MICROARRAY DATA ANALYSIS REPORT

Prepared by Ocimum Biosolutions
© Ocimumbio Solutions

July 2006

All rights reserved.
## CONTENTS

1.0 Executive summary .................................................4

2.0 Datasets Received ...................................................4

3.0 Analysis Requirements ............................................4

4.0 Analysis Performed ................................................5

5.0 Results ...................................................................9

6.0 References .............................................................9
1.0 Executive summary

This analysis was done using Genowiz™ our proprietary microarray and pathway analysis tool. Genowiz™ is a comprehensive gene expression analysis package that enables researchers to analyze microarray data in an intuitive bio-environment. It includes novel quantification matrices and algorithms that facilitate expression pattern analysis and gives an insight into metabolic, regulatory and disease pathways. It offers an easy to use customizable interface and allows for integration of biotools and laboratory information management systems.

Details of the microarray data analysis service are present at the specified location.

http://www.ocimumbio.com/web/research_services/microarray_analysis.asp

Apart from the microarray data analysis service, Ocimum also offers a number of other services too.

2.0 Datasets Received

- Hansen AXB MOE430A Array 1-30-04.CEL
- Hansen AXC MOE430A Array 2-27-04.CEL
- Hansen AXD MOE430A Array 5-7-04.CEL
- Hansen DXA MOE430A Array 3-19-04.CEL
- Hansen DXB MOE430A Array 1-30-04.CEL
- Hansen DXC MOE430A Array 2-18-04.CEL

3.0 Analysis Requirements

To mine the microarray data for characteristic expression patterns and subsequently extract biological significance of the genes with an up regulated differential expression pattern.

Analysis involves data pre-processing; elimination of outliers, non significantly expressed genes and false positives; and analysis of the gene lists in a biological perspective.

For any other analysis, please contact bdm@ocimumbio.com.
4.0 Analysis Performed

The following analysis was performed.

1. Data quantification and normalization.
2. Identification of differentially expressed genes.
3. Determination of biological significance

The details of the analysis is given below.

1) Data quantification and normalization: Data was quantified using RMA algorithm to make the samples comparable.

2) Identification of differentially expressed genes:

<table>
<thead>
<tr>
<th>Statistics</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correction</td>
<td>Benjamini and Hochberg False Discovery Rate</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>5% FDR</td>
</tr>
<tr>
<td>Fold Change</td>
<td>2</td>
</tr>
</tbody>
</table>

*Table 1: Parameters for finding differentially expressed genes*

The following groups were considered as replicates.

- **Group1 (AXgroup):**
  - Hansen AXB MOE430A Array 1-30-04.CEL
  - Hansen AXC MOE430A Array 2-27-04.CEL
  - Hansen AXD MOE430A Array 5-7-04.CEL

- **Group2 (DXgroup):**
  - Hansen DXA MOE430A Array 3-19-04.CEL
  - Hansen DXB MOE430A Array 1-30-04.CEL
  - Hansen DXC MOE430A Array 2-18-04.CEL

The dataset had over 39,000 transcripts. As there were two groups, each with three replicates, t-test was done to assess the statistical significance of differential expression. To eliminate false positives and to increase data accuracy and quality, Benjamini and Hochberg FDR method for multiple testing correction was performed on p-values derived from t-test. A False Discovery Rate of 5% was used as a cut off for statistical significance giving a list of 1025 differentially expressed genes. Fold change analysis was then performed on this reduced data, to obtain two fold up and down regulated
gene lists. Genes that show a two fold change in mean expression levels of each group are reported to be up regulated/down regulated. 861 genes were retained after fold change analysis, of which 742 were two fold up regulated and 119 were two fold down regulated (Fig 1.)

Analysis to identify genes with a similar expression profile, can also be performed similarly if required. A hierarchical clustering could be performed, where genes with a similar expression profile are present closer to each other than those with different profiles, and a dendrogram representation of the same can be viewed (Fig. 2). To get homogeneous clusters where within cluster variance is less and between cluster variance is more, partitional clustering algorithms like K-means, SOM (Self Organizing Maps) and Gene Shaving could be used. Visualizations like 2D PCA (Fig. 3), Pie Charts, Mean, Line (Fig. 4) and Bar graphs could be used for understanding and interpreting the data.
Figure 3 2D PCA view of clusters

Figure 4 Line graph showing genes up regulated in the first three samples and down regulated in the next three
3) Determination of biological significance

To determine the biological significance of the up regulated genes, functional classification was performed using Gene Ontology. This classifies the genes into relevant ontologies dealing with molecular function, biological processes and cellular components. An accompanying z-score report indicates which functions/processes/components are over represented in the gene list. The most significantly effected molecular function ontology identified was “Binding” with a z-score of 10.7. More specific terms related to “Binding” such as “ATP binding” were also identified. Some of the other functions found to be affected were “Chemokine activity”, “Cytokine activity” and “Growth factor activity”. “Regulation of cytokine biosynthesis”, “Chemotaxis”, “Signal Transduction” and “Immune Response” were the major processes effected.

A pathway analysis on the up regulated genes using pathways from KEGG revealed that the “T cell receptor pathway”, “MAPK pathway” and “JAK-STAT Signaling pathway” have been up-regulated.

Annotations for these genes of interest were also studied using a comprehensive annotation report from various publicly available databases. Additional information for these genes was mined from literature by connecting to PubMed using Genowiz™. The annotations and mined literature added to the understanding and interpretation of the functional aspects of the up regulated genes.

<table>
<thead>
<tr>
<th>Molecular Function Ontology</th>
<th>Up-regulated genes</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding</td>
<td>60</td>
<td>10.7</td>
</tr>
<tr>
<td>ATP binding</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Chemokine activity</td>
<td>22</td>
<td>9.8</td>
</tr>
<tr>
<td>Cytokine activity</td>
<td>16</td>
<td>7.3</td>
</tr>
<tr>
<td>Growth factor activity</td>
<td>13</td>
<td>7.2</td>
</tr>
</tbody>
</table>

**Table 1: Z-score reports showing ontologies most effected in the data A) Molecular function ontologies B) Biological process ontologies**
5.0 Results

The genes, which are significantly up or down regulated are listed in the attached, excel sheet. The intensity of the replicates has been merged to get a better picture of the expression. The number of genes that are two fold up regulated and down regulated are 742 and 119 respectively (See attached excel sheet) A comprehensive list of ontologies and pathways significantly up regulated has been provided in the supplementary material. A tab delimited annotation report has also been provided for the up regulated gene list. This analysis has identified ontologies and pathways that are significantly effected by two fold up regulated genes in the data. This helps in the understanding of gene interactions and their possible phenotypic effects.

6.0 References


Do not hesitate to contact us for any clarifications or suggestions. Please let us know how beneficial this analysis has been for your research.

Your feedback is important for us to be of a greater service to you!