

Transcriptome analysis of orofacial development in mouse embryo

Abstract

Gene expression profiles in fetal murine orofacial tissue were monitored using data from the Gene Expression Omnibus (GEO, GDS739). Of the more than 12,000 genes examined using the Affymetrix® GeneChip® U74Av2 Array, 411 genes were differentially expressed (at least 1.5 fold-change and less than a 5% false discovery rate) between gestation day 12 and 14. Partitioning around medoids (PAM) was used to partition the 411 genes into two discrete clusters. The biological process ontologies associated with the genes in each cluster were analyzed using a z-score report. This analysis showed that distinct biological themes were associated with each of the two patterns of gene expression.

Introduction

In this study GeneSifter was used to analyze microarray data generated from a time series examining orofacial development in mouse embryos. This analysis process can be broken down into three discrete tasks: identification of significantly regulated genes, identification of global patterns of gene expression, and the determination of the biological meaning of both individual genes and groups of genes. GeneSifter uses Gene Ontology (GO) Reports and z-scores to summarize the biological processes represented in a particular gene list. Z-scores can then be used to identify GO terms that are significantly over- or under-represented in a gene list. This study outlines the use of these methods to identify biological themes associated with discrete expression patterns in a mouse development time series.

The Data

The following time points were examined in this study:

Gestation Day 12 mouse orofacial tissue (Day 12)

Gestation Day 13 mouse orofacial tissue (Day 13)

Gestation Day 14 mouse orofacial tissue (Day 14)

Three biological replicates were prepared for each time point.

These samples were hybridized to the Affymetrix GeneChip Mouse U74Av2 array. Data from the CHP files was obtained for each sample from GEO (GDS739) and loaded into GeneSifter.

Filtering and visualization

Prior to visualization, the project was filtered to identify a subset of genes that were significantly differentially regulated during the development time series. The data set was first filtered to remove genes with “Absent” detection calls for all samples. A 1.5 fold-change cutoff was then applied and ANOVA was performed on the dataset ($p < 0.05$). Correction for multiple testing was performed using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995) to derive a false discovery rate estimate from the raw p-values. A false discovery rate of 5% was used as a cutoff for statistical significance. Of the more than 12,000 genes examined, 411 were significantly differentially expressed during orofacial development based on the filtering criteria described. Hierarchical clustering and PCA were used to visualize these 411 genes. Both methods identified two distinct patterns of gene expression within the 411 differentially expressed genes (figure 1).

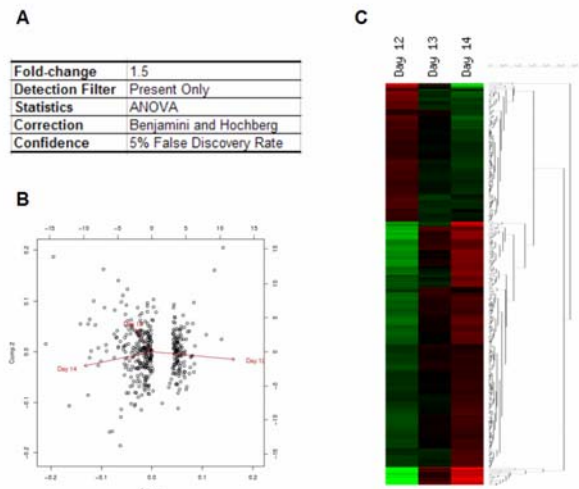


Figure 1: Filtering and visualization. A) Parameters used for filtering times series. These parameters produced a list of 411 differentially expressed genes. B) PCA analysis of filtered gene list. C) Hierarchical clustering of filtered gene list.

Partition clustering

The filtered data set was partitioned based on expression profiles using k-medoids clustering. For this analysis the partitioning around medoids (PAM) method was used, and several values for k were chosen, ranging from 2 to 6.

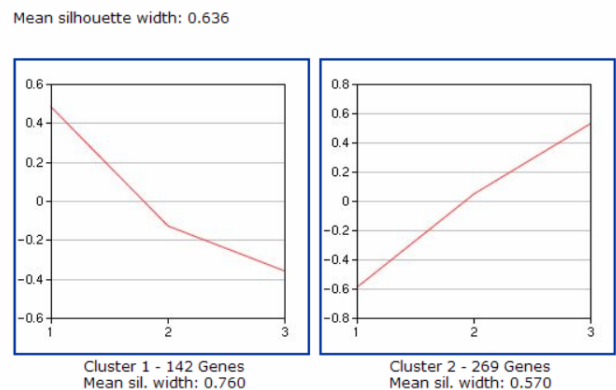


Figure 2: Partition clustering of the filtered gene list. PAM was used to separate 411 differentially expressed genes into two groups based on expression pattern. Clustering into two groups gave the largest mean silhouette width.

Application of the silhouette method (Kaufman and Rousseeuw, 1990) indicated that $k=2$ gave the best grouping of the filtered data. Figure 2 summarizes the expression patterns associated with each cluster.

