Advances in Bioresearch
Adv. Biores. 8 [Spl issue 1] 2016: 84-88
©2016 Society of Education, India
Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
ICV 86.92

Advances in Bioresearch

Full Length Article

A Network biology based approach for elucidation of Type II Diabetes Pathobiology using transcriptomics data

[1,2,*]Aditya Saxena, [3]Kumar Sachin and [1]Ashok Kumar Bhatia
[1]Department of Biotechnology, Institute of Applied Sciences & Humanities, GLA University, Mathura (U.P.)
[2]Uttarakhand Technical University, Dehradun (U.K.) INDIA
[3]Department of Biochemistry and Biotechnology, S.B.S. (PG) Institute of Biomedical Sciences & Research, Dehradun (U.K.) INDIA
*Corresponding Author: aditya.saxena@gla.ac.in , +91-8171630011
Full Address: GLA University, Mathura, 17 km stone, NH-2, P.O. Chaumuha, Mathura- 281 406 U.P. INDIA

ABSTRACT
Diabetes mellitus is a medical condition that is characterized by poor glycemic control, and low insulin responsiveness to target tissues that eventually leads to lack of insulin secretion by pancreatic β cells. In present study we used a network-based approach for elucidation of pathways and biological process implicated in diabetic liver through analyzing a publically available microarray dataset through open-source software Bioconductor and a web-based tool NetworkAnalyst. a microarray study was from Kanazawa University Hospital, Kanazawa, Japan was selected to analyze comprehensive gene-expression profile in hepatic tissues. Our network-based KEGG Analysis reported various pathways suggestive to type 2 diabetes pathobiology. As type 2 diabetes is a complex disease whose mechanism is still poorly known; indeed genes directly implicated to insulin-signaling pathways are generally inconspicuous in global gene-expression analysis above statistically-significant thresholds. So In our present study we tried to uncover T2D patho-biology using network-based approach and we will further warrant for development of new bioinformatics workflows in order to identify unknown aspects of this complex disease.
Keywords; *Type 2 Diabetes, NetworkAnalyst, Microarra, network biology*

Received 07/10/2016 Accepted 24/12/2016

How to cite this article:
A Saxena, K Sachin and A K Bhatia. A Network biology based approach for elucidation of Type II Diabetes Pathobiology using transcriptomics data. Adv. Biores., Vol 8 [Spl issue 1] 2016: 84-88. DOI:10.15515/abr.0976-4585.SI8488

NTRODUCTION

Diabetes mellitus type 2 is a medical condition that is characterized by poor glycemic control, and low insulin responsiveness to target tissues that eventually leads to lack of insulin secretion by pancreatic β cells [1]. Deregulation in insulin signaling at the molecular level is attributed to T2D that gives rise to different pathological phenotypes, viz. Normal Glucose Tolerance (NGT), Impaired Glucose Tolerance (IGT), Hyperinsulinemia (HI), and Insulin Resistance (IR) that eventually leads to overt diabetes. Various microarray studies have been conducted in past to study global gene-expression pattern in order to identify novel biomarkers, drug targets and deciphering new insights in diabetes pathobiology. Liver plays an important role in glucose homeostasis through modulation of sensitivity/resistance of peripheral tissues to insulin and deregulation in this process may leads to Type 2 diabetes.

In present study we used a network-based approach for elucidation of pathways and biological process implicated in diabetic liver through analyzing a publically available microarray dataset through open-source software Bioconductor and a web-based tool NetworkAnalyst.

MATERIALS AND METHODS

We have downloaded a microarray dataset GSE23343 from Gene Expression Omnibus (GEO) [2]. This study was conducted at Kanazawa University Hospital, Kanazawa, Japan to analyze comprehensive gene-expression profile in hepatic tissues of ten patients with Type 2 Diabetes and seven subjects with Normal

Glucose Tolerance using Affymetrix Human Genome U133 Plus 2.0 Array [3]. We have used 'R' an open source Statistical software environment (www.r-project.org), used for Microarray data analyses particularly using a special set of packages available at Bioconductor (www.bioconductor.org). Package GEOquery was used to fetch intensity-level *.cel* files in R Environment using function *getGEO*. Intensities values were $\log_2$ transformed, normalized and visualized as Box Plot (Fig. 1).

