Original Research Article

Microarray cDNA Dataset Analysis Reveals Potential Genes Associated with Type 2 Diabetes Mellitus for Further Treatment Exploration

ABSTRACT

Type 2 diabetes mellitus (T2DM) is an intricate and inadequately treatable metabolic disorder that requires modified treatment by identifying genetic variants as potential drug targets. In this study, we performed the system-level genetic analysis of the T2DM-related cDNAdataset and revealed 5 significant differentials expressed genes (DEGs) including ABRA, CYR61, NR4A1, KY, and TMEM131 as source genes.Among, these genes, 3 were down-regulated and 2 up-regulated. The biological function and gene ontology showed the association of these genes with cell apoptosis, cell communication, signal transduction, and insulin resistance. These genes are majorly expressed in multiple tissues specifically the brain, lungs, pancreas, and immune cells. The protein-protein network revealed the interaction of these source genes with important signature proteins including FOS, IGFN1, UBC, CTNB1, ITB5, JUN, HIF1A, p53, and other important genes. This study would be helpful to understand the etiology of T2DM and also improve the development of new drug treatments by identification of genes associated with T2DM.

KEYWORDS: DEGs, Cluster analysis, PPI network, T2DM

1. INTRODUCTION

Type 2 mellitus diabetes (T2DM) is a complex insulin-resistant metabolic disease. Overall, more than 285 million people are estimated to suffer from T2DM, representing about 90% of diabetes and Pakistan ranks seventh in the list of all countries(Muhammad et al., 2017).The survey in Pakistan shows that diabetes covers 5.1men and 6.8 percentof women in urban areas are suffering fromT2DM [2].The pancreas (β cells and adipose cells), kidneys, skeletal muscle, the brain, Langerhans islets, and adipose tissue are included in the organ for type 2 diabetes(Chatterjee et al., 2017). Although, many key factors involved which lead to the progression of insulin resistance are diet, obesity lifestyle, and exercise (Muhammad et al., 2017). Diabetes mellitus shows some effects like polyuria, blurring of vision, weight loss, and thirst (Organization, 1999).Several genes involve type 2 diabetes and each contributes to a greater chance of disease[5]and similar T2DM genes do not indicate proper insulin signaling pathways(Florez, 2008).

T2DM itself is a polygenicdisorder that develops because of the complex interaction of several genes with each other. It is still poorly understood and has not precisely defined how these genes interact to cause T2DM [7].Genomic expression and integrated pathways in insulin signaling may prove and could lead to clinically critical resistance to insulin and diabetes. The frameworks science approach possibly coordinates these natural systems and will help in uncovering key components associated with pathogenesis. As hereditary articulation is indispensable to the more likely comprehend the system of frameworks science, in this manner,cDNA microarray innovation is an important instrument for breaking down articulation dimensions of thousands of qualities in the meantime. The large number of expressional datasets in the open area gives a rich source of genome-wide T2DM data and offers the opportunity to find new genetic variants.In this analysis, we conducted a differential study to show the objective quality marks of insulin resistance and T2DM. A system-level analysis is presented to examine the complexity and potential function of the relationships between T2DM genes at the process level, as per genomic variations [8].In our investigation, we developed metabolic pathways to reveal new medication targets. We started the investigation by pointing to insulinexpression and related cell qualities, a characteristic and settled possibility for finding a mark set of qualities related to insulin obstruction or diabetes. The system set up in this paper is intended to center key inquiries: (1) Can natural procedures be perceived that is deregulated in metabolic pathways of insulin obstruction and diabetes (2) Can hereditary communication systems be useful to uncover new medication targets and biomarkers for advancing the treatment methodologies (Muhammad et al., 2017).

A quite valuable tool was present to find out variations in frequency and progress of T2DM by using the Microarray technique[9].Currently, many biomarkers for T2DM are established for diagnosis and treatment byemploying gene expression techniques.Data in public repositories are progressively increasing from gene expression studies in T2DM, which allows for the construction of a pooled dataset of gene expression comprising more people.An analysis of microarray differential expression will be worthwhile for future advances in T2DM diagnosis.This study, therefore, endorses the hypothesis that genes with different expressions can be useful biomarkers or therapeutic targets.

