www.infogenbio.com **Tel +44(0)7525 064993**

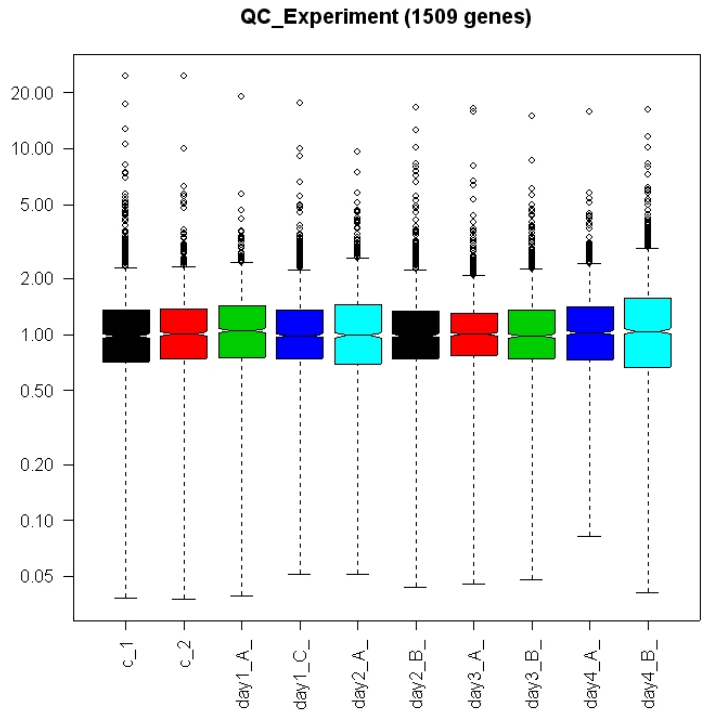
Analysis performed by Dr Ewan Hunter Infogen Bioinformatics Ltd

The next stage in the analysis is to analysis the quality controlled data. The aim of the project is to look at the effect of an agent on the transcription of a murine cell line.

The samples selected based on the quality control steps of the analysis are displayed in the table below:

Sample Name	Time days	Samples
no	no	no
N/A	N/A	N/A
rep2_day4_A_output_fused.xls	4	day4_A_
rep2_day3_A_output_fused.xls	3	day3_A_
rep2_day2_A_output_fused.xls	2	day2_A_
rep2_day1_A_output_fused.xls	1	day1_A_
rep2_control_C_output_fused.xls	ctrl	c_2
rep1_day4_B_output_fused.xls	4	day4_B_
rep1_day3_B_output_fused.xls	3	day3_B_
rep1_day2_B_output_fused.xls	2	day2_B_
rep1_day1_C_output_fused.xls	1	day1_C_
rep1_control_C_output_fused.xls	ctrl	c_1

BoxPlot of the Samples used in the analysis:



The table above also shows the parameters set up in this experiment, time and samples. The sample parameter is only used for checking the quality of the qc sample via the box plot tool. The main and only investigative parameter is time. The original aim was to use triplicates in this study but as the quality control step has suggested it is best to just use duplicates. This may influence the analysis as we have fewer measurements per condition but this is more beneficial over adding a bad sample.

The parameter time has 5 conditions:

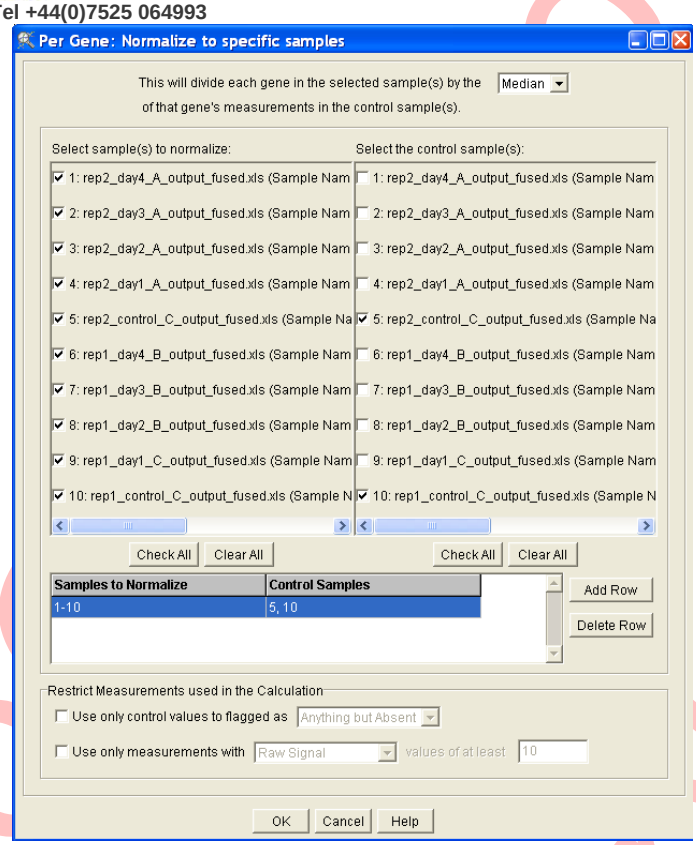
1. Control (vehicle control/no agent)
2. Day 1 post infection
3. Day 2 post infection
4. Day 3 post infection

All the above samples have been hybridized against a common reference pool so the normalization protocol adopted for this analysis is to normalize the post infection time points to the vehicle controls. This approach will blank out any vehicle effects and allow a direct comparison of the PI time points to uninfected cells.

Normalisation:

A Lowess curve was fit to the log-intensity versus log-ratio plot. 35.0% of the data was used to calculate the Lowess fit at each point. This curve was used to adjust the control value for each measurement. If the control channel was lower than 10 then 10 was used instead. Specific

samples were normalized to one another: sample(s) 1-10 were normalized against the median of the control sample(s) 5, 10. Each measurement for each gene in those specific samples was divided by the median of that gene's measurements in the corresponding control samples.



After the samples were normalized 2 interpretations were set up for the analysis protocol:

1. Default Interpretation (log data, In log scale (e))
2. Ratio Interpretation

Both of these interpretations are identical they only differ due to the scale of the data. The ratio data is the non logged version of the logged data. In both interpretations the view of the data in GeneSpring has been altered to take into effect that the samples have been normalized to the control samples. The control condition in the interpretations have been removed, this is purely visual as the data on the controls contribute to the ratios of the informative time points.

See Experiment>Analysis> Spearman_samples> For the Interpretations

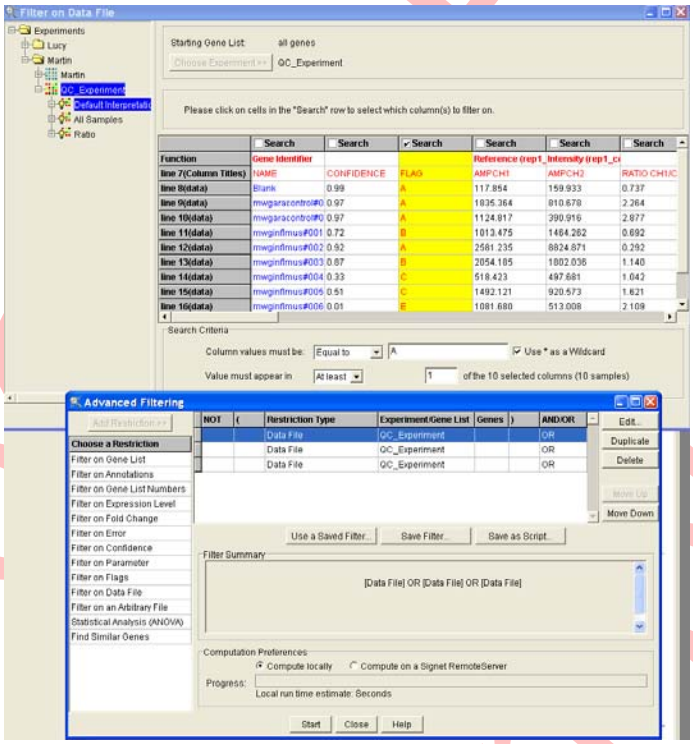
Quality Control of Genes

Before we can analysis our data statistical we need to make sure that the measurements for the genes of this experiment are reasonable. This is a step wise approach which will result in a quality controlled set of genes. Once a set of genes pass one filter they then proceed on to the next filter this filtering insures that we have a set of well measured genes.

1. Flag filter, use the feature extraction error calls
2. Standard Deviation filter on the genes

3. Remove constitutively expressed genes (non changing genes)
4. Quality controlled set of genes

The first step in this filtering schedule is to remove the genes with an unreliable measurement based on the BlueFuse feature extraction software. This software generates 5 flags A through to E. To filter on these values in GeneSpring the best tool to use is the Advanced Filter tool, Tools>Advanced Filter.



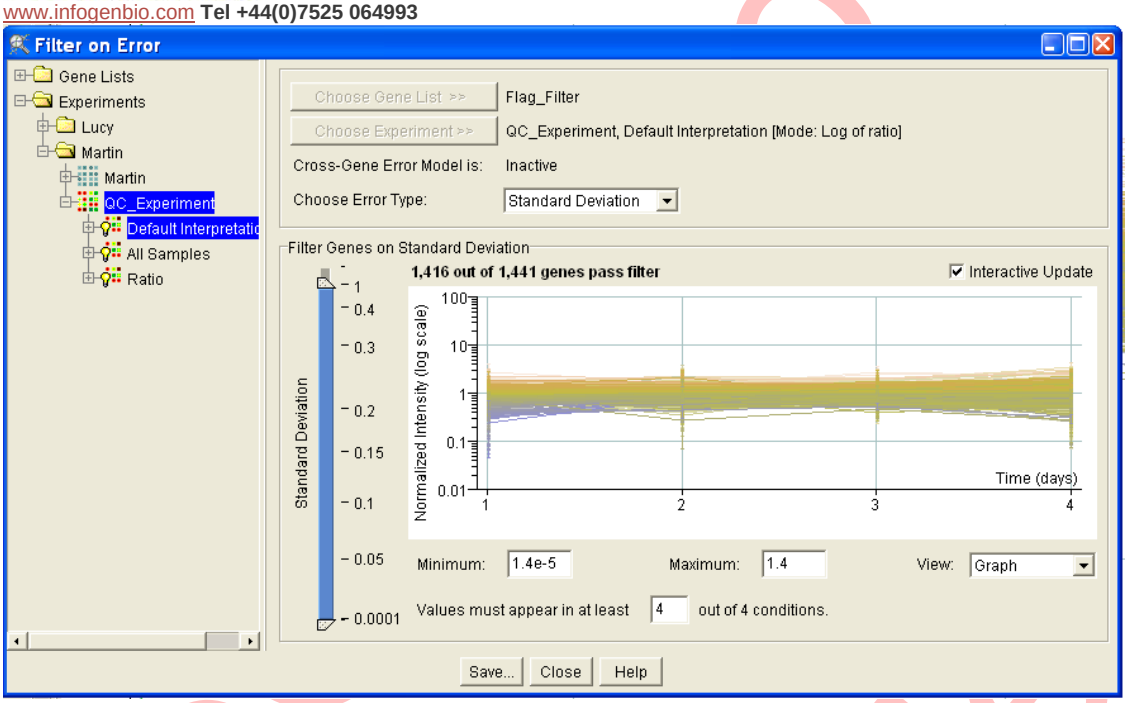
The above image captures the filtering steps used to generate the Flag filter for the Blue Fuse data files. The first 3 flags are the only ones to consider (A, B, and C). Combining these 3 filters by using an OR operator we are left with genes that are either A, B, or C at least once out of the 10 samples. In reverse this filter will exclude the genes which are either D or E in every sample.

