Interferome Help

Contents

Introduction

The Interferome v2.0 is an online database that enables the user to search a curated database for experimental evidence of interferon (IFN) regulation. The user submits a list of candidate genes and the database returns experimental evidence of their IFN regulation. The result set can be submitted to a range of secondary analyses to provide further biological insight into the interferon regulated gene sets. Analyses available include gene ontology analysis, transcription factor analysis, chromosomal location, and basal expression. Users are advised that the database is based on available evidence that any particular gene is regulated by IFN; the absence of an observed relationship in the Interferome should not be taken as evidence that the gene in question is never IFN regulated.

The Inteferome v2.0 is a substantial upgrade from the original Interferome database (Samarajiwa et al, 2008) and includes data from a substantially greater number of experiments. The original Interferome catalogued genes that were described in scientific literature as being interferon regulated; the bioinformatics and statistical method used to identify interferon regulated genes varied substantially between experiments. In contrast, the Interferome v2.0 utilises MIAME compliant microarray datasets which have been processed through a standard bioinformatics pipeline. Differences in the number of datasets and the statistical methods employed might lead to occasional discrepancies between the two Interferome databases, with genes being identified as being interferon regulated in one version of the database but not in the other.

Access:

The Interferome is available as an online resource at
http://Interferome.its.monash.edu.au/Interferome/home.jspx

Starting the Interferome

Opening the Interferome brings users to the *home* page (Figure 1: The Interferome home page.Figure 1).

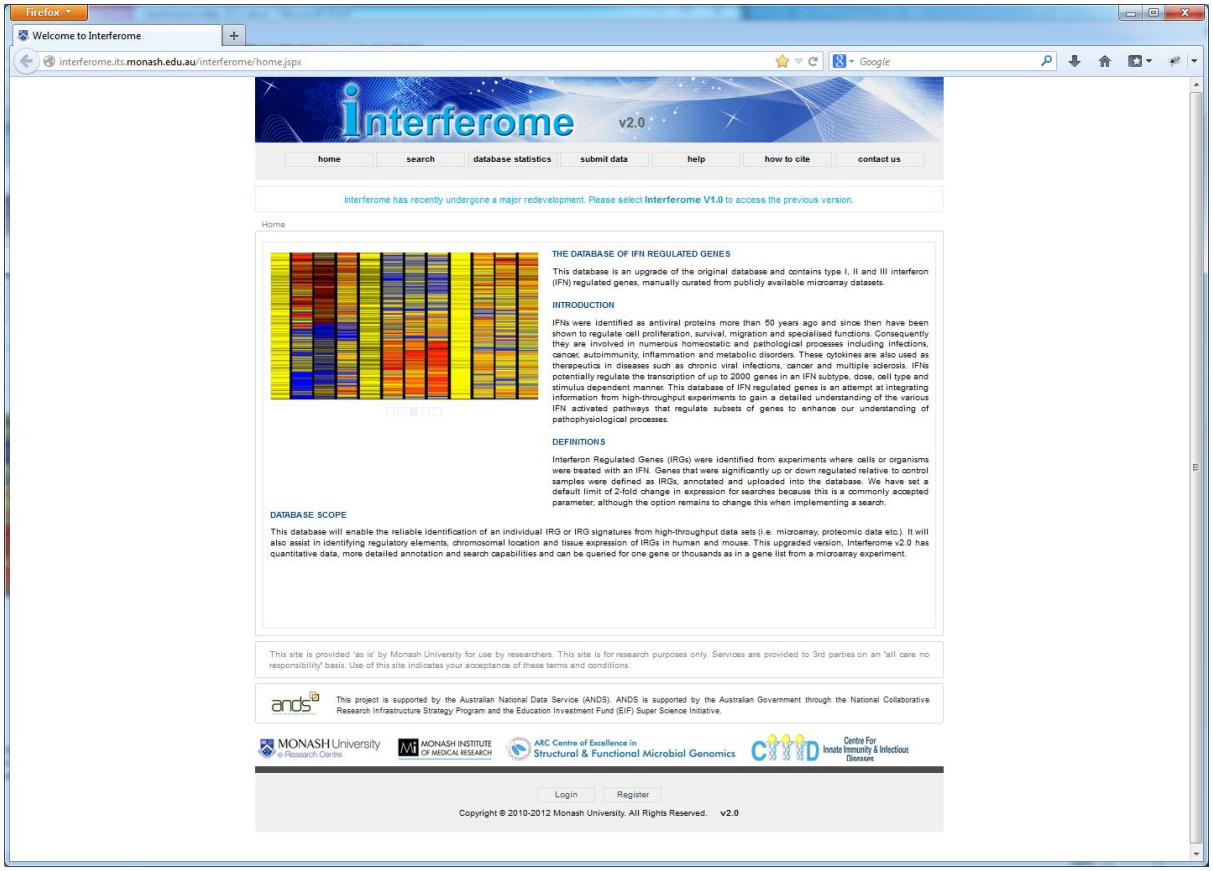


Figure 1: The Interferome home page.

Using the Interferome.

The *home* page includes some background details for the Interferome database. Across the top of the *home* page are several buttons that allow the user to start various activities. These buttons are present on all pages in the Interferome, and the user can use these buttons to start any processes from any location within the Interferome. The functions associated with each button are reviewed as follows:

1. **Home:** This button returns the user to the home page from any position within the Interferome.
2. **Search:** Clicking on this button opens up the search page (Figure 2). The search page button allows the user to start a search for genes that are regulated by interferon and, for most users, will represent the most used function. On this page the user will submit a list of genes to be queried for evidence of interferon regulation. The user can also narrow the search selecting a range of experimental conditions under which interferon regulation can occur. In the upper part of the page the user selects a range of biological conditions to be used for the search. In the lower part of the page the user specifies one or more genes that are to be searched for interferon regulation.
    a. ***Selecting experimental conditions***:
        i. **Interferon type and sub-type:** the user is presented with an option to search for responses from any type of interferon (default option) or to restrict the search to responses from type I, II or III. An associated option allows the user to narrow the search to a particular Interferon subtype. If the user selects Interferon type I they have the option of selecting the subtype IFNalpha, IFNbeta, or all (default). Selecting Interferon type II automatically restricts the subtype to IFNgamma, while selecting interferon type III restricts the search to IFNlambda. Selecting "any" from the Interferon type list also restricts the subtype options to "any".
        ii. **Treatment concentration:** The user is able to select data from any concentration range (default), or restrict responses to a defined window of concentrations. The user must submit the minimum and maximum concentrations in International Units per ml (IU/ml).
        iii. **Treatment time:** The user can define window of time after interferon treatment in which genes are observed to respond. The user can select any time period (default), or specify a maximum and minimum number of hours after the interferon treatment was applied.
        iv. **In vivo/in vitro:** Users can use a pull-down box to restrict their search to experiments conducted *in vivo*, *in vitro*, or both.
        v. **Species:** Users can use a pull-down box to restrict their search to experiments conducted in human, mouse, or both species.
        vi. **Biological system:** Users can use a pull-down box to restrict their search to a specific biological system. Options currently include Nervous System, Connective Tissue, Haemopoietic/Immune, Respiratory System, Gastrointestinal Tract, Urogenital/Reproductive, or all systems. The range of systems available is expected to increase as more datasets are entered into the Interferome.
        vii. **Organ**: Users can use a menu to restrict their search to experiments conducted on specific organs. Options currently include fetal brain, spleen, embryo, lung, liver, blood, umbilical vein, brain, ovary, or all organs.. The range of organs

available is expected to increase as more datasets are entered into the Interferome.