2. MATERIAL AND METHOD

2.1 Microarray data

Gene Expression Omnibus (http:/www.ncbi.nlm.nih.gov/geo/) has been accessed to CEL data set files containing 10 samples of disease and 40 standard samples. GPL570 chip [HG-U133 Plus 2] Affymetrix Human Genome U133 plus 2.0 Array (HG-U133 Plus 2), Inc., Santa Clara and CA, 95051, USA Technologie: In situ oligonucleotide) and genetic samples that can be used for the gene expression levels of the individual array are functional annotation facts (hgu133plus2).

2.2 Differential expression analysis

Collected and modified raw and pheno data files into R-recognizable, identifiable format, and errors were removed(Troyanskaya et al., 2001).To execute the normalization and quality control to process the probe intensity values using Robust Multi-Array Average (RMA) background and each probe indicates a gene, the R software with Bioconductor packages (the QualityMetrics)were used(Bolstad et al., 2003; Fujita et al., 2006; Obenchain et al., 2014).The RNA degradation analysis was conducted to monitor the quality of RNA by using AffyRNAdeg(Bolstad, 2008; Lu, 2004).After normalization, statistics of the software R have been performed to detect genetic variants (DEGs), by comparing normal to case samples, and several Benjamini-Hochberg method corrections have been carried out(Genovese et al., 2002).The substantial cutoffs were considered to be the false discovery rate (FDR) lower than 0.05, p-value 0.05, the average level of expression (AEL) in the order of 40 percent, and absolute logFC>1(Jin & Da, 2013).

2.3 Cluster and functional enrichment analysis

An expression data in a particular sample was used to validate changes in gene expression between T2DM tissue samples and normal samples(Muhammad et al., 2017).Enrichment analysis (Nam & Kim, 2008)investigated cellular functions(Nam & Kim, 2008).A significant cutoff has been determined by the functional annotation of these DEGs using the online FunRich tool (Huang et al., 2008) and the p-value EMBL-EBI databases of < 0.05.

2.4 Interactomic analysis

Proteins interact to perform biological functions with each other(Li et al., 2004).The application of the STRING(Kuhn et al., 2010) and HAPPI databases(Chen et al., 2009) were therefore used

to develop the protein-protein interaction (PPI) network using the Cytoscape software(Cline et al., 2007).

3. RESULTS

3.1 Normalization, RNA degradation plots, and differentially expressed genes (DEGs)

The sample level data show that the expression levels of the genes were standardized and each gene was located in a sample set with a 5′-end.The intensity of expression of each gene in the sample was reported in MA plots (Fig. 1a). M and A are shown as: M = log2(I1) – log2(I2), A = 1/2, where I1 is the array intensity studied and I2 is that of an array consisting of a median "pseudo" array across all arrays.

