Muddassir et al (2013). Journal of Infection and Molecular Biology. 1 (1): 15 – 17.
http://www.nexusacademicpublishers.com/journal/2

Research Article

Study of Genetical Genomics in *InVivo* and *InVitro* genetics

Muhammad Muddassir Ali[1*], Muhammad Ijaz[2], Hassaan Bin Aslam[3], Furqan Awan[2]

[1]Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore 54600, Pakistan; [2]Department of clinical medicine and surgery, University of Veterinary and Animal Sciences, Lahore 54600, Pakistan; [3]Department of Microbiology, University of Veterinary and Animal Sciences, Lahore 54600, Pakistan.
[*]Corresponding author: muddassir.ali@uvas.edu.pk, Cell: 00923336252189

ARTICLE HISTORY

Received: 2013-03-04
Revised: 2013-03-10
Accepted: 2013-03-12

Key Words: Genetical Genomics, Gene expression data, Biolayout, GeneNet, DAVID

ABSTRACT

Calculating gene clusters by various clustering approaches or looking at the statistically significant variations between genes are usual ways to study microarray gene expression data. In recent, in vitro experiments are being conducted to use widespread applications of genetical genomics approaches in non-model and model species. Nowadays, many bioinformatics tools are under trails to make the correlation based networks to explore the genetic pathways. The twenty samples were from five uncultured samples of wild-type mice, five uncultured samples of Knock-out mice, five cultured samples of Knock-out mice and five cultured samples of wild-type mice. Network analysis was carried out from two Bioinformatics tools BioLayout and GeneNet after getting the differentially expressed genes by the microarray cDNA experiments. David database was used to conduct Gene set enrichment analysis for differentially expressed genes. There were no significant overlapping functional categories between the cultured and uncultured dataset after making the comparison of functional enrichment analysis results which suggested that cultured samples are not able to provide the same biological results as uncultured data.

ARTICLE CITATION: Ali MM, Ijaz M, Aslam HB, Awan F (2013). Study of genetical genomics in *In-Vivo* and *In-Vitro* genetics. *J. Inf. Mol. Biol.* 1 (1): 15– 17.

INTRODUCTION

Microarray expression data derive thousands of differentially expressed transcripts from just from very small sample size. Different factors may be involved like changes in the interests of biological system and technical artifacts to affect this sample and data obtained. General approaches which are utilized to analyze the microarray data are to derive clusters by many explorative clustering techniques or by statistically significant variations between the groups of samples to analyze the microarray data. Correlation based networks are mode to analyze the gene expression data and to see co expression in expression profiles of data. This idea have already been used by many authors and researchers (Franke *et al.*, 2006; Zhang and Horvath, 2005) by previously establishes statistical approaches like Pearson's correlation coefficient. Nowadays in-vitro studies are used to establish and study the genetical genomics in broad ranges. In genetical genomics, gene expression and marker genotype studies are done in a segregating population (Jansen, 2003; Rockman and Kruglyak, 2006). Genetic pathways are studied in gentical genomics (Mitche et al., 2004). Genetic Pathways are made in graphical forms in which individual genes (nodes) interact and make functional linkages with other genes (graph edges). Many tools of bioinformatics are used to make these genetic pathways in form of graphical conventions. But every tool has its own limitations in giving complete biological inference because of built in computational algorithms. Nowadays different *in vivo* and *in vitro* studies are made in combination to check perform experiments in wide range and to see the similarity in results (Haley and Koning, 2007).

A comparison can be conducted on gene expression datasets by using network based tools GeneNet and BioLayout. In this study, cultured (*in vitro*) and uncultured (*in vivo*) osteocytes to be differentially expressed genes of cultured and uncultured cells. GeneNet is an R package and compared are taken from two types of mice, the knock out (KO) mice for the gene suppressor of cytokine signaling (*SOCS2*) and wild type mice. This gene does the regulation of the Insulin-like growth factor 1 (IGF-1) mediated cell signaling pathway. Insulin-like growth factor 1 attaches to the Insulin-like growth factor 1 receptor (IGF1R). This receptor is available in different body cells and also in osteocytes. After the binding of IGF-1 to this receptor a pathway is initiated this then provokes the growth of bone by cell division and multiplication (Dey et al., 1998). Aim of the study was to make the comparison between gene knock-out signature for the *SOCS2* cultured and uncultured data to find out the similarities by using the data-driven, network-based approaches. So, BioLayout and GeneNet was used to make correlation based networks and after getting the results from these tools, DAVID analysis was done to make the comparison of gene annotation results of cultured versus fresh samples to find similarities in the results.

