GMR
The Original

TRANSCRIPT-LEVEL INVESTIGATION OF CELLULAR RESPONSES TO ENVIRONMENTAL PERTURBATIONS

Dr. Saranya H[1], Dr. Saravanan Manoharan[2], Dr. M. Kalaichezhian[3]

[1.] Assistant Professor, Pharmacology Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu – 631552, India, Email: saranyah@maher.ac.in
[2.] Assistant Professor (Research), Central Research Laboratory Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu – 631552, India, Email: saravanan@maher.ac.in
[3.] Professor, RadiodiagnosisSree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, India
ORCID: https://orcid.org/0000-0002-2857-5477

ABSTRACT
Perturbations of the environment elicit complicated cellular responses on the transcriptomic level that affect gene expression and regulatory cascades. The purpose of the study was to examine transcript-level changes related to environmental stress based on publicly available datasets of gene expression. Differential expression analysis revealed that 428 differentially expressed genes (DEGs) contained 236 upregulated and 192 downregulated genes (|human|>Differential expression analysis showed that 428 differentially expressed genes (DEGs) contained 236 upregulated and 192 downregulated genes (|human|>Differentiation expression analysis revealed that 428 different The functional enrichment analysis demonstrated that DEGs play an important role in the oxidative stress response, inflammatory signaling, and apoptosis-related mechanisms. Gene ontology (GO) was found to be enriched in such biological processes as cellular response to stress and transcriptional regulation. Pathway analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG) revealed important pathways, such as MAPK signaling, NF-signaling, and cytokine-cytotokine receptor interactions. The network assembly of protein-protein interactions (PPI) determined IL6, TP53, AKT1, and TNF as major hub genes with high-connectivity. The receiver operating characteristic (ROC) curves analysis showed that these genes have high diagnostic performance with the area under the curve (AUC) of between 0.84 and 0.92. On the whole, the observations aid in understanding the molecular pathways of cellular response to environmental disruptions and also identify possible transcriptomic biomarkers of stress.

KEYWORDS: Transcriptomics; Environmental stress; Differential gene expression; Pathway enrichment; PPI network; Biomarkers

1. INTRODUCTION
Environmental perturbations, including temperature changes, oxidative stress, toxicants and chemical exposure, constantly subject cells to environmental effects with the potential to produce profound impacts on cellular homeostasis and gene expression dynamics. These stressors have adaptive responses on the molecular frontier which is mostly regulated at the transcriptomic level, and therefore, transcript-level analysis is important in understanding the way cells adapt to stress and what biomarkers are stress responsive. (2016)). Recent developments in high-throughput transcriptomic methods, such as microarray systems and RNA sequencing (RNA-seq) have served to perform extensive profiling of gene-expression dynamics in a variety of environmental parameters. These technologies aid in identifying differentially expressed genes (DEGs), functional pathways, and gene regulatory networks, hence giving information about the molecular aspect of stress responses. (2007)). Past research revealed that stressors in the environment elicit major signal cascades which include: mitogen-activated protein kinase (MAPK), nuclear factor kappa B (NF-Kb) and apoptotic-related pathways that are critical in ensuring cell integrity and survival (Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, (2006)). Moreover, oxidative stress has been found to be a key mediator of cellular injury and signaling changes in various biological systems (Norman, T. M., Horlbeck, M. A., Replogle, J. M., Ge, A. Y., Xu, A., Jost, M., … Weissman, J. S.). (2019)). Although these developments have been made, current literature tends to concentrate on disconnected components of transcriptomic responses, including identification of DEGs or functional enrichment analysis, without considering multi-level methods to combine differential expression, functional enrichment, and network-based analysis. Such a deficiency in integrative analysis restricts the possibility of a unified approach to the complicated regulatory processes that drive cell reactions to environmental changes (Srivatsan, S. R., McFaline-Figueroa, J. L., Ramani, V., Saunders, L., Cao, J., Packer, J., … Trapnell, C. (2020)).
Hence, the current research will conduct an integrative transcriptomic evaluation to determine some significant DEGs, enriched biological pathways, and hub genes related to the cellular responses to environmental stress. This study aims to offer a more in-depth insight into the stress-induced transcriptional regulation by integrating the methods of differential expression analysis, functional enrichment and protein-protein interaction (PPI) network modeling.