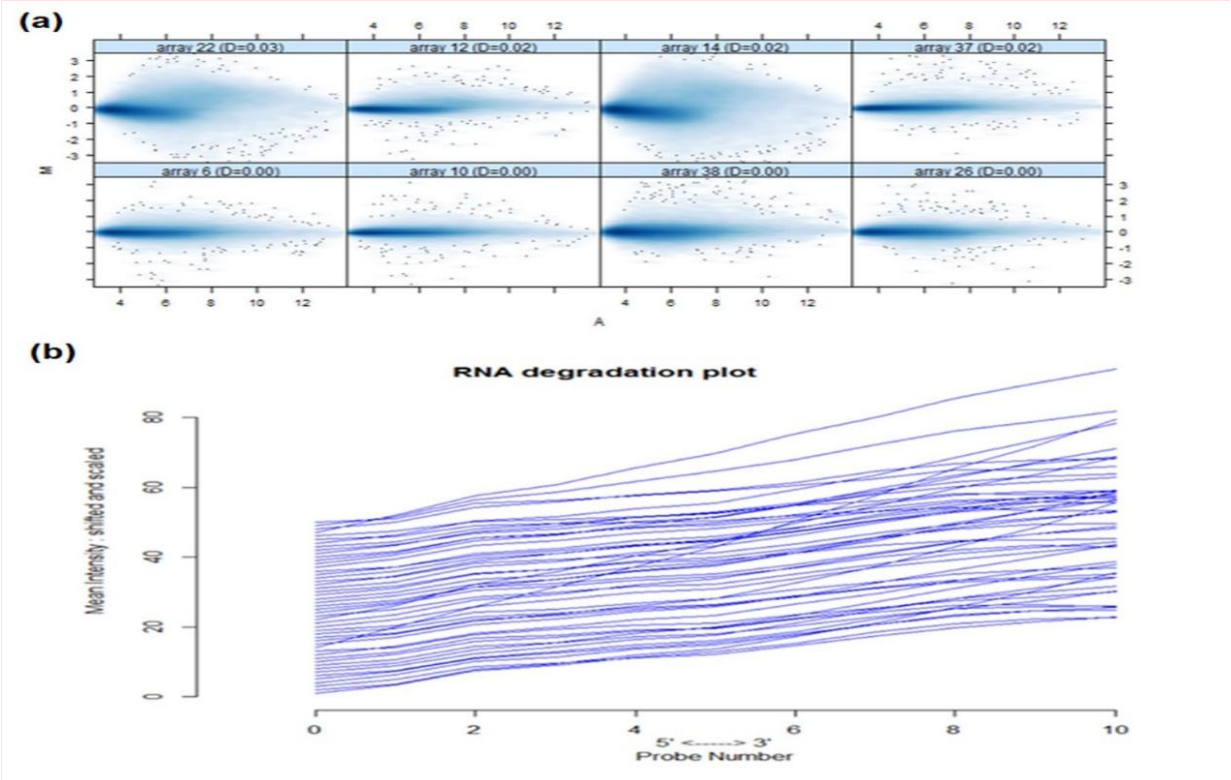

Fig. 1.(a) Genetic expression level of individual array quality (b) Side-by-side plot produced by plot AffyRNAdeg representing 5’ to 3’ trend.

The degradation of RNA is an integral part of cell mechanisms to monitor the removal of aberrant mRNAs or to generate mature 3′-5′ RNA processing transcripts. A plot was created side-by-side (Fig. 1b) for the AffyRNAdeg function plot, and the AffyRNAdeg function summary produces separate statistic summaries for a batch array, which evaluate the degradation severity and RNA level. There are 2 up- and 3 down-regular genes in the 5 differentially expressed genes substantially associated with T2DM.

3.2 Cluster Profiling and functional enrichment analysis

The clustering of functions and profiles is essential for the investigation of organisms ' genetic structure and metabolic function. Genetic actions, interactions, and control mechanisms of such profiles are defined and inferred.Differentiated from regular samples, the cluster profile for T2DM genetic expression showed significant changes between comparative groups measured from the distance between Euclidean (Fig. 2).

