International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Risk Assessment of Lead Induced Toxicity Profiling from Toxicogenomics Data of Human Cell Lines

Arpitha B M[1*]

[1]Department of Biotechnology, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India

Abstract: ***Heavy metal toxicity is a health menace and is posing a significant risk because of adulteration of drinking water and often associated with folk lore therapies. Lead induced toxicity poses considerable impact on the developing brains of fetuses and young children, including pregnant women. Gene expression profiling studies has made advancement in identifying key regulatory genes responsible for complex diseases. In this study, an effort has been made to assess the risk of lead induced toxicity by conducting toxico-genomics profile analysis that involved identification, characterization and mapping of significantly expressed genes to lead induced diseases. The data source used in this study was derived from Affymetrix cDNA Microarray based experiment on Lead acetate induced Peripheral blood mononuclear cells. The data was retrieved from public repositories, cleaned and further processed before statistical treatments. The results demonstrated significant expression of Metallothionein family of genes in lead acetate induced cell-lines. Further, annotation of these genes gave insights into the complex relationship of the genes with clinical outcomes observed during lead toxicity. Mapping of these genes by means of GENEMANIA produced a scale-free gene expression network. The study reveals an intricate relationships of heavy metal induced genes and their respective disease association in a scale free network.***

Keywords: Metallothionein, Heavy metal toxicity, Biological Networks, Gene expression profiling, Lead acetate

1. Introduction

Epidemiology studies involving estimation of the extent of heavy metal toxicity in population and its associated health effects has gained lot of momentum in recent years. In recent times many studies on causal relationships of increase in lead pollution and its consequential effects on people residing along the river basin has been carried out. (1) Many rivers such as Vrishabhavathi, tributary of Arkavathy River etc carries both treated and untreated effluents from plants of Bangalore water supply and Sewage board. (2) This has led to major deposition of organic and Inorganic contaminants in the river. Pb, Ni, Cd, Cu, Hg, Zn and Cr are at elevated levels in the river. (3,4) Information generation on neurological deficits particularly of those related to memory and cognition leading to deduced IQ values and Mental retardation among children is cause of concern. (5) Lead is a single tress heavy metal which is reported to precipitate into neurological deficits among the exposed population. More vulnerability is possible with respect to pediatric age group of children and women of reproductive age, since Lead is known to cross blood- brain barrier as well as placenta. (6) In this study the risk assessment of toxico-genomics profiling study has been conducted to get insights into the complex gene expression behavior which is a consequence of lead induced toxicity. This kind of study is expected to provide inferential features for extrapolation to other similarly defined study areas.

2. Methods

Data Source:

Microarray Affymetrix dataset of Lead acetate induced Human PBMC (Peripheral Blood Mononuclear Cells) was retrieved (7) from the NCBI GEO (Accession number: GSE37567**).**

Pre-processing of dataset for statistical treatments

Pre-processing of Microarray dataset was prerequisite for statistical treatments. "affy" library in R was used for normalization & Robust Multi-array Average (RMA) based normalization was performed. There are three processing steps in RMA method analysis which includes Background Correction, Normalization and Summarization. (8) Background correction was performed using RMA (Robust Multi-array Analysis) method of Bioconductor Package in R. (9) Background correction is done based on the distribution of perfect match values amongst the probes. Perfect match values are computed based on the specific binding of probe and array. Non linear correction method is followed in this method. Microarray data often contains the variations such as biases in dye coupling or hybridization efficiencies. (10) These variations are corrected by Normalization. RNA degradation plot and MVA plot (11) are key indicators of variation in data distribution. The presence of variations; the necessity for Normalization and the effect of Normalization can be determined by examining these plots.

