Tasks available for a gene list

- GO Enrichment
- Pathway Enrichment
- Filtering
- Applying Multiple Test Correction
- Plotting numeric data associated with a gene list
- Genome BrowserClustering

GO Enrichment

The Gene Ontology (GO) Enrichment p-value calculation uses either a Chi-Square or Fisher's Exact test to compare the genes included in the significant gene list to all possible genes present in the experiment or the background genes. For a microarray experiment, background genes consists of all genes on the chip/array; for a next generation sequencing experiment, all genes in the species transcriptome are considered background genes.

Because the calculation is essentially comparing overlapping sets of genes and does not use intensity values, GO Enrichment can be performed on an imported gene list even without any numerical values. GO Enrichment is available through the *Gene Expression* workflow.

If no annotation file has been specified for the gene list, GO Enrichment will use the full species transcriptome as the background genes. While suitable for next generation sequencing experiments, for microarray experiments, only the genes on the chip/array are appropriate. Please contact our technical support department for assistance with this step if needed.

Pathway Enrichment

Like GO Enrichment, Pathway Enrichment does not require numerical values, but instead operates on lists of genes - a list of significant genes vs. background genes. Consequently, Pathway Enrichment may be used with an imported list of genes even without any numerical values. The list of background genes is set to the species transcriptome by default, but can be set to a specific set of genes if the gene list has been associated with an annotation file.

Filtering

A gene list can be used to filter another spreadsheet. As an example, we will filter the results of an ANOVA on microarray data using a gene list. This will create a spreadsheet with ANOVA results for only the genes included in our gene list.

- Open the filtering gene list and target spreadsheets
- Select the target spreadsheet in the spreadsheet tree, in this example, genes are on rows in ANOVA result spreadsheet
- · Select Filter from the main toolbar
- Select Filter Rows Based on a List... from Filter Rows (Figure 1)

Edit Transform View Stat Filter	Tools W	indow Custor	n Help												
Filter Rows Interactive Filter											Workflows Choose				
Analysis X Filter Columns Clear Row Filter Based on Annotation					_										
			Sample Rows												
🦲 🗖 🚺 📷			Filter in Genor Randomly Shu												
1 (Gene List.txt)	^ Curre	nt Selection 175	Row Filter Ma												
2 (Down_Syndrome-GE) ANOVA-3way (ANOVAResults)		1. Column #	Filter Rows Ba	ased on a List	ene Symbol	5. Gene Title	6. RefSeq Transcript ID	7. p-value(Type)	8. p-value(Tissue)	9. p-value(Type * Tissue)	10. p-value(Down Syndrome vs. Normal)	11. Ratio(Down Syndrome vs. Normal)	12. Fold-Chan ge(Down Syndrome vs.		
	1.	212	200677_at	754	PTTG1IP	pituitary	NM_001286822	1.21194e-05	1.94366e-08	0.121414	1.21194e-05	1.55161	Normal) 1.55161		
	2.	275	200740_s_at	6612	SUMO3	small	NM_001286416	2.35057e-05	4.24553e-05	0.26166	2.35057e-05	1.60982	1.60982		
	3.	3169	203635_at	10311	DSCR3	Down	NM_006052 ///	3.16516e-05	0.0500067	0.0167661	3.16516e-05	1.31219	1.31219		
	4.	2284	202749_at	7485	WRB		NM_001146218	4.02563e-05	7.63954e-06	0.683564	4.02563e-05	1.86134	1.86134		
	5.	1860	202325_s_at	522	ATP5J	ATP synthase,	NM_001003696	4.60628e-05	0.00414671	0.0744811	4.60628e-05	1.80504	1.80504		
	6.	621	201086_x_at	6651	SON	SON DNA	NM_001291411	4.63436e-05	0.06715	0.0257103	4.63436e-05	1.40251	1.40251		
	7.	19790	220419_s_at	29761	USP25	ubiquitin	NM_001283041	6.29939e-05	0.00773066	0.0353252	6.29939e-05	1.49653	1.49653		
	8.	736	201201_at	1476	CSTB	cystatin B	NM_000100	6.76374e-05	0.000397141	0.613482	6.76374e-05	1.54478	1.54478		
	9.	20293	220922_s_at	30014 ///		sperm protein		7.13086e-05	7.05089e-05	0.00273324	7.13086e-05	0.844538	-1.18408		
	10.	1378	201843_s_at	2202	EFEMP1		NM_001039348	7.75104e-05	4.06033e-09	4.06976e-06	7.75104e-05	1.47381	1.47381		
	11.	1752	202217_at	8209 ///	C21orf33 ///	chromosome 21	NM_004649 ///	7.78656e-05	0.00482433	0.433046	7.78656e-05	1.47676	1.47676		
	12.	353	200818_at	539	ATP 5O	ATP synthase,		8.95525e-05	0.00859787	0.711013	8.95525e-05	1.61131	1.61131		
	13.	17758	218386_x_at	10600	USP 16	ubiquitin	NM_001001992	0.000115354	0.00102705	0.285254	0.000115354	1.59764	1.59764		
	14.	9059	209560_s_at	8788	DLK1	delta-like 1	NM_001032997	0.000118621	2.38791e-09	5.61723e-05	0.000118621	0.54551	-1.83315		
	15.	1177	201642_at	3460	IFNGR2	interferon	NM_005534 ///	0.000140321	0.000110691	0.0175371	0.000140321	1.34069	1.34069		
	16.	17930	218558_s_at	54148	MRPL39	mitochondrial	NM_017446 ///	0.000140577	0.0102415	0.499643	0.000140577	1.3735	1.3735		
	17.	17490	218118_s_at	100287932	TIMM23		NM_006327 ///	0.000151811	2.34518e-05	0.273614	0.000151811	1.23303	1.23303		
	18.	3956	204422_s_at	2247	FGF2	fibroblast	NM_002006	0.000175323	5.32835e-10	3.94422e-06	0.000175323	1.18386	1.18386		
	19.	21057	221689_s_at	51227	PIGP	phosphatidylino	NM_153681 ///	0.000188366	4.06409e-06	0.69133	0.000188366	1.69875	1.69875		
	20.	20134	220763_at	259217	HSPA12A		NM_025015 ///	0.000209582	0.0033191	0.131828	0.000209582	0.845449	-1.1828		
	Rows	22283 Columns:	22 <										>	j	

