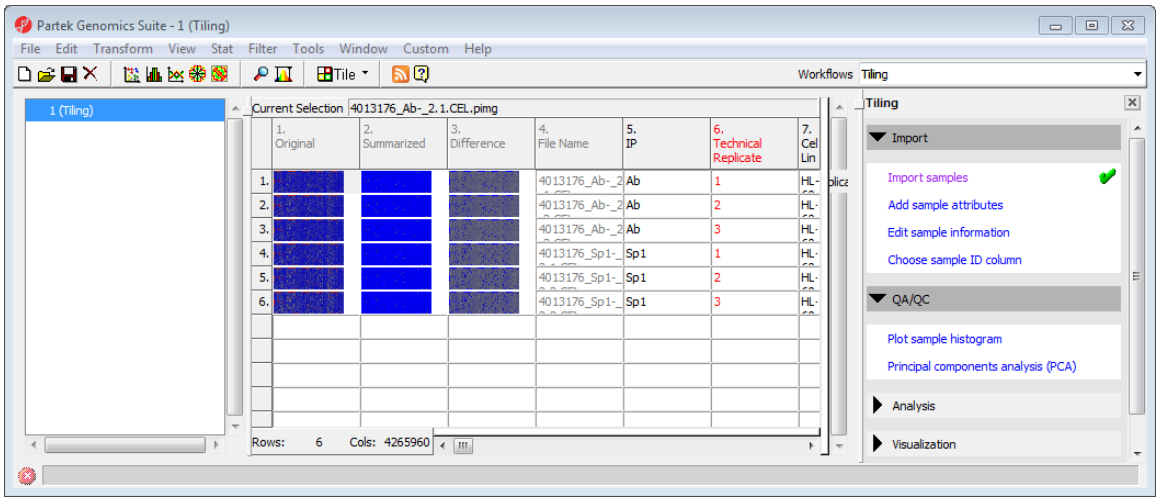


Figure 6: Viewing the spreadsheet

During import, we had associated sample information with the genomic data, we don't need to add more sample attributes, so I will skip the Add Sample Attributes step.

Note: The library file of this chip is Hs_PromPR_v02-3_NCBIv36.bpmap when this tutorial is written, we need to make sure the genome information is correctly specified.

*Choose **File>Properties**, make sure the species is Homo sapiens, click on **Edit Genome** to choose Genome version hg18*

In the future, if there is new library file distributed, you need to change the genome version correspondingly.

Quality Assurance / Quality Control (QA/QC)

- Select **Principal Components Analysis (PCA)** in the *QA/QC* section of the *Workflows* panel to invoke a PCA plot on the 4 million probes. A PCA plot will appear (Figure 7)

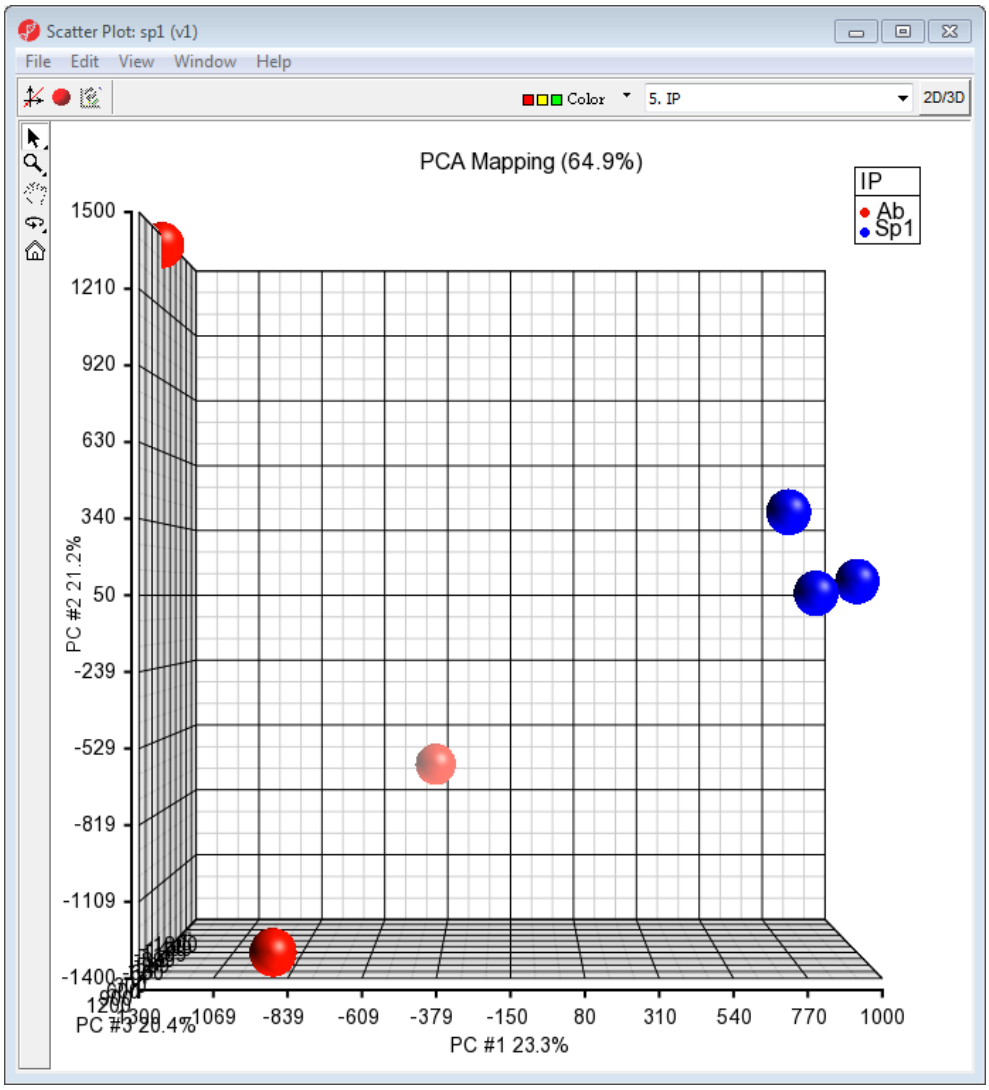


Figure 7: Viewing the PCA plot

This plot shows that the treated samples cluster more tightly than the Antibody minus samples. You can hold your middle mouse button and drag to rotate the plot.

- Close the PCA plot

You can also click on the **Sample Histogram** to examine the data using histogram; this part is for you to practice on your own.

Detecting Regions of Binding using the Pre-defined Workflow

When you have paired samples that have a negative control coupled with the experimental antibody pull down, you need to perform two-way ANOVA (which is equivalent to paired sample t-Test) to find the difference between treatment vs. control on each probe, and then we will use MAT algorithm to detect regions showing enrichment in treatment.