After running this filter you are left with 1441 genes from 1509 genes (Flag_Filter gene list in GeneSpring). These genes are then subjected to a standard deviation filter to remove the very deviant genes.

Standard Deviation Filter

Filter>Filter on Error>Default_Interpretation (log):

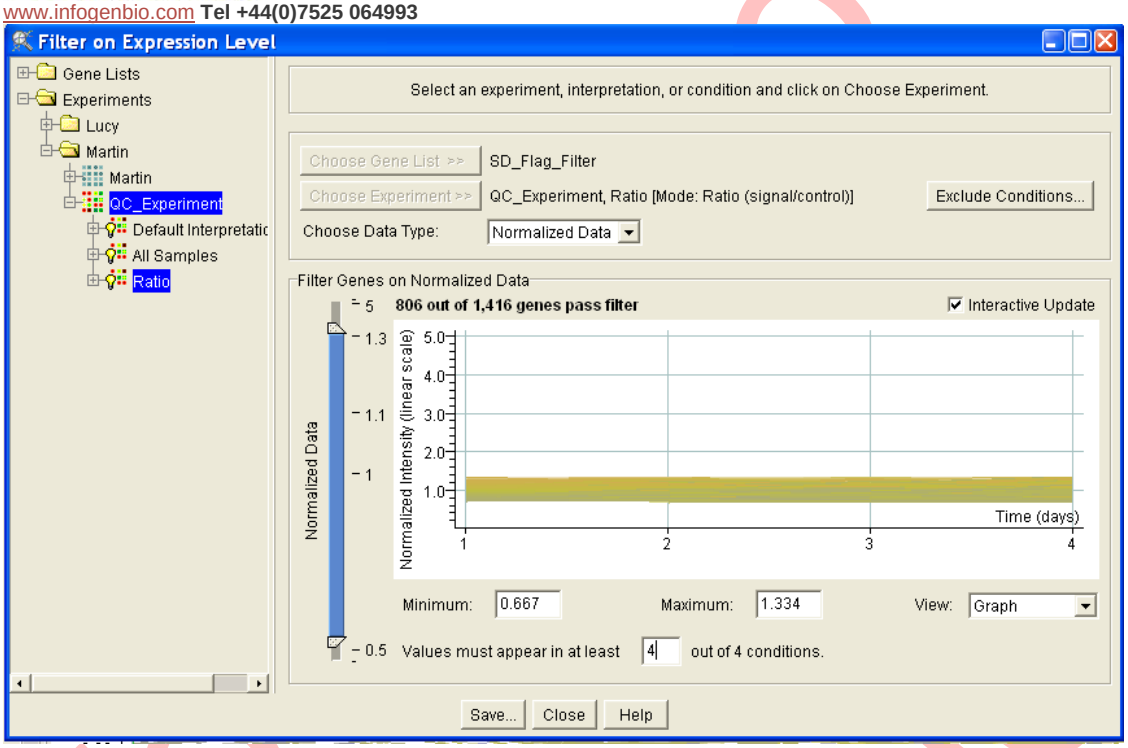
www.infogenbio.com **Tel +44(0)7525 064993**



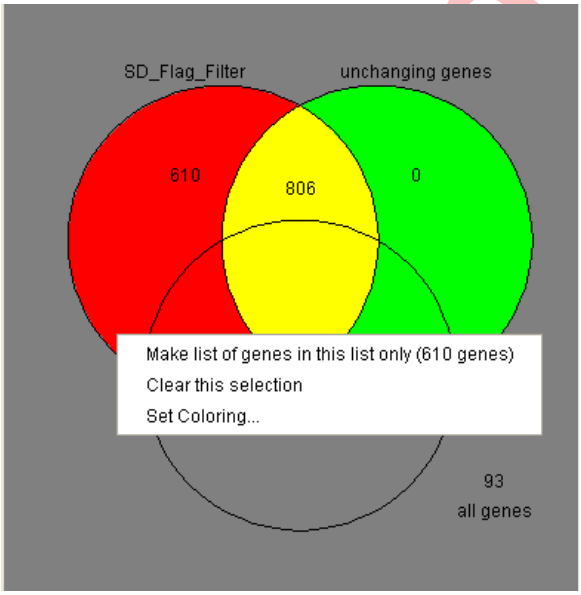
As we applying a statistical filter, we need to use the default interpretation as this is the log interpretation. To apply a SD filter we need to make sure that the data is a normally distributed as possible. For this filter a value of 1.4 SD was applied for the maximum value, the minimum value does not need to be altered as this represents the lowest SD value of a genes measurement. A maximum setting of 1.4 for 4 out of 4 conditions is a reasonably stringent filter. This filter required that each gene must be within 90% of a standard normal density plot. As such this filter is removing very deviant genes from the analysis. After running this filter 1416 genes are left from a starting list of 1441. The resultant gene list is saved as SD_Flag_Filter, this gene list is then used in the third an final step of gene measurement quality control.

Remove constitutively expressed genes (non changing genes)

This filter step will remove genes from the experiment that show no significant change in expression across the extent of this experiment, in other words genes that show no change in overall expression due to the action of the agent. In the majority of microarray experiment the minimum significant fold change accepted is 2 fold. This value of 2 fold is relative to the experiment and it greatly dependent on the number of replicates that you use in your experiment. Most 2-colour experiments cannot go below a 2 fold change due to the sensitivity of the technique used. The RATIO interpretation is the correct interpretation to use in this filter as we need to use arithmetic means for fold change comparisons to get the relatively expression of the genes. If you use the log interpretation the software will use geometric means.



The settings for this filter are evident in the image above, the minimum and maximum values and the filter applied to all 4 conditions results in genes that so no more than a 2 fold change across the experiment. The gene list used in this filter is the SD_Flag_Filter gene list and it produced 806 genes. These 806 genes are the non changing genes and these are the ones that are subtracted from the SD_Flag_Filter gene list to give the QC'd genes. The subtraction is performed by using a VENN diagram:

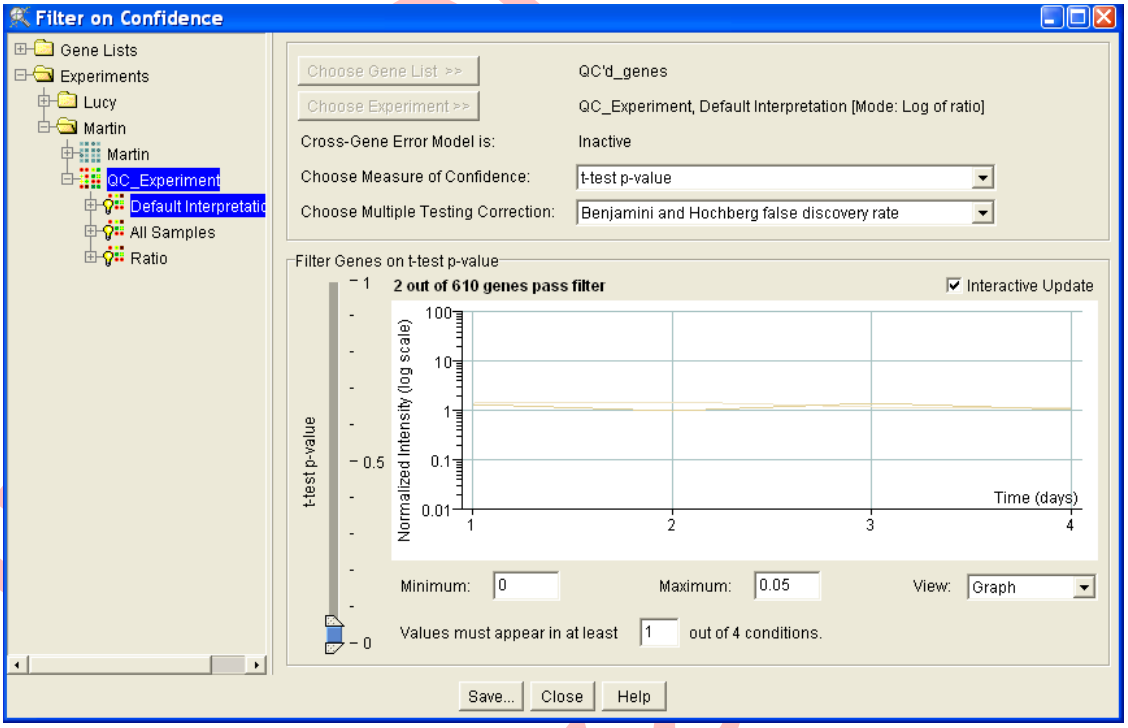


After removing the non changing genes (806), you are left with 610 QC'd genes. These remain genes have passed all the above filters and this gene list (QC'd genes) are then used in all statistical tests.

Statistics

The type of statistics applied to data depends greatly on the questions demanded of the data. I have adopted to apply the 1-sample test in this experiment as we have normalized all the experiment conditions to the control, and I believe it is more informative to find the genes which differ due to the control. We can extend this to a time series analysis by the type of filter we apply statistically and then subjecting the statistically significant genes to a clustering algorithm.

