
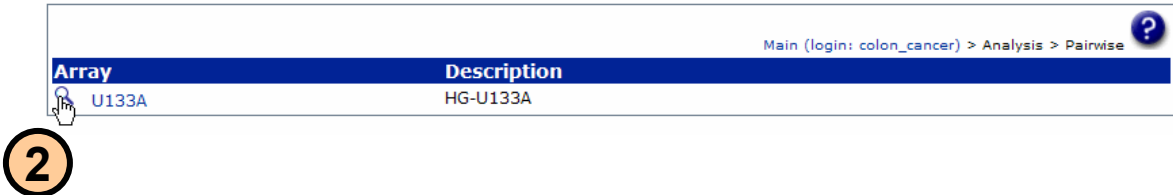
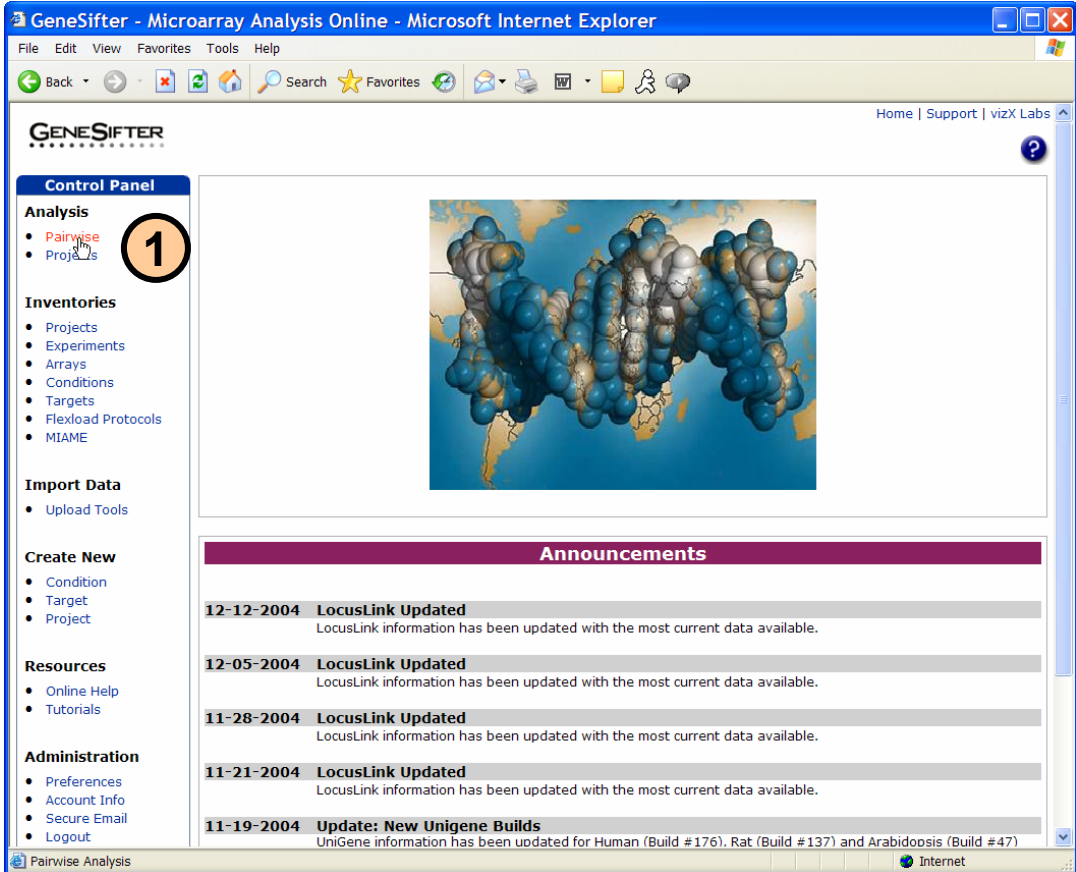


Colon Cancer Progression Tutorial

The following tutorial walks through a comparison of gene expression in a primary tumor colon cell line to that in a metastatic colon cancer cell line. Genes involved in distinct biological processes, including cell cycle and telomere maintenance, are differentially regulated in the progression from primary tumor growth to metastasis.

This dataset can be accessed through the GeneSifter Data Center.

1. Select **Pairwise** from the **Analysis** menu.
2. Select the magnifying glass icon () next to "U133A" in the list. The data presented here was generated using the Affymetrix® GeneChip® Human Genome U133A array. There are approximately 20,000 transcripts represented on this array.