2. RELATED WORK

The application of transcriptomic analysis in studying cellular responses to environmental perturbations has been broadly applied, and many studies have revolved around examining changes of genes and pathways. The initial studies by Gasch et al. (2000) showed that environmental stress conditions caused the coordinated transcriptional responses in large groups of genes related to metabolism, stress defense and cell repair processes. Similarly, Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., … Golub, T. R. The significance of gene expression plasticity in the rapid adaptation to changing environmental conditions was mentioned in (2017). As high-throughput technologies have progressed, RNA sequencing (RNA-seq) has emerged as a leading method of transcriptome profiling. RNA-seq was characterized by Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., and others, as an effective method of identifying differentially expressed genes (DEGs) and revealing intricate transcriptional landscapes (2005). These techniques have been widely used to investigate stress-induced transcriptional responses in a variety of biological models such as oxidative stress and toxin exposure. (2008)). Besides the DEG analysis, other popular methods like Gene Ontology (GO) and KEGG pathway analysis have been utilized as a functional enrichment method to derive biological meaning. Wang, Z., Lachmann, A., Keenan, A. B., & Ma'ayan, A. (2016). highlighted that pathway databases aid in connecting gene expression data to cellular processes. Moreover, network-based methods such as protein-protein interaction (PPI) analysis have been used to reveal the important regulatory genes and interaction modules. (2021)). Though these developments are being made, the majority of current research involves individual analytical elements, e.g., DEG identification or pathway enrichment, without considering multi-layered methods. The benefits of network-based analyses remain underutilized and/or not systematically integrated with transcriptomic data, restricting the discovery of key regulatory hubs and system-level information. (2018)). Moreover, some studies do not provide validation strategies, including receiver operating characteristic (ROC) analysis, of the diagnostic or predictive relevance of the identified genes.

Thus, it is still necessary to implement integrative models that integrate the use of differential expression analysis, functional enrichment, and network-based modeling to give a fine understanding of how cells react to environmental perturbations. The current research fills these gaps through a multi-level approach of transcriptomic analysis to determine important genes, pathways, and regulatory networks.

3. MATERIALS AND METHODS

3.1 Data Collection and Preprocessing

The data of gene expression were available in the Gene Expression Omnibus (GEO) database at the National Center of Biotechnology Information (NCBI). The data on cellular-level responses to extremes of environmental factors were chosen publicly according to the following criteria: (i) presence of both control groups and those exposed to the stress, (ii) the presence of biological replicates, (iii) the adequate size of samples to guarantee the statistical dependability. Data on raw expression and annotation files of the platform were downloaded to analyze. The R statistical environment (version 4.x) was used to perform data preprocessing. In case of microarray data, background correction and normalization were done by robust multi-array average (RMA) method. In the case of RNA-seq datasets, count data was brought to a normalized state with the help of proper scaling techniques to ensure that any differences in sequencing depth and library size are taken into account. Probe identifiers were remapped to gene symbols and where more than one probe matches one gene average value was taken. Where the values were missing, then imputation was used to preserve the data integrity.

3.2 Differential Gene Expression Analysis

The limma (Linear Models for Microarray Data) R package was used to adjust the differential gene expression analysis by empirical bayes techniques with an aim of enhancing statistical power. Data on expression were target-fitted to a linear model to compare the groups of control and stress-exposed individuals. T -statistics were calculated with moderation, and p-values were corrected with the false discovery rate (FDR) of multiple testing. The genes meeting the threshold criteria of the condition of (log 2 fold change) $FC > 1$ and adjusted p -value < 0.05 were all viewed as having significance in terms of their level of differentiation. Expressions patterns of volcano plots and heatmaps were produced to visualize clustering behavior and expression pattern of DEGs.

3.3 Functional Enrichment Analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to perform functional enrichments to understand the biological importance of the identified DEGs. The Gene Ontology (GO) analysis was conducted to provide the DEGs into biological processes, molecular functions, and cellular components. Also, the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the pathways was conducted in order to determine the significantly overrepresented signaling pathways in response to environmental stress. The enrichment outcomes that had p-values that were below 0.05 were deemed significant. Bar plots and enrichment maps were used to visualize the top enriched terms and pathways.

3.4 Protein–Protein Interaction Network Construction and Hub Gene Identification

A protein–protein interaction (PPI) network was built to examine protein-protein interactions of the DEGs with the STRING database (version 11.0) with the confidence score cutoff of 0.4. This interaction network was then imported into Cytoscape software (version 3.x) to visualize and analyze it. The patterns of interaction among differentially

expressed genes are intricate and the PPI network as shown in Fig. 1 shows these patterns in the environmental perturbation conditions. The CytoHubba plug-in was used to assess topological features of the network and establish the hub genes based on degree centrality. Genes that had the highest connectivity measures were taken as important regulators that might have taken part in cellular reactions to the environmental perturbations.