MATERIALS AND METHODS

Fresh and cultured osteocytes were used from five mice of knocked out gene (*SOCS2*) and from five wild-type mice. Total twenty samples were taken, five uncultured samples of wild-type mice, five uncultured samples of knock-out mice, five cultured samples of knock-out mice and five cultured samples of wild-type mice. The samples were subjected to microarray experiments and differentially expressed genes were obtained. There were 45,102 differentially expressed genes but only 1500 most significantly differentially expressed genes were used in the dataset by sorting on the basis of P-value in Microsoft Excel 2007.

The softwares GeneNet (Schäfer *et al.*, 2006) and BioLayout (Freeman *et al.*, 2007) were used to make correlation based regulatory networks for the 1500 most significantly

uses the graphical Gaussian models (GGMs) while BioLayout makes the clusters by Pearson correlation coefficient. The interaction dataset containing the gene IDs of both cultured and uncultured and genotype (wt or KO) were loaded into BioLayout. The correlation threshold 0.90 was taken to obtain the maximum number of clusters. Markov clustering algorithm (MCL) was run which is an inbuilt algorithm in this software and derive the cluster nodes by computing probability simulation. Then clusters and network was viewed by using the cluster viewer feature. Overlapping transcripts IDs among all clusters taken and subjected to the DAVID database for gene set enrichment analysis. Gene ontology tool was used to do the analysis in DAVID.

The same analysis was also conducted by using same dataset in GeneNet. Before running the GeneNet, installation of corpcor, locfdr, longitudinal and fdrtool was done (Efron, 2004). The dataset was formulated in pseudo-time series data in GeneNet R package by treating combinations of four genotype-treatments as a time point. Then the generated network topology text file from GeneNet was obtained and loaded into Cytoscape (Killcoyne *et al*., 2009) for making visualization of generated clusters and networks from GeneNet. Similarly DAVID analysis was conducted for the generated clusters in the same way as it was done in BioLayout. To get the significant annotations only P-values less than 0.05 was taken and finally a comparison was made between the significant annotations of the BioLayout and GeneNet results.

RESULTS AND DISCUSSION

General comparison between cultured and fresh dataset by using Microsoft Excel (2007)

When general comparison was conducted between cultured and uncultured datasets by using Microsoft Excel (2007), only 128 genes found to be overlapping among 1500 genes. DAVID analysis revealed following functional categories negative regulation of cell cycle, antral ovarian follicle growth, regulation of organelle organization. Negative regulation of nuclear division, regulation of transcription,DNA-dependent, response to hormone stimuli and endoplasmic reticulum.