Differential expression analysis

Differential expression analysis of the datasets was performed by employing "limma" package in R. "lmfit" and "eBayes fit" functions were employed to identify a set of up-regulated and down-regulated genes. "lmfit" function is used to fit linear models for each of the gene in the arrays. Two matrices are specified, a design matrix and contrast matrix. Design matrix is the RNA targets hybridized to arrays and contrast matrix is the coefficients defined by the design matrix. "eBayes fit" computes t-statistics, F- statistics and log-odds of differentially expression. It is mainly used when a distribution has to be computed from the datasets. (12)

GO Enrichment

Genomic and functional annotations of differentially expressed genes were performed. Gene ontology studies

were then performed based on Gene ontology database (www.*geneontology.org)*. (13) Information related to genes involved in biological process, cellular component and its molecular function were analyzed.

Clustering

Clustering facilitates grouping genes possessing similar functionality & provides insights of a sub-set of genes that are co-expressed as result of lead induced toxicity. "Agnes" module available for R package was used for the clustering of genes (14). It computes agglomerative hierarchical clustering.

Expression Network

The expression network of the differentially expressed genes was constructed using an open-source tool GENEMANIA. (15) The network was constructed based on co-expression and shared protein domains & an investigative analysis was performed.

3. Results

Data Source

The data sets were obtained from GEO database. The datasets included both control and lead treated samples. The study (7) from where the datasets were collected had the following experimental design: Cell-lines were treated with 10mkM lead acetate for Day1 and was grown in RPMI on Day 2. 102 datasets were used among which 51 were controls. The control sets were not treated with lead acetate. In this study 8 datasets were used among which 4 were samples and 4 were controls.

Pre-processing of dataset for statistical treatments

MA plots give the graphical overview of the distribution of datasets. In the Figure1, it can be observed that the data after normalization is evenly distributed across intensity values but whereas the plot before normalization shows the scattered datasets. The RNA degradation plot represents the complete RNA degradation. (Figure2)

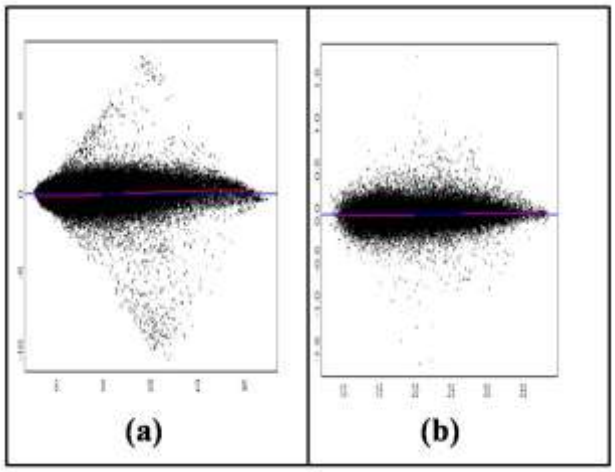

Figure 1: MVA PLOT

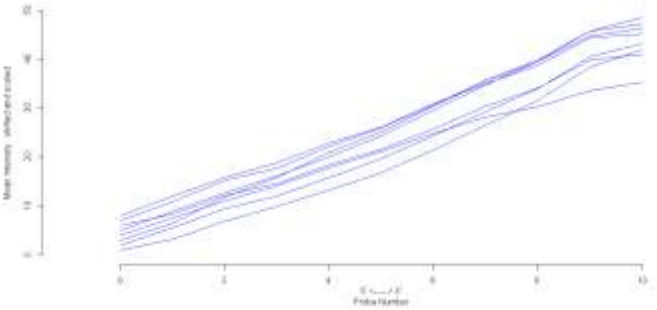

Figure 2: RNA Degradation plot

Differential Expression Analysis

In this study, 59 differentially expressed genes were identified such as MT1F, MT1G, MT1M, MT1H, MT1E, MT1X , MT2A, MT1M CRIM1, ALOX5, HPSE, THBD, CADM1, CDK5, BRCA2, CD93, CDK5 etc. All the genes were annotated at genomic level and their functional roles were identified. The main class of genes included Metallothioneins.