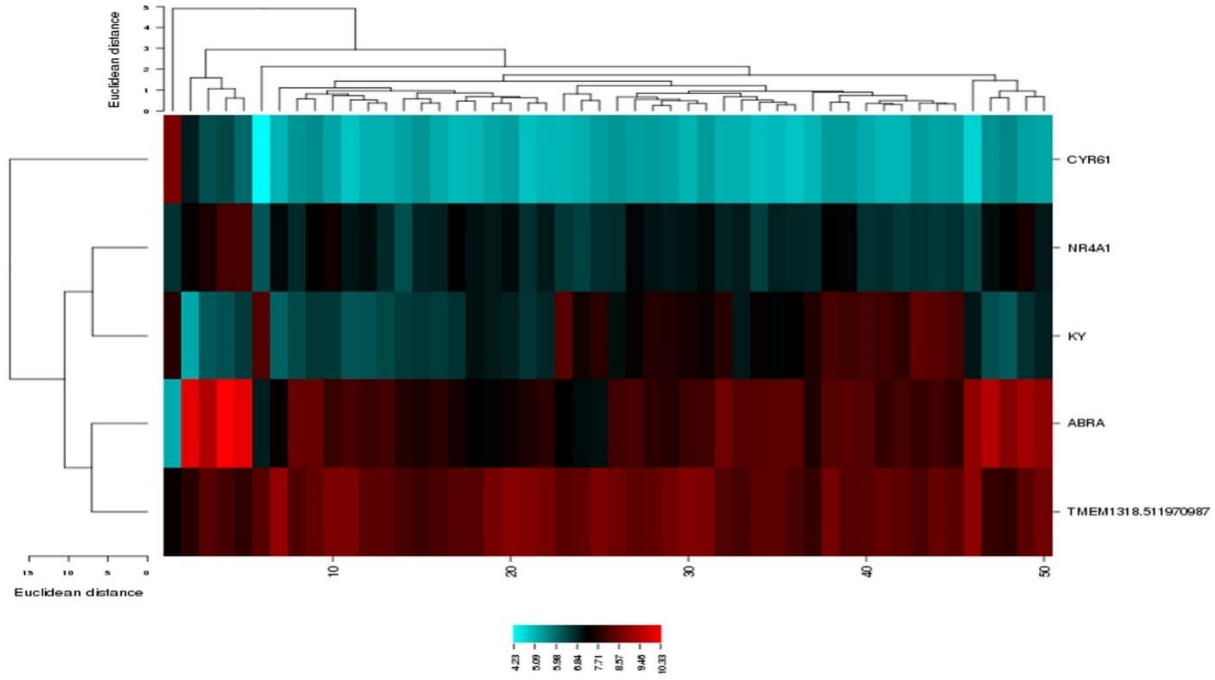

Fig. 2. Cluster analysis results for gene expression data. The expression values clustered in the red-shaded areas indicate over-expression and the blue-shaded areas indicate under-expression.

TMEM, ABRA, KY, and NR4A1 showed substantial genetic variations in diseased individuals. We analyzed DEG enrichment that showed enhanced terms for down and up-regulated genes. Cell communications, signal transduction, and metabolic functions are enriched terms. These genes are expressed in various tissues including immune cells, splenocytes, the pancreas, muscles, and the brain. Enrichment analysis has become the secondary analysis of genes identified by high-profile genomic methods because of their ability to give valuable insights into the collective biological function underlying a list of genes. By systematically mapping and by comparing their distribution into a gene set of interests with their biological annotations, the background distribution of the gene and protein ontology terms (GO). The enrichment analysis identified the GO terms as statistically underrepresented in the gene list.With their significant false discovery rates (FDR), their total number of enriched terms has been counted (Fig. 3).