Figure 5. Filtering rows based on a list

- Select the matching column of your target spreadsheet from the *Key column* drop-down menu; here we have selected **4. Gene Symbol** (Figure 2)
 Select the filtering gene list from the *Filter based on spreadsheet* drop-down menu; here we have selected **1 (Gene List.txt)**Select the matching column of your filtering gene list from the *Key column* drop-down menu; here we have selected **1. Symbol**

Select Rows of Spreadsheet 2/ANOVA-3way	×
Key column 4. Gene Symbol \checkmark of spreadsheet 2/ANOVA-3way PTTG1IP	
Filter based on spreadsheet 1 (Gene List.txt) V Key column 1. Symbol V ACVR1	
ОК	ancel

Figure 6. Selecting matching rows from filtering and target spreadsheets

• Select OK to apply the filter

The target spreadsheet will display the filtered rows (Figure 3). Note that the number of rows has gone from 22,283 prior to filtering (Figure 1) to 153 after filtering (Figure 3).

dit Transform View Stat Filter	TOOIS W	indow Custo	лп пер								Workflows Ch		
											worknows chi	Juse	
is X													
🧀 🗔 🔥 🔝 🌆	1			?									
(Gene List.txt)	Curren	nt Selection 346	59										
(Down_Syndrome-GE) ANOVA-3way (ANOVAResults)		1. Column #	2. Probeset ID	3. Entrez Gene	4. Gene Symbol	5. Gene Title	6. RefSeq Transcript ID	7. p-value(Type)	8. p-value(Tissue)	9. p-value(Type * Tissue)	10. p-value(Down Syndrome vs. Normal)	11. Ratio(Down Syndrome vs. Normal)	12. Fold-Cha ge(Down Syndrome vs.
	1.	1000	201465_s_at	3725	JUN	jun	NM_002228	0.00276405	0.00468229	0.113999	0.00276405	0.773342	Normal) -1.29309
	2.	8806	209305_s_at	4616	GADD45B	growth arrest	NM_015675	0.00316328	2.67727e-05	0.594121	0.00316328	0.785082	-1.27375
	3.	4931	205397_x_at	4088	SMAD3	SMAD family	NM_001145102	0.0072648	0.00156764	0.271038	0.0072648	0.839212	-1.19159
	4.	22241	AFFX-HUMISGF		STAT1	signal	NM_007315 ///	0.0117861	4.03509e-08	0.0527963	0.0117861	1.13528	1.13528
	5.	10700	211260_at	655	BMP7	bone	NM_001719	0.0130345	0.00821726	0.187847	0.0130345	0.901294	-1.1095
	6.	4265	204731_at	7049	TGFBR3	transforming	NM_001195683	0.0135102	0.000922318	0.694707	0.0135102	0.831716	-1.2023
	7.	4930	205396_at	4088	SMAD3	SMAD family	NM_001145102	0.0171952	0.263499	0.170553	0.0171952	0.871235	-1.1478
	8.	21042	221674_s_at	8646	CHRD	chordin	NM_001304472	0.0174254	0.209042	0.611109	0.0174254	0.927927	-1.0776
	9.	9242	209747_at	7043	TGFB3		NM_003239 ///	0.0237967	2.1632e-05	0.065672	0.0237967	0.874731	-1.1432
	10.	4824	205290_s_at	650	BMP2	bone	NM_001200 ///	0.0316624	0.0628251	0.542975	0.0316624	0.853061	-1.1722
	11.	3469	203935_at	90	ACVR1	activin A	NM_001105 ///		0.0142217	0.456649	0.0481196	1.24099	1.24099
	12.	7940	208434_at	2122	MECOM		NM_001105077		0.304578	0.0881724	0.0502184	0.922198	-1.08437
	13.	9147	209651_at	7041	TGFB1I1		NM_001042454	0.0514346	4.355e-09	0.297309	0.0514346	1.24158	1.24158
	14.	2222	202687_s_at	8743	TNFSF10		NM_001190942	0.0546648	0.0197439	0.0296169	0.0546648	1.09811	1.09811
	15.	10305	210838_s_at	94	ACVRL1	activin A	NM_000020 ///	0.0569891	0.0239596	0.0999182	0.0569891	0.907874	-1.1014
	16.	15065	215685_s_at	1746	DLX2	distal-less	NM_004405	0.0590513	0.0575855	0.304606	0.0590513	0.903221	-1.1071
	17.	10699	211259_s_at	655	BMP7	bone	NM_001719	0.0726008	0.00129412	0.353554	0.0726008	0.924517	-1.0816
	18.	8835	209335_at	1634	DCN	decorin	NM_001920 ///	0.0727882	4.35538e-09	0.000596523	0.0727882	1.13119	1.13119
	19.	5801	206268_at	10637	LEFTY1	left-right	NM_020997	0.0731086	0.0636334	0.728535	0.0731086	0.92232	-1.08422
	20.	4460	204926_at	3624	INHBA	inhibin beta A	NM_002192	0.0854303	0.331248	0.784966	0.0854303	0.952633	-1.04972
	Rows:	153 Columns:	22 «										