The statistical filter applied is under the Tools menu and within the Filter on Confidence:



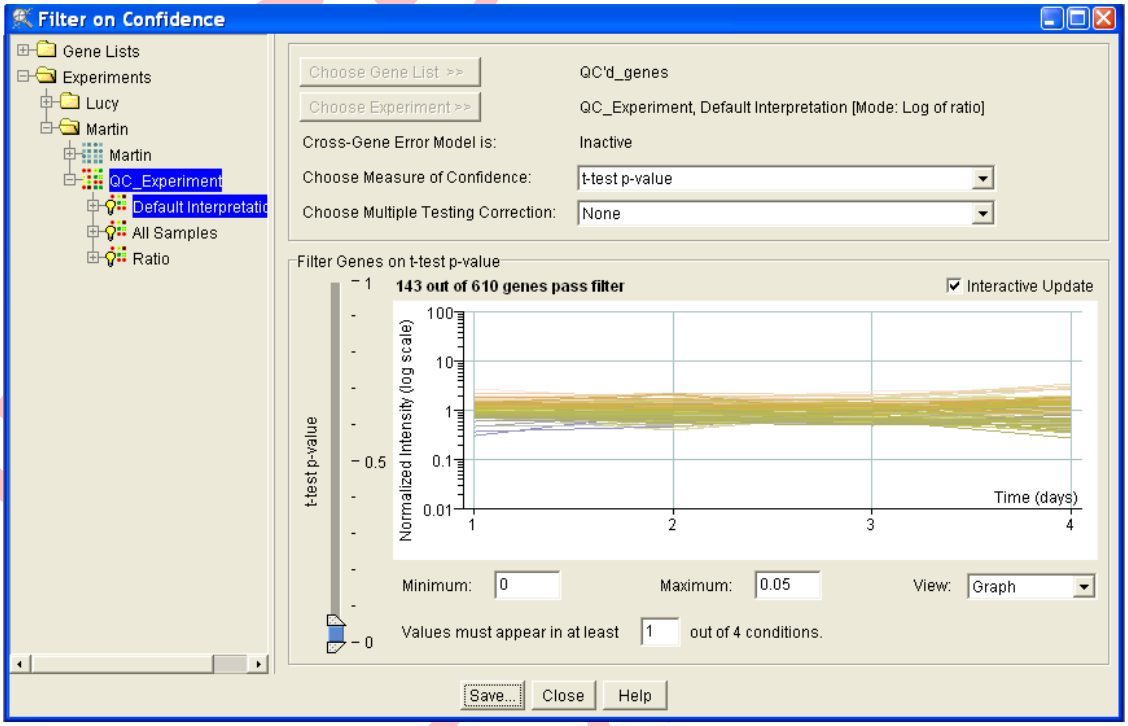
The filter above uses the QC'd genes and the default interpretation, the confidence cut off is set to 95% (0.05 error rate). A multiple test correction (MTC) has also be applied, the Benjamini and Hochberg false discovery rate (FDR) has been applied to correct for the error rate in the statistical test. Running this filter results in only 2 genes out of the 610 QC'd genes. The low return in genes could be due to 2 things lack of precision (low number of replicates) or variable measurements for the genes. Remember we have only applied a 1.4 SD filter on the genes going into the test.

Without running the FDR we would expect to see 30.5 genes pass the 1-sample t.test, as we are only 95% confident in the resultant genes that pass the statistical test. As we are using probability there is a 5% chance that of the total number of genes given into the, these will pass the statistical test purely due to the number of tests we are applying (610). As such the FDR test corrects for these errors, false positives or Type I errors. The FDR test adjusts the p.values of the genes that pass the statistical test, this result of this test depends on the number of genes that pass the statistical test and their resultant p.values. As the number of replicates in this experiment

is low, the value of T needed to get a significant p.value has to be large, so this is dependent on the individual measurements of the genes.

The best approach for this data type is to remove the FDR test and except that some of the genes passing the test are false positives. The reciprocal of false positives are false negatives (Type II errors), these are biologically significant genes but are determine insignificant due to the statistical test.

In this filter I have also only applied a filter in where the genes only have to be significantly changing in 1 condition. This is the correct filter as we only wish to find the genes which are changing from the control situation (conditions have been normalized to this control). If we applied more conditions we would only get genes which are either constantly on or off relatively to the control situation. Applying the 1 out of 4 condition filter will give the statistically significantly changing genes from the control. Once these genes have been found, common trends in gene expression can be gleamed from these genes by subjecting the statistically significant genes to a clustering algorithm.



The statistical genes for the clustering steps were produced using the above settings:

QC'd genes, Default Interpretation, cutoff of 0.05 (No MTC) and 1 condition out of 4

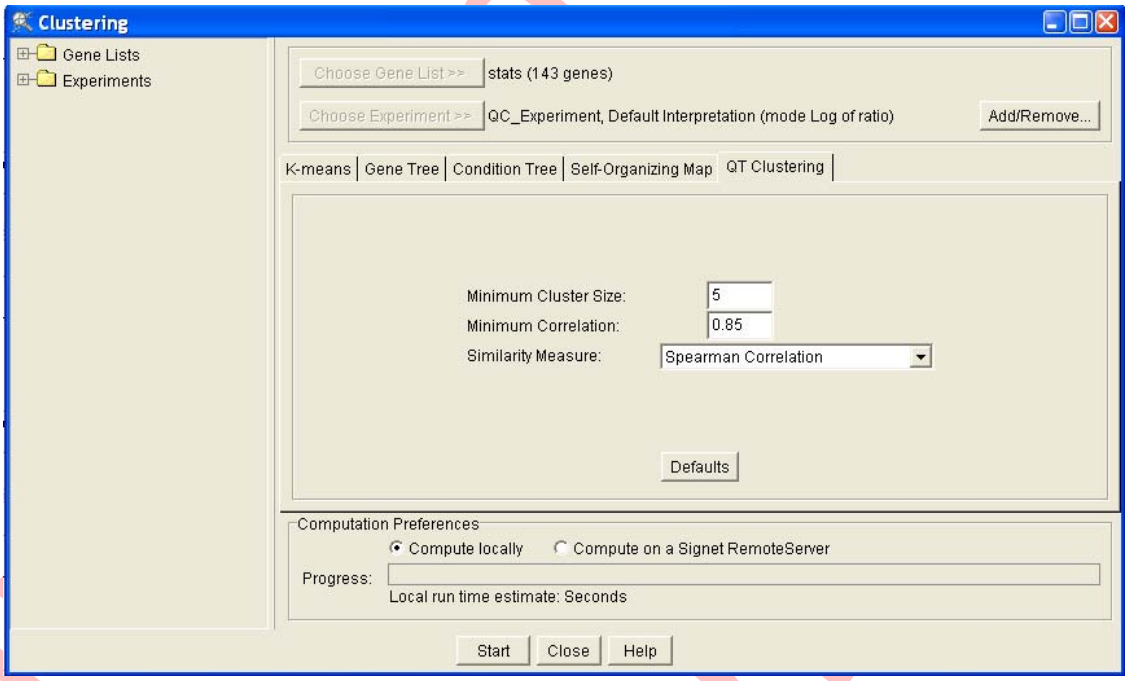
The filter produced 143 genes out of 610, in theory 31 genes could be false positive. However when you apply a FDR correction you only get 2 genes, so by doing so you in theory lose 110 genes which are informative (143-2-31). The resultant genes from this filter have been saved as the stats gene list in GeneSpring.

www.infogenbio.com **Tel +44(0)7525 064993**

Clustering

There are numerous clustering algorithms in GeneSpring and the best algorithm to use on the statistically significant genes is Quality Threshold (QT) clustering. QT clustering is a very powerful tool to use when you have a small set of refined genes. The gene list we created is very small and is ideally suited for this tool. The experiment is also suited for this clustering tool as there are not very many conditions and as such the trends in the experiment are not too complicated.

Settings for QT Clustering:



.

The above image describes the settings for the QT clusters that were then used further in the analysis process. The default setting for this tool were not used because the minimum cluster size is 10, this states that the minimum number of genes in formed cluster is 10. As we only have 143 genes this only produce around 5 sets, leaving the majority of the genes unclustered.

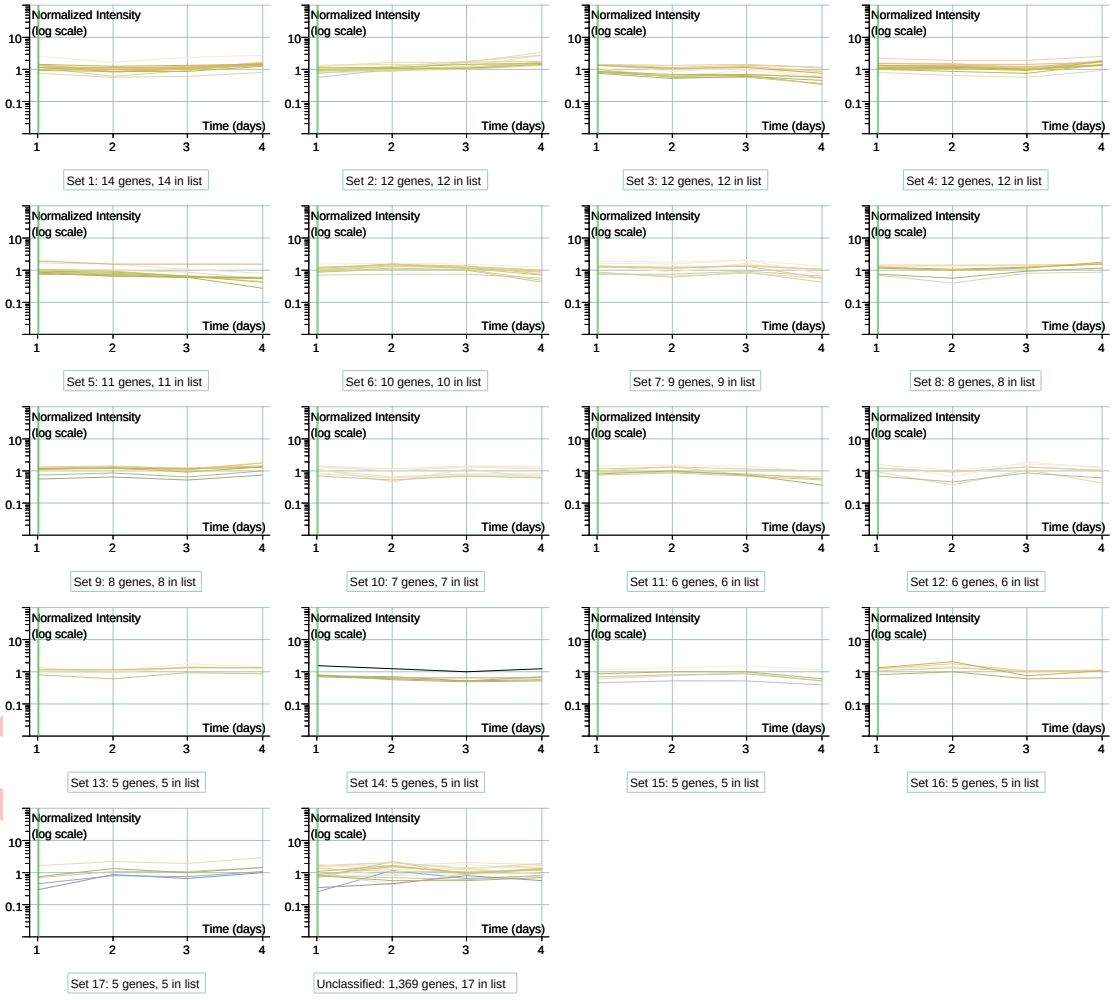
The algorithm works by comparing the relationship between genes based on their vectored correlations. Genes with similar expression trends will be clustered together, this depends on the trend of the gene and the correlation metric applied in the test. The default correlation for GeneSpring is Standard Correlation (cosine angle of separation around zero), this setting for this experiment was also not ideal, as it didn't clustering the data very thoroughly. Correlation metric like Pearson and Standard correlation are not ideal for this experiment as they are more sensitive to the relative expression of the genes in the experiment. The main reason for stating this comment is due to the type of normalization, I have normalized the post-infection conditions to the control samples and as such the normalized values are not true reflections of the cellular expression rather the trend of expression within the experiment.