viii. **Cell type:** Users can use a menu to restrict their search to experiments conducted on specific cell types. Options include microglial cell, fibroblast, CD8+ DC, alveolar basal epithelial cell, hepatocytes, epithelium, blood monocyte derived macrophages, neurones, primary hippocampal neuron, whole tissue, and monocyte derived dendritic cells.

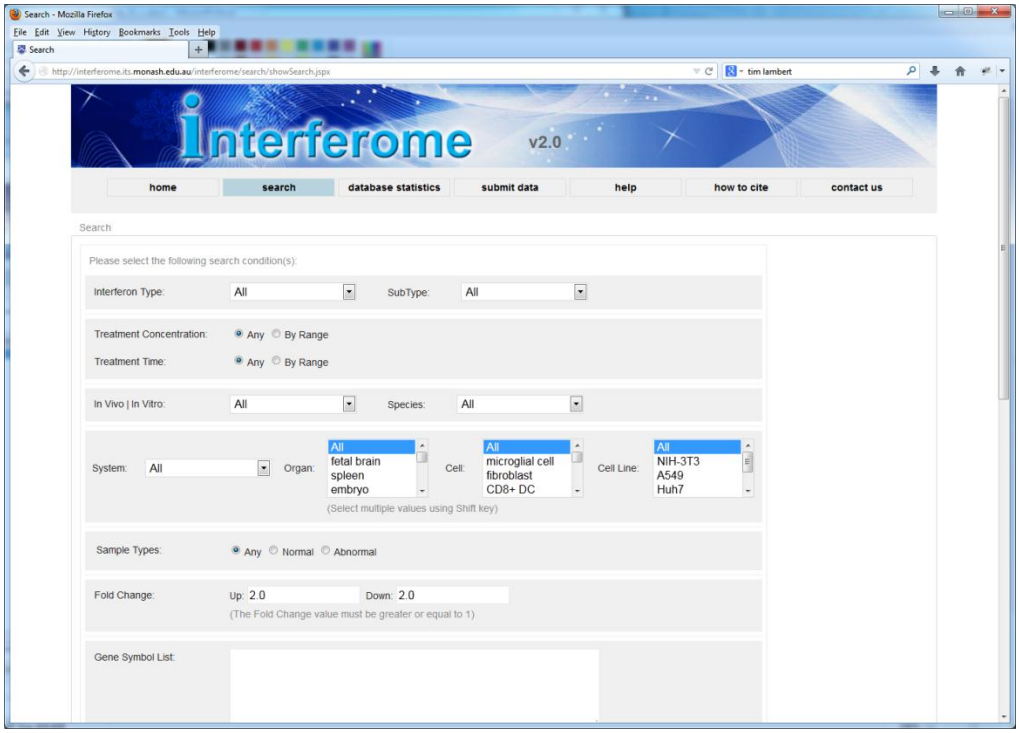
ix. **Cell Line:** Users can restrict their search to selected cell lines. Options include NIH-3T3, A549, Huh7 and BE(2)-C.

x. **Normal/abnormal:** Users can restrict the search to normal or abnormal conditions term "abnormal" represents a genetic variant, a disease condition, or a specific pre-treatment prior to stimulation with interferon.

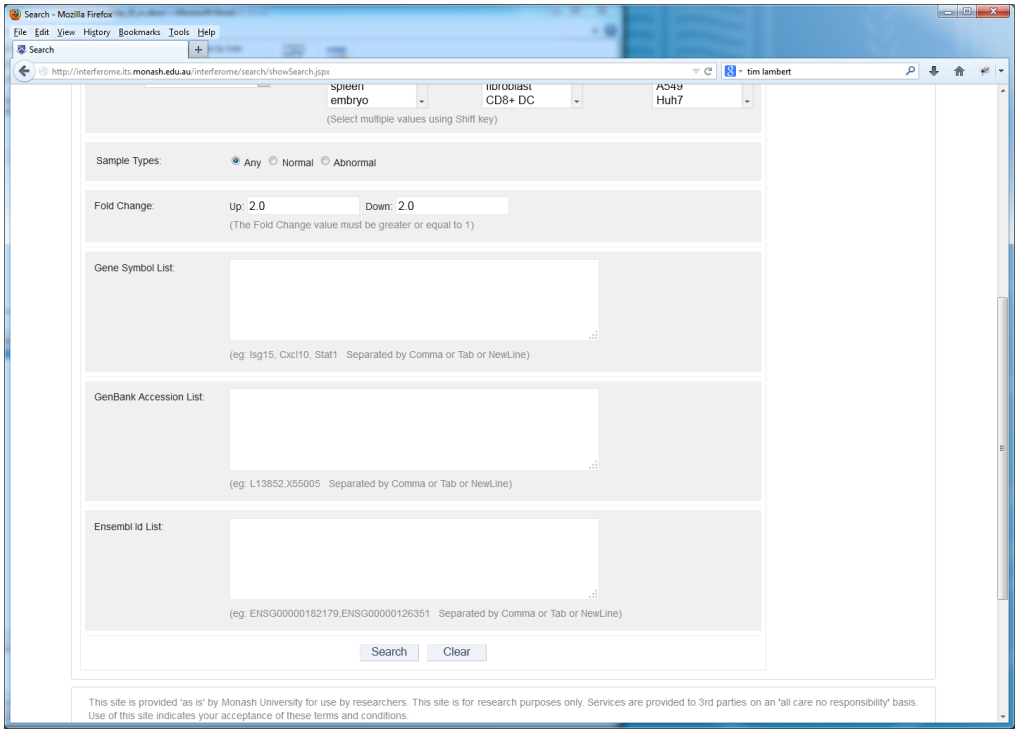
xi. **Fold change:** The user may search for genes that show responses above a specified threshold level. The default threshold of 2 fold change is set to select genes that show an interferon response that is equal to or exceeds fold up- or down-regulation. Advanced users might opt to raise the threshold to restrict the search to genes that show high changes in expression following interferon stimulation, and might be associated with major biological effects. Alternatively, users may reduce the thresholds to include genes that show low levels of response to interferon. While all gene included in the database show a statistically significant ($p<0.05$) response to interferon, **users are cautioned that no corrections for false discovery rate have been included**, and that selecting a low threshold value might increase the chance of returning false positives along with genes that have little biological effect following stimulation with interferon.

xii. **Genes:** At the bottom of the search page the user must specify which genes to query for interferon responses. The user can identify genes using gene symbol, Genbank identifiers, Ensembl identifiers, or any combination of these terms. Ensembl identifiers have the advantage that they are a curated and non-redundant description of genes within the two organisms.

b. ***Executing the search.*** The user must click on the "search" button at the foot of the page once all the search conditions have been completed. The results are then displayed on the "Gene summary" page .