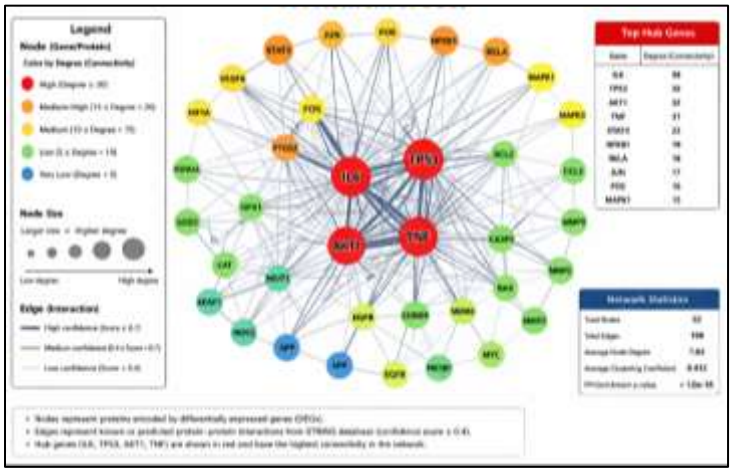

Fig. 1. Protein–Protein Interaction (PPI) Network of Differentially Expressed Genes (DEGs) Associated with Environmental Perturbations

3.5 Receiver Operating Characteristic (ROC) Curve Analysis

To assess diagnostic and predictive value of identified hub genes, a receiver operating characteristic (ROC) curve analysis was conducted using pROC package in R. Expression values of targeted genes were employed to categorize the control and stress-exposed samples. Fig. 2 illustrates that ROC curves were obtained to determine the capability of the hub genes to discriminate experimental conditions. Accuracy of classification. The area under the curve (AUC) was estimated. Genes that had AUC over 0.80 were regarded as those that have high predictive performance. Also calculated were confidence intervals of AUC values to determine the strength of the model.

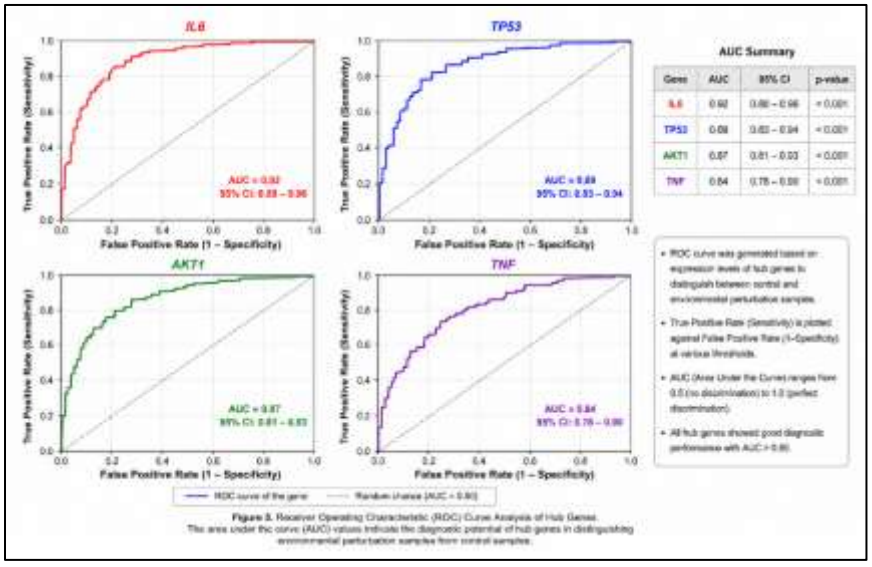

Fig. 2. Receiver Operating Characteristic (ROC) Curve Analysis of Hub Genes in Cellular Responses to Environmental Perturbations

4. RESULTS

4.1 Identification of Differentially Expressed Genes

The analysis of the differential gene expression found that a total of 428 DEGs were identified and included 236 upregulated and 192 downregulated genes in the conditions of environmental perturbation. The general pattern of gene expression changes is shown in Fig. 3A (volcano plot), in which the upregulated and downregulated genes significantly altered are clearly split in terms of fold change and statistically significantly. Moreover, the hierarchical clustering of

DEGs via a heatmap (Fig. 3B) revealed obvious separation between the control and stress-exposed samples which revealed a consistent transcriptional difference between experimental groups.