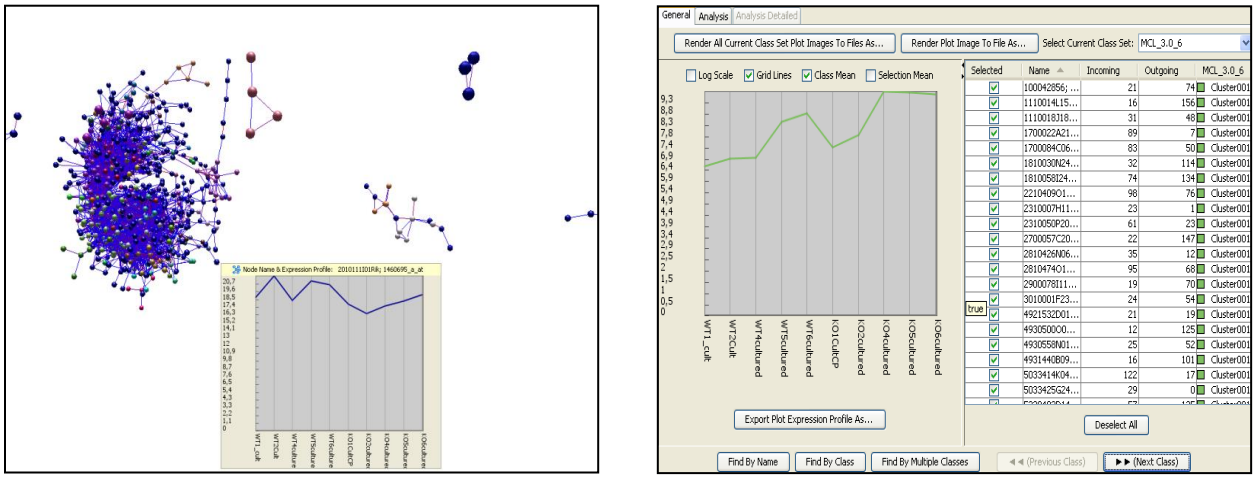

Figure 2: BioLayout, SnapShot. At left side, network snapshot in BioLayout, a main large network having nodes and edges is represented. At right side, a graph along with list of genes in form of cluster is shown.

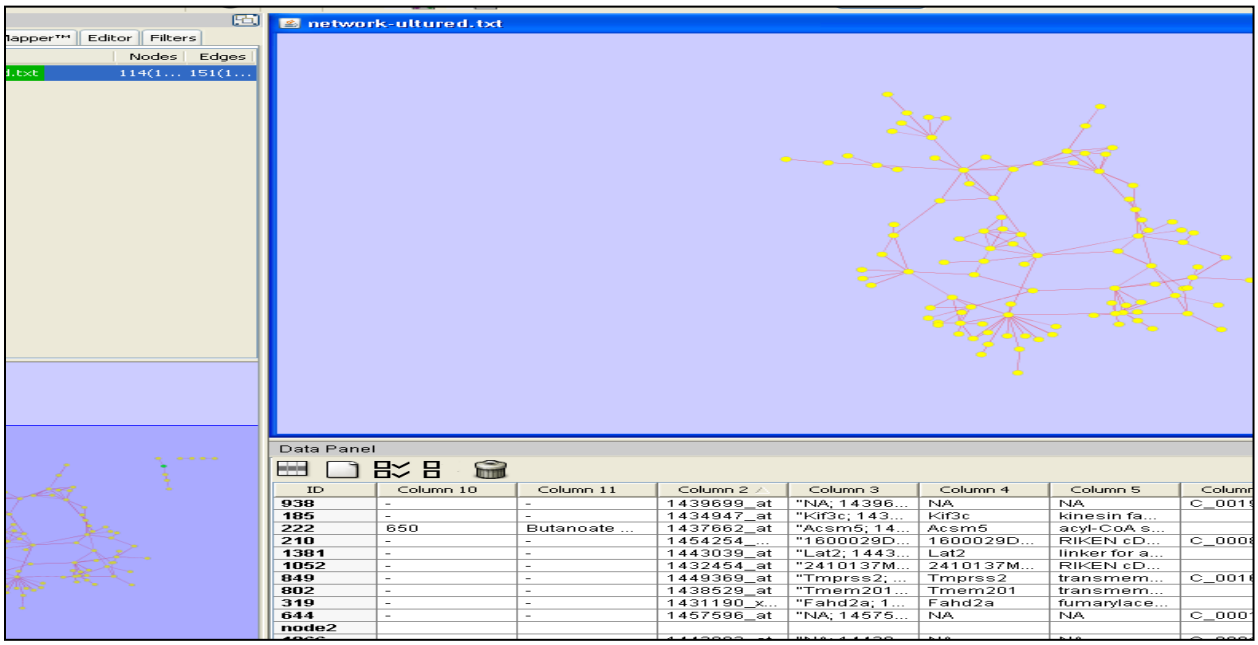

Figure 3: Snap shot of Cytoscape results. Network showing the generated nodes and edges in GeneNet.