- From the *Analysis* panel, select **ANOVA.** The message in Figure 8 will appear. You will need to specify a contrast of the two groups when you do ANOVA, so choose to include T statistics of the contrast in the result output. This is necessary for the next step in the analysis

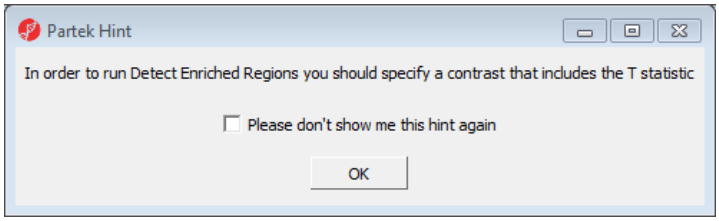


Figure 8: Reviewing the Partek hint to specify T statistics in the output

- After selecting **OK** to dismiss the reminder, the ANOVA dialog will be invoked. For this data set, select both factors located in the *Experimental Factors* panel, and move them to the *ANOVA Factor(s)* panel by selecting the **Add Factor >** button (Figure 9)

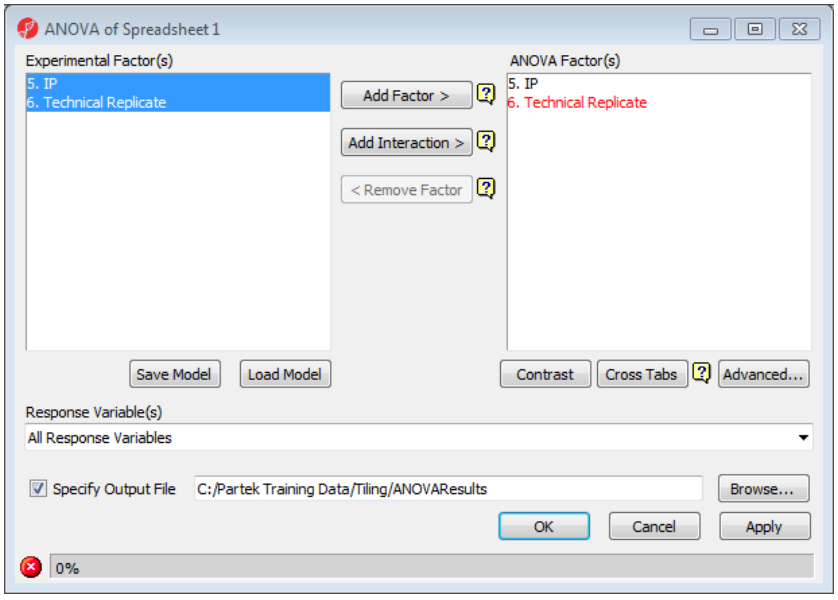


Figure 9: Configuring the ANOVA dialog

- Select the **Contrast** button on the *ANOVA* dialog to setup a contrast between *Sp1* and *Ab*
- Add **Sp1** to the *group1* panel to the right, and **Ab** to the *group2* panel to the right (Figure 10)

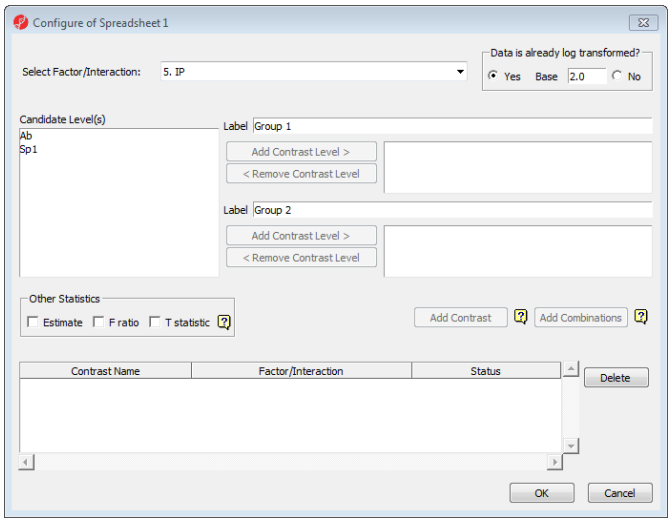


Figure 10: Setting up contrasts between Sp1 and Ab

- Select the **T statistic** check box
- Select the **Add Contrast** button to add it to the *Contrasts* panel at the bottom of the dialog (Figure 11). Note that sequence of steps is important.

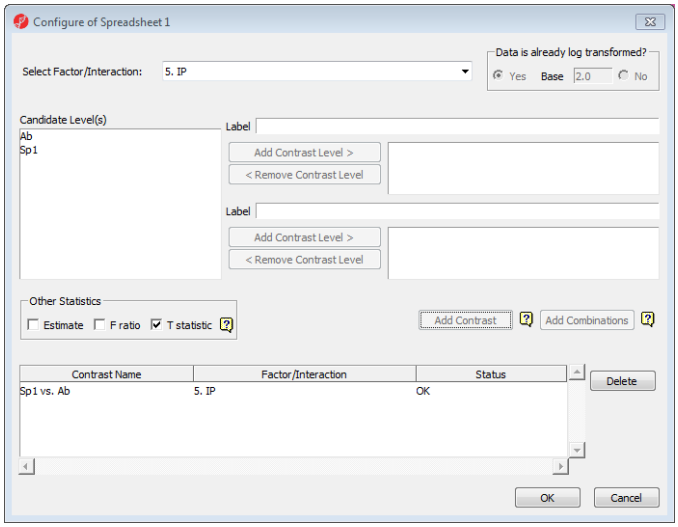
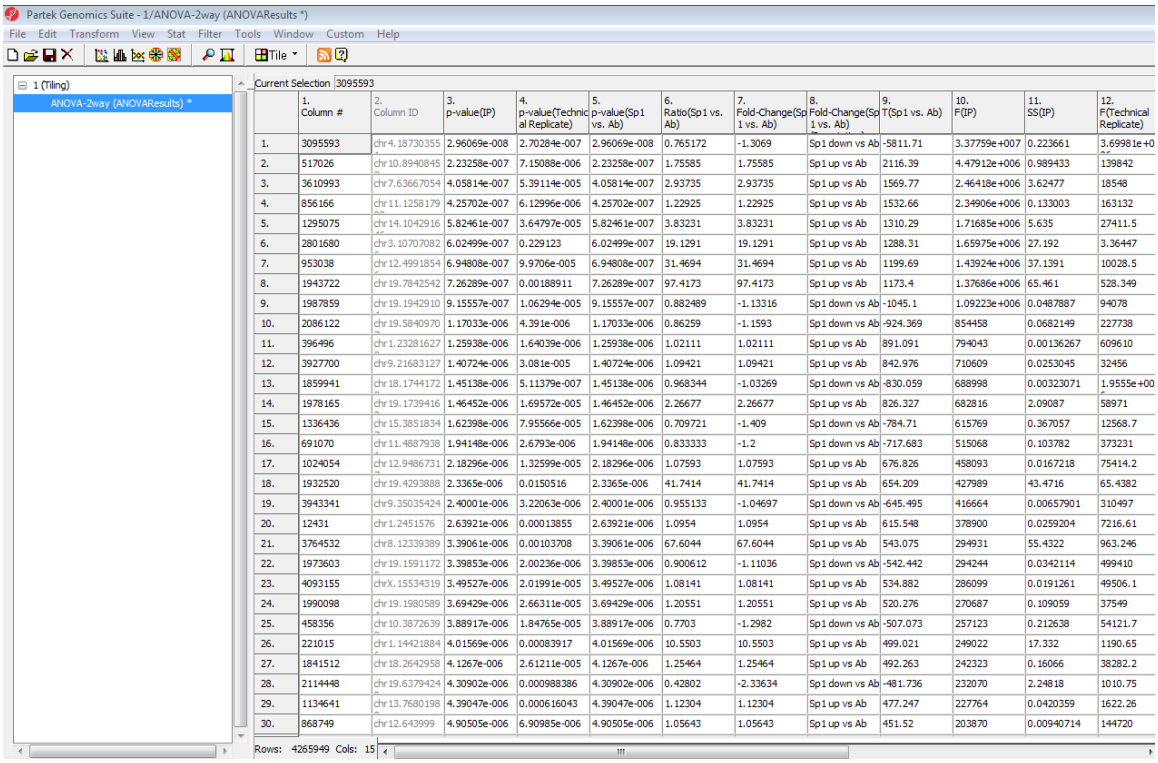