GO Enrichment

Gene ontology analysis was performed for the identified differentially expressed genes. This was carried out using the Gene Ontology database. Most of the genes such as Metallothionein 1 pseudogene 2, Metallothionein 1E, Metallothionein 1F, Metallothionein 1G, were involved in the binding process with the heavy metals such as lead, zinc and cadmium.

Clustering

Clustering analysis performed in R gave a hierarchical agglomerative clustering of complete linkage. The first cluster in the Figure 3 consisted of mainly Metallothionein groups and the whole cluster was mainly involved in the binding of heavy metals.

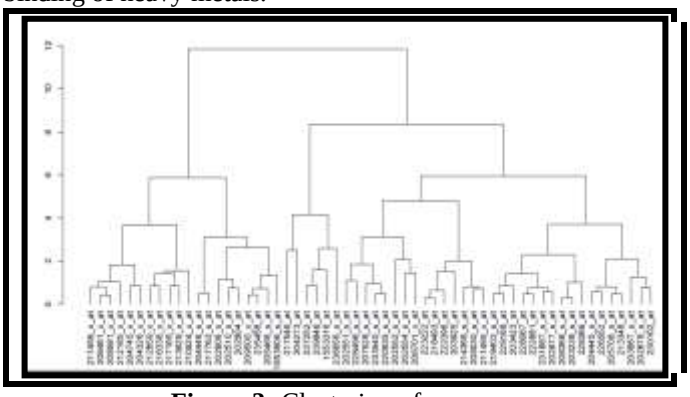

Figure 3: Clustering of genes

Expression Network

Gene Expression Network Analysis with GENEMANIA gave the following results: 7 genes were identified as a response to cadmium ion binding; 9 genes were identified in cellular response to metal ion; 7 genes in response to zinc ion binding; 9 genes response to inorganic substance binding; 11 genes in negative regulation of growth; 9 genes in response to metal ion binding; 11 genes in regulation of growth & 6 genes in response to copper ion binding (Figure 4). It was observed that Metallothionein family genes formed the core part of the network.

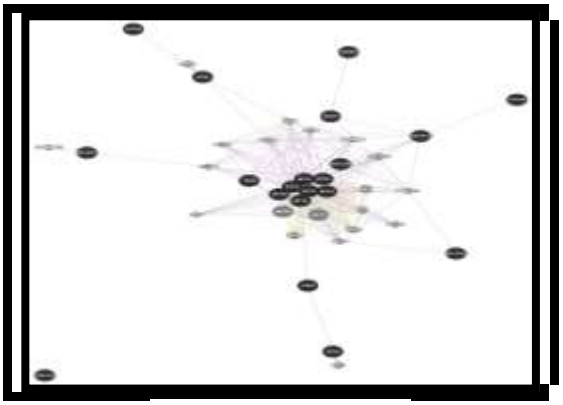

Figure 4: Gene network

4. Discussion

Elevated Lead levels in the water have a major impact in the health condition on the population along the river basin. A risk assessment of lead toxicity was done taking gene expression data of lead acetate induced cell-lines. The Gene expression data generated from a study on the lead acetate induced cell-lines was used as a data-source in our study. The data was preprocessed using RMA method and differentially expressed genes were identified using "Limma" package in R. The identified genes were annotated at genomic level and functionally. Gene ontology studies of the genes were performed and they were clustered using "Agnes" package in R. Clustering analysis showed that the Metallothionein family genes were co expressed. Expression network was built using GENEMANIA .The analysis of this network demonstrated that Metallothionein family of genes formed core part of the network. Disorders that are hallmark of Lead induced toxicity are nervous system damage, renal impairment, cognitive dysfunction and hypertension.

5. Conclusion

The study points out a strong correlation between clinical outcomes of lead toxicity & Metallothionein family of genes that formed core of the gene expression network, as delineated by Network Biology approach. Studies on similar lines involving other heavy metals such as cadmium, nickel, mercury can be useful in exposing the multifaceted machinery participating in impairment of gene regulatory network and its probable connotation in clinical settings.