2.a.



2.b.

Figure 2. The upper (2a) and lower (2b) sectionsof the search page.

Search Results

a. ***Gene summary:*** The gene summary page (Figure 3) summarises the results from the gene search. A sentence at the top of the page , identifies how many genes were

retrieved from the search, the number of search terms (Ensembl, Genbank or gene symbol terms) that were used, and the species that were searched, as defined on the search page. Below is a table that summarise genes returned. Columns include the Ensembl identifier, the gene symbol, a description of the gene, Entrez, Genbank and unigene identifiers. The Ensembl, Entrez and Genbank identifiers are links to external databases where further information is available for each gene. The user may choose to submit the data for downstream analyses within the Interferome. The downstream analyses may be performed in any order through a number of buttons at the top of the page. Each analysis option will be discussed subsequently. Alternatively, the user may download the list of genes as a text file.

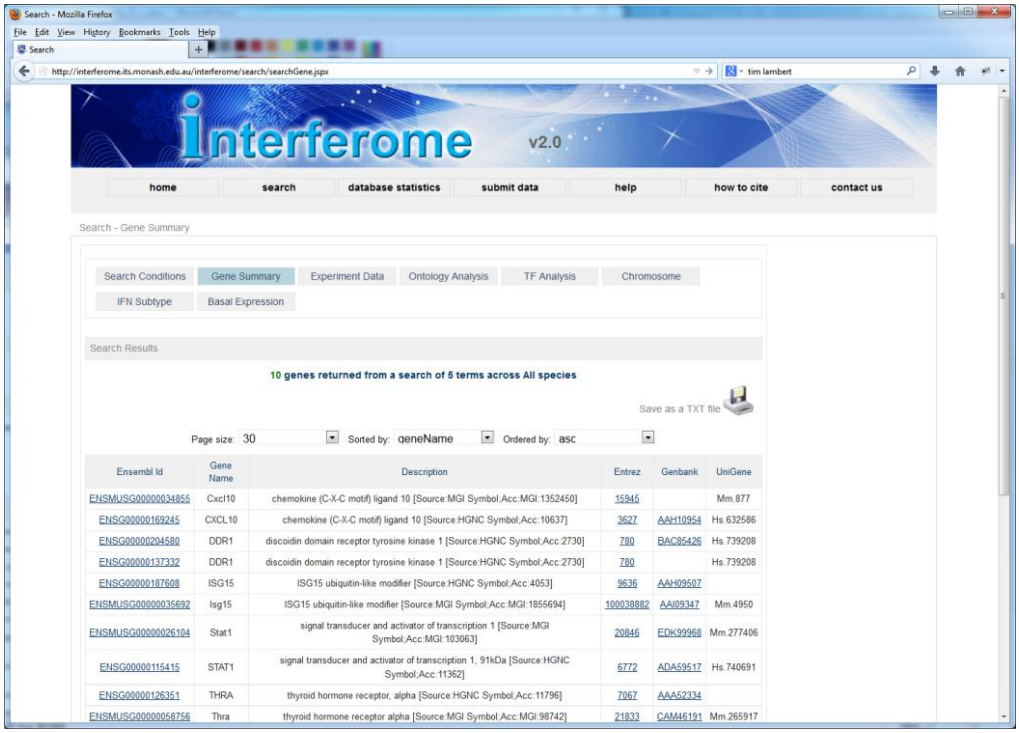


Figure 3. The Gene Summary page that summarises the genes that are retrieved from the Interferome using the conditions specified on the search page.

b. ***Search Conditions:*** This option returns the user to the search page (described above) where they are able to make changes to the search criteria used.

c. ***Experimental data:*** This page presents a summary of all the probes that show statistically significant responses to the specified search conditions (Figure 4). The table presents each significant response from a microarray probe on a separate line. The first column presents the dataset from the Interferome from which the response was obtained. This data is presented in the form of a link, clicking on which will bring up detailed annotations of the experiment performed. (Figure 5).

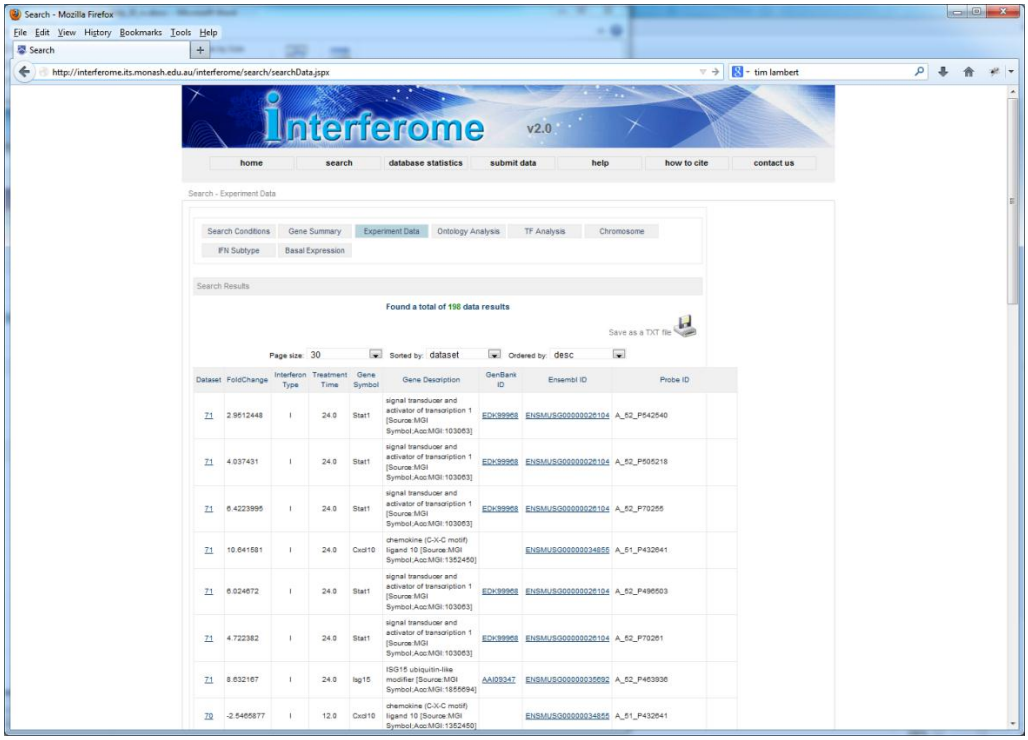


Figure 4. The Experiment Data page which summarises all the genes that respond to the terms specified in the search conditions.