Figure 11: Adding the contrast to the Contrast panel

- Select **OK**
- Select **OK** again to run the ANOVA

In the resulting spreadsheet, each row is a probe, and the columns contain the ANOVA results, like p-value, fold change, T statistic, etc. (Figure 12). When the T statistic value is positive, the SP1 group has higher average intensity than the Ab group. When the T statistic is negative, the SP1 group has lower average intensity than that of the Ab group.



	1. Column #	2. Column ID	3. p-value(IP)	4. p-value(Technical Replicate)	5. p-value(Sp1 vs. Ab)	6. Ratio(Sp1 vs. Ab)	7. Fold-Change(Sp1 vs. Ab)	8. Fold-Change(Sp1 vs. Ab)	9. T(Sp1 vs. Ab)	10. F(IP)	11. SS(IP)	12. F(Technical Replicate)
1.	3095593	chr4.18730355	2.96069e-008	2.70284e-007	2.96069e-008	0.765172	-1.3069	Sp1 down vs Ab	-5811.71	3.37759e+007	0.223661	3.69981e+0
2.	517026	chr10.8940845	2.23258e-007	7.15088e-006	2.23258e-007	1.75585	1.75585	Sp1 up vs Ab	2116.39	4.47912e+006	0.989433	139842
3.	3610993	chr7.63667054	4.05814e-007	5.39114e-005	4.05814e-007	2.93735	2.93735	Sp1 up vs Ab	1569.77	2.46418e+006	3.62477	18548
4.	856166	chr11.1258179	4.25702e-007	6.12996e-006	4.25702e-007	1.22925	1.22925	Sp1 up vs Ab	1532.66	2.34906e+006	0.133003	163132
5.	1295075	chr14.1042916	5.82461e-007	3.64797e-005	5.82461e-007	3.83231	3.83231	Sp1 up vs Ab	1310.29	1.71685e+006	5.635	27411.5
6.	2801680	chr3.10707082	6.02499e-007	0.229123	6.02499e-007	19.1291	19.1291	Sp1 up vs Ab	1288.31	1.65975e+006	27.192	3.36447
7.	953038	chr12.4991854	6.94808e-007	9.9706e-005	6.94808e-007	31.4694	31.4694	Sp1 up vs Ab	1199.69	1.43924e+006	37.1391	10028.5
8.	1943722	chr19.7842542	7.26289e-007	0.00188911	7.26289e-007	97.4173	97.4173	Sp1 up vs Ab	1173.4	1.37686e+006	65.461	528.349
9.	1987859	chr19.1942910	9.15557e-007	1.06294e-005	9.15557e-007	0.882489	-1.13316	Sp1 down vs Ab	-1045.1	1.09223e+006	0.0487887	94078
10.	2086122	chr19.5840970	1.17033e-006	4.391e-006	1.17033e-006	0.86259	-1.1593	Sp1 down vs Ab	-924.369	854458	0.0682149	227738
11.	396496	chr1.23281627	1.25938e-006	1.64039e-006	1.25938e-006	1.02111	1.02111	Sp1 up vs Ab	891.091	794043	0.00136267	609610
12.	3927700	chr9.21683127	1.40724e-006	3.081e-005	1.40724e-006	1.09421	1.09421	Sp1 up vs Ab	842.976	710609	0.0253045	32456
13.	1859941	chr18.1744172	1.45138e-006	5.11379e-007	1.45138e-006	0.968344	-1.03269	Sp1 down vs Ab	-830.059	688998	0.00323071	1.9555e+00
14.	1978165	chr19.1739416	1.46452e-006	1.69572e-005	1.46452e-006	2.26677	2.26677	Sp1 up vs Ab	826.327	682816	2.09087	58971
15.	1336436	chr15.3851834	1.62398e-006	7.95566e-005	1.62398e-006	0.709721	-1.409	Sp1 down vs Ab	-784.71	615769	0.367057	12568.7
16.	691070	chr11.4887938	1.94148e-006	2.6793e-006	1.94148e-006	0.833333	-1.2	Sp1 down vs Ab	-717.683	515068	0.103782	373231
17.	1024054	chr12.9486731	2.18296e-006	1.32599e-005	2.18296e-006	1.07593	1.07593	Sp1 up vs Ab	676.826	458093	0.0167218	75414.2
18.	1932520	chr19.4293888	2.3365e-006	0.0150516	2.3365e-006	41.7414	41.7414	Sp1 up vs Ab	654.209	427989	43.4716	65.4382
19.	3943341	chr9.35035424	2.40001e-006	3.22063e-006	2.40001e-006	0.955133	-1.04697	Sp1 down vs Ab	-645.495	416664	0.00657901	310497
20.	12431	chr1.2451576	2.63921e-006	0.00013855	2.63921e-006	1.0954	1.0954	Sp1 up vs Ab	615.548	378900	0.0259204	7216.61
21.	3764532	chr8.12339389	3.39061e-006	0.00103708	3.39061e-006	67.6044	67.6044	Sp1 up vs Ab	543.075	294931	55.4322	963.246
22.	1973603	chr19.1591172	3.39853e-006	2.00236e-006	3.39853e-006	0.900612	-1.11036	Sp1 down vs Ab	-542.442	294244	0.0342114	499410
23.	4093155	chrX.15534319	3.49527e-006	2.01991e-005	3.49527e-006	1.08141	1.08141	Sp1 up vs Ab	534.882	286099	0.0191261	49506.1
24.	1990098	chr19.1980589	3.69429e-006	2.66311e-005	3.69429e-006	1.20551	1.20551	Sp1 up vs Ab	520.276	270687	0.109059	37549
25.	458356	chr10.3872639	3.88917e-006	1.84765e-005	3.88917e-006	0.7703	-1.2982	Sp1 down vs Ab	-507.073	257123	0.212638	54121.7
26.	221015	chr1.14421884	4.01569e-006	0.00083917	4.01569e-006	10.5503	10.5503	Sp1 up vs Ab	499.021	249022	17.332	1190.65
27.	1841512	chr18.2642958	4.1267e-006	2.61211e-005	4.1267e-006	1.25464	1.25464	Sp1 up vs Ab	492.263	242323	0.16066	38282.2
28.	2114448	chr19.6379424	4.30902e-006	0.000988386	4.30902e-006	0.42802	-2.33634	Sp1 down vs Ab	-481.736	232070	2.24818	1010.75
29.	1134641	chr13.7680198	4.39047e-006	0.000616043	4.39047e-006	1.12304	1.12304	Sp1 up vs Ab	477.247	227764	0.0420359	1622.26
30.	868749	chr12.643999	4.90505e-006	6.90985e-006	4.90505e-006	1.05643	1.05643	Sp1 up vs Ab	451.52	203870	0.00940714	144720