Colon Cancer Tutorial

(continued)

3. Pairwise analysis is used to identify differentially expressed genes in two groups. There are three replicates for each of the two groups in this study. Select the three replicates for the primary cell line as group 1. Select the three replicates for the metastatic line for group 2.

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

Normalization: All Median
Normalizes each array to its median intensity.

Statistics: t-test
Performs a two-sample, unpaired t-test for each gene that passes the quality and fold-change cutoffs.

Quality: 1
Filters out genes that received absent calls in both groups.

Threshold: 1.5
Filters out genes with less than a 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: Log Transform Data
Log base2 transforms the signal values.

5. Select the **Analyze** button

3

The screenshot shows a web interface for 'Pairwise Analysis: U133A'. At the top, there is a breadcrumb trail: 'Main (login: colon_cancer) > Analysis > Arrays > Pairwise'. Below this is a table with columns for 'Group', 'Experiment', 'Target', and 'Condition'. The 'Group' column has sub-columns '1' and '2'. The table contains 7 rows of data. Below the table is an 'Advanced Analysis Settings' section with several dropdown menus and checkboxes. At the bottom right of the settings section are 'Analyze' and 'Reset' buttons.

Group		Experiment	Target	Condition
1	2			
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GSM21712	SW480-1	Primary
<input checked="" type="checkbox"/>	<input type="checkbox"/>	GSM21713	SW480-2	Primary
<input checked="" type="checkbox"/>	<input type="checkbox"/>	GSM21714	SW480-3	Primary
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GSM21715	SW620-1	Metastatic
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GSM21716	SW620-2	Metastatic
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GSM21718	SW620-3	Metastatic

Advanced Analysis Settings

Normalization: All Median

Statistics: t-test

Quality: 1

Threshold: Lower: 1.5 Upper: None

Correction: Benjamini and Hochberg

Show genes that are:

- Up-regulated
- Down-regulated

Data Transformation:

- No Transformation
- Log Transform Data
- Data Already Log Transformed

Analyze Reset

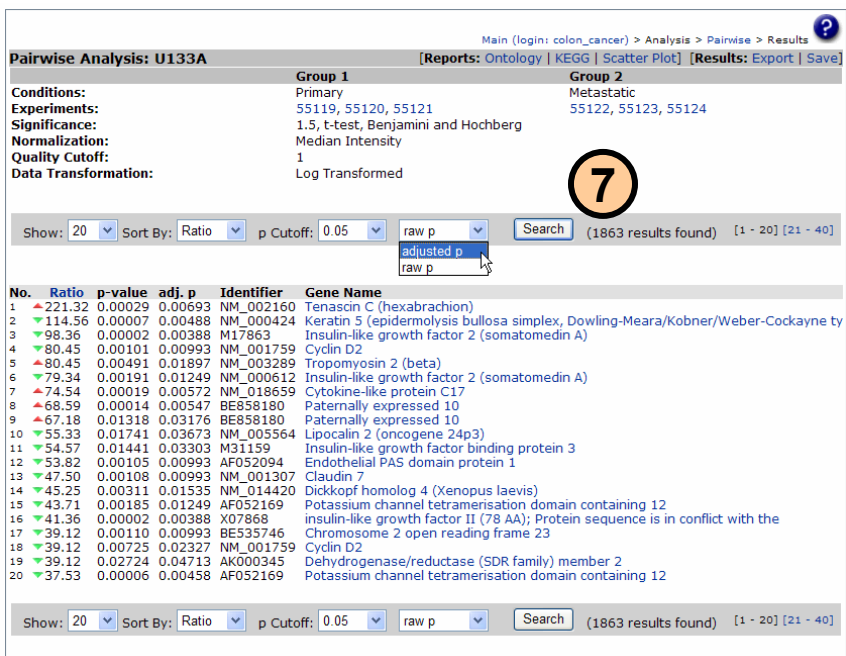
4

5

Colon Cancer Tutorial (continued)

- After the analysis is performed a gene list will be returned. This list contains the genes that are differentially expressed based on the pairwise analysis setting selected. 1863 genes passed the filtering criteria – a 1.5 fold or greater change in expression, present calls in at least one of the groups and a raw p-value of at least 0.05 from the t-test. The genes are sorted by fold change and the first 20 genes in the list are displayed.
- To filter the list using the adjusted p value (false discovery rate), select “adjusted p” from the pull-down menu and the select the **Search** button.
- The list filtered on the adjusted p value contains 1534 genes with a false discovery rate less than 5%.
- To view data and gene summary for any gene in the list select the **Gene Name**.

