J Human Gen Genom. 2021 December; 5(2): A-10-349-1 doi: 10.52547/jhgg.5.2.67
Published online 2022 November 29 Research Article

Top-ranked expressed genes in human pancreatic islets: a bioinformatic analysis

Hamed Dabiri [1, 2], Majid Sadeghizadeh [1*], Ensiyeh Hajizadeh-Saffar [3,4], Mahdi Habibi-Anbouhi [5], Vahab Ziaei [5], Zahra Moghadasi [5]

1Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran.
2Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran.
3Department of Regenerative Medicine, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran.
4Advanced Therapy Medicinal Product Technology Development Center (ATMP-TDC), Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran
5National Cell Bank of Iran, Pasteur Institute of Iran, Tehran, Iran.

*Corresponding author: Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran. E-mail: sadeghma@modares.ac.ir

Received 2022 September 02; **Revised** 2022 October 05; **Accepted** 2022 October 10.

Abstract

Background: The Islets of Langerhans include Alpha, Beta, Delta, and Epsilon cells whose secret hormones play important roles in glucose metabolism as well as some physiological processes in our own body.
Method: In this study, we selected a microarray row data of transplanted pancreatic Islets from Gene Expression Omnibus (GEO) database. The row data of 10 individual samples was analyzed with R programing software. Top expressed genes in human pancreatic islets were chosen and the gene stable IDs were returned to gene names and descriptions by BioMart tools. The selected genes were categorized into biological processes by Protein Analysis Through Evolutionary Relationships (PANTHER) online database. Also, sub-cellular localization of their proteins was investigated by the protein atlas database.
Results: The results showed that the quality controls of all 10 individual samples of microarray chips such as signal intensity and uniformity of the images were passed. From 336 genes, 284 proteins were classified by PANTHER online database. They are categorized into "Translational proteins", "Metabolism enzymes", and "Protein modifying enzymes" biological processes, respectively.
Conclusion: In this study, we presented 500 top-ranked expressed genes in human pancreatic islets. We also represented calcification and sub-cellular localization of these high expressed genes in separate supplementary files. Research data can be used for pancreatic research as well as potential drug design for type I diabetes or pancreatic cancers.

Keywords: Microarray data analysis, Islets of Langerhans, Gene expression profile

Background

The parts of the pancreas that contain the endocrine (hormone-producing) cells are known as pancreatic islets or islets of Langerhans (1). 1-2% of the pancreas' volume and 10-15% of its blood flow go to the pancreatic islets (2). The pancreatic islets play a crucial role in the metabolism of glucose and are distributed throughout the human pancreas in density patterns (3).

A healthy adult person has roughly 1 million islets, each of which has an average diameter of about 0.2 mm, scattered throughout the pancreas in the shape of density pathways (4). (At least) five different kinds of cells discharge hormones from

[DOI: 10.52547/jhgg.67]

the pancreatic islets right into the bloodstream (4). Endocrine cell types are distributed as follows in rat islets: 20% of all islet cells produce glucagon in alpha cells, beta cells (70%) that produce insulin and amylin, 10% of delta cells produce somatostatin, epsilon cells (1%) that produce ghrelin, pancreatic polypeptide is produced by PP cells (gamma or F cells) in small amounts (5%) (5).

Type 1 diabetes (T1D) is an autoimmune disease. The destruction of pancreatic beta-cells by mostly autoreactive T cells was initiated and developed in type I diabetes. The event leads to need on exogenous insulin for personal life and health (6). The development of type 1 diabetes is complicated by immune regulation and immune response, accompanied by cellular immunity, which plays a key role. More specifically, pancreatic cell destruction is caused by invading CD4(+) and CD8(+) T cells, B cells, natural killer cells, dendritic cells and other immune cells, resulting in type 1 diabetes treatment of autoimmune diseases, but are only partially effective and not curative (6). They cause a general decreasing of the immune response and increase susceptibility to infections and malignancies. In most cases, the functional balance or the number of regulatory T cells is disturbed in T1D (7).