Network Inference using BioLayout

By conducting BioLayout analysis a total of 738 nodes, 7625 edges and 38 clusters were obtained for the interaction dataset containing the gene IDS for both cultured and uncultured datasets. Graphs were also generated in BioLayout that were showing the gene expression pattern in clusters of the cultured KO and wild-type

data. In the graphs, the genes which were more expressed in wild type or KO mice had a peak whereas low expressed genes had a downward pattern in the graphs. Network was showing nodes and edges and clusters were representing the co-expressed genes in list form. There were peaks for the *Socs2* gene in wild-type mice representing a higher expression while there was lower expression in KO mice (Figure 2). The results of present study clearly showed high prevalence of MRSA in hospitals of district Kohat. The antibiotics used in this study are frequently prescribed by clinicians without carrying out anti-biogram assays on samples of their patients. Moreover, common people have easy access to these antibiotics available to them at local pharmacies without any prescriptions from doctors. Therefore, the rise in resistance is clearly due to irrational use of antibiotics and non existence of protocols and guidelines for safe practice in terms of treatment of MRSA.

Network Inference using GeneNet

GeneNet generated 148 nodes and 150 edges. The network file was loaded in to Cytoscape for making visualization and this generated information in form of network and cluster. In Cytoscape a big network was observed. The genes which were having the functional interactions with each other were in one network (Figure 3).

FUNCTIONAL ENRICHMENT STUDY FOR COMPARISON OF CULTURED AND FRESH DATASETS USING DAVID

Comparison of interaction dataset for BioLayout and GeneNet

Only two terms, "cytoplasm and cytoskeleton", were overlapped in all categories having a P-value less than 0.05 for the interaction dataset. The P-values for cytoskeleton and cytoplasm were 0.0047 and 0.035 in BioLayout, whereas in Gene Net they were 0.027 and 0.025, respectively. Similarly, it was also observed that there were different genes involved in same processes of both tools as it was observed in comparison of fresh and cultured dataset in BioLayout (Table 1&2).

BioLayout have the ability to make the large graphs along with hundreds of edges and nodes by using the pairwise Pearson correlation coefficients and by sophisticated layout algorithm. GeneNet, on the other hand, calculates the nodes and edges by partial correlation matrices and uses the Benjamini and Hochberg's approach (Benjamini and Hochberg's, 1995).

In GeneNet a very low number of edges and nodes are observed due the use of partial correlation measures because it calculates the linear correlation between two genes by removing the effect of any distinct correlation effect of any other genes. A higher number of edges and nodes are obtained in BioLayout due to the use of Pearson correlation which calculates the correlations among genes by even directing the indirect correlations of all other genes.

From the results it was obvious that general comparison of the cultured and uncultured data from the DAVID database without using the network approach revealed that there were significant functional categories like regulation of organelle organization, regulation of cell cycle, regulation of transcription, response to hormone stimulus etc. From previous literature it is clear that *SOCS2* have a major role in many pathways like regulation of growh hormones (Turnely, 2005), activation of transcription 5b target in liver (Vidal *et al*., 2007) and regulation of organization in the growing skeleton (Macrae *et al*., 2009) and in growth regulation (Greenhalghand, 2005) which supports our findings.

CONCLUSION

Our findings showed that there were no significant common and overlapping functional categories when comparing the cultured and uncultured datasets. Based on these findings it is suggested that cultured samples cannot be used as an alternative to uncultured samples in getting the useful biological information for studying genetical genomics. Still, more verifications and repetitions are needed to confirm these findings by using different combination of bioinformatics tools and by using different kind of *in vivo* and *in vitro* datasets.

Table 1: Genes involved in cytoskeleton activity for both cultured and fresh dataset in Bio Layout and GeneNet

Genes involved in cytoskeleton activity in interaction dataset in Biolayout	Genes involved in cytoskeleton activity in interaction dataset in GeneNet
1 ADP-ribosylation factor-like 2 2- CDC42 effector protein (Rho GTPase binding) 1 3- Sfi1 homolog, spindle assembly associated (yeast). 4- centrosomal protein 164.	1- DNA methyltransferase 3B. 2- SMEK homolog 2, suppressor of mek1 (Dictyostelium). 3- centrosomal protein 250 4- pericentriolar material 1,

Table 2: Genes involved in cytoplasm activity for both cultured and fresh dataset in Bio Layout and GeneNet