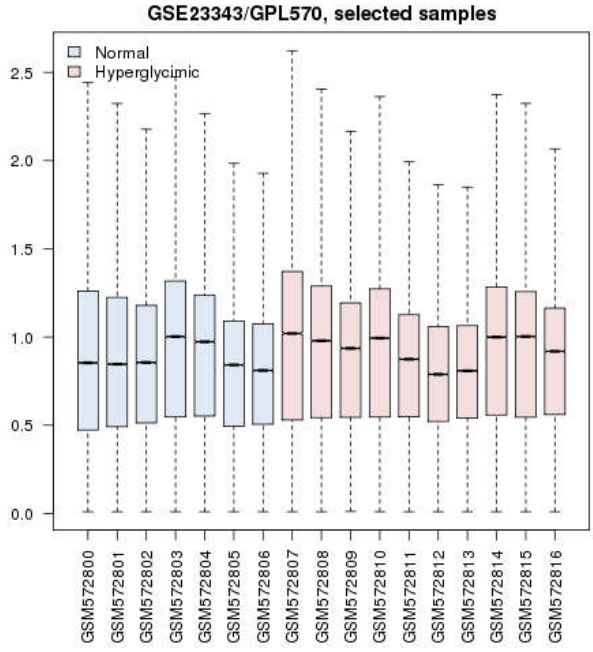

Figure 1 Box Plot of samples

Differential gene expression between two phenotypes was estimated using package *limma* [4] that implements an empirical Bayes moderation approach based modified *t-statistics*. A list of top 250 genes

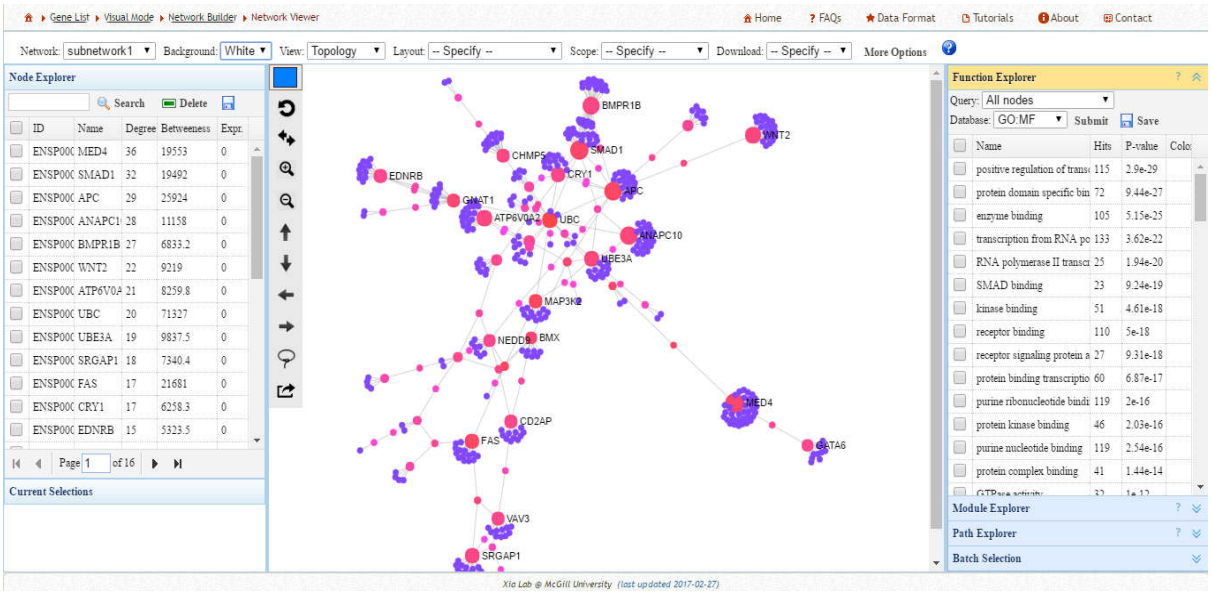

Figure 2 Network viewer step in NetworkAnalyst

was prepared containing summary-level Meta data (*Log Fold change, P-value, Adjusted P-value, and t statistics*) pertaining to Differential Gene Expression in which 46 affymetrix ids were failed to map to any official gene symbol. To infer the biological meaning of this list, we have adopted a network-biology approach because it promise to include additional genes in the functional analysis that might have been implicated in the disease phenotype but failed to be reported in original gene list due to either low expression value, absence on microarray chip, experimental error or failure to cross statistical thresholds.

We used NetwrokAnalyst, a comprehensive web-based tool designed to support statistical, visual and network-based meta-analysis of gene expression data. [5] For a single list of seed genes, it offers functionalities for the analysis of Protein-protein Interactions, Gene-miRNA Interactions, TF-gene Interactions, Protein-drug Interactions, and Protein-chemical Interactions. Seed Genes were mapped to STRING v10 interactome [6] and a first-degree Protein-Protein Interaction Network was built with a confidence score cutoff of 900. Largest network of 451 nodes was selected containing 45 seed gene and downloaded in SIF format. NetworkAnalyst calculate two estimates of node centrality: *Degree Centrality* that measure the number of connections that a node has to other nodes and *Betweenness Centrality* that measure the numbers of shortest paths passing through the node. (See Appendix 2 for table showing degree, and *betweenness centrality* of network nodes).