Figure 12: Viewing the ANOVA result spreadsheet

Next, the MAT algorithm will be used to detect enriched regions of the treatment.

- Select **Detect Enriched Regions** from the *Workflows* panel. Use the default settings and make sure the *Column to Test* contains the t-statistic from the ANOVA contrast result (Figure 13)

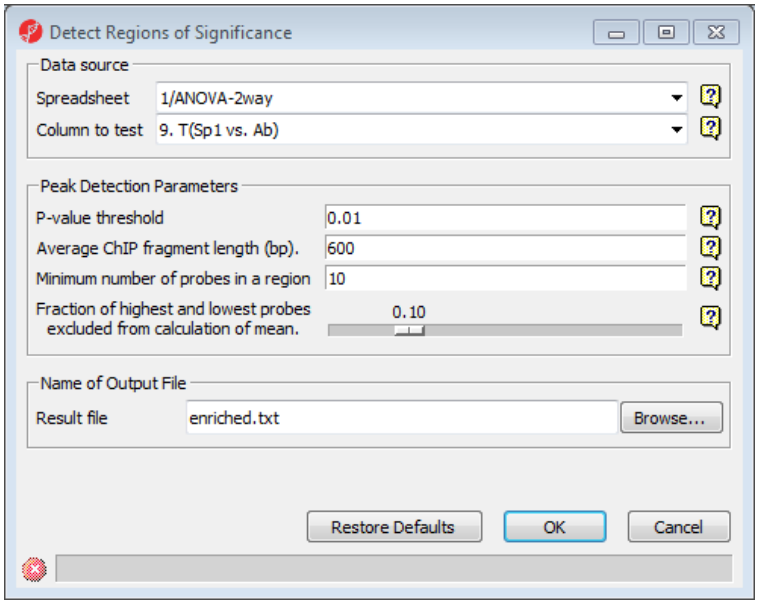
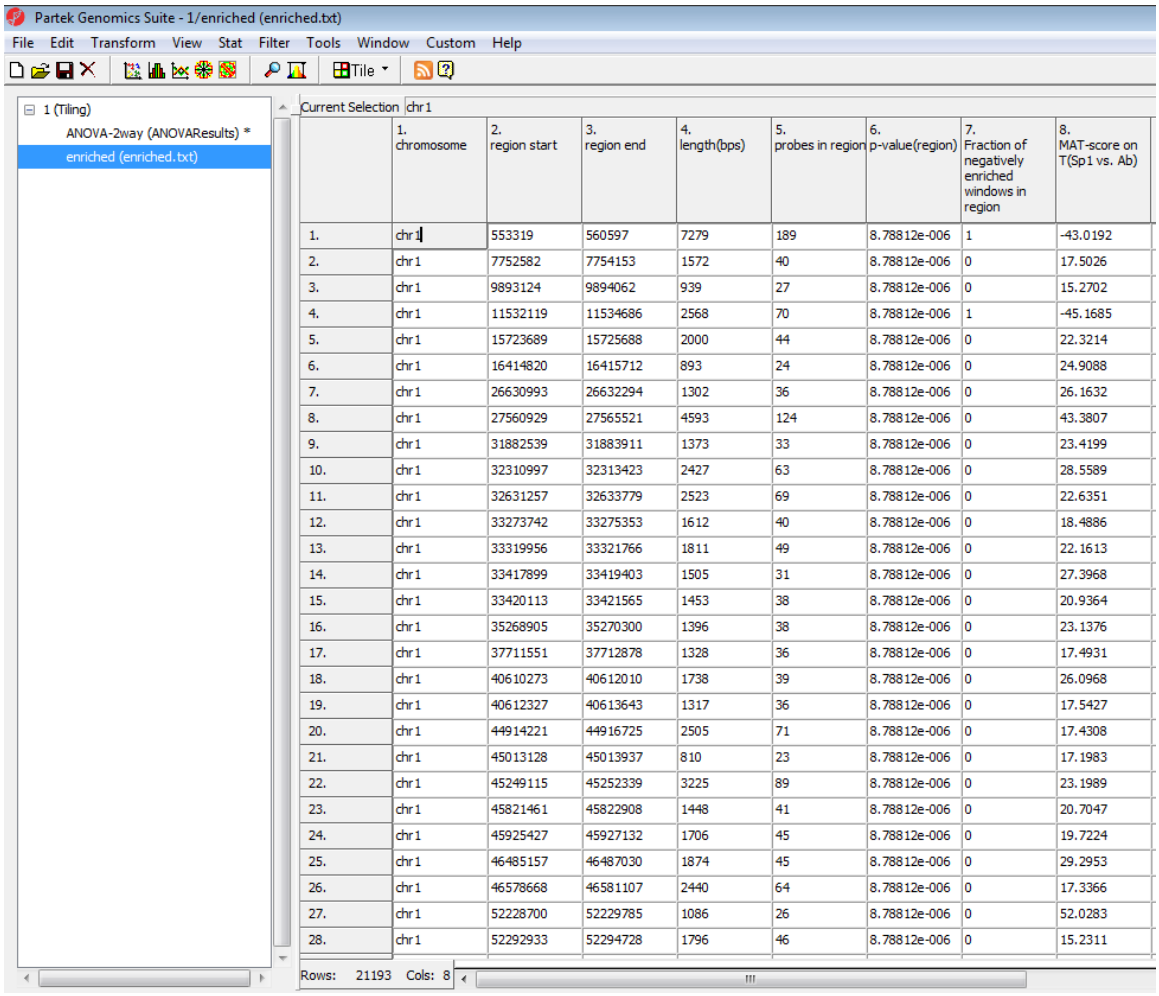