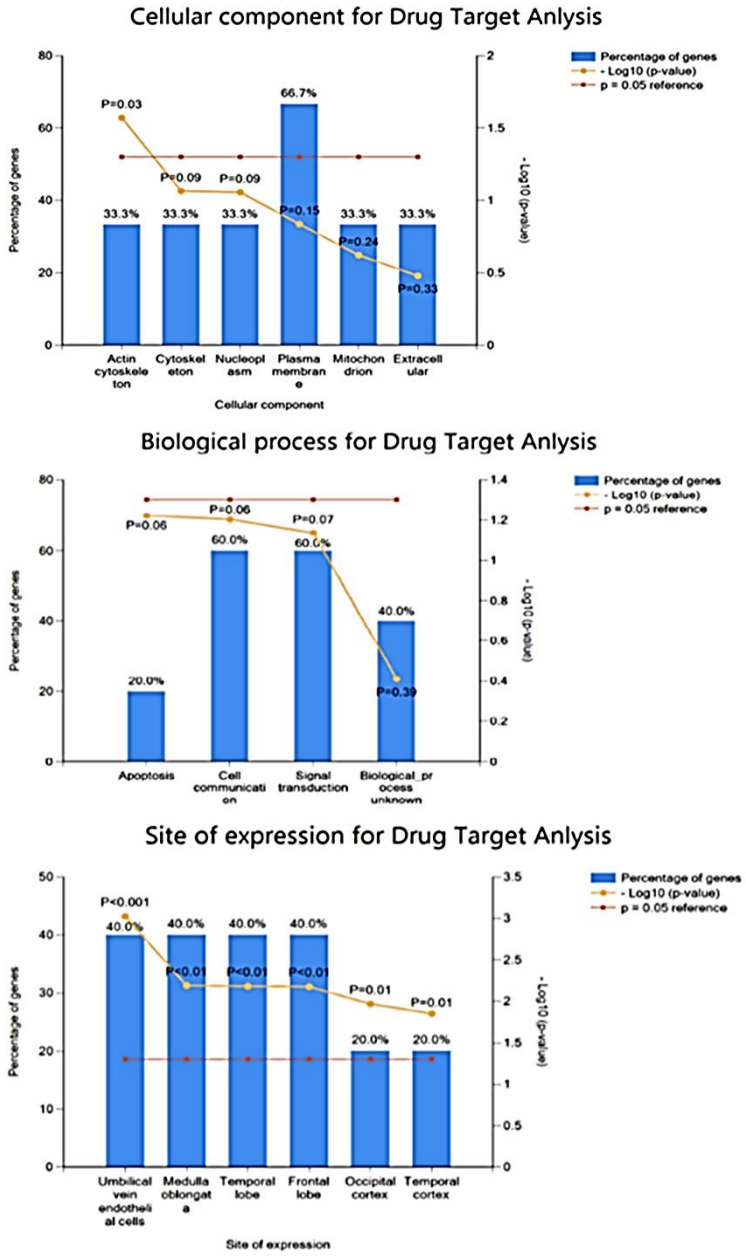

Fig. 3. (a) Cellular component analysis of ABRA, CYR61, NR4A1, KY, and TMEM131 for drug target (b) biological process of each gene (c) Site expression analysis of ABRA, CYR61, NR4A1, KY, and TMEM131

3.3PPI network analysis

The protein-protein interaction networks have been critical to identifying links between the various proteins associated with T2DM, the nodes of which represent proteins and whose undirected and potentially weighted edges connect interactive protein pairings. For incorporating reliability information related to the corresponding interactions, Edge weights were used.The interaction between the protein-to-protein in T2DM was analyzed with differential genes includingABRA (ABRA_Human), CYR61 (CYR61_Human), NR4A1 (NR4A1_Human), KY (KY_Human), and TMEM131 (TM131_Human). These interacting five source genes were obtained from HAPPI and STRING databases and their interaction network was generated. These source genes are interacting withFOS_HUMAN (proto-oncogene c-Fos for cytokine mediated, insulin resistance and SMAD protein signal transduction), SRF_HUMAN (serum response factor as transcriptional factor of many developmental and regulatory pathways), MKL1_HUMAN(myocardin-related transcription factor A associated with SRF), JUN_HUMAN (transcription factor AP-1 associated withcAMP signaling pathway, insulin resistance and T2DM), MYC_HUMAN (myc proto-oncogene protein linked with the transcription of growth related genes and activates myc-GLUT4 expression leads to T2DM), HIF1A_HUMAN (hypoxia-inducible factor 1-alpha alleviates hepatic oxidative stress and insulin resistance), ACTB_HUMAN (actin, cytoplasmic 1 responsible for kinase binding), UBC_HUMAN (polyubiquitin-C activates MAPK expression, cytokine mediated signaling, Wnt signaling and insulin resistance)and other related gene signatures (Fig. 4).The source genes have been found in direct and indirect association with FOS, UBC, JUN, CTNB1, RXRA, IGFN, and SODM proteins responsible for insulin resistance and diabetes (Fig. 5).