As such I have adopted to use the Spearman correlation for the metric in the QT clustering and I selected 5 as the minimum gene cluster size, as I wanted to cluster the data as well as possible.

www.infogenbio.com **Tel +44(0)7525 064993**

When I ran the QT Clustering algorithm with these settings and the default correlation (0.9), 17 clusters were produced, not all genes were clustered 17 were left in the unclassified set. When I dropped the correlation percentage down to 0.85 this did not cluster these 17 genes so no further settings were tried. I only produced with the data that was clustered

See image below for the QT Cluster output (0.9, 5, Spearman Correlation)



The image above is represents the final step in the analysis of this data statistically the next step was to run the stats genes against the ontological data and the KEGG pathways.

Ontology Analysis

Ontology gene list were created by first running the GeneSpider on the CUSTOM mouse genome and then splitting out the Gene Ontology classifications into their 3 main classifications, GO Biological Process, GO Cellular Component and GO Molecular Function.

www.infogenbio.com **Tel +44(0)7525 064993**

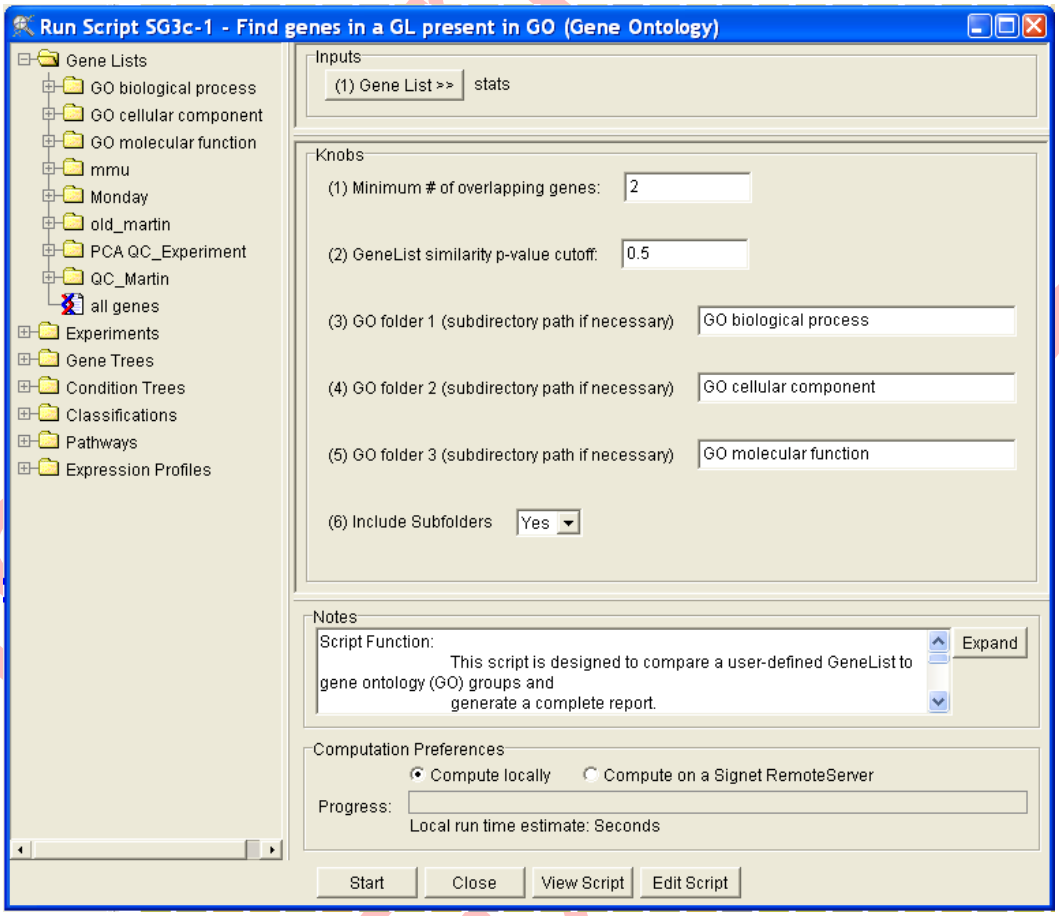
The aim of this analysis is to see if there was any enrichment of know gene ontologies within the genes found in the statistical significant list. This analysis was also performed for the KEGG pathways, KEGG pathways were imported into GeneSpring and then split out as gene lists.

To run these 2 analyses, 2 scripts from the BioScript Library were used:

GO script: SG3c-1 - Find genes in a GL present in GO (Gene Ontology)

KEGG script: SG3b-1 - Find genes in a GL present in PATHWAYS

GO Script settings:



This script compares the overlap of the stats genes and the ontology folders, the ontology folders contain all the biologically classified gene lists. This reflects the settings, the minimum gene overlap before a new gene list forms is 2 and then these lists are then filtered based on the significance of the overlap. The statistical test is a hypergeometrical comparison of the gene membership, as such this is based on the level of annotation within the genome. The filter of 0.5 states that we are 50% confident that the overlap that generated the new gene list is statistically significant result, but keep in mind that the genes used (stats) are significant based on expression. See next page for the outcome of this script:

www.infogenbio.com **Tel +44(0)7525 064993**

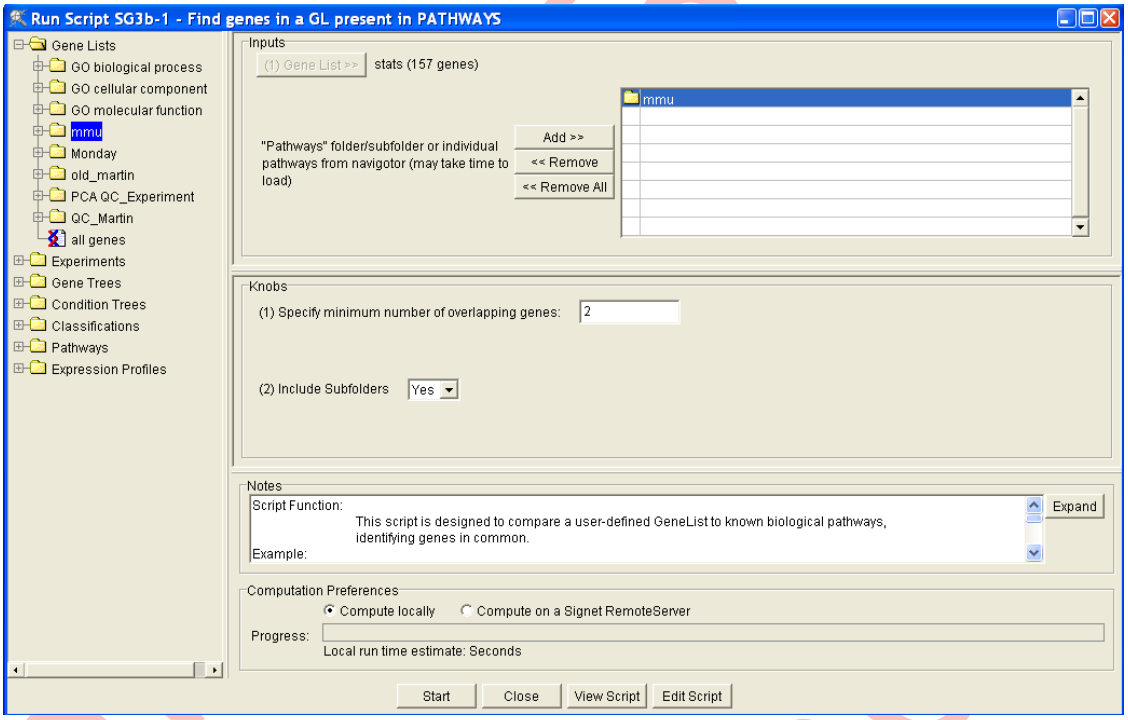
Molecular Function

Sorted stats and transcriptional repressor activity	
Sorted stats and hematopoietin or interferon-class (D200-domain) cytokine rec	
Sorted stats and transcriptional activator activity	0.0871
Sorted stats and chemokine activity	0.189
Sorted stats and transcription factor activity	0.199
Sorted stats and calmodulin binding	0.258
Sorted stats and interleukin receptor activity	0.315
Sorted stats and receptor activity	0.32
Sorted stats and oxidoreductase activity	0.32
Sorted stats and receptor binding	0.331
	0.36
Cellular Component	0.473

Sorted stats and adherens junction	
Sorted stats and spindle	
Sorted stats and integral to plasma membrane	0.0561
Sorted stats and intracellular	0.0871
Sorted stats and plasma membrane	0.282
Sorted stats and extracellular space	0.305
Sorted stats and mitochondrion	0.305
Sorted stats and transcription factor complex	0.318
	0.36
Biological Process	0.471

Sorted stats and B-cell differentiation	0.0301
Sorted stats and learning	0.0561
Sorted stats and negative regulation of cell cycle	0.0767
Sorted stats and metabolism	0.122
Sorted stats and homophilic cell adhesion	0.159
Sorted stats and inflammatory response	0.164
Sorted stats and positive regulation of transcription, DNA-dependei	0.199
Sorted stats and electron transport	0.239
Sorted stats and defense response	0.258
Sorted stats and development	0.288
Sorted stats and anti-apoptosis	0.32
Sorted stats and protein folding	0.36
Sorted stats and skeletal development	0.36
Sorted stats and organogenesis	0.363
Sorted stats and cell cycle	0.383
Sorted stats and chemotaxis	0.431
Sorted stats and intracellular signaling cascade	0.431
Sorted stats and cell surface receptor linked signal transduction	0.435
Sorted stats and regulation of cell growth	0.437
Sorted stats and transmembrane receptor protein tyrosine kinase s	0.473

KEGG Script and Result:



This script is simpler than the previous as we only have to set the minimum gene overlap, and like before I have set the overlap to 2.

KEGG

Genes overlapping with Pathways	
Sorted stats and Arginine and proline metabolism	0.0767
Sorted stats and Lysine degradation	0.122
Sorted stats and Apoptosis	0.619
Sorted stats and Nicotinate and nicotinamide metabolism	0.807

The genelists for the GO analysis and KEGG analysis can be found under the main Gene List folder within the folder Martin and then in the subfolder called QC. You can also visualize the KEGG pathways under the Pathway folder in the GeneSpring navigator.

This concludes the analysis of the data in GeneSpring.

I have also performed some addition analysis on this experiment by analyzing the genes that were clustered by the QT clustering tool. I extracted all the 17 clusters into excel with the genes common name, and then using the common name I ran the Agilent Literature search tool on each individual cluster set.

www.infogenbio.com **Tel +44(0)7525 064993**

Agilent Literature Search is a meta-search tool for automatically querying multiple text-based search engines (both public and proprietary) in order to aid biologists faced with the daunting task of manually searching and extracting associations among genes/proteins of interest.