At the time of diagnosis, however, about a third of the beta cells in the pancreas are intact and able to produce insulin (8). In the event that inhibiting the immune system attacks the cells, it may lead to preventing in the disease progression through the results of repairing beta-cell mass through the beta-cell duplication and differentiation of other pancreatic cells to insulin-producing (9). several clinical trials have been used polyclonal regulatory T cells to modulate the immune system of autoimmune diseases that have acceptable patient outputs (10). In this regard, it is highly desirable to induce T cell tolerance to the specific beta cell antigen in order to prevent the development of further disease, particularly in individuals who are at high risk of developing diabetes or who have recently developed diabetes. Interestingly, immunosuppressive bispecific antibodies and CAR-Treg cells are two tools for this goal (11-12).

Overall, the hub genes recognized using bioinformatic strategies can help find the molecular mechanisms underlying the development of pancreatic cancer and offer potential targets for the prognosis and therapy of this disease (13). Moreover, bioinformatic evaluation may be represented the genes had been appreciably related to T1D. Therefore, the hub genes, elevate the understanding of the development of T1D, and specific genes is probably used as candidate target molecules to diagnosis as well as treatment of T1D (14).

In this study, we selected microarray row data of transplanted pancreatic Islets from the Gene Expression Omnibus (GEO) database. The row data was analyzed with R programing software. Top 500 expressed genes in human pancreatic islets were selected to categorize in the Protein Analysis Through Evolutionary Relationships (PANTHER) online database. Also, sub-cellular localization of the proteins was examined by the protein atlas comprehensive database.

Materials and methods

Data extraction and pre-processing

The GEO database (https://www.ncbi.nlm.nih.gov/geo) was used to extract microarray data of islets of Langerhans(18). For this aim, the word "human islet" was searched in the GEO database and the results were filtered to reach only affymetrix data. After the investigation of the results, one GEO dataset was selected for subsequent analysis. The affymetrix row data of individual samples was downloaded separately.

Affymetrix data of 10 individual microarray samples were readied and analyzed using the R program (R v4.1.1 for Windows). The quality control of the array chips was carried out by study of the image uniformity and box blot alignment of the samples. Non-uniform images and boxplots that were not aligned were ignored as outlier data.

The signal intensity of all 10 samples was normalized using the "mas5" function of the "Affy" package in the R programming language.

Filtering genes without significant signal intensity

Genes without significant signal intensity were filtered by the "mas5calls" function of the "Affy" package in the R programming language. Briefly, the "collapse" command code was used to filter the genes that have signal intensity below the threshold (signal intensity = 1) and background noises. The remaining genes with signal intensity > =1 were analyzed in the next steps.

Select top-ranked expressed genes

For selecting the top-ranked expressed genes in the human pancreatic islets, the mean of signal intensity of 10 individual samples was calculated by a "mean" function for each gene on the filtered table in the R programming language. The microarray IDs probes were sorted by the mean value of signal intensity. Five hundred IDs probes with the most values in mean of signal intensity were chosen as top-ranked expressed genes in human pancreatic islets. The top 500 microarray IDs probes were returned to gene names and descriptions by BioMarts of ensembles database (http://asia.ensembl.org/biomart/martview/538efc456306c68c761e3c8569eac1b9)(15).

Protein Analysis Through Evolutionary Relationships (PANTHER)

PANTHER classification system (http://pantherdb.org) is an online database for gene ontology to facilitate high-throughput analysis in gene and protein research. Top-ranked expressed genes in human pancreatic islets were analyzed in The PANTHER classification system using "biological processes"(16).

Protein Atlas

For investigation of subcellular localization of selected proteins, the list of subcellular localization proteins was

[DOI: 10.52547/jhgg.67]

downloaded from the protein atlas database. The common gene stable ID between the top-ranked expressed genes and the grouped genes from the protein atlas were found by Listdiff (http://www.listdiff.com/) online tool. The stable ID of outputs from each subcellular localization were returned to gene names by BioMarts of ensembles database (http://asia.ensembl.org/biomart).