A

Biological Process Ontology	List	Array	z-score
phosphate transport	10	36	8.69
cell adhesion	26	265	6.59
ossification	7	39	5.46
muscle development	12	106	5.03
humoral defense mechanism (sensu Vertebrata)	6	38	4.61
complement activation	5	28	4.59
negative regulation of cell proliferation	5	30	4.37
negative regulation of Wnt signaling pathway	2	6	4.34
extracellular structure organization and biogenesis	3	21	3.02
organogenesis	32	651	2.98
immune response	18	314	2.88
proteolysis and peptidolysis	16	305	2.32

Molecular Function Ontology	List	Array	z-score
extracellular matrix structural constituent	13	43	10.59
calcium ion binding	34	295	8.89
structural molecule activity	31	363	6.46
growth factor binding	6	30	5.52
structural constituent of cytoskeleton	9	73	4.75
cyclin-dependent protein kinase inhibitor activity	2	8	3.68

B

Biological Process Ontology	List	Array	z-score
response to X-ray	2	2	11.35
chromatin assembly or disassembly	4	13	8.6
transcription, DNA-dependent	34	921	5.81
mitotic cell cycle	11	161	5.56
induction of apoptosis by intracellular signals	2	12	4.28
development	31	1058	4.09
embryonic development	5	93	3.05
immune cell activation	4	68	2.94

Molecular Function Ontology	List	Array	z-score
DNA binding	37	1009	6.47
ATP-dependent helicase activity	5	46	5.39
protein binding	37	1468	3.96
transcription regulator activity	17	608	2.95

Figure 3: Distinct biological themes are associated with each cluster. Each table lists highly over-represented biological process and molecular function gene ontologies for that cluster. A) Ontologies for cluster 2 (up-regulated genes). B) Ontologies for cluster 1 (down-regulated genes).

Biological significance

The biological process and molecular function gene ontologies associated with the genes in each cluster were examined using a z-score report. This report identifies ontologies that are significantly over-represented in a gene list. The z-score report identified distinct biological themes associated with each cluster. Figure 3 lists the predominant biological process and molecular function ontologies over-represented in each cluster. See supplemental material to view a comprehensive list of ontologies identified for each cluster.

To identify potential regulators of the changes in gene expression, 26 genes with the molecular function ontology “transcription factor activity” were selected from the filtered gene list. Hierarchical clustering of these genes showed two groups of genes with transcription factor activity, one with increased expression across the time series and one with decreased expression (figure 4). These 26 genes are candidates for regulators of fetal orofacial development. See

supplemental material to view annotation for each of these 26 genes.

Summary

Transcriptome analysis of orofacial development in mouse embryos using the Affymetrix GeneChip U74av2 array identified 411 genes that are differentially expressed during the time period examined. Among these 411 genes there was a significant enrichment of genes involved in cell adhesion, phosphate transport, transcription and the mitotic cell cycle. 26 genes with transcription factor activity were identified as potential regulators of the gene expression changes that occur during murine orofacial development.

In this report, we have demonstrated that the use of GeneSifter combines statistical analysis, pattern recognition and the determination of biological significance, allowing users to rapidly identify biological themes associated with a pattern of gene expression and to understand the biology underlying a gene cluster.

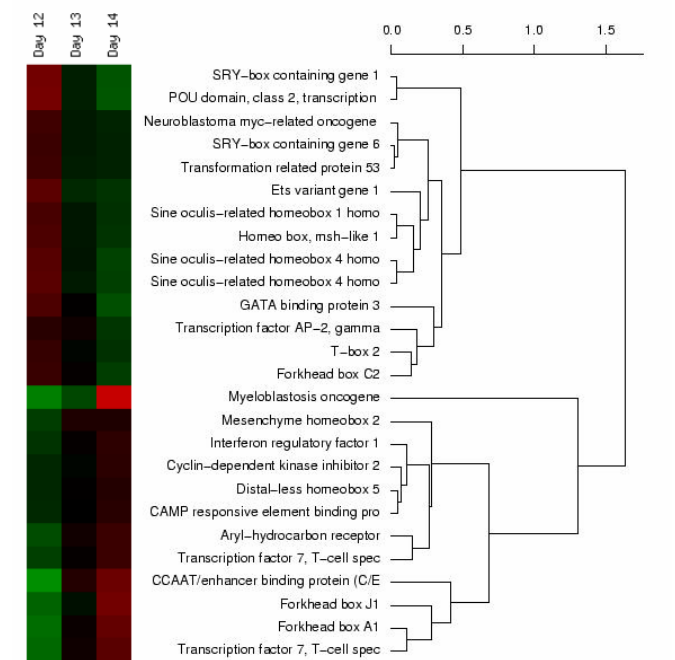


Figure 4: Hierarchical clustering of 26 genes with the molecular function ontology “transcription factor activity” identified in the list of 411 differentially expressed genes.

Supplemental Material

Raw data is available from Gene Expression Omnibus (GEO), dataset GDS739.

Access to this data and tutorials are available through the GeneSifter Data Center (www.genesifter.net/web/DC).

References

Reiner, et al. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19(3):368-375
 Kaufman L, Rousseeuw PJ: *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley; 1990
 Doniger, et al. 2003. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology* 4:R7