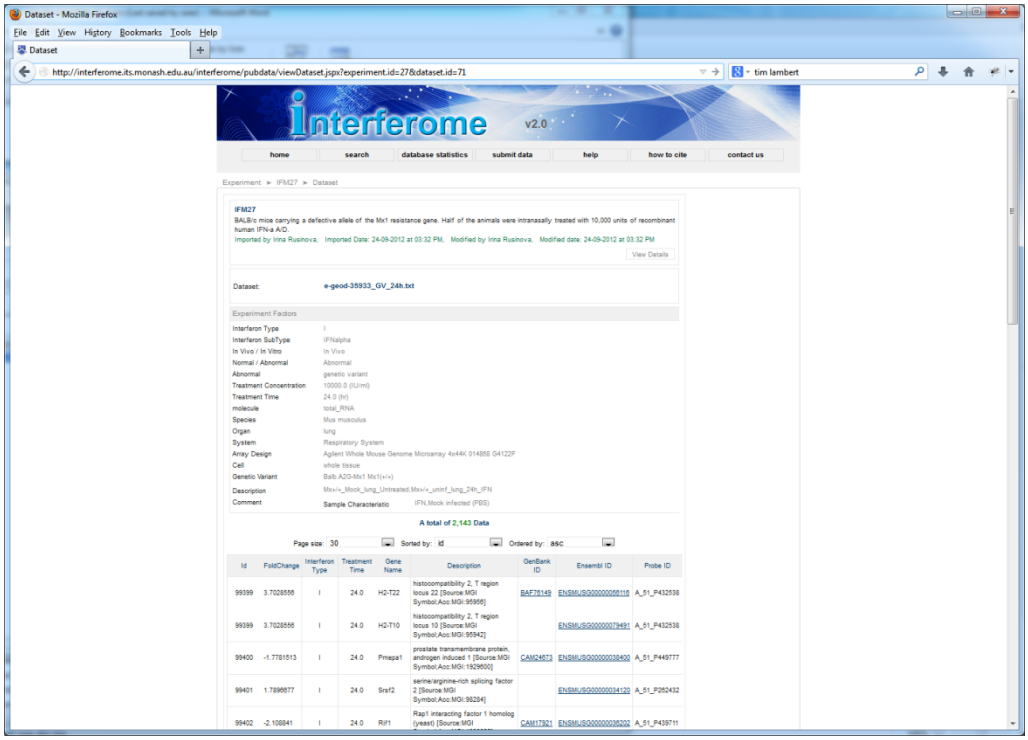


Figure 5. The Dataset page. Clicking on the "dataset" link on the "experiment page" opens the dataset page which summarises experiment and treatment conditions using MIAME compliant standards.

d. ***Ontology analysis:*** This analysis summarises the gene ontology (GO) terms that are represented in the gene list (**Figure 6**). Ontological results are presented separately for molecular function, cellular component and biological process. Results are presented in the form of a word cloud that shows the most highly represented ontologies present, and a table that presents all ontologies represented. The ontologies in the word cloud present a link which, when clicked, opens a page from the AMIGO website that summarises that ontological term. The table presents each ontological term, ranked in decreasing order of their numerical representation in the result set (the most abundant at the top of the table). The first column presents the unique GO identifier for that term; the second presents the GO terms name, and the third a brief description. The fourth column presents the number of times that term is represented in the result set. The final column presents a test (hypergeometric mean) for the statistical significance of the result. There is an option to save the results of the ontology analysis as a text file.

Figure 6. The summary of the Gene Ontology (GO) terms represented in the result set.

e. ***Transcription factor (TF) analysis:*** This analysis presents a linear representation of 1500 base pairs of sequence upstream (5') of the transcription start site of the IRG, with coloured blocks depicting the predicted transcription factor (TF)-binding elements in the promoter (Figure 7). Predictions are based on the MATCH algorithm using TRANSFAC 2012 professional matrices applying minimum false positive cut-off (http://www.biobase-international.com/product/transcription-factor-binding-sites). Holding the mouse cursor over each coloured block presents details of the associated transcription factor. There is an option for the user to download the data as a text file, and to save the image in svg format. See the section on *saving images in SVG format* for more details on this function.

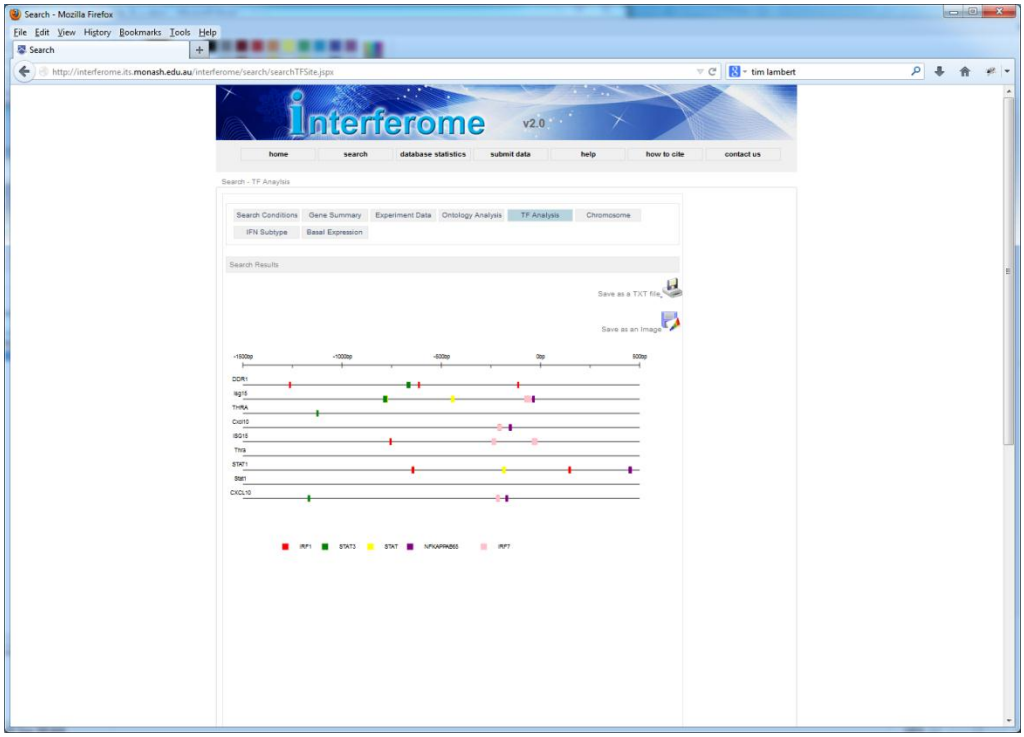


Figure 7. The transcription factor (TF) page. Each line represents 1500 bases upstream of a gene from the result set. The coloured boxes represent the location of a transcription factor binding site predicted by TRANSFAC2012.

f. ***Chromosome:*** This analysis presents a drawing of the human or mouse chromosomes with the location of the genes in the result set plotted in red (Figure 8). Images presented depend on the search conditions; images for both human and mouse are presented when both species are selected in the search, but only single images (human or mouse) are presented if only one species is selected for the search. The user has the option to save the data as a text file, or to download the image in svg format. See the section on *saving images in SVG format* for more details on this function.