6



Pairwise Analysis: U133A [Reports: Ontology | KEGG | Scatter Plot] [Results: Export | Save]

Conditions: Group 1: Primary, Group 2: Metastatic
 Experiments: 55119, 55120, 55121, 55122, 55123, 55124
 Significance: 1.5, t-test, Benjamini and Hochberg
 Normalization: Median Intensity
 Quality Cutoff: 1
 Data Transformation: Log Transformed

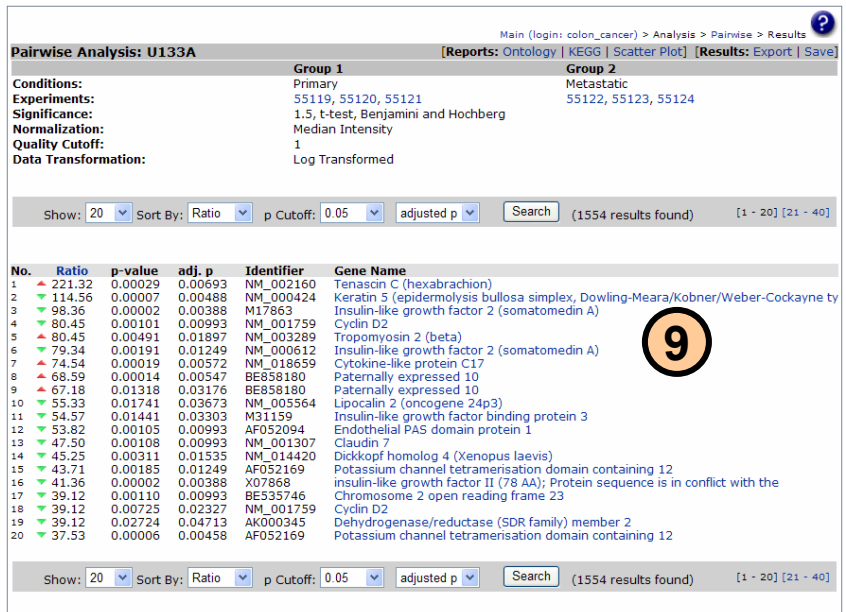
Show: 20 Sort By: Ratio p Cutoff: 0.05 raw p Search (1863 results found) [1 - 20] [21 - 40]

No.	Ratio	p-value	adj. p	Identifier	Gene Name
1	221.32	0.00029	0.00693	NM_002160	Tenascin C (hexabrachion)
2	114.56	0.00007	0.00488	NM_000424	Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne ty
3	98.36	0.00002	0.00388	M17863	Insulin-like growth factor 2 (somatomedin A)
4	80.45	0.00101	0.00993	NM_001759	Cyclin D2
5	80.45	0.00491	0.01897	NM_003289	Tropomyosin 2 (beta)
6	79.34	0.00191	0.01249	NM_000612	Insulin-like growth factor 2 (somatomedin A)
7	74.54	0.00019	0.00372	NM_018659	Cytokine-like protein C17
8	68.59	0.00014	0.00547	BE858180	Paternally expressed 10
9	67.18	0.01318	0.03176	BE858180	Paternally expressed 10
10	55.33	0.01741	0.03673	NM_005564	Lipocalin 2 (oncogene 24p3)
11	54.57	0.01441	0.03303	M31159	Insulin-like growth factor binding protein 3
12	53.82	0.00105	0.00993	AF052094	Endothelial PAS domain protein 1
13	47.50	0.00108	0.00993	NM_001307	Claudin 7
14	45.25	0.00311	0.01535	NM_014420	Dickkopf homolog 4 (Xenopus laevis)
15	43.71	0.00185	0.01249	AF052169	Potassium channel tetramerisation domain containing 12
16	41.36	0.00002	0.00388	X07868	insulin-like growth factor II (78 AA); Protein sequence is in conflict with the
17	39.12	0.00110	0.00993	BE595746	Chromosome 2 open reading frame 23
18	39.12	0.00725	0.02327	NM_001759	Cyclin D2
19	39.12	0.02724	0.04713	AK000345	Dehydrogenase/reductase (SDR family) member 2
20	37.53	0.00006	0.00458	AF052169	Potassium channel tetramerisation domain containing 12