Figure 7. Filtered spreadsheet. The black bar on the right-hand side of the spreadsheet shows the fraction of filtered-out samples in black vs. the retained samples in yellow.

To use this filtered list for downstream analysis, we can save it.

- Right-click the open spreadsheet in the spreadsheet tree
- Select Clone...
- Use the Clone Spreadsheet dialog to name the new spreadsheet and choose its place in the spreadsheet hierarchy
- Select OK

The new spreadsheet will open. If you want to use the new spreadsheet again in the future, be sure to save it.

Applying Multiple Test Correction

If your imported data contains a list of p-values, you can use any of the available multiple test corrections.

• Select Stat from the main toolbar

- Select Multiple Test
- Select Multiple Test Corrections to launch a dialog with available options (Figure 4), it will add corrected p-value column(s) to the right of the selected p-value column(s)

Ø Multiple Test Correction of Spreadsheet 1							
Method □ Bonferroni □ Dunn-Sidak ☑ FDR Step Up □ FDR Step) Down	FDR q-Value					
Candidate Column(s)	->	Selected Column(s) 5. p-value					
		OK Canc	el				

Figure 8. Options available for Multiple Test Corrections

Plotting numeric data associated with a gene list

A variety of profile plots can be used to visualize the numerical data associated with your imported gene list.

- Select View from the main toolbar
- Select any applicable option

Genome Browser

If you have imported numerical data associated with genes (like p-values or fold-changes), you can visualize these values in the *Genome Browser* once an annotation file is associated to the spreadsheet, and there is genomic location information in the annotation file.

- Right-click on a row header in the imported gene list spreadsheet
- Select Browse to location

If the annotations have been configured properly, you should see a *Regions* track for the first column of numerical data, a cytoband track, and an annotation track. You can also add another track to display a second column of numerical data.

- Select New Track
- Select Add a track from spreadsheet
- Select Next >

A new track titled *Regions* will be added.

- Select Regions in the track preferences panel to edit it
- Select the other numerical column in the Bar height by drop-down menu

Clustering

For a gene list with expression values on each sample, clustering can be performed. Access the clustering function through the toolbar, not from a workflow. The workflow implementations assume that the data to be clustered are found on a parent spreadsheet and the list of genes is in a child spreadsheet.

- Select **Tools** form the main toolbar
- Select Discover then Hierarchical Clustering

Hierarchical Clustering assumes that samples are rows and genes are columns so consider transposing your data if this is not the case. If you have only one column or row of data, cluster only on the dimension with multiple categories by deselecting either *Rows* or *Columns* from *What to Cluster* in the *Hierar chical Clustering* dialog.

« Adding annotations to a gene list Starting with a list of genomic regions »

Additional Assistance

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