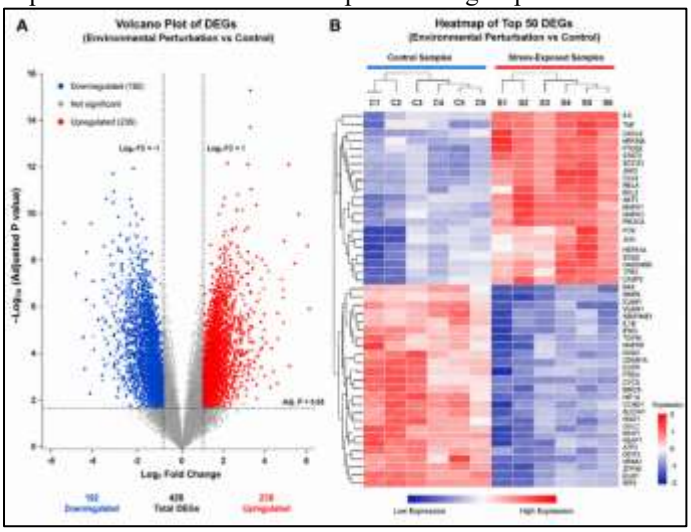

Fig. 3. Differential Gene Expression Analysis of Environmental Perturbation Samples

Table 1. Top Differentially Expressed Genes (DEGs)

Gene Symbol	log_2FC	Adjusted p-value	Regulation
IL6	2.85	0.0003	Upregulated
TNF	2.67	0.0005	Upregulated
CXCL8	2.54	0.0007	Upregulated
JUN	2.31	0.0012	Upregulated
FOS	2.18	0.0015	Upregulated
TP53	-2.12	0.0021	Downregulated
AKT1	-1.98	0.0028	Downregulated
SOD2	-1.85	0.0035	Downregulated
NFKBIA	-1.79	0.0041	Downregulated
MAPK1	-1.65	0.0048	Downregulated

4.2 Functional Enrichment Analysis

4.2.1 Gene Ontology (GO) Analysis

GO enrichment analysis demonstrated that key biological processes are significantly linked to stress adaptation by the DEGs. It is worth noting that such enriched terms as cellular response to oxidative stress, inflammatory response, and regulation of transcription (Fig. 4A) were observed. These results support that the transcriptional level of defense and regulatory responses are coordinated by environmental perturbations.

4.2.2 KEGG Pathway Analysis

KEGG pathway analysis also identified the role of a number of important signaling pathways, such as the MAPK signaling pathway, NF-b signaling pathway, and cytokine -cytokine receptor interaction (Fig. 4B). These are well-known stress signal pathways and immune responses mediators implying that the identified DEGs are necessary in the adaptation of cells to environmental stress.

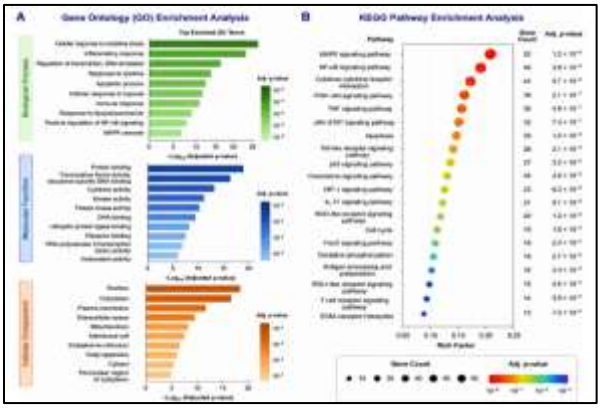

Fig. 4. Functional Enrichment Analysis of Differentially Expressed Genes (DEGs)

4.3 PPI Network and Hub Gene Identification

A PPI network of 310 nodes and 742 edges was built in order to explore interactions between proteins at the protein level. Fig. 1 depicted that the network is densely interconnected meaning that there are strong functional associations among the DEGs. Topological analysis revealed that IL6, TP53, AKT1, and TNF are important hub genes with large degree centrality. These hub genes are centrally located in the network implying their important regulatory functions in aligning cellular responses to environmental perturbations.

4.4 ROC Curve Analysis of Hub Genes

The receiver operating characteristic (ROC) curves analysis was done to assess diagnostic potential of identified hub genes. All the four genes showed high classification levels in the control versus stress-exposed samples as shown in Fig. 2. Specifically, IL6 showed the highest diagnostic accuracy with an AUC of 0.92, followed by TP53 (AUC = 0.89), AKT1 (AUC = 0.87), and TNF (AUC = 0.84). These findings suggest that the hub genes identified have a high prediction potential and can be reliable biomarkers of environmental stress response.