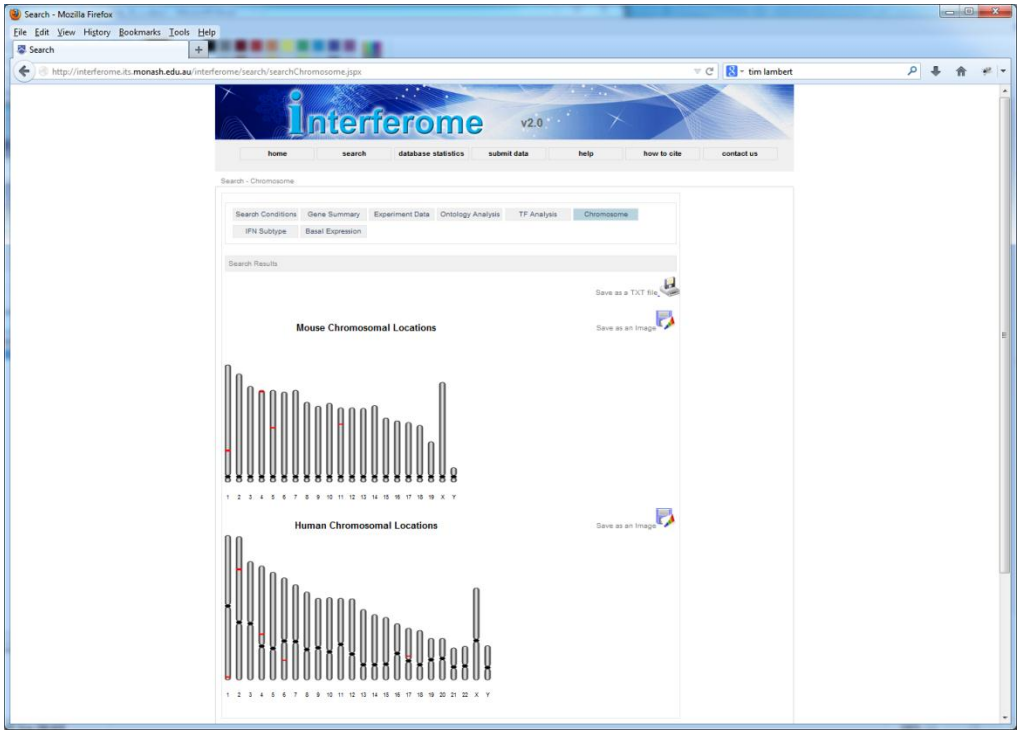


Figure 8. The chromosome image, showing the location of the interferon regulated genes on the chromosomes of the species selected for the search.

g. ***IFN subtype:*** This analysis displays a Venn diagram representing how many genes are regulated by each type of IFN and the overlaps of genes regulated by two or all three IFN types (Figure 9). It should be noted that there are far fewer data sets of genes regulated by type II or type III IFNs and this may introduce bias or false negatives in regard to genes regulated by these subtypes. Thus, caution is encouraged in the interpretation of negative results from this analysis. The user has the option to save the data as a text file, or to download the image in svg format. See the section on *saving images in svg format* for more details on this function.

Figure 9. A venn diagram showing how many genes in the result set are regulated by type I, II or III interferons, either individually or in combination

h. ***Basal expression:*** this analysis displays a heat map of the expression of IRGs in their unstimulated state, across various tissues and cells (Figure 10). The human and mouse expression data were obtained from the tissues and cell lines data in the BioGPS portal (http://biogps.org). Genes from the Interferome result set are linked to the BioGPS data through gene symbol, and the Basal Expression allows the user to peruse the expression of every gene symbol in unstimulated tissues from all species that were searched. For example, if a search of both species found IFN regulation of gene *xxx* in human but not mouse, this analysis would still allow the user to inspect the unstimulated expression of that gene in both species. A blue-red colour scheme is used to present the intensity of the basal expression for each gene, with red indicating high expression and blue low expression. Genes with a pale or near white colour indicates genes that have an expression close to the median value for that tissue type. The user has the option to save the data as a text file, or to download the image in svg format. See the section *on saving images in svg format* for more details on this function.

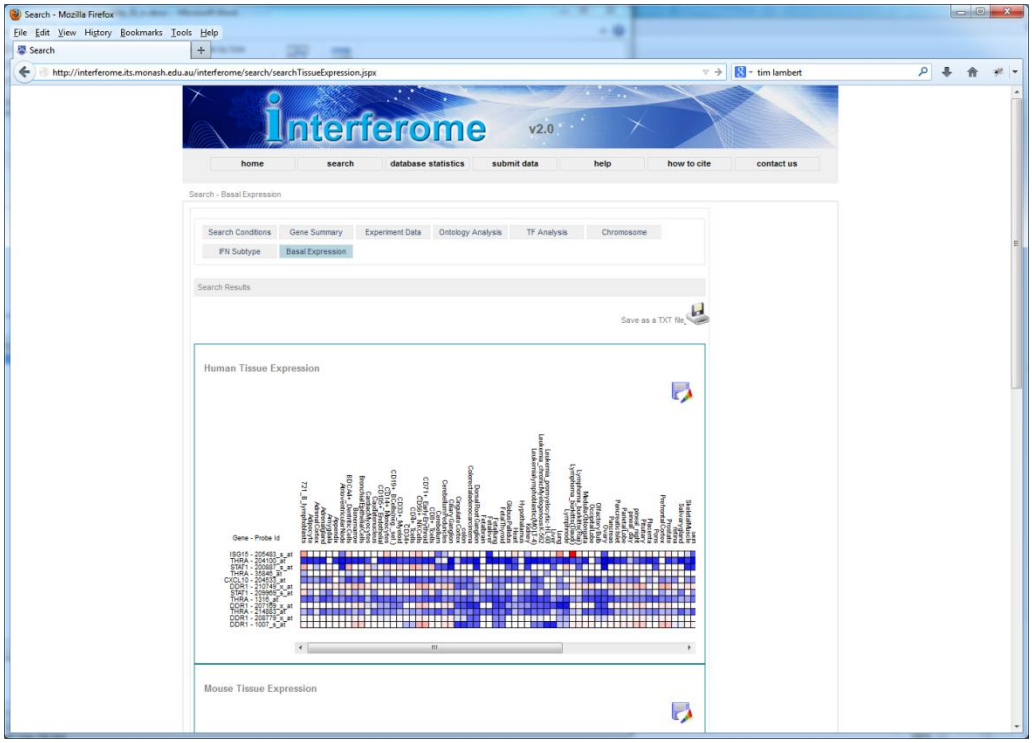


Figure 10. A heat map showing the intensity of expression of each gene in the result set from a range of cells and tissues in resting or unstimulated conditions. Deep red indicates very high expression, pale red indicates moderately high expression, pale blue indicates moderately low expression, and deep blue indicates very low expression. Boxes that appear to be white have expression values very close to the median for that tissue.