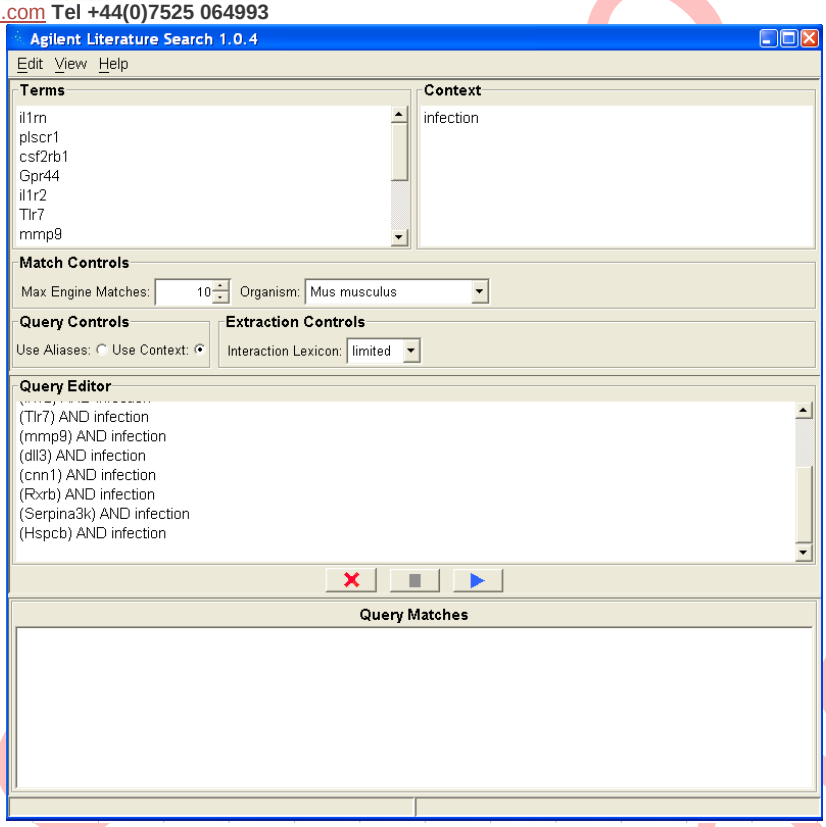
Agilent Literature Search can be used in conjunction with Cytoscape v2.1, which provides a means of generating an overview network view of gene/protein associations.

Agilent Literature Search Features

- Meta-search engine combining Information Retrieval (IR) & Knowledge Extraction (KE)
- uses PubMed, OMIM, and USPTO search engines
- User context-based symbol normalization
- Symbol identification, interaction extraction
- Putative network generation from literature

Agilent Literature Search provides an easy-to-use interface to its powerful search capabilities. A set of queries can be defined interactively, then submitted to multiple user-selected search engines. The retrieved results (documents) are fetched from their respective sources and each document is then parsed into sentences and analyzed for protein-protein associations. *Agilent Literature Search* uses a set of *lexicons* for defining protein names (and aliases) and association terms (verbs) of interest. An association is extracted for every sentence containing at least two protein names and one verb. Associations are then converted into *interactions*, which are further grouped into a network. The sentences and source hyperlinks for each association are further stored as attributes of the corresponding *interactions* .

Agilent Literature Search Window:



The terms I have used are the common names, in the example above I have used the genes clustered from Set2. I have also used a context, Infection, this helps in building associations between the genes used, I only ran these searches against PubMed.

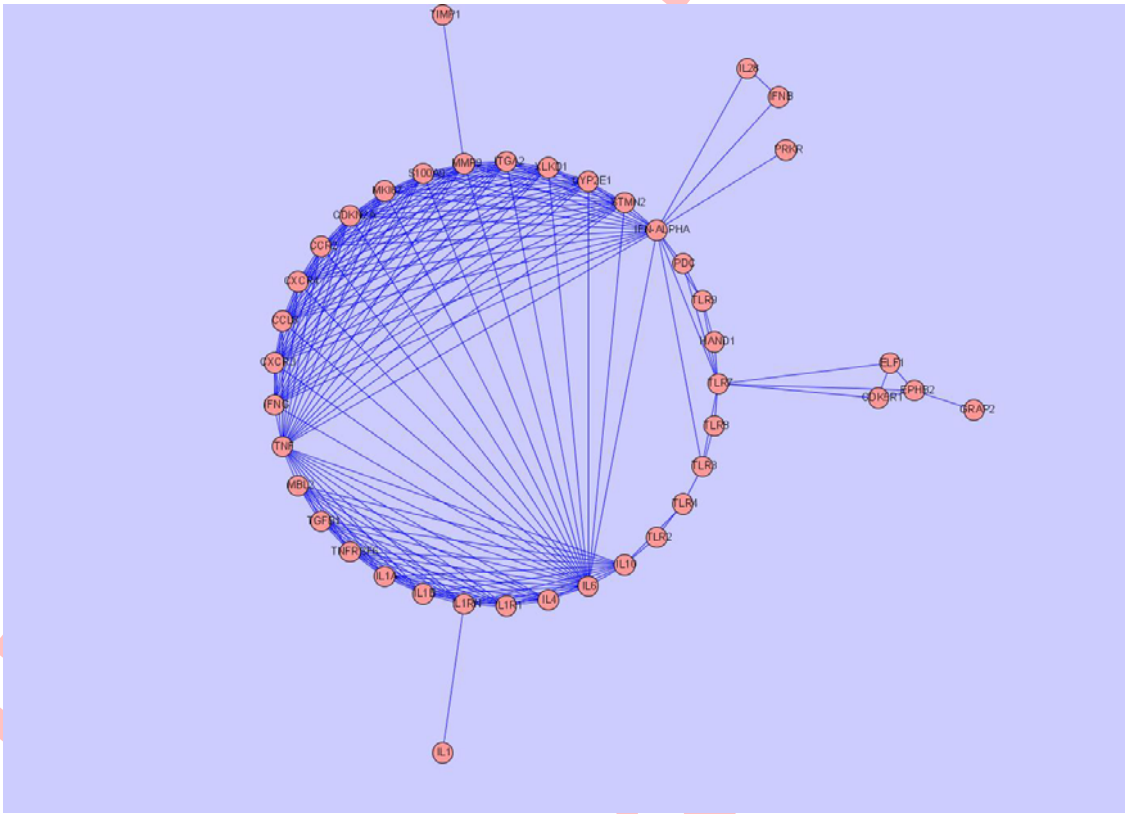
The outcome if this analysis is a list of abstracts linking the genes and the association and a linkage map of gene interaction (Cytoscape output). As I mentioned I have ran all 17 sets using this tool and I have generated the results as excel sets for each set. Below is an example of one result, see the folder Agilent Literature Search for the other result.

Set_2

mwginflmus#081	il1rn	interleukin 1 receptor antagonist; il1rn
mwginflmus#133	plscr1	phospholipid scramblase 1; plscr1
mwgmouse30K#A:03451	csf2rb1	colony stimulating factor 2 receptor, beta 1, low-affinity (granulocyte-macrophage)
mwgmouse30K#A:04150	Gpr44	G protein-coupled receptor 44
mwgmouse30K#A:08211	il1r2	Mus musculus interleukin 1 receptor, type II (Il1r2)
mwgmouse30K#B:01167	Tlr7	Toll-like receptor 7
mwgmouse30K#B:01590	mmp9	Mus musculus matrix metalloproteinase 9 (Mmp9)
mwgmouse30K#B:01655	dll3	delta-like 3 (Drosophila)

www.infogenbio.com **Tel +44(0)7525 064993**

mwgmouse30K#B:04125	cnn1	calponin 1
mwgmouse30K#B:06158	Rxrb	Retinoid X receptor beta
mwgmouse30K#B:07326	Serpina3k	Serine (or cysteine) proteinase inhibitor, clade A, member 3M
mwgmouse30K#B:07715	Hspcb	Heat shock protein 1, beta



Results

1. Polymorphisms in immunoregulatory genes and the risk of histologic chorioamnionitis in Caucasoid women: a case control study (by Annells MF,Hart PH,Mullighan CG,Heatley SL,Robinson JS,McDonald HM).

BACKGROUND: Chorioamnionitis is a common underlying cause of preterm birth (PTB). It is hypothesised...
Source: [PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15723707

2. Extremely high interleukin-6 blood levels and outcome in the critically ill are associated with tumor necrosis factor- and interleukin-1-related gene polymorphisms (by Watanabe E,Hirasawa H,Oda S,Matsuda K,Hatano M,Tokuhisa T).

OBJECTIVE: To determine the allelic frequencies of interleukin (IL)-6-, IL-1-, and tumor necrosis fa...

Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15644653

3. Polymorphism in intron 2 of the interleukin-1 receptor antagonist gene, local midtrimester cytokine response to vaginal flora, and subsequent preterm birth (by Genc MR,Onderdonk AB,Vardhana S,Delaney ML,Norwitz ER,Tuomala RE,Paraskevas LR,Witkin SS,MAP Study Group).

OBJECTIVE: This study investigated the association between polymorphism in intron 2 of the interleuk...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15507961

4. Cytokine and Chemokine Gene Polymorphisms Among Ethnically Diverse North Americans With HIV-1 Infection (by Wang C,Song W,Lobashevsky E,Wilson CM,Douglas SD,Mytilineos J,Schoenbaum EE,Tang J,Kaslow RA).

: Twenty-four common single nucleotide polymorphisms (SNPs) in 10 cytokine and chemokine genes were ...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15021309

5. Variation within genes encoding interleukin-1 and the interleukin-1 receptor antagonist influence the severity of meningococcal disease (by Read RC,Cannings C,Naylor SC,Timms JM,Maheswaran R,Borrow R,Kaczmarski EB,Duff GW).

BACKGROUND: Genetically determined variation in proinflammatory cytokine release influences severity...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12667023

6. Association analysis of polymorphisms at the interleukin-1 locus in essential hypertension (by Lin RC,Morris BJ).

Infection with microorganisms such as Helicobacter pylori and Chlamydia pneumoniae has been associat...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11840488

7. Influence of interleukin-1 receptor antagonist gene polymorphism on disease (by Witkin SS,Gerber S,Ledger WJ).

Interleukin-1 receptor antagonist (IL-1RA) is a naturally occurring competitive inhibitor of interle...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11740709

8. An interleukin-1 genotype is associated with fatal outcome of meningococcal disease (by Read RC,Camp NJ,di Giovine FS,Borrow R,Kaczmarski EB,Chaudhary AG,Fox AJ,Duff GW).

To determine whether known variants of the interleukin-1 (IL-1) and tumor necrosis factor (TNF) gene...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11023482

9. Resistance to human cytomegalovirus infection may be influenced by genetic polymorphisms of the tumour necrosis factor-alpha and interleukin-1 receptor antagonist genes (by Hurme M,Helminen M).

To examine whether there are genetic differences between cytomegalovirus (CMV)-seronegative and CMV-...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=10066041

10. IL-1 receptor antagonist (IL-1Ra) plasma levels are co-ordinately regulated by both IL-1Ra and IL-1beta genes (by Hurme M,Santtila S).