Figure 13: Configuring the Detect Regions of Significance dialog

- The result will be displayed in a child spreadsheet (Figure 14)



	1. chromosome	2. region start	3. region end	4. length(bps)	5. probes in region	6. p-value(region)	7. Fraction of negatively enriched windows in region	8. MAT-score on T(Sp1 vs. Ab)
1.	chr1	553319	560597	7279	189	8.78812e-006	1	-43.0192
2.	chr1	7752582	7754153	1572	40	8.78812e-006	0	17.5026
3.	chr1	9893124	9894062	939	27	8.78812e-006	0	15.2702
4.	chr1	11532119	11534686	2568	70	8.78812e-006	1	-45.1685
5.	chr1	15723689	15725688	2000	44	8.78812e-006	0	22.3214
6.	chr1	16414820	16415712	893	24	8.78812e-006	0	24.9088
7.	chr1	26630993	26632294	1302	36	8.78812e-006	0	26.1632
8.	chr1	27560929	27565521	4593	124	8.78812e-006	0	43.3807
9.	chr1	31882539	31883911	1373	33	8.78812e-006	0	23.4199
10.	chr1	32310997	32313423	2427	63	8.78812e-006	0	28.5589
11.	chr1	32631257	32633779	2523	69	8.78812e-006	0	22.6351
12.	chr1	33273742	33275353	1612	40	8.78812e-006	0	18.4886
13.	chr1	33319956	33321766	1811	49	8.78812e-006	0	22.1613
14.	chr1	33417899	33419403	1505	31	8.78812e-006	0	27.3968
15.	chr1	33420113	33421565	1453	38	8.78812e-006	0	20.9364
16.	chr1	35268905	35270300	1396	38	8.78812e-006	0	23.1376
17.	chr1	37711551	37712878	1328	36	8.78812e-006	0	17.4931
18.	chr1	40610273	40612010	1738	39	8.78812e-006	0	26.0968
19.	chr1	40612327	40613643	1317	36	8.78812e-006	0	17.5427
20.	chr1	44914221	44916725	2505	71	8.78812e-006	0	17.4308
21.	chr1	45013128	45013937	810	23	8.78812e-006	0	17.1983
22.	chr1	45249115	45252339	3225	89	8.78812e-006	0	23.1989
23.	chr1	45821461	45822908	1448	41	8.78812e-006	0	20.7047
24.	chr1	45925427	45927132	1706	45	8.78812e-006	0	19.7224
25.	chr1	46485157	46487030	1874	45	8.78812e-006	0	29.2953
26.	chr1	46578668	46581107	2440	64	8.78812e-006	0	17.3366
27.	chr1	52228700	52229785	1086	26	8.78812e-006	0	52.0283
28.	chr1	52292933	52294728	1796	46	8.78812e-006	0	15.2311

Figure 14: Viewing the result spreadsheet

In the result spreadsheet, each row represents a region of significant enrichment. The columns have information about the regions, such as genomic location, number of probes in the region, and statistics results.
The p-value of the region is the empirical p-value of the most significant MAT score included within this region.

The fraction of negatively enrichments of the region represents the proportion of false positive probes included in this region.

MAT score of the region is the maximum MAT score for this region. Positive score means SP1 (group 1 in the contrast) is enriched comparing to Ab (group 2 in the contrast).

- Make sure the enriched spreadsheet is the active spreadsheet, and select **Create region list** on the workflow
- Select the **Specify New Criteria** button on the left panel of *List Creator*

- Type **A** as name of the criteria
- Choose MAT-score(region) from the column drop-down list in the dialog
- Specify to include values greater than 0 only (Figure 15) since positive MAT score mean SP1 is enriched comparing to Ab

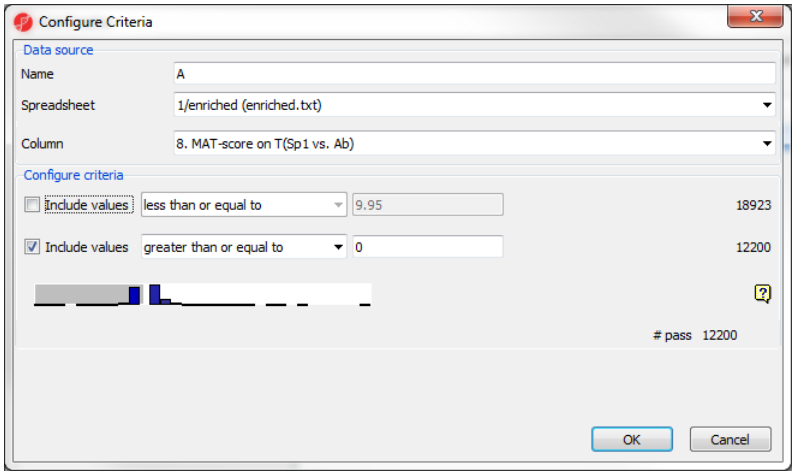


Figure 15: Setting region list criteria to choose regions whose MAT score are positive

- Select **OK,** the criteria will appear on the Criteria panel of List Creator (Figure 16)