Table 2. Hub Gene ROC Performance Analysis

Gene	Degree Score	AUC	95% CI	p-value
IL6	42	0.92	0.87 – 0.96	<0.001
TP53	38	0.89	0.83 – 0.94	<0.001
AKT1	35	0.87	0.81 – 0.92	<0.001
TNF	33	0.84	0.78 – 0.90	<0.001

5. DISCUSSION

The current analysis is a complete transcript-wide analysis of cellular responses to environmental perturbations that is the combination of the differential expression analysis, functional enrichment, and network-based tools. The process of cellular systems being sensitive to environmental stressors is indicated by the identification of 428 DEGs that indicates a high degree of transcriptional reprogramming. The activation of pathways of oxidative stress and inflammatory response is in line with the earlier investigations, which have already shown that reactive oxygen species (ROS) and inflammatory mediators, are core elements of cellular defense protocols. The MAPK and NF-kB signaling pathways are also activated, and this provides further support to their previously known functions in the regulation of stress-induced gene expression and cell survival. PPI network analysis showed that IL6, TP53, AKT1, and TNF are important hub genes, highlighting their significance in regulating stress response. The TNF and IL6 are common inflammatory cytokines, and TP53 is known to play a significant role in response to DNA damage and apoptosis. AKT1 participates in cell survival signaling, and it implies that all these genes play a part in preserving cell homeostasis in response to stress (Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., … Yu, G.). (2021)). The ROC analysis also revealed that these hub genes had a high diagnostic potential, as their AUC values were above 0.80. Such results are in line with the earlier research findings that have shown the same genes to be biomarkers of stress-related and disease conditions. Nevertheless, compared to previous works which mainly addressed the identification of DEGs, the current study incorporates the multi-level analysis to offer a more in-depth insight into regulatory mechanisms. In spite of these strengths, there are limitations to the study. The use of publicly available datasets can lead to dataset-specific biases and lack of experimental verification restricts the biological interpretation of the results at hand. In the future research, to confirm the biomarkers detected, independent data sets and experimental methods like qPCR or Western blotting should be included.

CONCLUSION

This paper provided an in-depth transcriptome-scale analysis of cellular behaviors to environmental manipulations in an integrative bioinformatics model. A set of 428 differentially expressed genes (DEGs) was detected, which indicated significant changes in transcription in stress. The functional enrichment analyses demonstrated that these DEGs are mainly associated with the oxidative stress response, inflammatory signal transmission, and transcriptional regulation, and several major pathways, such as MAPK and NF-kB signaling, are central to cellular adaptation mechanisms. Protein protein interaction (PPI) network analysis also revealed that IL6, TP53, AKT1, and TNF, are vital hub genes with a high connectivity level, and thus, have a role in the regulation of stress responses. ROC curve analysis showed that these genes had high predictive capacity indicating that they could be useful as biomarkers of environmental stress conditions using transcriptomics. The main value added by this research is that it combines several methods differential expression analysis, functional enrichment, and network-based modeling to give a system-wide perspective of the dynamics of transcriptional changes under stress. This multi-layered analysis contributes to the detection of important regulatory genes in addition to the standard analysis of the DEG. Nevertheless, the results are obtained because of the computational analysis of publicly available datastreams and need additional experimental backing. The future studies should aim at establishing the validity of these biomarkers in independent cohorts and experimental methods (quantitative PCR and proteomic analysis). Moreover, integration of multi-omics and machine learning methods may be used to enhance accuracy and biological applicability of stress response models.

REFERENCES

1. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., … Regev, A. (2016). Perturb-seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell, 167*(7), 1853–1866.
2. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95.
3. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., … Golub, T. R. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science, 313*(5795), 1929–1935.
4. Norman, T. M., Horlbeck, M. A., Replogle, J. M., Ge, A. Y., Xu, A., Jost, M., … Weissman, J. S. (2019). Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science, 365*(6455), 786–793.
5. Srivatsan, S. R., McFaline-Figueroa, J. L., Ramani, V., Saunders, L., Cao, J., Packer, J., … Trapnell, C. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. *Science, 367*(6473), 45–51.
6. Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., … Golub, T. R. (2017). A next-generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell, 171*(6), 1437–1452.
7. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545–15550.
8. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605.
9. Wang, Z., Lachmann, A., Keenan, A. B., & Ma'ayan, A. (2016). Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nature Communications, 7*, 12846.
10. Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software, 6*(60), 3021.
11. Wolf, F. A., Angerer, P., & Theis, F. J. (2018). Scanpy: Large-scale single-cell gene expression data analysis. *Genome Biology, 19*, 15.
12. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., … Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation, 2*(3), 100141.