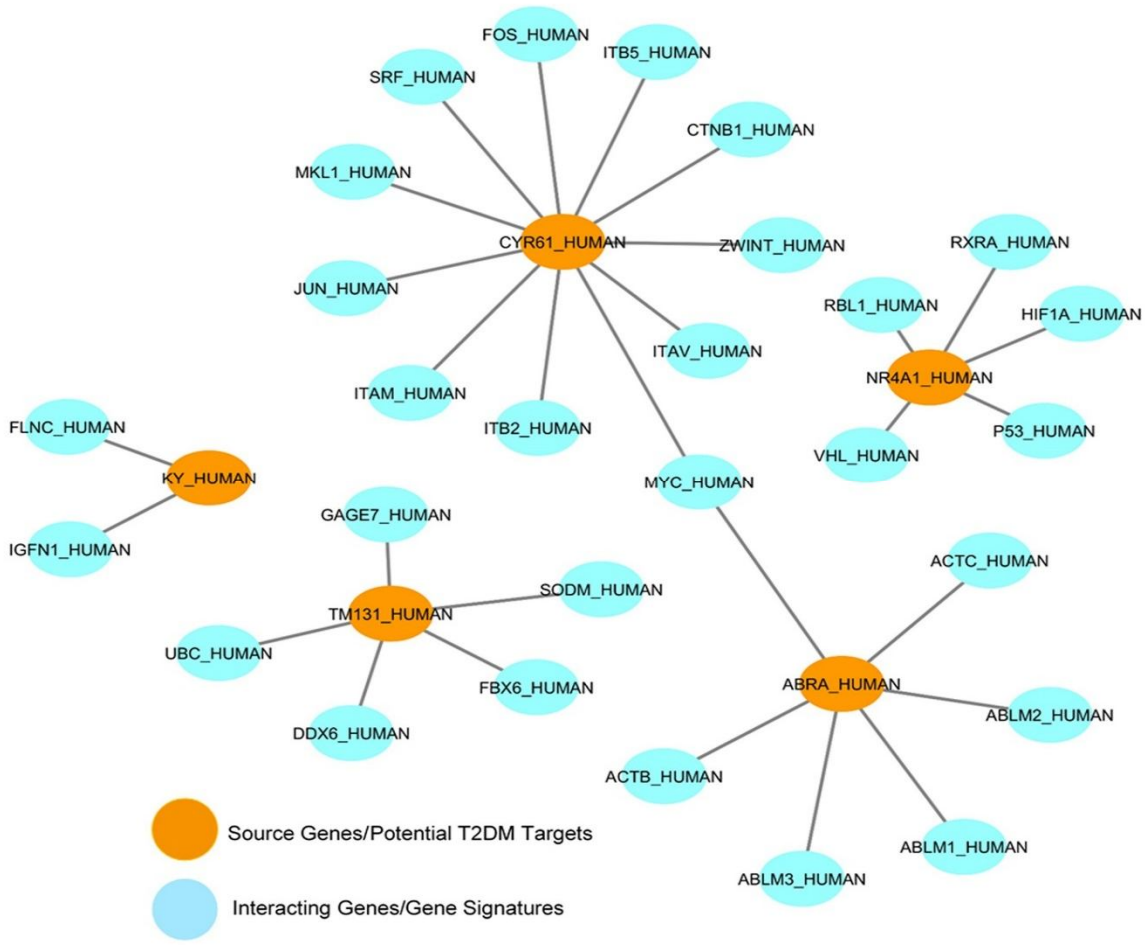

Fig. 4. Protein-Protein Interaction network of ABRA (Uniprot ID: ABRA_Human), CYR61 (Uniprot ID: CYR61_Human), NR4A1 (Uniprot ID: NR4A1_Human), KY (Uniprot ID: KY_Human) and TMEM131 (Uniprot ID: TM131_Human). Each dark color shows genes and lines showing their interaction with other genes.

