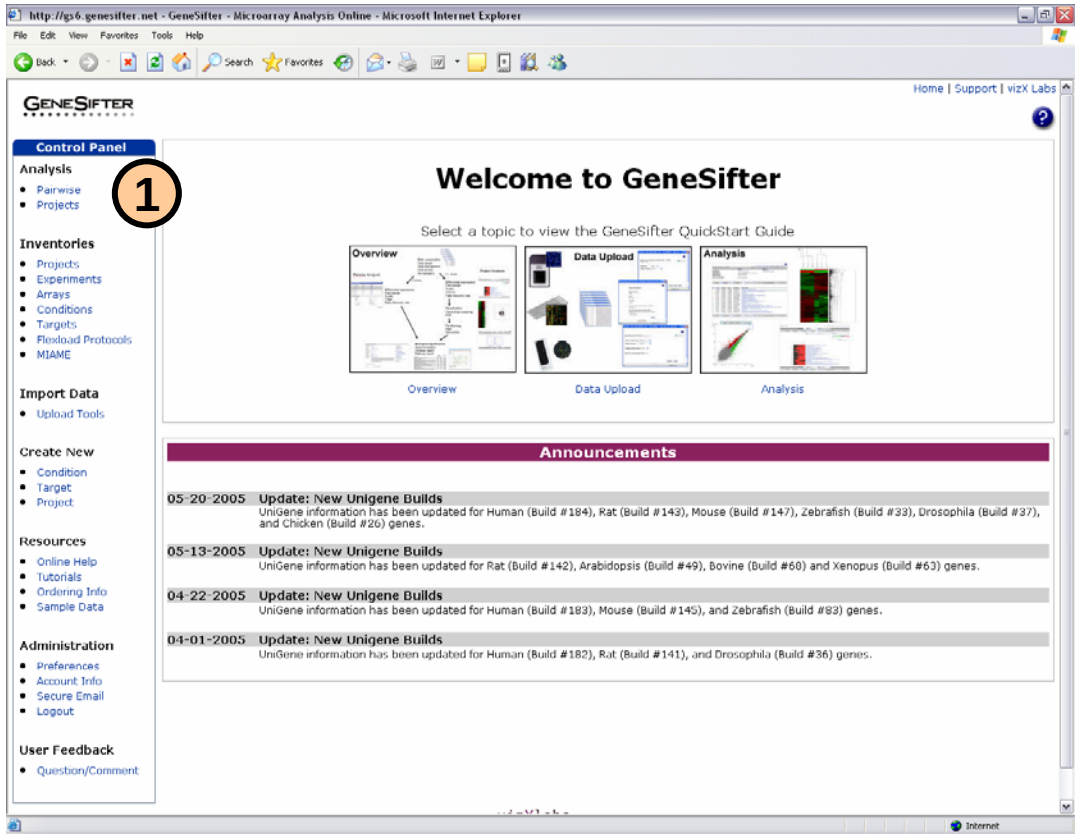
Androgen-independent prostate cancer

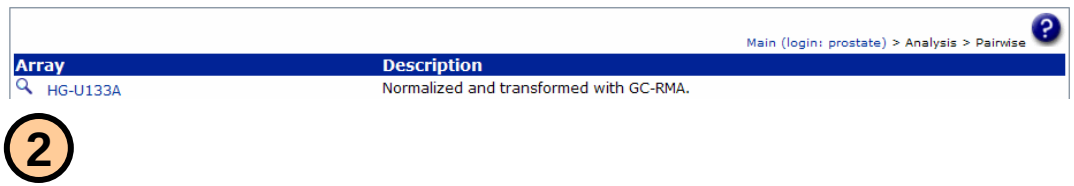
The following tutorial walks through the identification of biological themes in a microarray dataset examining androgen-independent prostate cancer.

Visit the GeneSifter Data Center (www.genesifter.net/web/dataCenter.html) to register for free access to the dataset.

1. From the **Control Panel** select “Pairwise” from the **Analysis** section.

2. Click on the magnifying glass icon next to “**HG-U133A**” to begin the analysis. The data presented here was generated using the Affymetrix® GeneChip® Human Genome U133A array. There are approximately ~22,000 transcripts represented on this array.