The genes in the IL-1 complex code for three proteins, IL-1alpha, IL-1beta and the IL-1 receptor ant...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=9710237

11. Toll-like receptor signaling inhibits hepatitis B virus replication in vivo (by Isogawa M,Robek MD,Furuichi Y,Chisari FV).

Toll-like receptors (TLR) play a key role in innate immunity. To examine the ability of diverse TLRs...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15890966

12. Inhibition of toll-like receptor 7- and 9-mediated alpha/beta interferon production in human plasmacytoid dendritic cells by respiratory syncytial virus and measles virus (by Schlender J,Hornung V,Finke S,Gunthner-Biller M,Marozin S,Brzozka K,Moghim S,Endres S,Hartmann G,Conzelmann KK).

Human plasmacytoid dendritic cells (PDC) are key sentinels alerting both innate and adaptive immune ...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15827165

13. IFN-alpha enhances TLR3-mediated antiviral cytokine expression in human endothelial and epithelial cells by up-regulating TLR3 expression (by Tissari J,Siren J,Meri S,Julkunen I,Matikainen S).

TLRs play a critical role in early innate immune response to virus infection. TLR3 together with TLR...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15778392

14. Disparate expression of IL-12 by SJL/J and B10.S macrophages during Theiler's virus infection is associated with activity of TLR7 and mitogen-activated protein kinases (by Petro TM).

Differences in components of innate anti-viral immune responses may account for the contrast in susc...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15777634

15. Identification and quantification of innate immune system mediators in human breast milk (by Armogida SA,Yannaras NM,Melton AL,Srivastava MD).

Breast-feeding decreases the risk of breast cancer in mothers and infection, allergy, and autoimmuni...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15603202

16. Plasmacytoid dendritic cell precursors/type I interferon-producing cells sense viral infection by Toll-like receptor (TLR) 7 and TLR9 (by Ito T,Wang YH,Liu YJ).

Plasmacytoid dendritic cell (pDC) precursors, also called type I IFN (alpha/beta/omega)-producing ce...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15592841

17. TLRs govern neutrophil activity in aspergillosis (by Bellocchio S,Moretti S,Perruccio K,Fallarino F,Bozza S,Montagnoli C,Mosci P,Lipford GB,Pitzurra L,Romani L).

Polymorphonuclear neutrophils (PMNs) are essential in initiation and execution of the acute inflamma...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15585866

18. Replication-dependent potent IFN-alpha induction in human plasmacytoid dendritic cells by a single-stranded RNA virus (by Hornung V,Schlender J,Guenthner-Biller M,Rothenfusser S,Endres S,Conzelmann KK,Hartmann G).

Plasmacytoid dendritic cells sense viral ssRNA or its degradation products via TLR7/8 and CpG motifs...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15528327

19. Ethanol suppresses cytokine responses induced through Toll-like receptors as well as

innate resistance to Escherichia coli in a mouse model for binge drinking (by Pruett SB,Zheng Q,Fan R,Matthews K,Schwab C).

Toll-like receptors (TLRs) recognize molecular patterns associated with pathogens and initiate vario...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15528012

20. Three different neutrophil subsets exhibited in mice with different susceptibilities to infection by methicillin-resistant Staphylococcus aureus (by Tsuda Y,Takahashi H,Kobayashi M,Hanafusa T,Herndon DN,Suzuki F).

Neutrophils (PMN) have been described as critical effector cells in the host's antibacterial innate ...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15308102

21. Overexpression of MMP9 and tissue factor in unstable carotid plaques associated with Chlamydia pneumoniae, inflammation, and apoptosis (by Stintzing S,Heuschmann P,Barbera L,Ocker M,Jung A,Kirchner T,Neureiter D).

Tissue remodeling by matrix metalloproteinases (MMPs) and plasminogen activators such as tissue fact...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15818457

22. [Connective tissue degradation in human amniochorion after stimulation with choriodecidual lymphocytes infected with group B streptococci] (by Estrada Gutierrez G,Reyes Trejo R,Maida Claros R,Beltran Montoya J,Vadillo Ortega F)

OBJECTIVE: To identify whether soluble products from choriodecidual blood cells stimulated with grou...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15813471

23. Upregulation of MMP-9/TIMP-1 enzymatic system in eosinophilic meningitis caused by Angiostrongylus cantonensis (by Chen KM,Lee HH,Chou HL,Liu JY,Tsai BC,Lai SC).

Proteolysis depends on the balance between the proteases and their inhibitors. Matrix metalloprotein...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15810979

24. Matrix metalloproteinase-9 expression is associated with the presence of Chlamydia pneumoniae in human coronary atherosclerotic plaques (by Arno G,Kaski JC,Smith DA,Akiyu JP,Hughes SE,Baboonian C).

OBJECTIVE: To investigate the association between Chlamydia pneumoniae and matrix metalloproteinase-...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15772219

25. Efficacy of albendazole-GM6001 co-therapy against Angiostrongylus cantonensis-induced meningitis in BALB/c mice (by Lai SC,Jiang ST,Chen KM,Hsu JD,Shyu LY,Lee HH).

Angiostrongylus cantonensis causes a form of parasitic meningitis in humans. Albendazole kills the n...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15716044

26. Matrix metalloproteinase-2 and -9 in the granulomatous fibrosis of rats infected with Angiostrongylus cantonensis (by Hsu LS,Lee HH,Chen KM,Chou HL,Lai SC).

The histomorphology of granuloma formation and gelatinase production were investigated in the brains...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15701257

27. Ultrastructural localization of matrix metalloproteinase-9 in eosinophils from the cerebrospinal fluid of mice with eosinophilic meningitis caused by Angiostrongylus cantonensis (by Tseng YK,Tu WC,Lee HH,Chen KM,Chou HL,Lai SC).

Matrix metalloproteinase-9 (MMP-9) has been implicated in the pathogenesis of eosinophilic meningiti...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15667715

28. Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection (by Bieche I,Asselah T,Laurendeau I,Vidaud D,Degot C,Paradis V,Bedossa P,Valla DC,Marcellin P,Vidaud M).

The molecular mechanisms of acute hepatitis C virus (HCV) infection, end-stage hepatitis (cirrhosis)...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15661146

29. Excessive matrix metalloproteinase-9 in the plasma of community-acquired pneumonia (by Yang SF,Chu SC,Chiang IC,Kuo WF,Chiou HL,Chou FP,Kuo WH,Hsieh YS).

BACKGROUND: It has been shown that matrix metalloproteinase-9 (MMP-9) is involved in the pathogenesi...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15653116

30. The use of real time rtPCR to quantify inflammatory mediator expression in leukocytes from patients with severe sepsis (by Kalkoff M,Cursons RT,Sleigh JW,Jacobson GM).

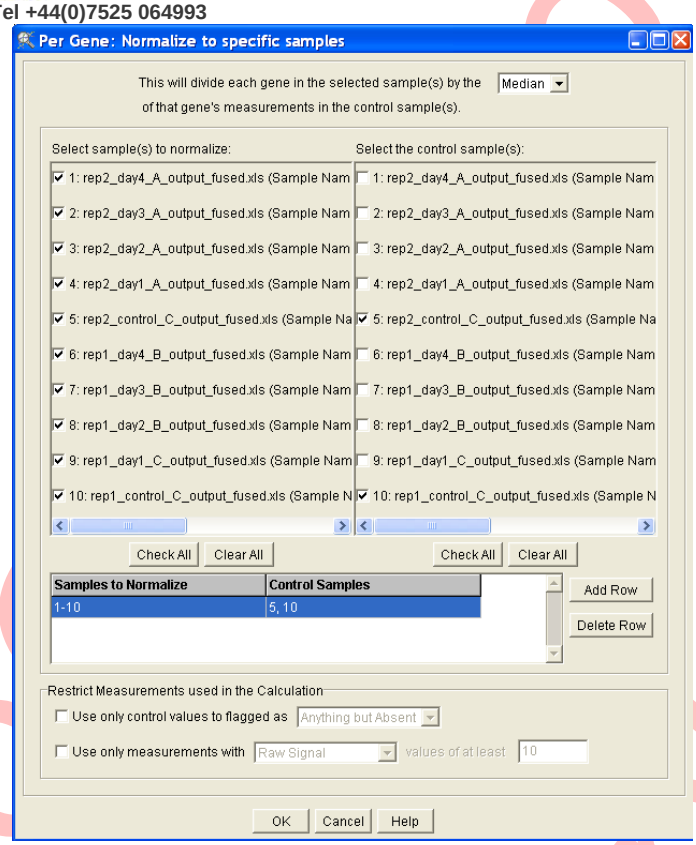
Real-time reverse transcriptase polymerase chain reaction (RT rtPCR) was used to quantify the patter...
Source:
[PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15648982

Analysis of Specific Time points referenced to the control samples

To refresh the memory all the time point have been normalized to the control samples using the following normalization protocol:

Normalisation:

A Lowess curve was fit to the log-intensity versus log-ratio plot. 35.0% of the data was used to calculate the Lowess fit at each point. This curve was used to adjust the control value for each measurement. If the control channel was lower than 10 then 10 was used instead. Specific samples were normalized to one another: sample(s) 1-10 were normalized against the median of the control sample(s) 5, 10. Each measurement for each gene in those specific samples was divided by the median of that gene's measurements in the corresponding control samples.



After the samples were normalized 2 interpretations were set up for the analysis protocol:

3. Default Interpretation (log data, ln log scale (e))
4. Ratio Interpretation

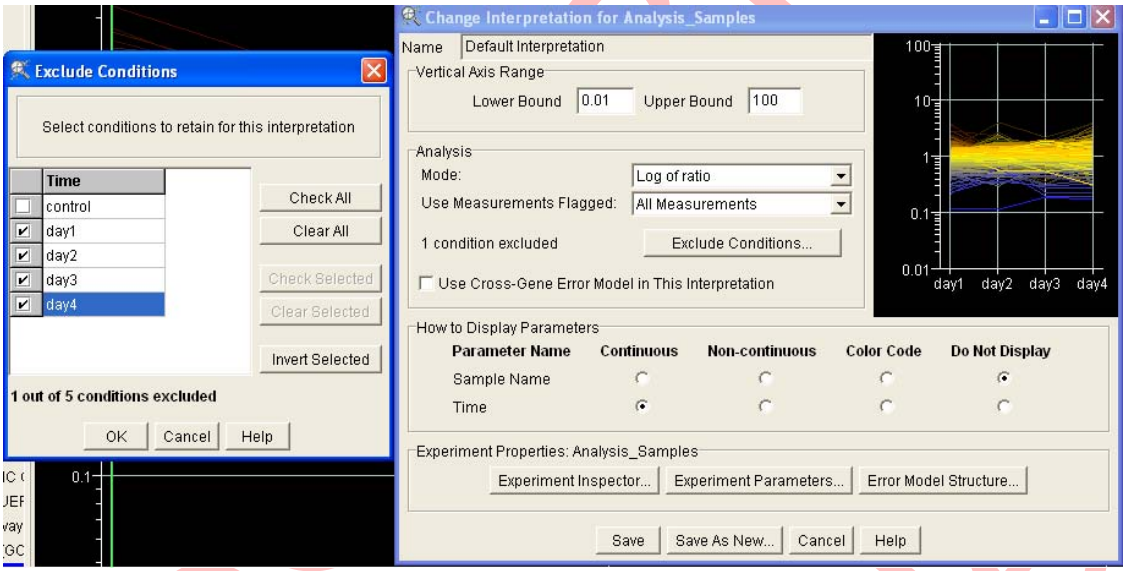
Both of these interpretations are identical they only differ due to the scale of the data. The ratio data is the non logged version of the logged data. In both interpretations the view of the data in GeneSpring has been altered to take into effect that the samples have been normalized to the control samples. The control condition in the interpretations have been removed, this is purely visual as the data on the controls contribute to the ratios of the informative time points.