Show: 20 Sort By: Ratio p Cutoff: 0.05 raw p Search (1863 results found) [1 - 20] [21 - 40]

7

8



Pairwise Analysis: U133A [Reports: Ontology | KEGG | Scatter Plot] [Results: Export | Save]

Conditions: Group 1: Primary, Group 2: Metastatic
 Experiments: 55119, 55120, 55121, 55122, 55123, 55124
 Significance: 1.5, t-test, Benjamini and Hochberg
 Normalization: Median Intensity
 Quality Cutoff: 1
 Data Transformation: Log Transformed

Show: 20 Sort By: Ratio p Cutoff: 0.05 adjusted p Search (1534 results found) [1 - 20] [21 - 40]

No.	Ratio	p-value	adj. p	Identifier	Gene Name
1	221.32	0.00029	0.00693	NM_002160	Tenascin C (hexabrachion)
2	114.56	0.00007	0.00488	NM_000424	Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne ty
3	98.36	0.00002	0.00388	M17863	Insulin-like growth factor 2 (somatomedin A)
4	80.45	0.00101	0.00993	NM_001759	Cyclin D2
5	80.45	0.00491	0.01897	NM_003289	Tropomyosin 2 (beta)
6	79.34	0.00191	0.01249	NM_000612	Insulin-like growth factor 2 (somatomedin A)
7	74.54	0.00019	0.00372	NM_018659	Cytokine-like protein C17
8	68.59	0.00014	0.00547	BE858180	Paternally expressed 10
9	67.18	0.01318	0.03176	BE858180	Paternally expressed 10
10	55.33	0.01741	0.03673	NM_005564	Lipocalin 2 (oncogene 24p3)
11	54.57	0.01441	0.03303	M31159	Insulin-like growth factor binding protein 3
12	53.82	0.00105	0.00993	AF052094	Endothelial PAS domain protein 1
13	47.50	0.00108	0.00993	NM_001307	Claudin 7
14	45.25	0.00311	0.01535	NM_014420	Dickkopf homolog 4 (Xenopus laevis)
15	43.71	0.00185	0.01249	AF052169	Potassium channel tetramerisation domain containing 12
16	41.36	0.00002	0.00388	X07868	insulin-like growth factor II (78 AA); Protein sequence is in conflict with the
17	39.12	0.00110	0.00993	BE595746	Chromosome 2 open reading frame 23
18	39.12	0.00725	0.02327	NM_001759	Cyclin D2
19	39.12	0.02724	0.04713	AK000345	Dehydrogenase/reductase (SDR family) member 2
20	37.53	0.00006	0.00458	AF052169	Potassium channel tetramerisation domain containing 12

Show: 20 Sort By: Ratio p Cutoff: 0.05 adjusted p Search (1534 results found) [1 - 20] [21 - 40]

9

Colon Cancer Tutorial

(continued)

10. Selecting a gene from the list will bring up a data summary and a One-Click Gene Summary™ for the gene. The One-Click Gene Summary provides a synopsis of current UniGene and LocusLink information for the gene.
11. Go back to the gene list by clicking the “Back” button in your browser.
12. Select the **Ontology** link to view a summary of the gene Ontology terms associated with the genes in the list. See the online help system for information about the other reports.

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

10

» **One-Click Gene Summary™**

Probe Set ID: 200953_s_at
 Accession No.: NM_001759
 Cluster ID: Hs.376071
 UG Title: Cyclin D2
 Gene ID: CCND2
 Homologene: -
 Chromosome: 12
 CytoBand: 12p13
 Seq Count: 599
 LocusLink: 894
 Gene Name: cyclin D2
 OMIM: 123833
 RefSeq mRNA: NM_001759 (FASTA)
 RefSeq Prot: NP_001750 (FASTA)
 Summary: The protein encoded by this gene belongs to the highly conserved cyclin family, whose members are characterized by a dramatic periodicity in protein abundance through the cell cycle. Cyclins function as regulators of CDK kinases. Different cyclins exhibit distinct expression and degradation patterns which contribute to the temporal coordination of each mitotic event. This cyclin forms a complex with and functions as a regulatory subunit of CDK4 or CDK6, whose activity is required for cell cycle G1/S transition. This protein has been shown to interact with and be involved in the phosphorylation of tumor suppressor protein Rb. Knockout studies of the homologous gene in mouse suggest the essential roles of this gene in ovarian granulosa and germ cell proliferation. High level expression of this gene was observed in ovarian and testicular tumors.