Androgen-independent prostate cancer

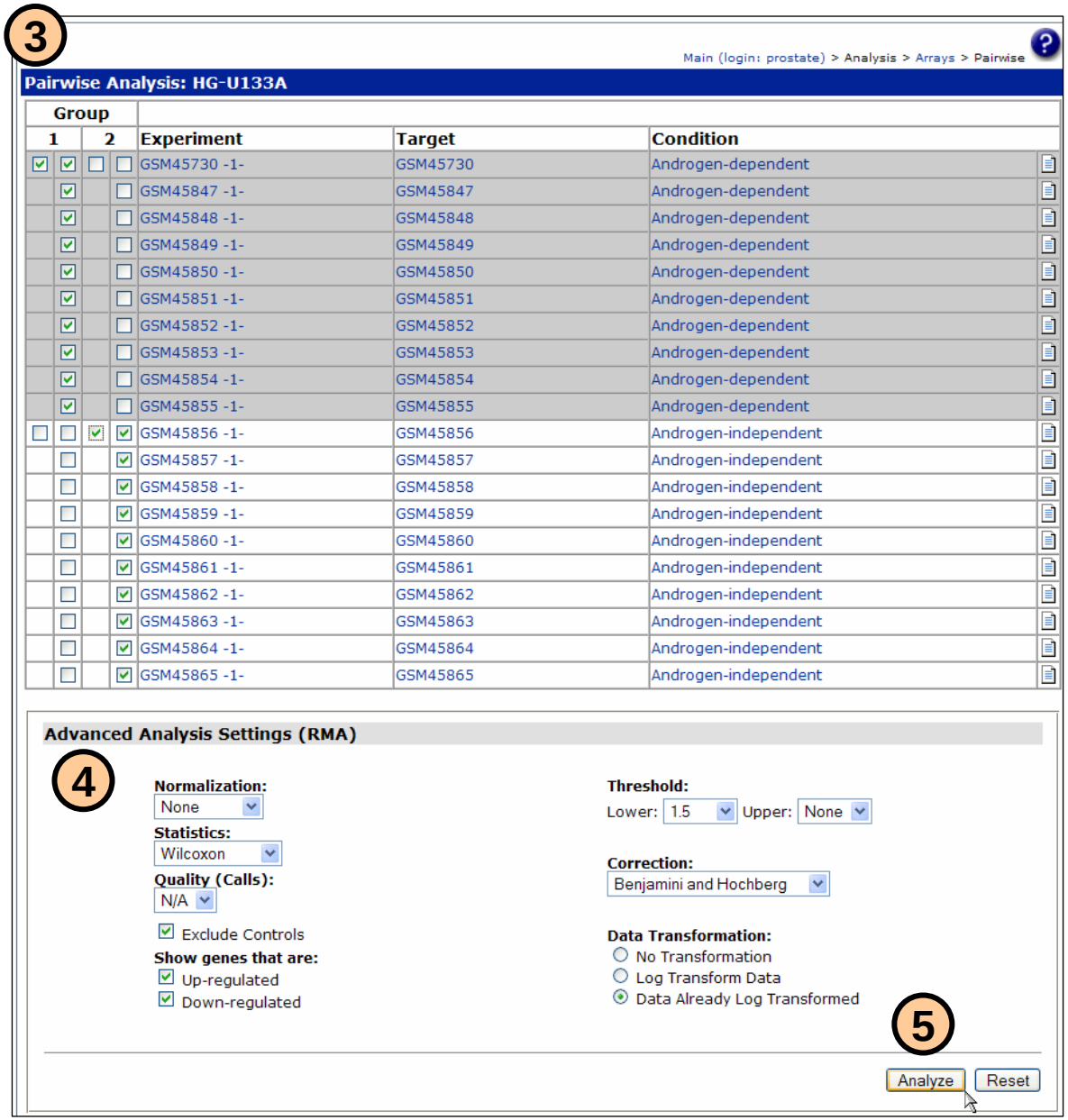
(continued)

3. At the top of the page is a list of the different experiment groups contained in this analysis. We'll be comparing expression between the androgen-dependent and -independent samples. Select the ten "**Androgen-dependent**" arrays to place them in Group 1. Select the ten "**Androgen-independent**" samples for Group 2.