42 out of top 52 nodes were found to be seed nodes pointing toward the fact genes in the original list were infect functionally involved with each other and it is highly likely that deregulation in this interaction result in the T2D phenotype. Nodes with Degree-centrality > 4 were selected for Gene Ontology based Biological Process, Molecular Function and Cellular Component analysis using WEB-based Gene Set Analysis Toolkit [7].

Majority of these genes were membrane proteins and involved in protein binding related to regulatory processes. To identify enriched functions of the resulting network, Function Explorer was used for KEGG

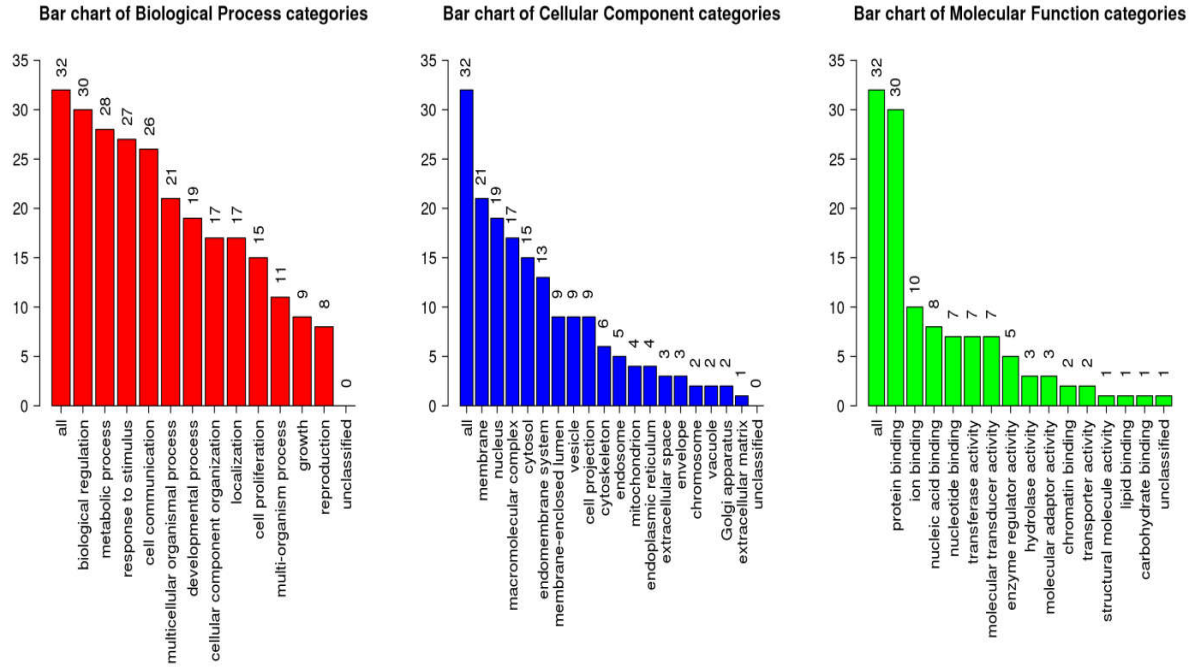

Figure 3 GO-BP, CC, and MF analysis of network genes with degree-centrality > 4

Pathways. Table 1 shows analysis results for top 10 KEGG pathways, ranked as per their associated *P-values* after due correction for False Discovery Rate.

RESULTS AND DISCUSSION

To interpret list of genes obtained through microarray experimentation or any high throughput method is a challenging task, firstly microarray data is noisy; it contains technical and biological variability. So it is a better option to look at differentially expressed genes in the context of protein-protein interaction network. Because node centrality of a gene/protein might be a better measure of its significance then its numerical expression values as low-expressed gene are also found to play important roles in disease pathogenesis through their connected partners.

Our network-based KEGG Analysis reported top most pathway “Wnt signaling pathway”; the role of Wnt signaling in metabolic homeostasis and its implication in development of diabetes and other metabolic diseases has recently been uncovered. Wnt signaling involves bipartite transcription factor beta-catenin/TCF7L2 and TCF7L2 has been ascribed as T2D susceptible gene through various genome-wide association studies. [8].