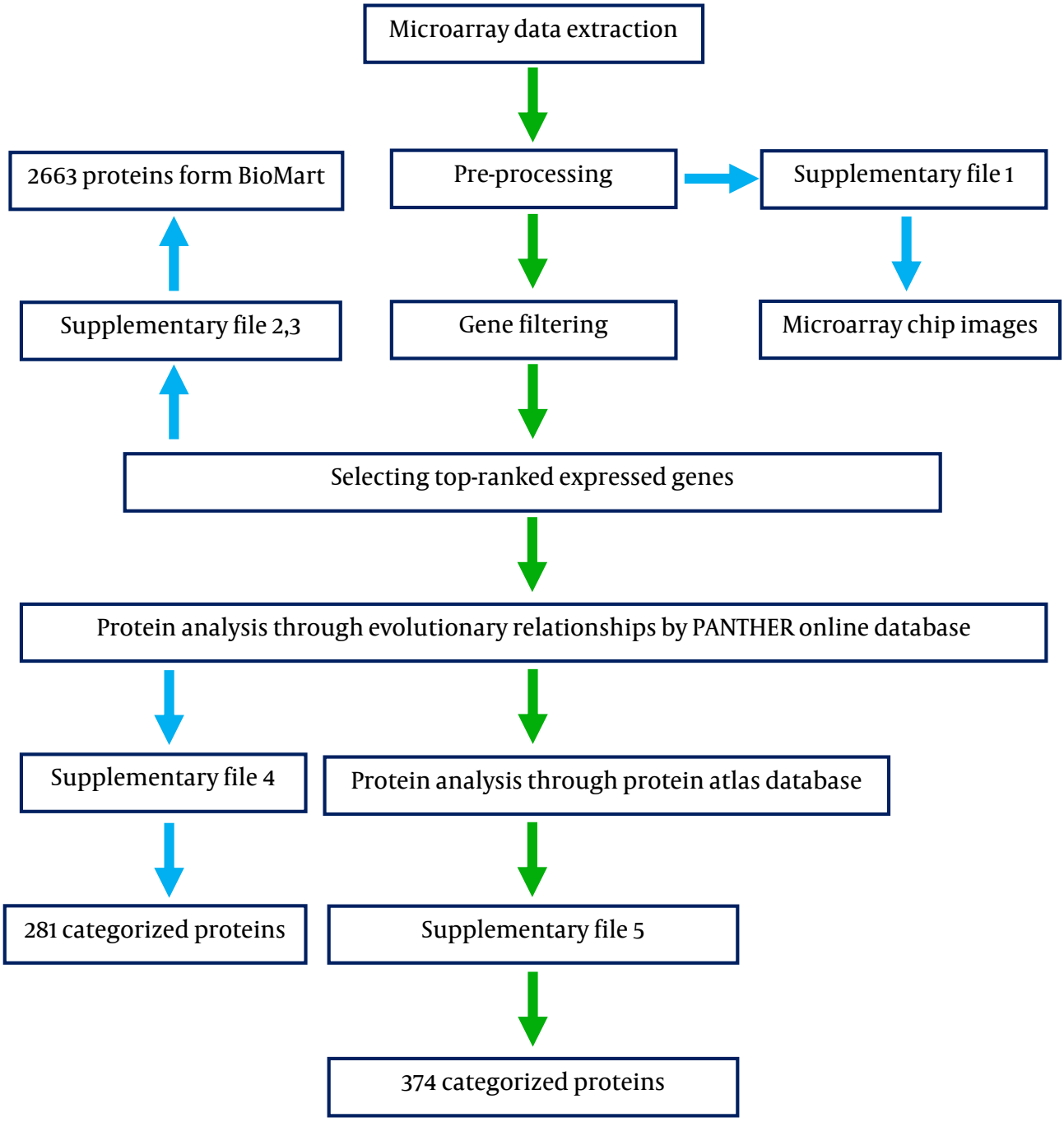


Figure 1. The workflow of the study steps

[DOI: 10.52547/jhgg.67]

Results

Extracting and reading Affy Data

Kurian et al. analyzed the gene expression in 59 human islet preparations by microarray technique. Correlation with diabetes reversal after transplantation of islets in diabetic mice was recorded in their study. Among 59 individual samples, 32 samples decreased the diabetic symptoms of the mouse models. Therefore, we selected and downloaded 10 individual microarray samples from their deposited raw data that were associated with "diabetes reversal status: Yes" characteristics (Table 1).