4. Pairwise analysis combines a fold-change cutoff and comparison statistics to generate a list of differentially expressed genes. Select the following analysis parameters:

   Normalization: None
   Data was already normalized with GC-RMA during upload.
   Statistics: Wilcoxon
   Performs a Wilcoxon rank sum test for each gene that passes the fold change cutoff.
   Quality (Calls): N/A
   Quality calls are not generated by GC-RMA.
   Check **Exclude Controls**.
   Threshold: 1.5
   Filters out genes with less than 1.5 fold change in expression.
   Correction: Benjamini and Hochberg
   Calculates a false discovery rate from the raw p-values using the method of Benjamini and Hochberg.
   Data Transformation: Data Already Log Transformed.
   Data was logged during GC-RMA normalization.

5. Click **Analyze**.

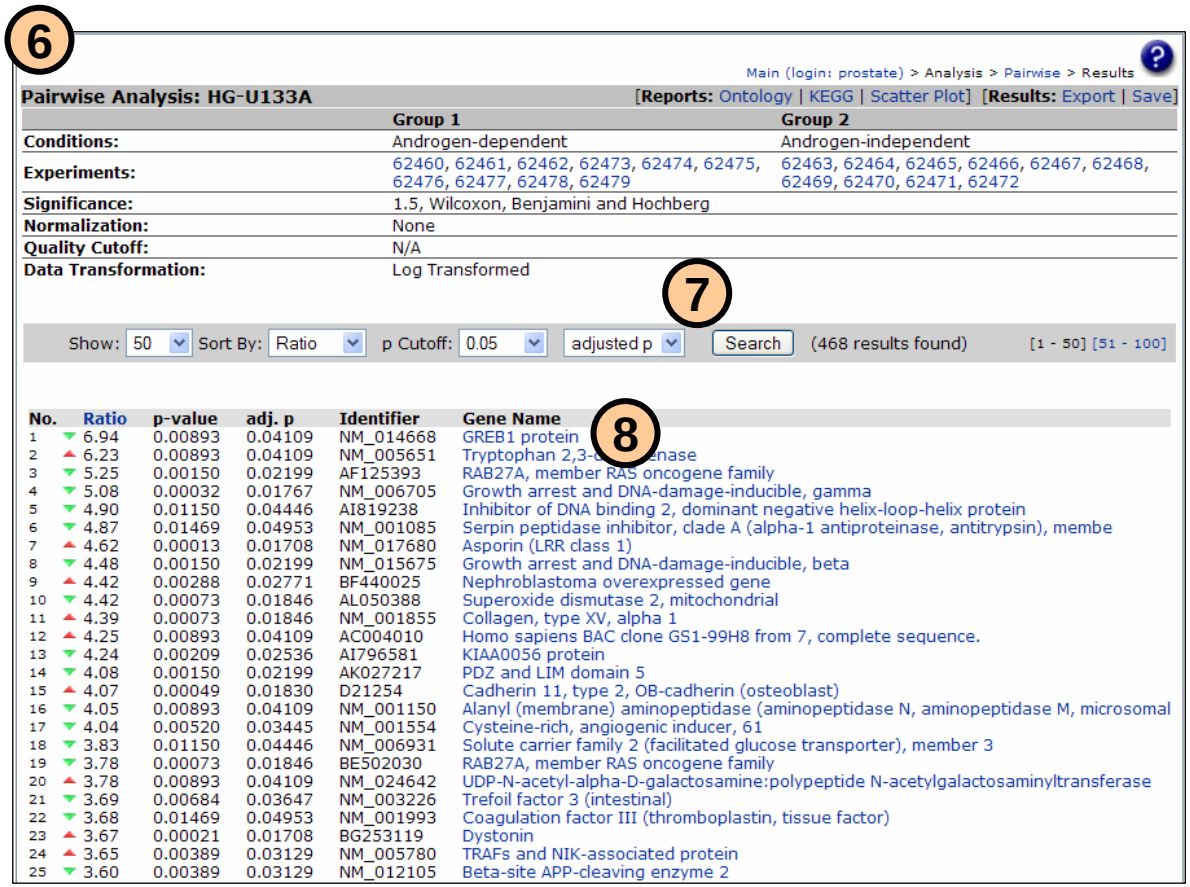


Androgen-independent prostate cancer
(continued)

6. After the analysis is performed, a gene list will be returned. The **Pairwise Analysis Results Page** lists the genes that are differentially expressed, based on the selected parameters. 785 genes passed the initial filtering criteria. The genes are sorted by fold change, and the first 50 genes in the list are displayed.