[Perform Sequence Analysis]

CCND2

- Gene Ontologies:**
- Biological Process
 - regulation of cell cycle
 - cytokinesis
 - Cellular Component
 - nucleus
 - KEGG Pathways:**
 - Cell cycle

11

Main (login: colon_cancer) > Analysis > Pairwise > Results ?

Pairwise Analysis: U133A [Reports: Ontology | KEGG | Scatter Plot] [Results: Export | Save]

Conditions:	Group 1	Group 2
Experiments:	Primary	Metastatic
Significance:	55119, 55120, 55121	55122, 55123, 55124
Normalization:	1.5, t-test, Benjamini and Hochberg	
Quality Cutoff:	Median Intensity	
Data Transformation:	1	
	Log Transformed	

Show: 20 Sort By: Ratio p Cutoff: 0.05 adjusted p Search (1554 results found) [1 - 20] [21 - 40]

No.	Ratio	p-value	adj. p	Identifier	Gene Name
1	221.32	0.00029	0.00693	NM_002160	Tenascin C (hexabrachion)
2	114.56	0.00007	0.00488	NM_000424	Keratin 5 (epidemolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne ty
3	98.36	0.00002	0.00388	M17863	Insulin-like growth factor 2 (somatomedin A)
4	80.45	0.00101	0.00993	NM_001759	Cyclin D2
5	80.45	0.00491	0.01897	NM_003289	Tropomyosin 2 (beta)
6	79.34	0.00191	0.01249	NM_000612	Insulin-like growth factor 2 (somatomedin A)
7	74.54	0.00019	0.00572	NM_018659	Cytokine-like protein C17
8	68.59	0.00014	0.00547	BE858180	Paternally expressed 10
9	67.18	0.01318	0.03176	BE858180	Paternally expressed 10
10	55.33	0.01741	0.03673	NM_005564	Lipocalin 2 (oncogene 24p3)
11	54.57	0.01441	0.03303	M31159	Insulin-like growth factor binding protein 3
12	53.82	0.00105	0.00993	AF052094	Endothelial PAS domain protein 1
13	47.50	0.00108	0.00993	NM_001307	Claudin 7
14	45.25	0.00311	0.01535	NM_014420	Dickkopf homolog 4 (Xenopus laevis)
15	43.71	0.00185	0.01249	AF052169	Potassium channel tetramerisation domain containing 12
16	41.36	0.00002	0.00388	X07868	insulin-like growth factor II (78 AA); Protein sequence is in conflict with the
17	39.12	0.00110	0.00993	BE535746	Chromosome 2 open reading frame 23
18	39.12	0.00725	0.02327	NM_001759	Cyclin D2
19	39.12	0.02724	0.04713	AK000345	Dehydrogenase/reductase (SDR family) member 2
20	37.53	0.00006	0.00458	AF052169	Potassium channel tetramerisation domain containing 12

Show: 20 Sort By: Ratio p Cutoff: 0.05 adjusted p Search (1554 results found) [1 - 20] [21 - 40]

12

Colon Cancer Tutorial (continued)

- 13. The Ontology Report lists the Gene Ontology terms associated with the 1534 genes in the pairwise results gene list. See the help documents for this page for more information about the Ontology Report.
- 14. Click on **z-score report**.
- 15. The z-score report lists the biological process ontologies that are significantly over or under-represented in the gene list. Select the red arrow in the z-score column to sort the list by z-score for the up-regulated genes.

13

GeneSifter - Microarray Analysis Online | Analysis - Microsoft Internet Explorer

Group 1: Primary
Group 2: Metastatic

[Biological Process | Cellular Component | Molecular Function]

[Ontology Report | Z-score Report]

Ontology	Genes	GO	List	Array	z-score
physiological process	789	512	277	7479	2.23 0.07
cellular process	531	342	189	4856	1.89 1.06
development	102	61	41	1364	-3.42 -1.47
regulation of biological process	39	22	17	380	-0.64 0.82
biological_process unknown	34	20	14	277	0.43 1.22
behavior	4	2	2	84	-1.56 -0.64
viral life cycle	2	1	1	33	-0.83 -0.20
obsolete biological process	-	0	0	6	-0.65 -0.48

cellular process (35.38%)
physiological process (25.8%)
development (6.54%)
regulation of biological process (2.00%)
biological_process unknown (2.27%)
behavior (0.27%)
viral life cycle (0.13%)