3. **Database statistics**. This page presents an overview of the data held within the database, including the number of experiments and datasets that are stimulated by type I, II and III interferons, for both mouse and human, and the number of genes that respond significantly to these stimulations. Some experiments include stimulations with more than one type of interferon (eg interferon I and II applied in combination), thus the individual entries on a line might sum to more than the recorded total. The table is generated dynamically, and will change automatically whenever the database is updated with more experiments and datasets.

4. **Submit data.** The Interferome v2.0 team is committed to the ongoing curation and improvement of the database. The prime activity is the continual addition of new data. Users are encouraged to submit data for inclusion in the Interferome v2.0 via the new 'Submit data' tab. We request that data are MIAME compliant and a form be uploaded from the link (http://www.mged.org/Workgroups/MIAME/miame.html), which is provided on the website. This will accompany the data file upload which will be reviewed by the Interferome DB administrators prior to statistical analysis using a standard pipeline (See *How data was processed* below). Data sets that show a significant response to IFN treatments will be loaded into the database. Relevant supporting experimental data must be made available with the gene expression data to ensure that all required metadata and experimental factors are loaded into the database.
5. **Help.** An Interferome User Guide is made available in pdf format to help users make the best use of the Interferome database.
6. **How to cite.** An explanation on how to cite the Interferome database is provided.
7. **Contact us**. A form for submitting questions and queries to the Inteferome team is provided.

Common Use Case Examples

1. Identifying if a gene is interferon regulated

a. Paste the gene symbol, Genbank accession or ENSEMBL identifier into the appropriate box on the search page, define the search parameters (eg the Interferon type, system, organ etc), then click search at the bottom of the page (Figure 11).
b. The Interferome will return a summary of any interferon regulated genes that fit the search (Figure 12).
c. Clicking on the Experiment tab will show all the experiments where this gene was identified as being regulated by interferon (Figure 13). Data presented include the fold change observed, the interferon type, the treatment time (in hours), a brief description of the gene, links to Genbank and Ensembl entries for the gene or genes, and the probes from the microarray from which the data was derived. This data can be saved in a text file.
d. Clicking on the dataset column brings up details of the experiment dataset, including a MIAME compliant experimental description and a summary of all genes found to be significant in that dataset (as discussed in the *Identifying if a gene is interferon regulated* section above).
e. The user may submit the result set to the full range of secondary analyses that are available (eg ontology analysis, transcription factor analysis, chromosome etc).

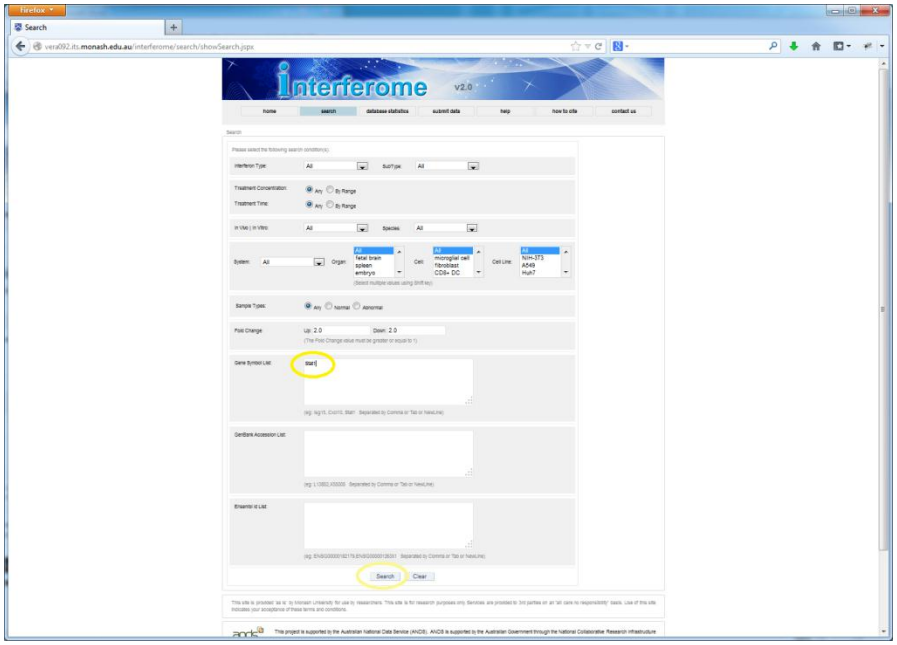


Figure 11. Completing the search page to query a single gene for interferon regulation.

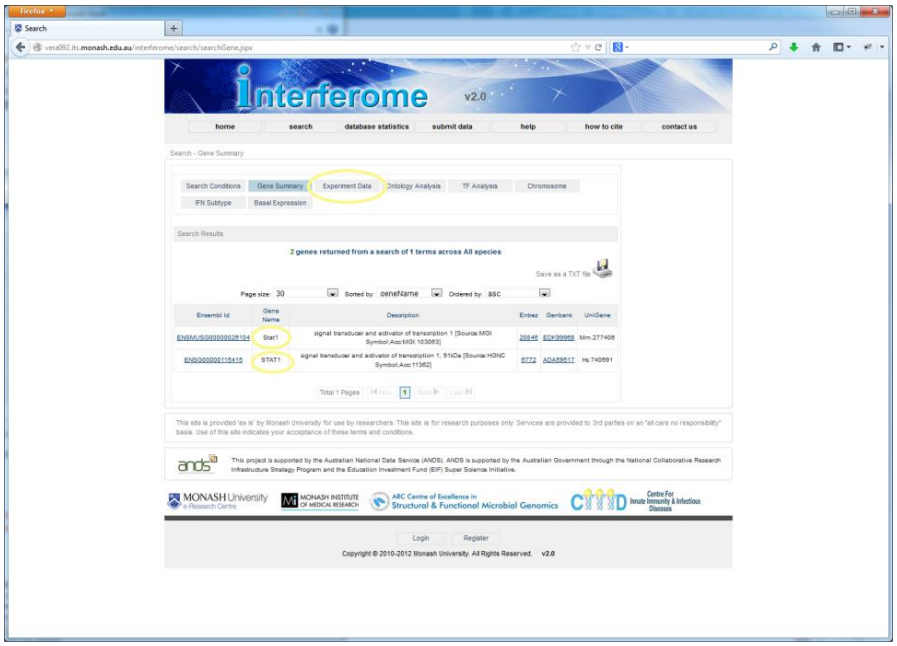


Figure 12. Results from the gene search.