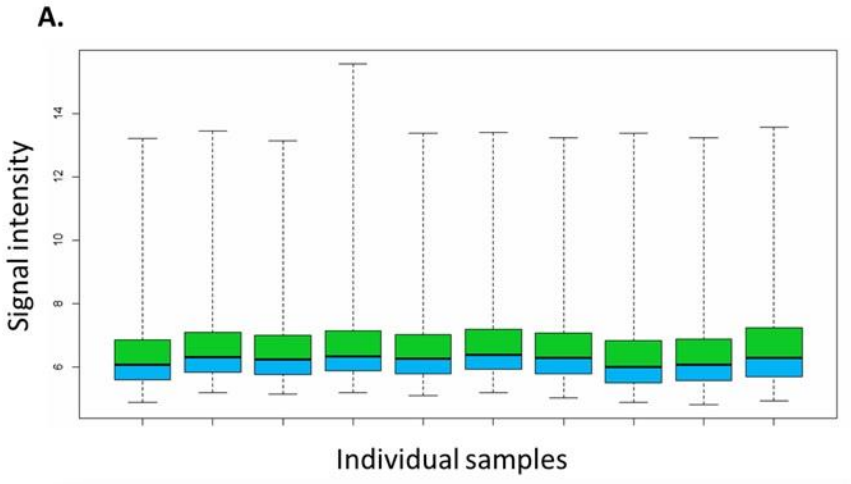
The results of reading Affy data illustrated that each sample had 54675 probe IDs that could be processed. The platforms of the arrays were HG-U133_Plus_2. The quality control of all samples was carried out by observing the complete image of each sample separately (Supplementary File 1). The array chip images revealed that all samples had the uniform normal signals without any noises. Indeed, there were not any non-uniform images and boxplots that were not aligned with outlier data. Therefore, all 10 samples are entered into the subsequent analysis process.

Table 1. Some details of input datasets.

Sample	Description	Source name	Diabetes reversal status	Scan protocol
GSM1941673	SAMPLE 1	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941675	SAMPLE 3	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941680	SAMPLE 8	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941681	SAMPLE 9	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941684	SAMPLE 12	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941685	SAMPLE 13	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941686	SAMPLE 14	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941689	SAMPLE 17	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941690	SAMPLE 18	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4
GSM1941691	SAMPLE 19	Human Islet Cells	Yes	Affymetrix EukGE-WS2v4

Signal normalization of Affy data

During gene expression analysis using microarray, the total signal intensity in different samples will not be the same. In order to analyze the microarray data, compare between different samples, and also get the average for the desired genes in all studied samples, it is necessary to normalize the signal in all samples. This step was done with the "mas5" function in the R programming language. The results showed the boxplot of all sample signals were uniformed along a straight line after the run of the "mas5" function for normalization (Figure 2). Also, the histogram of 10 sample signals became the Bell diagram after the normalization process (Figure 3).



[DOI: 10.52547/jhgg.67]

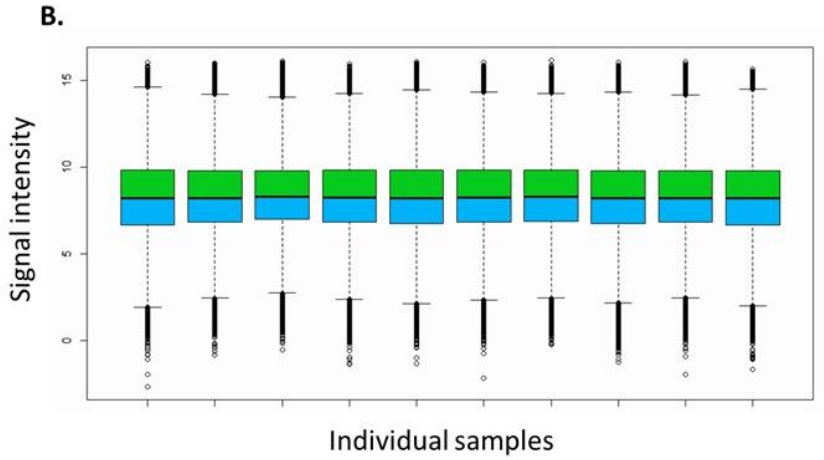


Figure 2. Boxplot of 10 individual samples before and after normalization processes. **A**. Before normalization, the signal intensity of microarray chips was disordered. B. After normalization, the signal intensity of microarray chips was ordered in one line. The green areas refer to signal intensities more than the mean and the blue areas refer to signal intensities less than the mean signal intensities.