14

[Biological Process | Cellular Component | Molecular Function]

Group 1: Primary
Group 2: Metastatic

[Ontology Report | **Z-score Report**]
Export Report

Ontology	Genes	GO	Totals		z-score	
			List	Array	▲	▼
physiological process	789	512	277	7479	2.23	0.07
metabolism	547	364	183	5141	2.18	-0.80
cell growth and/or maintenance	360	242	118	2889	4.71	1.34
nucleobase, nucleoside, nucleotide and nucleic acid metabolism	242	173	69	2094	3.53	-1.12
cell communication	226	138	88	2463	-2.32	-0.39
cell proliferation	131	97	34	899	5.35	0.14
development	102	61	41	1364	-3.42	-1.47
cell cycle	97	72	25	588	5.72	0.74
phosphate metabolism	79	54	25	629	2.09	0.38
phosphorus metabolism	79	54	25	629	2.09	0.38
response to external stimulus	75	45	30	1035	-3.09	-1.45
cell organization and biogenesis	74	53	21	503	3.67	0.58
morphogenesis	70	42	28	847	-2.01	-0.63
phosphorylation	68	48	20	514	2.59	0.24
organogenesis	58	32	26	739	-2.58	-0.27
cell surface receptor linked signal transduction	56	34	22	839	-3.11	-1.73
protein transport	55	33	22	359	2.03	2.49
intracellular transport	54	32	22	352	1.93	2.59
regulation of cell cycle	53	38	15	330	3.68	0.83

15

Colon Cancer Tutorial

(continued)

16. The list is now sorted by z-score for the up-regulated genes. The most over or under represented ontologies are at the top of the list.

17. The z-score report shows that there is a significant enrichment of genes involved several biological processes, including cell cycle, RNA processing and telomere maintenance. The arrow highlights the “telomerase-dependent telomere maintenance” ontology which is significantly over-represented in the up-regulated gene list. Select the icon in the **Genes** column to view a list of the genes with this ontology.

Z-score reports can be generated for the Molecular Function and Cellular Component ontologies as well



[**Biological Process** | Cellular Component | Molecular Function]

Group 1: Primary
Group 2: Metastatic

[[Ontology Report](#) | **Z-score Report**]
[Export Report](#)

16

17



Ontology	Genes	GO	Totals			z-score		
			List	▲	▼	Array	▲	▼
G2/M transition of mitotic cell cycle			5	5	0	7	6.92	-0.52
cell cycle			97	72	25	588	5.72	0.74
cytokinesis			21	18	3	85	5.45	-0.08
RNA processing			49	40	9	274	5.43	-0.37
cell proliferation			131	97	34	899	5.35	0.14
cysteinyl-tRNA aminoacylation			2	2	0	2	5.33	-0.28
positive regulation of cell cycle			2	2	0	2	5.33	-0.28
positive regulation of mitosis			2	2	0	2	5.33	-0.28
retrograde transport, endosome to Golgi			2	2	0	2	5.33	-0.28
mitotic cell cycle			47	37	10	255	5.17	0.19
RNA metabolism			51	41	10	296	5.12	-0.30
cell growth and/or maintenance			360	242	118	2889	4.71	1.34
response to DNA damage stimulus			31	27	4	186	4.40	-1.13
telomerase-dependent telomere maintenance			5	5	0	14	4.40	-0.73
response to endogenous stimulus			31	27	4	189	4.31	-1.16
traversing start control point of mitotic cell cycle			3	3	0	6	4.29	-0.48
G-protein coupled receptor protein signaling pathway			18	10	8	494	-4.21	-2.52
DNA replication, synthesis of RNA primer			2	2	0	3	4.20	-0.34
mitotic anaphase			2	2	0	3	4.20	-0.34
myoblast differentiation			3	2	1	3	4.20	2.72
telomere maintenance			5	5	0	15	4.18	-0.76
S phase of mitotic cell cycle			24	19	5	119	4.15	0.29
mRNA metabolism			23	20	3	134	3.92	-0.90
RNA modification			12	10	2	49	3.91	0.14
DNA replication			23	18	5	117	3.86	0.33