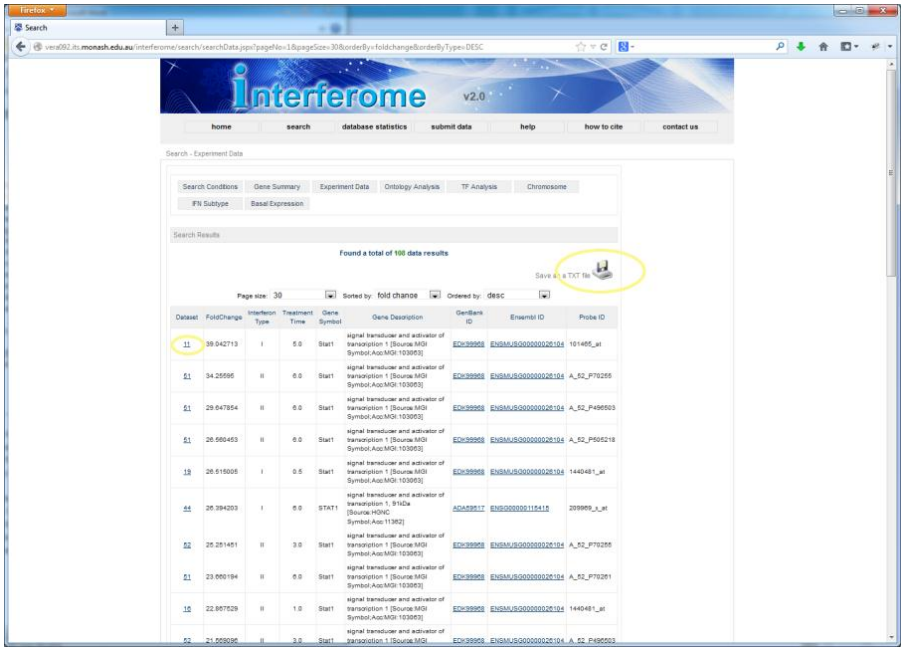


Figure 13. A table of experiments where the selected gene(s) showed a significant response to interferon stimulation.

2. Looking for interferon regulation within a gene signature.

a. Paste the list of gene identifiers representing the gene signature (gene symbol, Genbank Accession or ENSEMBL identifier) into the appropriate box on the search page and select the species to be interrogated and click the "Search" button) (Figure 14).

b. The Interferome will return a summary of any interferon regulated genes present exist in the submitted gene list (Figure 15).

c. Clicking on the "experiment" button brings up a summary of the experimental evidence for Interferon regulation of each of the genes submitted, including fold change, interferon type, treatment time, links to external Genbank and Ensembl identifiers, and the probes used to establish the Interferon response (Figure 16). Clicking on the *dataset* link for any gene will bring up MIAME compliant description of the experiment where the response was created (as discussed in the *Identifying if a gene is interferon regulated* section above).

d. The result set can be used for the range of secondary analyses. For example, clicking on the chromosome button will display the location of each of the genes across the genome (Figure 17).

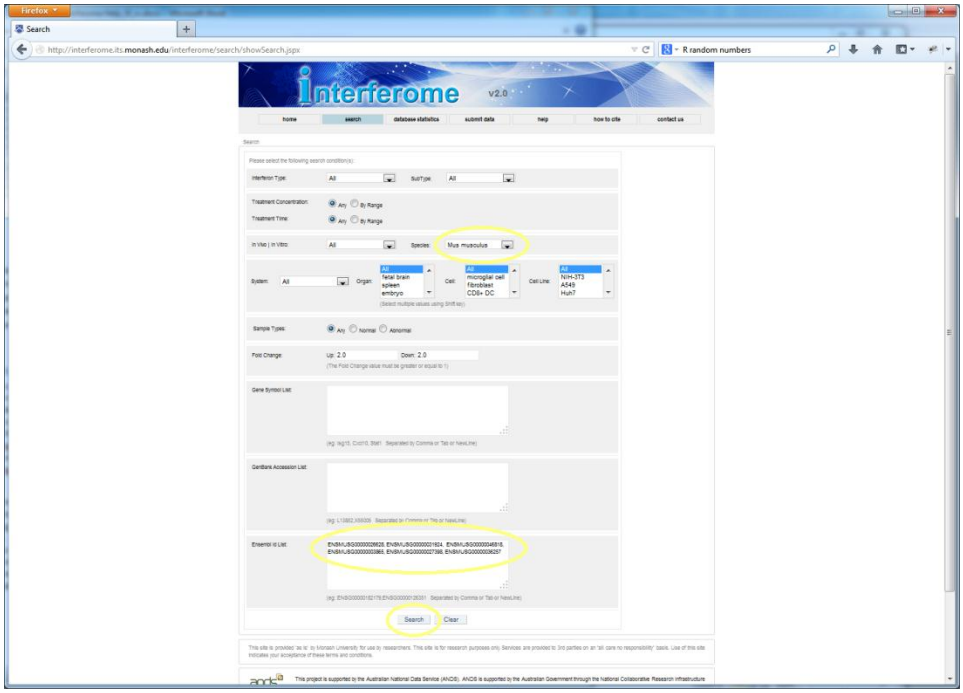


Figure 14. Search page for a gene signature.

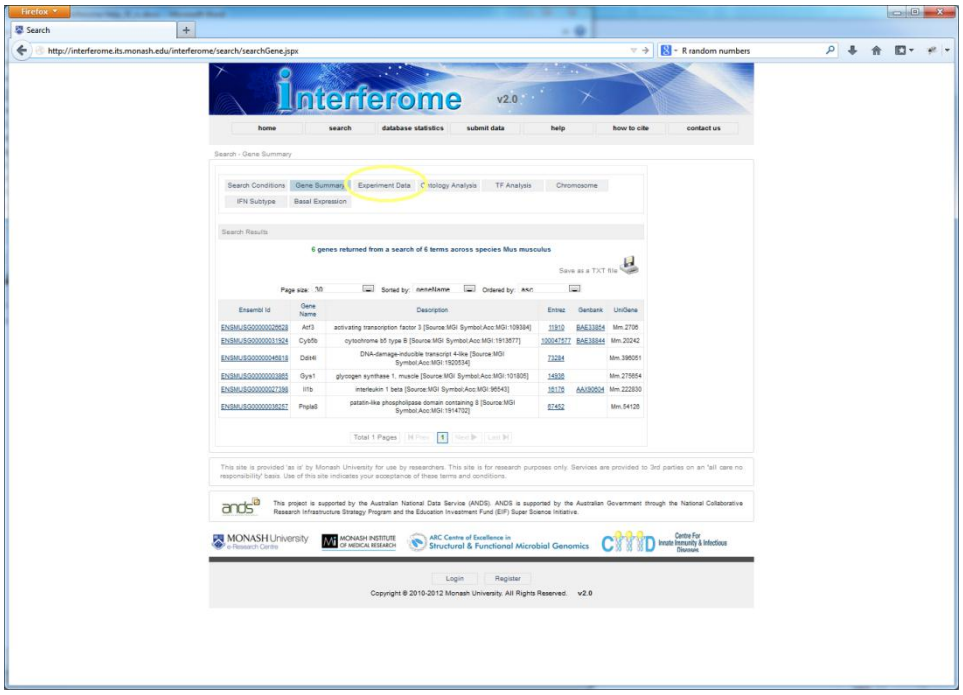


Figure 15. Result set from a gene signature search.