Table 1; pathways enriched in Network based analysis using KEGG

S. No.	Pathway	Total	Expected	Hits	P. Value	FDR
1	Wnt signaling pathway	144	8.83	57	2.26E-33	4.91E-31
2	Pathways in cancer	310	19	76	1.81E-28	1.96E-26
3	HTLV-I infection	199	12.2	57	7.48E-25	5.41E-23
4	Basal cell carcinoma	47	2.88	29	2.96E-24	1.61E-22
5	Melanogenesis	101	6.19	40	1.89E-23	8.20E-22
6	TGF-beta signaling pathway	84	5.15	35	1.77E-21	6.40E-20
7	Circadian rhythm - mammal	22	1.35	16	1.43E-15	4.43E-14
8	Chagas disease (American trypanosomiasis)	89	5.46	30	1.68E-15	4.55E-14
9	Hedgehog signaling pathway	56	3.43	24	2.30E-15	5.55E-14
10	Cell cycle	124	7.6	33	1.41E-13	3.05E-12

Another pathway that was observed among top pathways in all tissue studies was 'TGF-β signaling pathway'. TGF-β signaling is involved in the regulation of insulin gene transcription, pancreatic β cells function, and glucose tolerance and energy homeostasis.[9-12] *in-vitro* studies have confirmed its role in inhibition of insulin-stimulated tyrosine phosphorylation of insulin receptor substrate-1, [13] moreover; it also inhibits insulin/insulin-like growth factor-1 (IGF-1)-dependent adipose cell differentiation (adipogenesis) process. TGF-β exerts this effect through induction of extracellular matrix (ECM) synthesis, which in turn inhibits adipogenesis and leads towards metabolic syndrome.

Another pathways "pathways in cancer" was also found linked with T2D pathophysiology. Several studies have analyzed the association between diabetes and cancer development and T2DM has been reported to acts as a predictor of mortality from cancer of the colon, pancreas, female breast, male liver and bladder. [15] Other enriched pathways "Basal cell carcinoma" and "Cell cycle" also corroborate to this fact.

CONCLUSION

Type 2 diabetes is a complex disease whose mechanism is still poorly known; indeed genes directly implicated to insulin-signaling pathways are generally inconspicuous in global gene-expression analysis above statistically-significant thresholds. So In our present study we tried to uncover T2D patho-biology using network-based approach and we will further warrant for development of new bioinformatics workflows in order to identify unknown aspects of this complex disease.

REFERENCES

1. Global report on diabetes, (2016). World Health Organization, ISBN 978 92 4 1565257.
2. Edgar R, Domrachev M, Lash AE. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. Jan 1;30(1):207-10
3. Misu H, Takamura T, Takayama H, Hayashi H et al. (2010) .A liver-derived secretary protein, selenoprotein P, causes insulin resistance. Cell Metab; 12(5):483-95. PMID: 21035759.
4. Ritchie, M.E., Phipson, B., Wu, D. et. al.(2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 43(7): e47.
5. Xia J, Gill E, and Hancock REW (2015) "NetworkAnalyst for Statistical, Visual and Network-based Approaches for Meta-analysis of Expression Data" Nature Protocols 10, 823–844
6. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan; 43(Database issue):D447-52. doi: 10.1093/nar/gku1003.
7. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. [2013] Jing Wang, Dexter Duncan, Zhiao Shi, Bing Zhang. Nucleic Acids Res. ; 41(Web Server issue): W77–W83. doi: 10.1093/nar/gkt439
8. Wilfred Ip, Yu-ting Alex Chiang, Tianru Jin [2012] The involvement of the wnt signaling pathway and TCF7L2 in diabetes mellitus: The current understanding, dispute, and perspectiveCell Biosci. 2: 28.
9. Tan CK, Leuenberger N, Tan MJ, *et. al.* (2011). Smad3 deficiency in mice protects against insulin resistance and obesity induced by a high-fat diet. Diabetes. 60: 464–76.
10. Yadav H, Quijano C, Kamaraju AK, *et. al.* (2011). Protection from obesity and diabetes by blockade of TGF-b/Smad3 signaling. Cell Metab. 14: 67–79.
11. Brown ML, Schneyer AL. (2010). Emerging roles for the TGF beta family in pancreatic beta-cell homeostasis. Trends Endocrinol Metab. 21: 441–48.
12. Lin HM, Lee JH, Yadav H, *et al.* (2009). Transforming growth factor-beta/Smad3 signaling regulates insulin gene transcription and pancreatic islet beta-cell function. J Biol Chem. 284: 12246–57.
13. Gagnon AM, Chabot J, Pardasani D *et. al.* (1998). Extracellular matrix induced by TGFβ impairs insulin signal transduction in 3T3-L1 preadipose cells. Journal of Cellular Physiology. 175(3): 370–378.

14. Coughlin SS, Calle EE, Teras LR, Petrelli J, Thun MJ. (2004). Diabetes mellitus as a predictor of cancer mortality in a large cohort of US adults. Am J Epidemiol. 159:1160–7.