7. To filter the list using the adjusted p-value (false discovery rate), select "adjusted p" from the pull-down menu, and then click the **Search** button. Using the adjusted p-value reduces the gene list to 468 results with a false discovery rate of less than 5%.

8. To view data and a gene summary for any gene in the list, select the **Gene Name**.



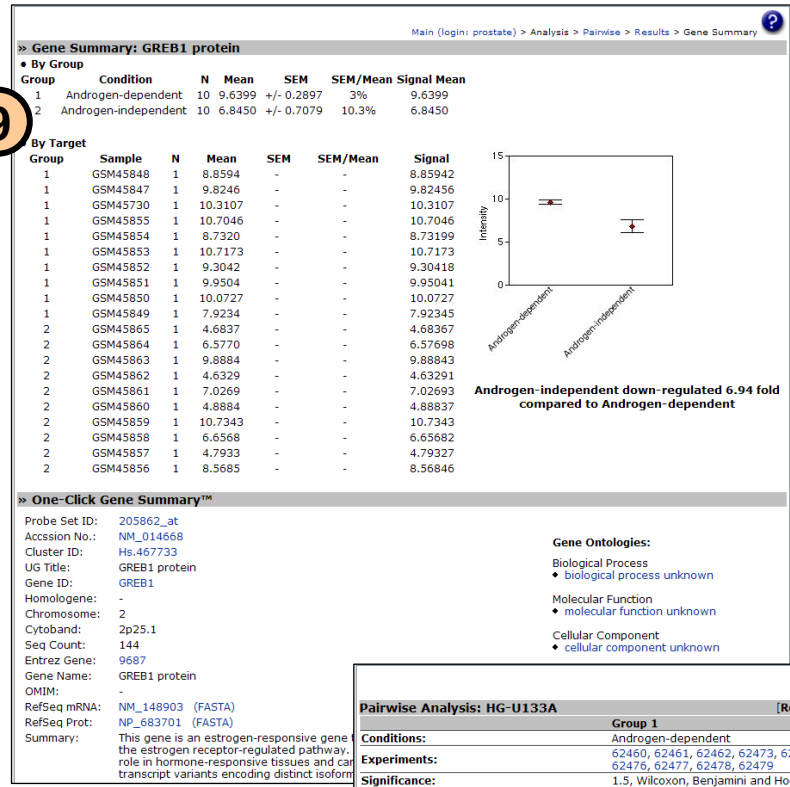
Androgen-independent prostate cancer
(continued)

9. Selecting a gene from the list will bring up the data summary and a One-Click Gene Summary™ for the gene. The One-Click Gene Summary provides a synopsis of current UniGene and Entrez Gene information for the gene.

   Go back to the gene list by clicking the **Back** button in your browser.

10. Select the **Ontology** link to view a summary of the Gene Ontology terms associated with the genes in the list. See the online help for more information about the other reports.

   Note: To view page-specific help documents for any page, select the question mark icon (?) in the upper right corner of each page.



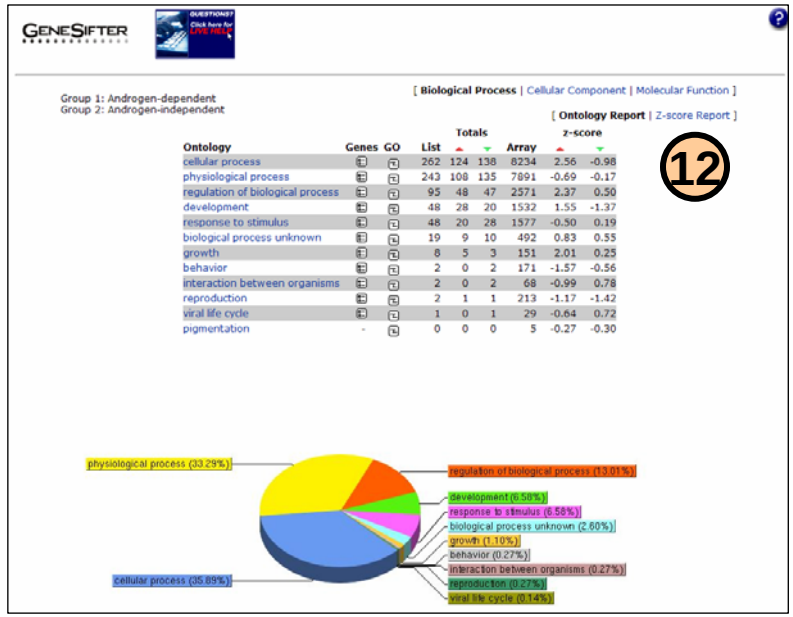
Androgen-independent prostate cancer
(continued)