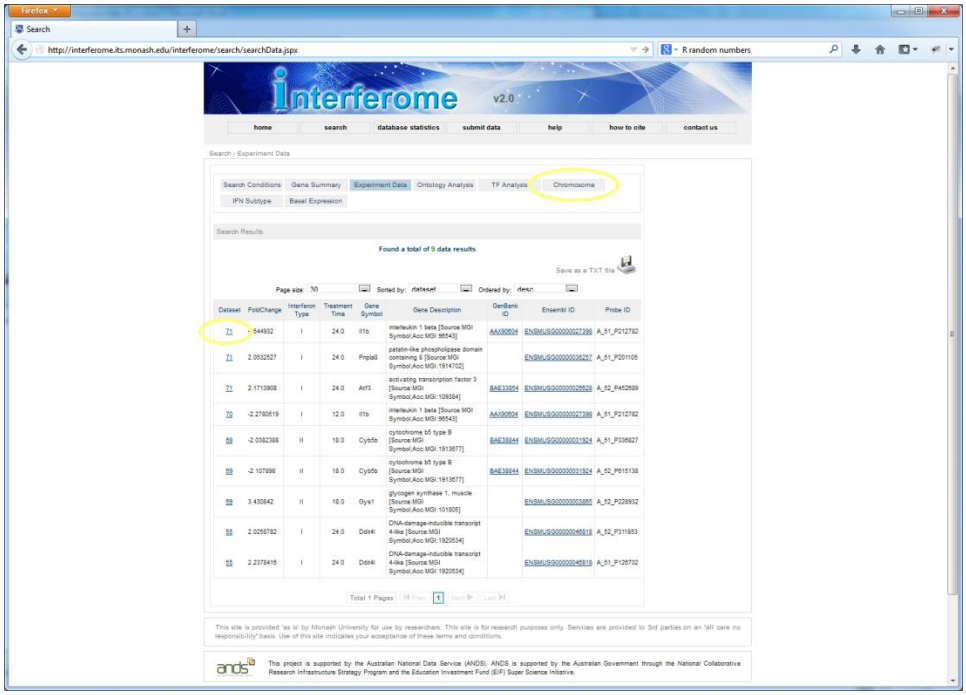


Figure 16. Experiment page from the gene signature search.

Figure 17. The chromosome page following a gene signature analysis, showing the location of each gene across the genome.

Saving images in SVG Format

Various pages present results of analyses in graphic form eg the TF analysis and Chromosome pages. The user is presented with an option to save the image in svg format. The svg format is an image format that allows lossless scaling. In other words, the image can be zoomed in or out at any scale without experiencing the loss of quality typically seen with other image formats eg jpeg, bmp. The icon above the image allows the user to save the image in the svg format. Some software allow svg files to be imported directly eg Adobe illustrator, Photoshop and most web browsers. Some software are not yet capable of directly importing svg files, and the user might want to convert the svg into a different format (eg png or bmp format). Options to make the conversion include the following:

1. Open the saved svg image in Internet Explorer, then right click on the image and choose to save the image in a bmp or png format.
2. Use image processing software to convert from svg to other image file formats, such as Photoshop, or Gimp.
3. Submit the svg image to a number of online resources which will convert the svg image into a different image format eg http://www.fileformat.info, http://www.online-convert.com. Please note that these are external sites and the Interferome team take no responsibility for their availability, quality or performance.

How data was processed

Microarray datasets were manually selected from EBI Array express (http://www.ebi.ac.uk/arrayexpress/) and GEO (http://www.ncbi.nlm.nih.gov/geo/) databases along with in-house experiments. Selection criteria were the type of cells or tissues treated with interferon (all interferon types are accepted), and details of any additional treatments. The samples are classified as normal or abnormal, with the term abnormal representing a genetic variant, a disease condition, or a specific pre-treatment prior to stimulation with interferon

Data was collected from various microarray platforms and array designs, including Agilent and Affymetrix.

Microarray data was stored and processed using the runs BioArray Software Environment (BASE 2). BASE2 plugins were used for data pre-processing and analysis. Pre-processing included normalisation, filtering of low intensity probes, and log transformation. Affymetrix microarrays were normalised using the Robust Multi-Array Averaging (RMA) method. Agilent arrays were normalised using methods from the LIMMA Bioconductor package, including filtering of non-uniform spots, background subtraction, and normalisation to the 75^{th} percentile. Probes were tested for significant differences between Interferon treatments and an appropriate control using the Welch t-test. Probes with a p-value<0.05 were deemed statistically significant. Fold change values are provided on a linear scale.

All relevant supporting data is loaded with each experiment and dataset, including the type and sub-type of IFN treatment, concentration, time of observation after treatment, cell line, cell type, tissue type, organ, system (eg respiratory, nervous), and normal/abnormal status.

References

Bolstad, B.M., Irizarry R. A., Astrand, M., and Speed, T.P. (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. Bioinformatics 19(2):185-193.

Dudoit, S., Y.H. Yang, M.J. Callow, and T. Speed (2000).Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical report 2000 Statistics Department, University of California, Berkeley.

Irizarry, RA, Hobbs, B, Collin, F, Beazer-Barclay, YD, Antonellis, KJ, Scherf, U, Speed, TP (2002) Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Biostatistics 4:249-64.

Pan, W. (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics 18: 546-554.

Rafael. A. Irizarry, Benjamin M. Bolstad, Francois Collin, Leslie M. Cope, Bridget Hobbs and Terence P. Speed (2003), Summaries of Affymetrix GeneChip probe level data Nucleic Acids Research 31(4):e15.

SA Samarajiwa, S Forster, K Auchettl, and PJ Hertzog INTERFEROME: the database of interferon regulated genes. *Nucleic Acids Research*2009 Jan 37(Database Issue):D852-7 doi:10.1093/nar/gkn732.

Smyth, G. K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420

Smyth, G. K., and Speed, T. P. (2003). Normalization of cDNA microarray data. Methods 31, 265-273. Vallon-Christersson J, Nordborg N, Svensson M, Hakkinen J. (2009). BASE--2nd generation software for microarray data management and analysis. BMC Bioinformatics 10:330.