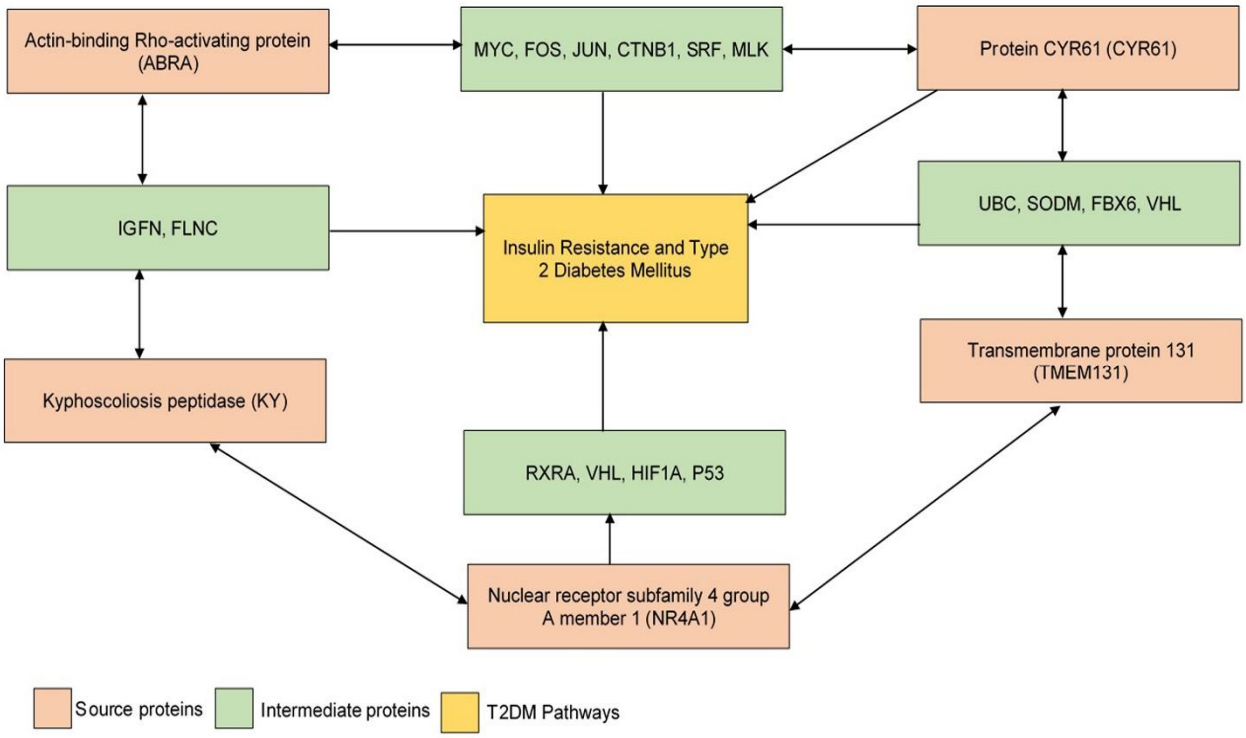

Fig. 5. Pathway analysis indicates the association of source differentially expressed genes with insulin resistance and type 2 diabetes mellitus through direct and indirect (intermediate) pathways.

4. DISCUSSION

The current study shows the significant link between gene expression and genetic variation and its functional role in disease.This provides a valid list of potential T2DM genes based on their expression in skeletal muscles, which shows their interaction with and correlation to the metabolic pathways associated with T2DM(Muhammad et al., 2017).The expanding databases of nucleotides and proteins require more sophisticated information management methods[24].Microarray data set containing 50 samples of normal and diseased patients of T2DM which were used to show differentially expressed genes (DEGs). In a single DEG having 50 patients, 40 normal persons and 10 are diseased. There are 5 genes used for the data analysis of which 3 genes are downregulated and 2 are upregulated whichshows contribution towards diabetes. These genes which show progression toward T2DM are apoptosis, cell communication,

signal transduction, and biological process(Malik et al., 2012);(Oduaran, 2015). Retinoic acid receptor-mediated signaling involving in the biological pathway of T2DM (Stumvoll et al., 2005). In the DEG analysis, we used 5 genes ABRA, CYR61, NR4A1,KY, and TMEM131 in which ABRA, CYR61, and NR4A1 are down-regulated genes, and KY, and TMEM131 are up-regulated genes which showed that these are all involving in the T2DM.These differentials expressed regulated genes may act as therapeutic goals or diagnostic instruments in a number of T2DM and these results can provide useful directions for future use.The genetic networks have extended transcript analysis to predict interactive genesignatures with a role in pathophysiological diabetes.

The reduction of beta cell contribution results in beta cell failure and causes apoptosis with increased lipids and hyperglycemia. The Pancreatic beta cells have a highly developed endoplasmic reticular (ER) for folding, exporting, and processing newly synthesized insulin(Laybutt et al., 2007). ABRA_HUMAN gene functions as actin binding Rho activating protein which mediates a growth factor, CYR61_HUMAN gene function as a cysteine-rich angiogenic inducer 61(CYR61), a novel biomarker that reflects myocardial injury (Klingenberg et al., 2017), NR4A1_HUMAN gene functions ascell proliferation, differentiation, and apoptosis (Wansa et al., 2002), KY_HUMAN gene functions as slowly progressive congenital myopathy and TMEM131 gene serves many biological functions.

The most significant gene interaction networks have been designed to analyze the interaction (ABRA, CYR61, NR4A1, KY, and TMEM131).These genes have been shown to interact with FOS_HUMAN, SRF_HUMAN, ZWINT_HUMAN, MKL1_HUMAN, JUN_HUMAN, MYC_HUMAN, HIF1A_HUMAN, P53_HUMAN, RXRA_HUMAN, VHL_HUMAN, ACTB_HUMAN, UBC_HUMAN,and other genes are reported that all of are involved in T2DM.This study shows several genes associated with T2DM that could be potential targets of development of new drugs.

5. CONCLUSION

It has been observed that many genetic variations are involved in the progression of T2DM. The system-level functional analysis showed that up and down-regulated differentially expressed genes are related to the etiology of the disease. These genetic variants including ABRA, CYR61,

NR4A1, KY, and TMEM131 are highly associated with insulin resistance and pathogenesis. This study would be helpful to understand the association of these genes with T2DM,and disease mechanism, and also the provide therapeutic options for T2DM by developing new drug treatment.