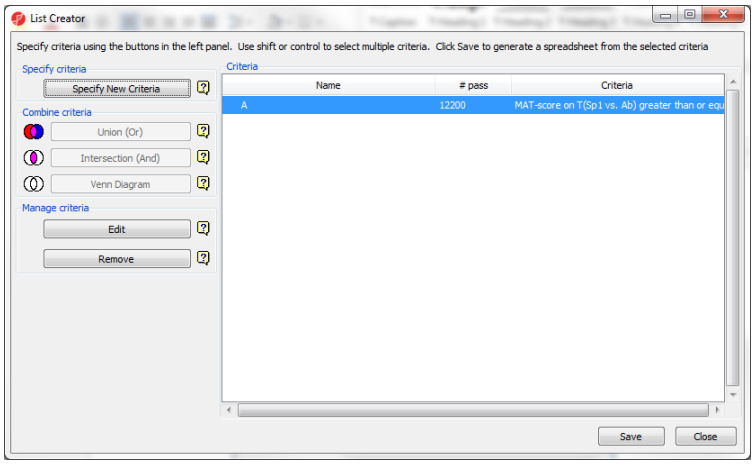


Figure 16: Select a criteria and save a list of the regions meet the criteria

- You can create multiple criteria, use Ctrl+left click on each one of them to select and click on Union or Intersection button to create new criteria
- Select one a criteria to highlight it, click **Save** to save a list that meet the selected criteria
- Make sure the criteria you want to use is checked, and select **OK** (Figure 17)

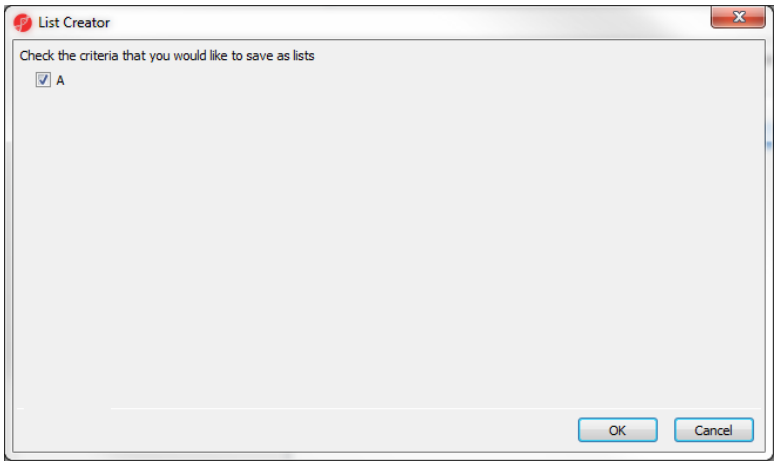


Figure 17: Naming the result list by using the default name

After selecting OK, a new spreadsheet will be created containing the regions that meet the criteria on the rows, which is a subset of the enriched result spreadsheet (Figure 18). The *List Creator* dialog can be closed at this point.

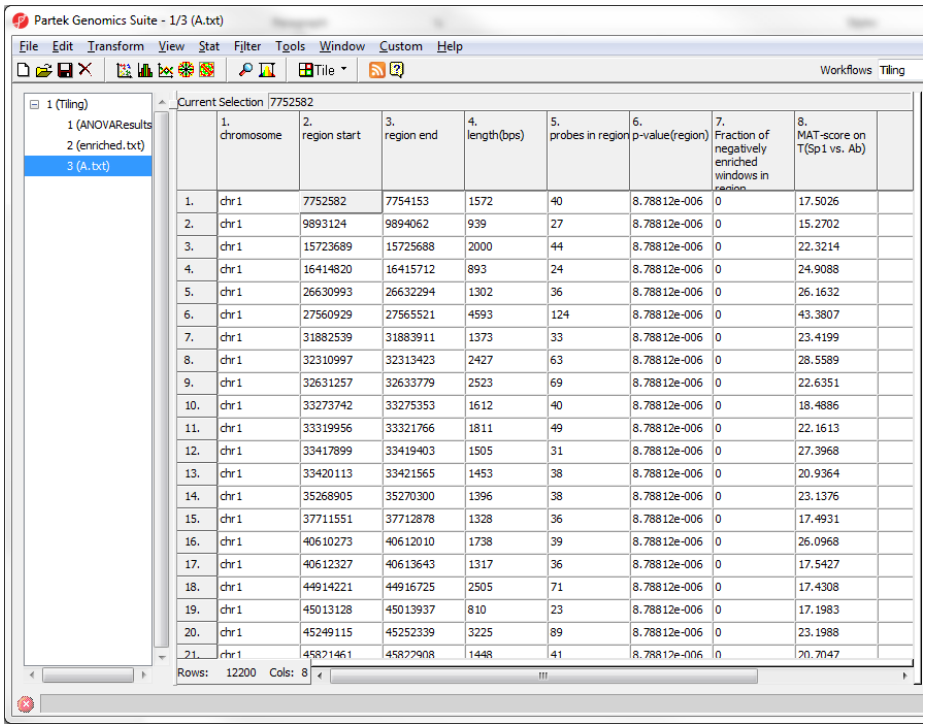


Figure 18: Viewing the significant region list spreadsheet

- Select **Find overlapping genes** from the workflow while the newly created region spreadsheet is active (From File>Properties, make sure the genome build is correctly specified, hg18).
- In the *Output Overlapping Features* dialog select **RefSeq Transcripts – 2014-04-29**. Select **OK** (download if needed) (Figure 19).

- A new spreadsheet (*gene-list*) will be created by annotating the enriched regions with RefSeq Transcripts with each row of the new spreadsheet corresponds to a transcript.