6. Future Perspectives

Metallothionein genes are also involved in the binding of lead with high affinity and can be key molecular markers for early diagnosis of lead intoxicated subjects. Similar type of studies conducted on other heavy metal induced toxicity can bring to light the complex mechanism of dysfunction of gene regulatory network and its possible clinical outcomes.

References

[1] John Fawell and Mark J Nieuwenhuijsen (2003) Contaminants in drinking water Environmental pollution and health. *Br Med Bull*, 68 (1): 199-208

[2] Jayadev , E.T. Puttaih (2013) Studies on heavy metals contamination in Vrishabhavathi river water and ground water of the surrounding river. *IJSER*, 4(1), 1-9

[3] F.L.Kaushal , H.K. Parwana and S.P. Varma (1993) Effect of waste water irrigation on soilproperties. *Indian journal of Environmental protection* 13: 374-378

[4] Erin N. Kelly, David W. Schindler, Peter V. Hodson, Jeffrey W. Short, Roseanna Radmanovich, and Charlene C. Nielsen (2010) Oil sands development contributes elements toxic at low concentrations to the Athabasca River and its tributaries. *PNAS*, 107: 16178-16183

[5] Matthew K. Reuer, Nathan W. Bower, Jed H. Koball, et al., (2012), Lead, Arsenic, and Cadmium Contamination and Its Impact on Children's Health in La Oroya, Peru, *ISRN Public Health*, 2012: 1-2

[6] Howard Hu, Martha María Téllez-Rojo, David Bellinger, Donald Smith, Adrienne S. Ettinger, Héctor Lamadrid-Figueroa, Joel Schwartz, Lourdes Schnaas, Adriana Mercado-García, Mauricio Hernández-Avila (2006) Fetal Lead Exposure at Each Stage of Pregnancy as a Predictor of Infant Mental Development. *Environ Health Perspect*, 114(11): 1730–1735

[7] Bruce S Gillis, Zarema Arbieva and Igor M Gavin (2012) Analysis of lead toxicity in human cells. *BMC Genomics*, 13(344): 1471-2164

[8] Roel GW Verhaak, Frank JT Staal, Peter JM Valk, Bob Lowenberg, Marcel JT Reinders and Dick de Ridder (2006) The effect of oligonucleotide microarray data pre-processing on the analysis of patient-cohort studies. *BMC Bioinformatics* , 7 : 105

[9] Federico M Giorgi, Anthony M Bolger, Marc Lohse and Bjoern Usadel (2010) Algorithm-driven Artifacts in median polish summarization of Microarray data. *BMC Bioinformatics*, 11:553

[10] YH Yang, S Dudoit, P Luu, DM Lin, V Peng, J Ngai, TP Speed. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30 (4) pp. e15

[11] W. Gregory Alvord, Jean A. Roayaei, Octavio A. Quiñones, and Katherine T. Schneider (2007) A microarray analysis for differential gene expression in the soybean genome using Bioconductor and R. *Brief Bioinform*, 8: 415-431.

[12] K.Smyth G (2004) Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat Appl Genet Mol Biol*, 3

[13] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al. (2000) Gene ontology: tool for the unification of biology The Gene Ontology Consortium. *Nat Genet*, 25(1):25-29

[14] Struyf, A., Hubert, M. and Rousseeuw, P.J. (1997) Integrating Robust Clustering Techniques in S-PLUS. *Computational Statistics and Data Analysis*, 26:17–37

[15] David Warde-Farley, Sylva L. Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucl. Acids Res*, 38 (suppl 2): W214-W220

Author Profile

Arpitha B M is Assistant Professor in Dayananda Sagar College of Engineering. She has received the BE degree in 2012 and M. Tech degree in Bioinformatics in 2014. Her research interests are Structural Bioinformatics, Biostatistics and QSAR.