REFERENCES

Bolstad, B. (2008). Preprocessing and normalization for affymetrix genechip expression microarrays. *Methods in microarray normalization*, *41*.

Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, *19*(2), 185-193.

Chatterjee, S., Khunti, K., & Davies, M. J. (2017). Type 2 diabetes. *The Lancet*, *389*(10085), 2239-2251.

Chen, J. Y., Mamidipalli, S., & Huan, T. (2009). HAPPI: an online database of comprehensive human annotated and predicted protein interactions. *BMC genomics*, *10*(1), 1-11.

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., & Gross, B. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocol*, *2*(10), 2366.

Florez, J. (2008). Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia*, *51*(7), 1100.

Fujita, A., Sato, J. R., de Oliveira Rodrigues, L., Ferreira, C. E., & Sogayar, M. C. (2006). Evaluating different methods of microarray data normalization. *BMC bioinformatics*, *7*(1), 469.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15*(4), 870-878.

Huang, J., Liu, Y., Hou, H., & You, T. (2008). Simultaneous electrochemical determination of dopamine, uric acid and ascorbic acid using palladium nanoparticle-loaded carbon nanofibers modified electrode. *Biosensors and Bioelectronics*, *24*(4), 632-637.

Jin, Y., & Da, W. (2013). RETRACTED ARTICLE: Screening of key genes in gastric cancer with DNA microarray analysis. *European Journal of Medical Research*, *18*(1), 37.

Klingenberg, R., Aghlmandi, S., Liebetrau, C., Räber, L., Gencer, B., Nanchen, D., Carballo, D., Akhmedov, A., Montecucco, F., & Zoller, S. (2017). Cysteine-rich angiogenic inducer 61 (Cyr61): a novel soluble biomarker of acute myocardial injury improves risk stratification after acute coronary syndromes. *European Heart Journal*, *38*(47), 3493-3502.

Kuhn, M., Szklarczyk, D., Franceschini, A., Campillos, M., von Mering, C., Jensen, L. J., Beyer, A., & Bork, P. (2010). STITCH 2: an interaction network database for small molecules and proteins. *Nucleic acids research*, *38*(suppl_1), D552-D556.

Laybutt, D., Preston, A., Åkerfeldt, M., Kench, J., Busch, A., Biankin, A., & Biden, T. (2007). Endoplasmic reticulum stress contributes to beta cell apoptosis in type 2 diabetes. *Diabetologia*, *50*(4), 752-763.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D. J., Chesneau, A., & Hao, T. (2004). A map of the interactome network of the metazoan C. elegans. *Science*.

Lu, C. (2004). Improving the scaling normalization for high-density oligonucleotide GeneChip expression microarrays. *BMC bioinformatics*, *5*(1), 103.

Malik, A., Korol, A., Weber, M., Hankeln, T., Avivi, A., & Band, M. (2012). Transcriptome analysis of the spalax hypoxia survival response includes suppression of apoptosis and tight control of angiogenesis. *BMC genomics*, *13*(1), 615.

Muhammad, S. A., Raza, W., Nguyen, T., Bai, B., Wu, X., & Chen, J. (2017). Cellular Signaling Pathways in Insulin Resistance-Systems Biology Analyses of Microarray Dataset Reveals New Drug Target Gene Signatures of Type 2 Diabetes Mellitus. *Frontiers in physiology*, *8*, 13.

Nam, D., & Kim, S.-Y. (2008). Gene-set approach for expression pattern analysis. *Brief Bioinform*, *9*(3), 189-197.

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., & Morgan, M. (2014). VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics*, *30*(14), 2076.

Oduaran, O. H. (2015). *A bioinformatics approach to the identification of type 2 diabetes susceptibility gene variants in Africans*

Organization, W. H. (1999). *Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus*.

Stumvoll, M., Goldstein, B. J., & van Haeften, T. W. (2005). Type 2 diabetes: principles of pathogenesis and therapy. *The Lancet*, *365*(9467), 1333-1346.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520-525.

Wansa, K. S. A., Harris, J. M., & Muscat, G. E. (2002). The activation function-1 domain of Nur77/NR4A1 mediates trans-activation, cell specificity, and coactivator recruitment. *Journal of Biological Chemistry*, *277*(36), 33001-33011.