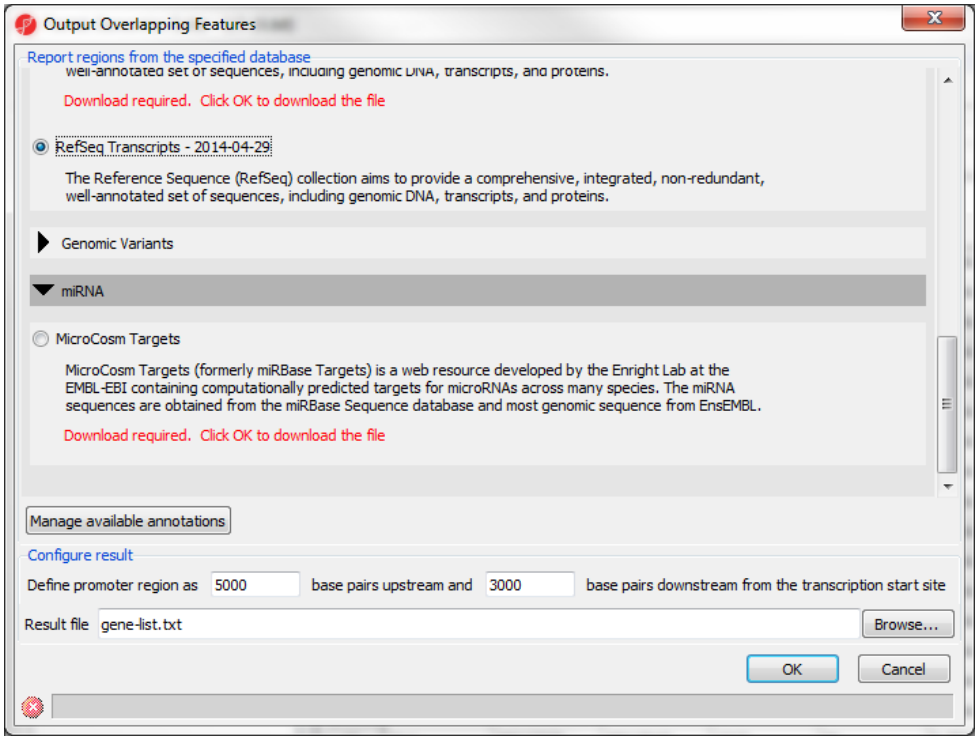
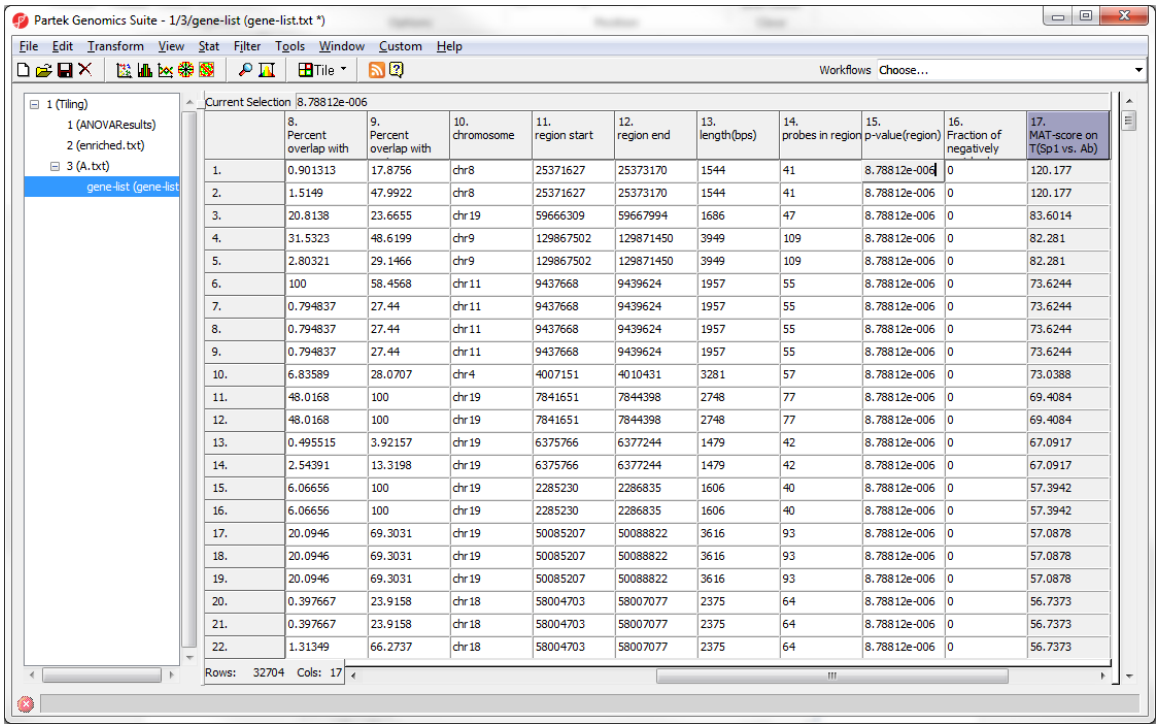


Figure 19: Configuring the Output Overlapping Features dialog

- Right click on the MAT-score column header, and select the **Sort Descending** option; this will place the most significant region on the top of the list (Figure 20)



	8. Percent overlap with	9. Percent overlap with	10. chromosome	11. region start	12. region end	13. length(bps)	14. probes in region	15. p-value(region)	16. Fraction of negatively	17. MAT-score on T(Sp1 vs. Ab)
1.	0.901313	17.8756	chr8	25371627	25373170	1544	41	8.78812e-006	0	120.177
2.	1.5149	47.9922	chr8	25371627	25373170	1544	41	8.78812e-006	0	120.177
3.	20.8138	23.6655	chr19	59666309	59667994	1686	47	8.78812e-006	0	83.6014
4.	31.5323	48.6199	chr9	129867502	129871450	3949	109	8.78812e-006	0	82.281
5.	2.80321	29.1466	chr9	129867502	129871450	3949	109	8.78812e-006	0	82.281
6.	100	58.4568	chr11	9437668	9439624	1957	55	8.78812e-006	0	73.6244
7.	0.794837	27.44	chr11	9437668	9439624	1957	55	8.78812e-006	0	73.6244
8.	0.794837	27.44	chr11	9437668	9439624	1957	55	8.78812e-006	0	73.6244
9.	0.794837	27.44	chr11	9437668	9439624	1957	55	8.78812e-006	0	73.6244
10.	6.83589	28.0707	chr4	4007151	4010431	3281	57	8.78812e-006	0	73.0388
11.	48.0168	100	chr19	7841651	7844398	2748	77	8.78812e-006	0	69.4084
12.	48.0168	100	chr19	7841651	7844398	2748	77	8.78812e-006	0	69.4084
13.	0.495515	3.92157	chr19	6375766	6377244	1479	42	8.78812e-006	0	67.0917
14.	2.54391	13.3198	chr19	6375766	6377244	1479	42	8.78812e-006	0	67.0917
15.	6.06656	100	chr19	2285230	2286835	1606	40	8.78812e-006	0	57.3942
16.	6.06656	100	chr19	2285230	2286835	1606	40	8.78812e-006	0	57.3942
17.	20.0946	69.3031	chr19	50085207	50088822	3616	93	8.78812e-006	0	57.0878
18.	20.0946	69.3031	chr19	50085207	50088822	3616	93	8.78812e-006	0	57.0878
19.	20.0946	69.3031	chr19	50085207	50088822	3616	93	8.78812e-006	0	57.0878
20.	0.397667	23.9158	chr18	58004703	58007077	2375	64	8.78812e-006	0	56.7373
21.	0.397667	23.9158	chr18	58004703	58007077	2375	64	8.78812e-006	0	56.7373
22.	1.31349	66.2737	chr18	58004703	58007077	2375	64	8.78812e-006	0	56.7373

Figure 20: Sorting the gene-list spreadsheet based on MAT score

Figure 20 shows the **Find overlapping genes** results. For a detailed look at the most significant region, right click on the first row header, and select **Browse to Location**.