11. The Ontology Report lists the Gene Ontology terms associated with the 468 genes from the pairwise results list. See the page-specific help for a complete description of the fields in the Ontology Report.

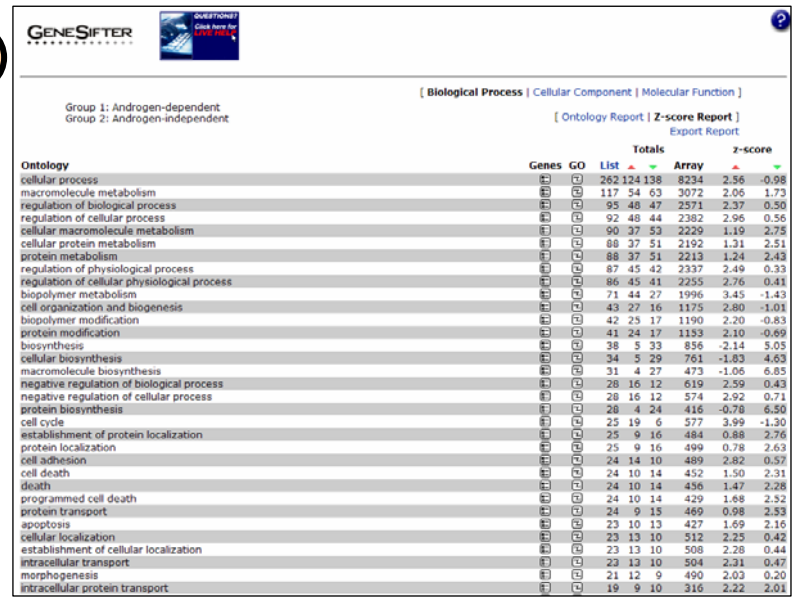
12. Click on **Z-score Report**.

13. The z-score report lists the biological process ontology terms that are significantly over or under-represented in the gene list (z-score greater than 2, or less than -2, respectively). Select the red arrow in the z-score column to sort the list so that the most over-represented genes are at the top.

    A z-score report can also be generated for the molecular function and cellular component categories.



Ontology	Genes	GO	Totals List			Array	z-score	
cellular process			262	124	138	8234	2.56	-0.98
physiological process			243	108	135	7891	-0.69	-0.17
regulation of biological process			95	48	47	2571	2.37	0.50
development			48	28	20	1532	1.55	-1.37
response to stimulus			48	20	28	1577	-0.50	0.19
biological process unknown			19	9	10	492	0.83	0.55
growth			8	5	3	151	2.01	0.25
behavior			2	0	2	171	-1.57	-0.56
interaction between organisms			2	0	2	68	-0.99	0.78
reproduction			2	1	1	213	-1.17	-1.42
viral life cycle			1	0	1	29	-0.64	0.72
pigmentation	-		0	0	0	5	-0.27	-0.30