Genes involved in cytoplasm activity in interaction dataset in GeneNet	Genes involved in cytoplasm activity in interaction dataset in Biolayout
1 Aldehyde, dehydrogenase family 1 2- subfamily A2, RAN GTPase activating protein 1 3- RAN GTPase activating protein 1 4- RAN binding protein 17 5- ADP-ribosylation factor 3 6- progesterone receptor, testis-specific serine kinase 2 7- ring finger protein 17 8- retinoic acid induced 14	1- Threonyl-tRNA synthetase-like 2 2- CDC42 effector protein (Rho GTPase binding) 1 3- centrosomal protein 250 3- annexin A6, centromere protein J 4- mindbomb homolog 2 (Drosophila) 5- rho/rac guanine nucleotide exchange factor (GEF) 18

REFERENCES

Benjamini and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B. 57: 289–300

Dey BR, Spence SL, Nissley P and Furlanetto RW (1998). Interaction of human suppressor of cytokine signaling (SOCS)-2 with the insulin-like growth factor-I receptor. J. Biol. Chem. 273: 24095–101.

Efron B (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. J. Am. Statist.Assoc. 99: 96–104.

Enright AJ, Kunin V and Ouzounis CA (2003). Protein families and TRIBES in genome sequence space. Nucleic Acids Res. 31: 4632–4638.

Enright AJ, Van Dongen S and Ouzounis CA (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30: 1575–1584.

Franke L, Van Bakel H, Fokkens L, De Jong ED, Egmont-Peterson M, and Wijmenga C (2006). Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am. J. Hum Genet. 78: 1011 – 1025.

Freeman TC, Goldovsky L, Brosch M, Van Dongen S, Mazière P, Grocock RJ, Freilich S, Thornton J, and Enright AJ (2007). Construction, Visualisation and Clustering of Transcription Networks from Microarray Expression Data. PLoS Comput Biol. 3(10): 206.

Greenhalgh CJ, Rico-Bautista E, Lorentzon M, Thaus AL, Morgan PO, Willson TA, Zervoudakis P, Metcalf D, Street I, Nicola NA, Nash AD, Fabri LJ, Norstedt G, Ohlsson C, Flores-Morales A, Alexander WS and Hilton DJ (2005). SOCS2 negatively regulates growth hormone action in vitro and in vivo. J Clin Invest.115: 397-406.

Haley CS and de Koning DJ (2007). Towards in vitro genetics. Trends Genet.10: 1016.

Haley CS, de Koning DJ and Cabrera CP (2007). Genetical Genomics: Combining Gene Expression with Marker Genotypes in Poultry. Poult Sci. 86: 1501–1509.
Jansen RC (2003). Studying complex biological systems using multifactorial perturbation. Nat. Rev. Genet. 4: 145–15.
Killcoyne SGW, Carter JS, and Boyle J (2009). Cytoscape: a community based framework for network modeling. Methods Mol. Biol. 563: 219-239.
Li L, Stoeckert CJ Jr and Roos DS (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13: 2178–2189.
Macrae VE, Horvat S, Pells SC, Dale H, Collinson RS, Pitsillides AA, Ahmed SF and Farquharson C (2009). Increased bone mass, altered trabecular architecture and modified growth plate organization in the growing skeleton of SOCS2 deficient mice. J Cell Physiol. 18: 276-84.
Raponi M, Belly RT, Karp JE, Lancet JE, Atkins D and Wang Y (2004). Microarray analysis reveals genetic pathways modulated by tipifarnib in acute myeloid leukemia. BMC Cancer. 4: 56.
Rockman MV and Kruglyak L (2006). Genetics of global gene expression. Nat. Rev. Genet. 7: 862–872.
Schäfer J, Rainer O and Strimmer K (2006). Reverse Engineering Genetic Networks using the GeneNet Package. R News. 6: 50-53.
Theocharidis A, Van Dongen S, Enright AJ and Freeman TC (2009). Network visualization and analysis of gene expression data using BioLayout *Express* 3D. Nature Protocols. 10: 1535.
Turnley AM (2005). Role of SOCS2 in growth hormone actions. Trends Endocrinol Metab. 16: 53-8.
Vidal OM, Merino R, Rico-Bautista E, Fernandez-Perez L, Chia DJ, Woelfle J, Ono M, Lenhard B, Norstedt G, Rotwein P and Flores-Morales A (2007). In vivo transcript profiling and phylogenetic analysis identifies suppressor of cytokine signaling 2 as a direct signal transducer and activator of transcription 5b target in liver. Mol Endocrinol. 21: 293-311.
Zhang B and Horvath S (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4: 17.