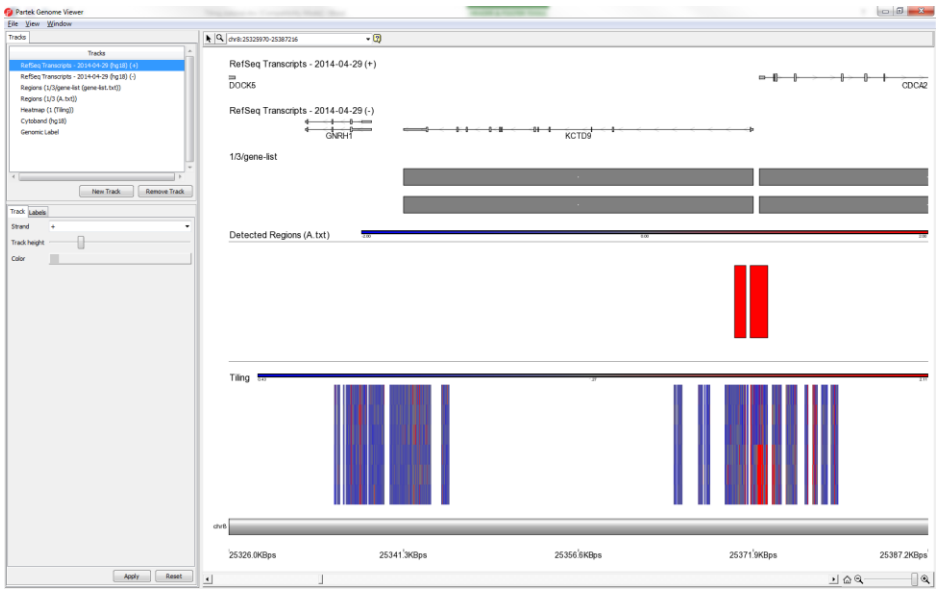


Figure 21: Viewing the region view for a region on the top of the result spreadsheet

In the *Chromosome View* (Figure 21), the region selected in the first row of the gene-list spreadsheet is the left most one shown in the *Detected Regions* track.

Furthermore, the top track provides gene isoform information as present in the chosen transcriptome database (*RefSeq* in this example), followed by information retrieved from the *gene-list* spreadsheet, *enriched* spreadsheet, and the original data spreadsheet (*Tiling*), respectively. To help make this plot easier to understand, we will delete some tracks on the plot, since the gene-list track is redundant information, it shows the boundary of the transcript which is provided by the top two tracks (transcriptome database RefSeq track), we will remove it:

- Select **Regions (1/enriched/gene-list(gene-list.txt))** in the list of tracks on the left (Figure 22)

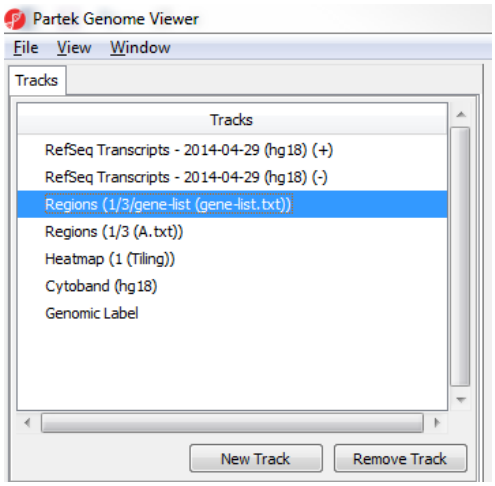


Figure 22: Configuring the list of tracks

- Delete the *Regions (1/enriched/gene-list(gene-list.txt))* track by selecting the **Remove Track** button on the corresponding line; the dialog will look like in Figure 23

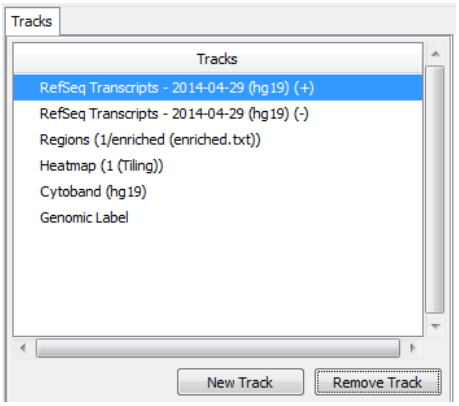


Figure 23: Configuring the list of tracks: After deleting the 3nd track, leave the tracks showing RefSeq transcripts, enriched regions, the heatmap of the samples, cytobands of the current chromosome and genomic coordinates (label)

- Select the **Heatmap** track (which is the third track now in the list of tracks), go to the **Labels** tab, set the *Sample label* drop-down list to point to the **IP** column of the top level spreadsheet, and select the **Label all samples** radio button.
- Select **Apply**

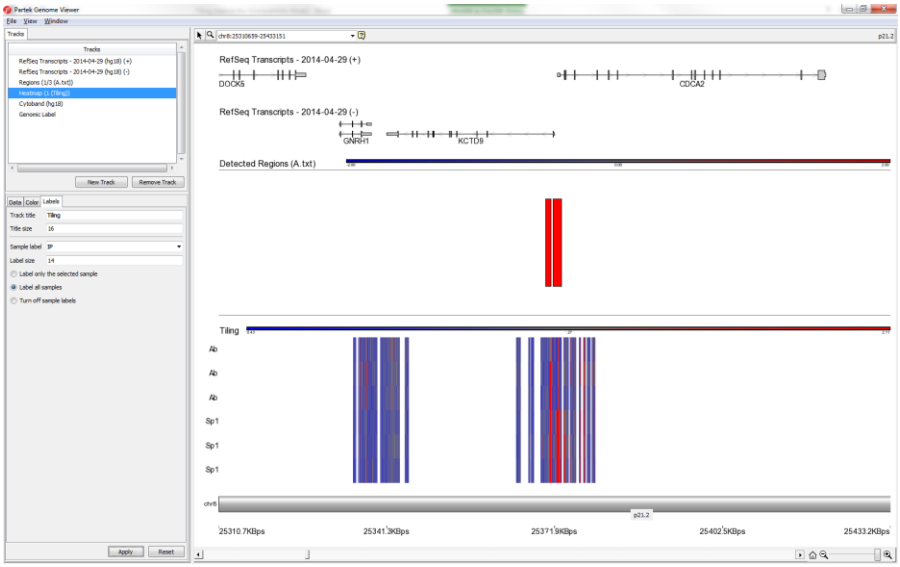


Figure 24: Viewing the promoter region and the genes

Figure 24 shows the region that is at the top of the *gene-list* spreadsheet (the left-most one). The top track is the isoform track with RefSeq transcripts, and the second track shows regions detected as significant binding region. The 3rd track is the heap map of the six samples, and the bottom track is the chromosome cytoband with genomic coordinates.

End of Tutorial

This is the end of the tiling tutorial. If you need additional assistance with this data set, you can call our technical support staff at +1-314-878-2329 or email support@partek.com.