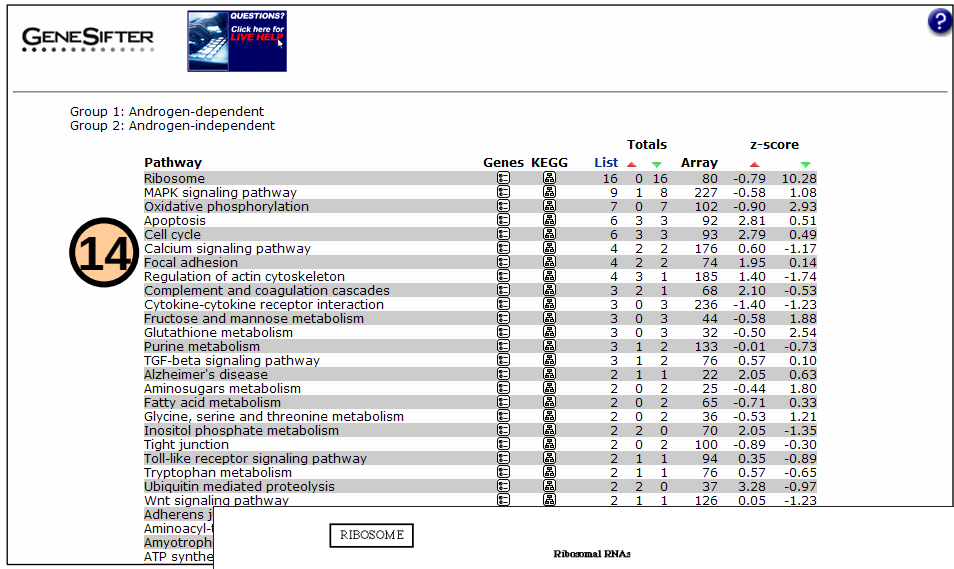


Ontology	Genes	GO	Totals List			Array	z-score	
cellular process			262	124	138	8234	2.56	-0.98
macromolecule metabolism			117	54	63	3072	2.06	1.73
regulation of biological process			95	48	47	2571	2.37	0.50
regulation of cellular process			92	48	44	2382	2.96	0.56
cellular macromolecule metabolism			90	37	53	2229	1.19	2.75
cellular protein metabolism			88	37	51	2192	1.31	2.51
protein metabolism			88	37	51	2213	1.24	2.43
regulation of physiological process			87	45	42	2337	2.49	0.33
regulation of cellular physiological process			86	45	41	2255	2.76	0.41
biopolymer metabolism			71	44	27	1996	3.45	-1.43
cell organization and biogenesis			43	27	16	1175	2.80	-1.01
biopolymer modification			42	25	17	1190	2.20	-0.83
protein modification			41	24	17	1153	2.10	-0.69
biosynthesis			38	5	33	856	-2.14	5.05
cellular biosynthesis			34	5	29	761	-1.83	4.63
macromolecule biosynthesis			31	4	27	473	-1.06	6.85
negative regulation of biological process			28	16	12	619	2.59	0.43
negative regulation of cellular process			28	16	12	574	2.92	0.71
protein biosynthesis			28	4	24	416	-0.78	6.50
cell cycle			25	19	6	577	3.99	-1.30
establishment of protein localization			25	9	16	484	0.88	2.76
protein localization			25	9	16	499	0.78	2.63
cell adhesion			24	14	10	489	2.82	0.57
cell death			24	10	14	452	1.50	2.31
death			24	10	14	456	1.47	2.28
programmed cell death			24	10	14	429	1.68	2.52
protein transport			24	9	15	469	0.98	2.53
apoptosis			23	10	13	427	1.69	2.16
cellular localization			23	13	10	512	2.25	0.42
establishment of cellular localization			23	13	10	508	2.28	0.44
intracellular transport			23	13	10	504	2.31	0.47
morphogenesis			21	12	9	490	2.03	0.20
intracellular protein transport			19	9	10	316	2.22	2.01

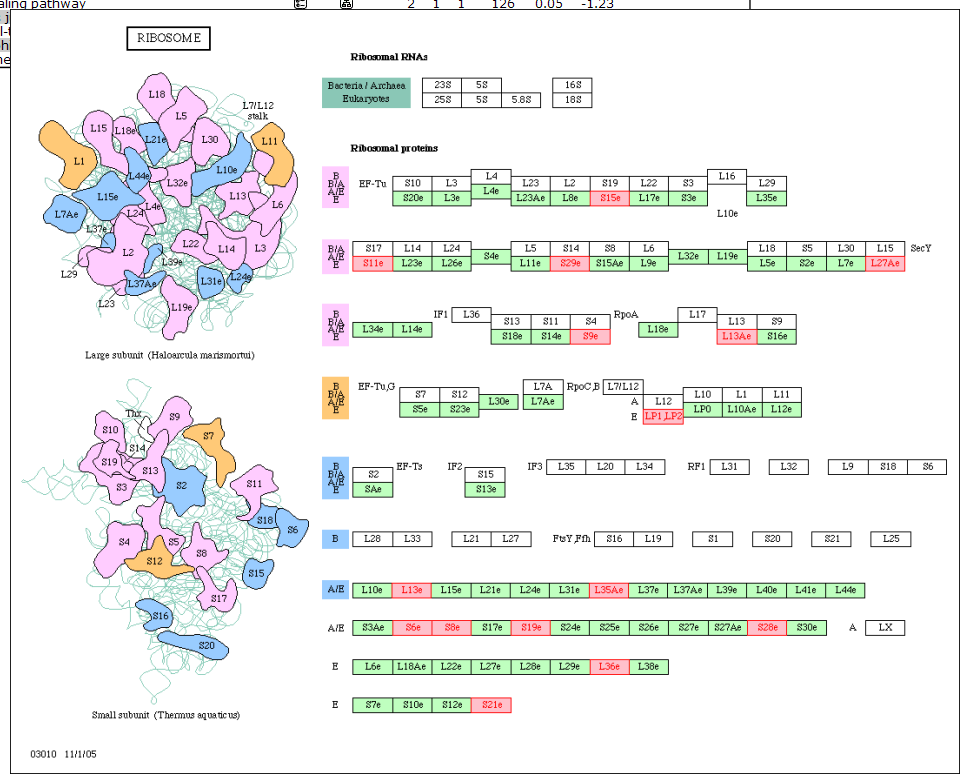
Androgen-independent prostate cancer
(continued)