See Experiment>Analysis>Analysis_Samples> For the Interpretations

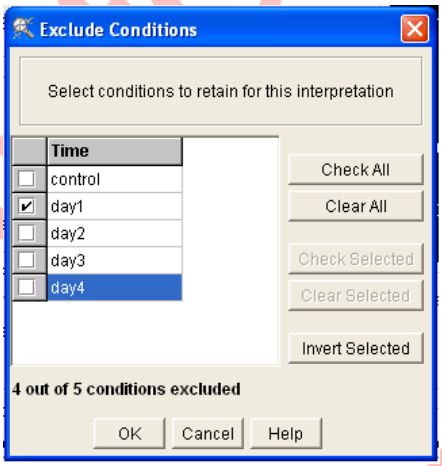
The specific asked was to find the genes which are significantly change at each time referenced to the control sample. As the samples have been normalized to the control sample this analysis is easily performed with the Filter on Confidence tool within GeneSpring.

The interpretation selected for this analysis is the log interpretation as we are performing a statistically test using a parametric analysis. In this study the log interpretation is the default interpretation, before this interpretation can be used the conditions visible have to be altered as this test is only asking to find the significant genes at each time point, independently. Whereas the default interpretation visualizes all the time points, so the interpretation needs to be modified.

To modify the default interpretation right click and select inspect and the Change Interpretation window appears. To restrict the interpretation to just 1 time point click the “Exclude Condition” button. Once you have done this you will get the image below appear:



To unselect Time Points just uncheck the relevant time points:



In example above the default interpretation will just represent time point 1 day, once you have selected your time point just press OK and then press SAVE.

Filter on Confidence

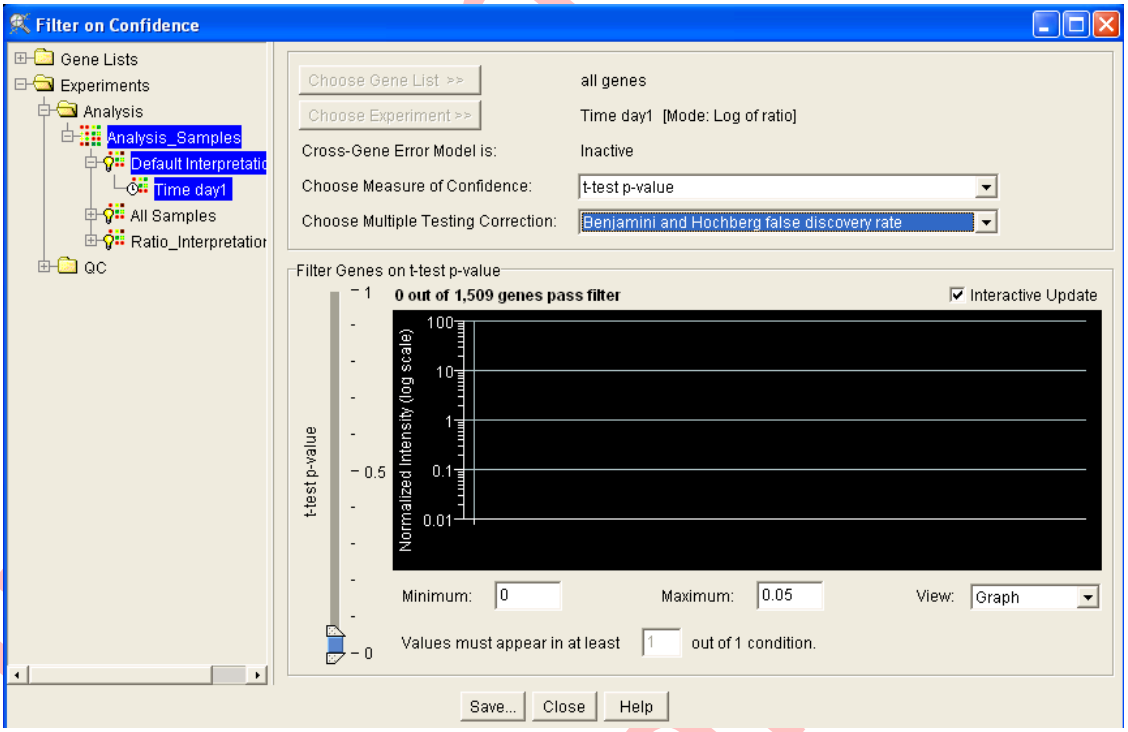
Filter>Filter on Confidence

The gene list selected for this analysis was the “All Genes” genes list as this test will find all the genes are significant changing from 1. This test evaluates whether a gene is differentially expressed for a condition. Asks “Is mean intensity statistically different from 1 (0 for log mode)?, the filter uses 1-sample t-test.

$$t = \frac{\overline{X} - 1}{\frac{s_X}{\sqrt{n}}}$$

The all genes gene list was selected, as quality controlling the gene prior to this test would no affect the outcome of this test.

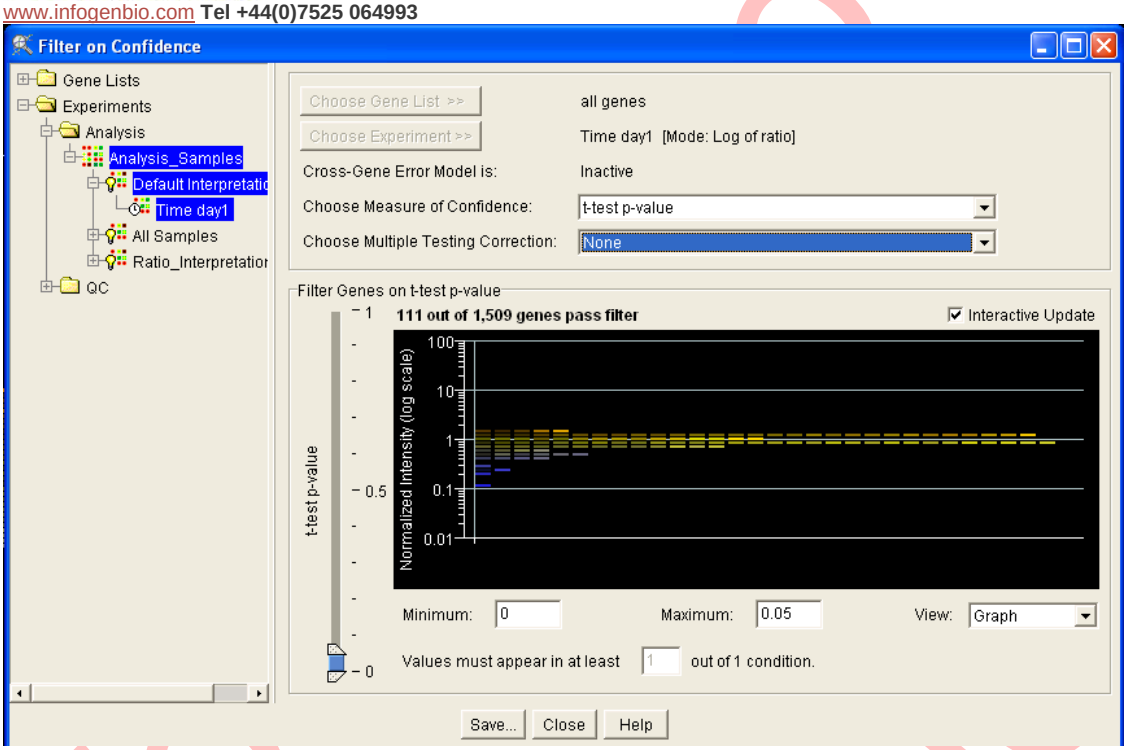
All Genes, p0.05 and Benjamini & Hochberg FDR



The example above displays the settings for all genes at day 1 statistical filter with a p-value cutoff at 0.05 and FDR multiple testing correction, the outcome of this test was zero genes.

When this filter was applied to quality controlled genes the same outcome was found, zero genes. Therefore the multiple testing correction was removed from the filter.

All Genes, p0.05 and no multiple testing correction



The example above displays the settings for all genes at day 1 statistical filter with a p-value cutoff at 0.05, the outcome of this test was 111 genes. These genes were saved as d1_0.05_noMTC and can found in the Day 1 folder under the main gene list folder in GeneSpring.

The same settings were used for all the subsequent time points day_2, day_3, and day_4. As previously stated the default interpretation was modified in order that the test was only applied to one time point.

Gene List	No of Genes in List
d1_0.05_noMTC	111
d2_0.05_noMTC	71
d3_0.05_noMTC	91
d4_0.05_noMTC	78

In a quick analysis there is no gene that is found in all four gene lists. In the first three gene list there were only 2 genes in common

1110019C06Rik	RIKEN cDNA 1110019C06 gene
Col1a2	Procollagen, type I, alpha 2

In the last three gene list there was only 1 common gene, Gdf9 (Growth Differentiation Factor 9), all four gene list seem contain unique sets of genes.

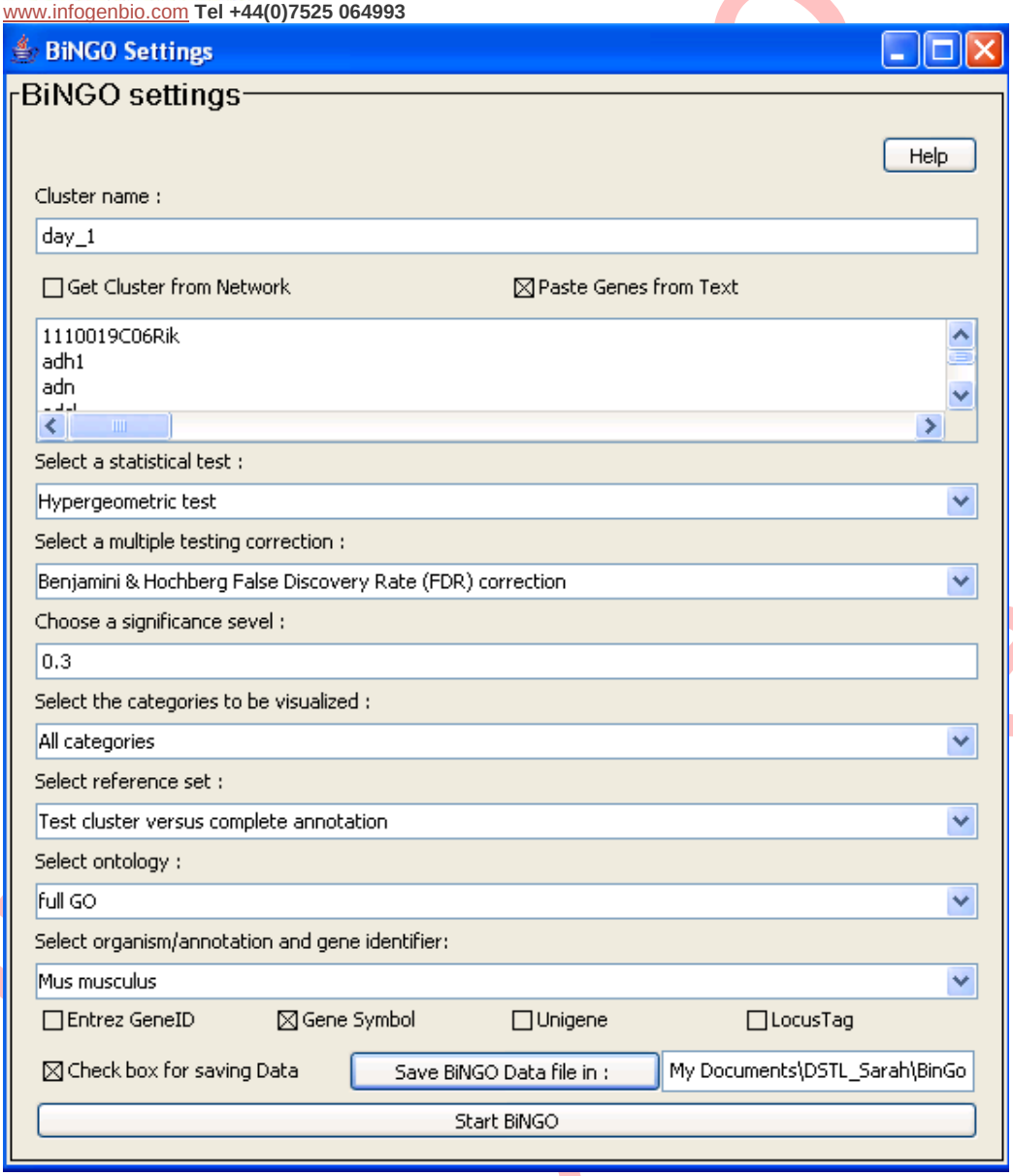
www.infogenbio.com **Tel +44(0)7525 064993**

Ontology Comparison of Significant Genes

Once the statistical gene are found the next step is to classify these genes, this can be done in 2 ways by expression (clustering algorithms) or by ontology (know biological information, Gene Ontology Consortium (GO) classifications). The next step in this analysis was to perform the ontology analysis on the statistically significant genes, for all the four days. This analysis was performed using a piece of software called BinGO run within a software called Cytoscape. More information about BinGO can be found at http://www.psb.ugent.be/cbd/papers/BiNGO and more information on Cytoscape can be found at http://www.cytoscape.org

The significant genes are removed from GeneSpring (right click on the gene list, select inspect, the Gene Inspector appears, press the Configure columns button and then check the Common names option. Then copy to Clipboard and then paste into EXCEL) and then the common names for these gene are then used within BinGO to perform the ontology comparison with the GO classifications.

www.infogenbio.com **Tel +44(0)7525 064993**



The image above sets out the options selected for the analysis performed on both significant gene lists. Both gene lists were compared to the Full GO ontology for Mus musculus, the outcome of this analysis are file that can be opened by EXCEL.

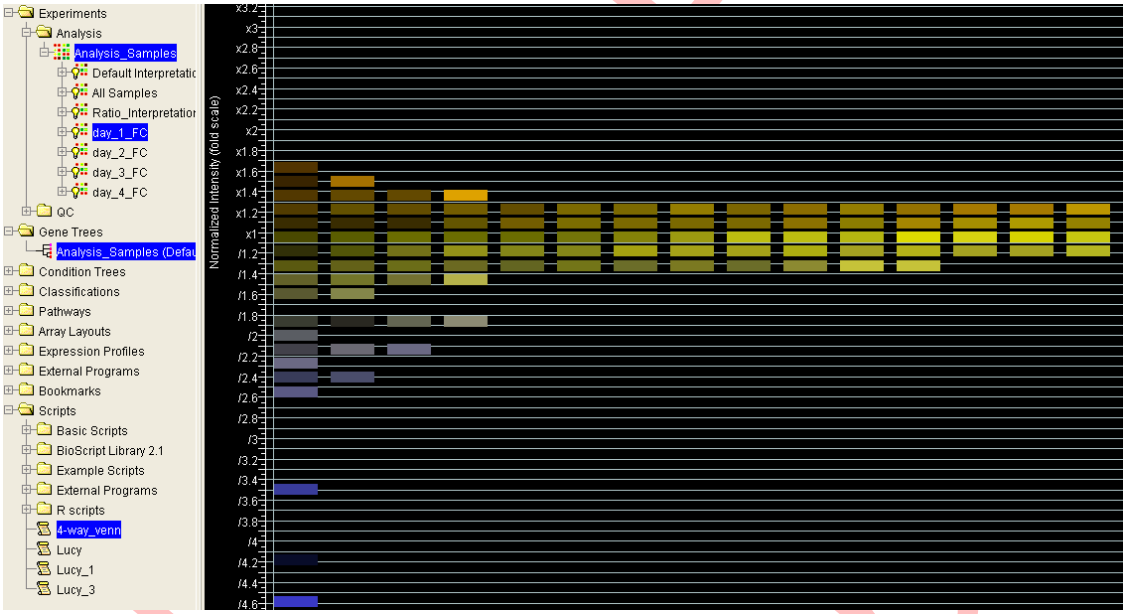
BinGo\day_1.bgo (link to file)

BinGo\day_2.bgo

BinGo\day_3.bgo

Fold Change analysis of Time Point to Control

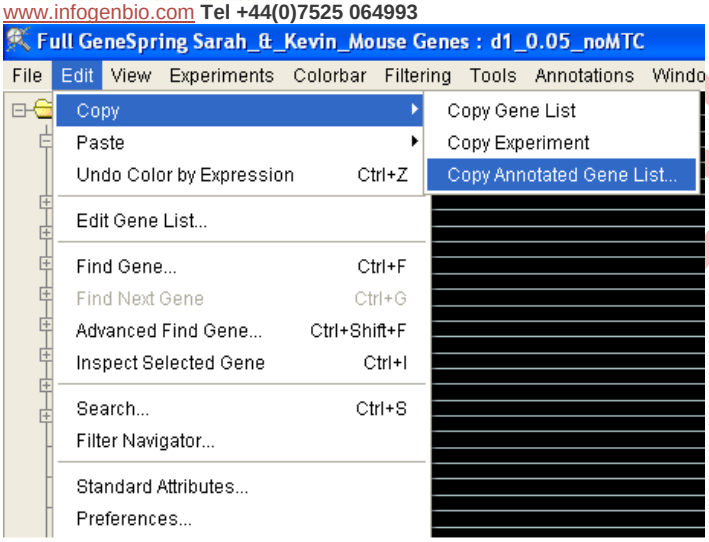
The next step in analysis was to look at the fold change within the significant genes, this is simply performed by saving a Fold Change Interpretation for each of the specific days.



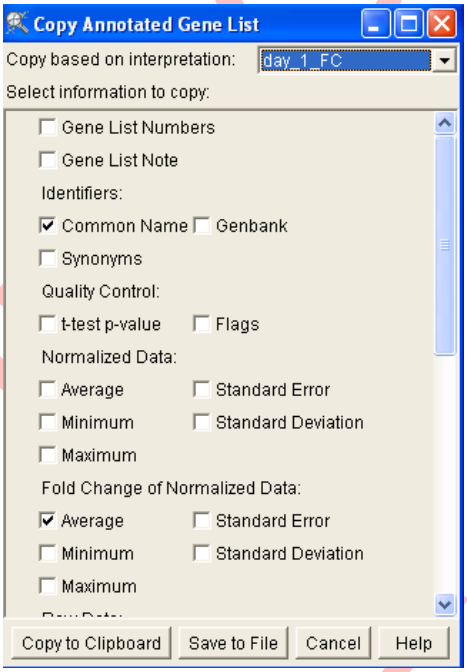
The example above is the fold change interpretation for day 1, select the day_1 genes from the gene list and then select the day_1_FC interpretation. On the vertical axis you can see the difference in fold change for each of these genes. This difference in expression is specific to the control.

Once you are in the Fold Change interpretation you can export the fold change differences by using the copy annotated gene list option:

www.infogenbio.com **Tel +44(0)7525 064993**



Once you select this option you will see this window:



Just select the average of the “Fold Change of Normalized Data” and unselect everything else except the Common names, after you have done this just press Copy to Clipboard, open up EXCEL and then just paste the gene list into the spreadsheet.

www.infogenbio.com **Tel +44(0)7525 064993**

The table below is just the first few genes of the day_1 significant gene list, to convert the numbers from GeneSpring fold change output to meaningful fold change just add 1 to the number from GeneSpring. Numbers which begin with – are under expressed, so -1.506 actually means the gene is under expressed at day 1 by 2.5 times relative to the control.

0.29 means the gene is over expressed at day 1 by 1.29 times relative to the control

Day_1 genes	Output FC		Add 1 to the value
mwgmouse30K#A:01045	-1.50601	1110019C06Rik	down X_2.5
mwgmouse30K#A:08104	-0.08848	adh1	down_X_1.09
mwgmouse30K#A:06244	-0.32721	adn	down_X_1.33
mwgmouse30K#A:11781	-0.85795	adsl	down_X_1.85
mwgmouse30K#A:04005	-0.60239	ahsg	down_X_1.60
mwgmouse30K#A:03045	-0.13642	Ap1s2; EST1; 1500012A13Rik	
mwgmouse30K#A:07496	-0.08658	ap2m1	Down_X_1.08
mwgmouse30K#A:04408	0.163458	areg	up_X_1.16
mwgmouse30K#A:03903	-0.4275	asgr1	Down_X_1.42
mwgmouse30K#B:05064	-0.16915	atpi	Down_X_1.16
mwgmouse30K#A:04257	0.157951	Bambi	Up_X_1.157
mwgmouse30K#A:10683	0.292698	Birc3	up_X_1.29
mwgmouse30K#A:07930	-4.56687	Bmp6	Down_X_5.56
mwgmouse30K#B:00848	-0.35871	Casp7	Down_X_1.36

This analysis can be done for all the other time points, just remember to select the right interpretation.