14. Return to the **Pairwise Analysis Results Page** and click on the **KEGG** link. This will bring up a z-score report for the KEGG pathway terms associated with the gene list.

15. Click on the KEGG logo (▣) for **Ribosome** to show the KEGG pathway diagram. Differentially regulated genes are shown in red.



Pathway	Genes	KEGG	Totals List	▲	▼	Totals Array	z-score ▲	z-score ▼
Ribosome	E	▣	16	0	16	80	-0.79	10.28
MAPK signaling pathway	E	▣	9	1	8	227	-0.58	1.08
Oxidative phosphorylation	E	▣	7	0	7	102	-0.90	2.93
Apoptosis	E	▣	6	3	3	92	2.81	0.51
Cell cycle	E	▣	6	3	3	93	2.79	0.49
Calcium signaling pathway	E	▣	4	2	2	176	0.60	-1.17
Focal adhesion	E	▣	4	2	2	74	1.95	0.14
Regulation of actin cytoskeleton	E	▣	4	3	1	185	1.40	-1.74
Complement and coagulation cascades	E	▣	3	2	1	68	2.10	-0.53
Cytokine-cytokine receptor interaction	E	▣	3	0	3	236	-1.40	-1.23
Fructose and mannose metabolism	E	▣	3	0	3	44	-0.58	1.88
Glutathione metabolism	E	▣	3	0	3	32	-0.50	2.54
Purine metabolism	E	▣	3	1	2	133	-0.01	-0.73
TGF-beta signaling pathway	E	▣	3	1	2	76	0.57	0.10
Alzheimer's disease	E	▣	2	1	1	22	2.05	0.63
Aminosugars metabolism	E	▣	2	0	2	25	-0.44	1.80
Fatty acid metabolism	E	▣	2	0	2	65	-0.71	0.33
Glycine, serine and threonine metabolism	E	▣	2	0	2	36	-0.53	1.21
Inositol phosphate metabolism	E	▣	2	2	0	70	2.05	-1.35
Tight junction	E	▣	2	0	2	100	-0.89	-0.30
Toll-like receptor signaling pathway	E	▣	2	1	1	94	0.35	-0.89
Tryptophan metabolism	E	▣	2	1	1	76	0.57	-0.65
Ubiquitin mediated proteolysis	E	▣	2	2	0	37	3.28	-0.97
Wnt signaling pathway	E	▣	2	1	1	126	0.05	-1.23
Adherens j								
Aminoacyl-								
Amyotroph								
ATP synthe								



Androgen-independent prostate cancer
(continued)

16. Return to the **Pairwise Analysis Results Page** and click **Scatter Plot**.

17. This will bring up a scatter plot of the results. Upregulated genes are shown in red, while downregulated genes are green. The gray spots are those genes that did not pass the analysis criteria (statistics, fold change, etc.). Drag the blue box to an area, and click **Zoom** to see more detail.

18. Click on data points in the detail to bring up the One-Click Gene Summary for a specific gene.

    Only a few specific aspects of the data set have been demonstrated here. Feel free to examine the data further on your own.