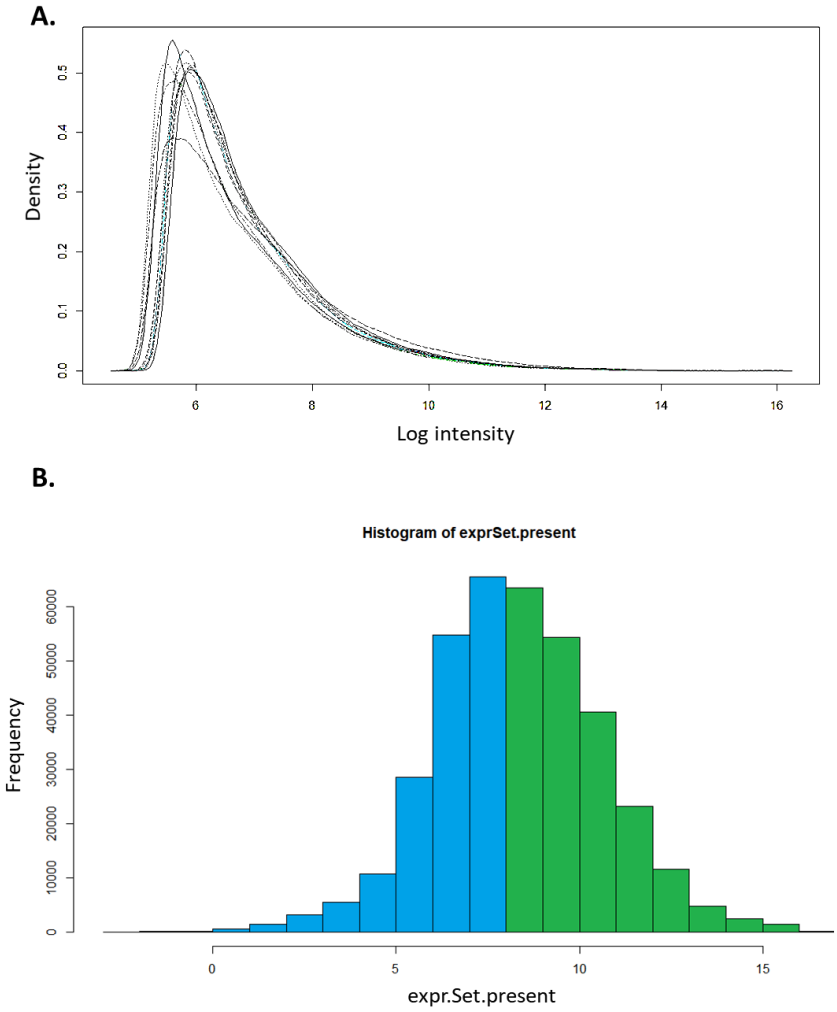


Figure 3. Histogram of 10 individual samples before and after normalization processes. **A.** Before normalization, the signal intensity of microarray chips was shifted to the left. **B**. After normalization, the signal intensity of microarray chips is converted to a Bell diagram. The green areas refer to signal intensities more than the mean and the blue areas refer to signal intensities less than the mean signal intensities.

[DOI: 10.52547/jhgg.67]

Filtering genes without significant signal intensity

Genes with poor signal intensity were removed by the "mas5calls" function. The genes that have signal intensity below the threshold (signal intensity = 1) or background noises were removed. Among 54675 IDs to be processed, the remaining genes with signal intensity > = 1 are 37190 that were analyzed in the next steps. Indeed, signal intensity of 17485 genes was under the threshold. Therefore they were removed.

Select top-ranked expressed genes

Five hundred probe IDs with the highest average signal in 10 samples were selected as the genes with the highest expression levels in Langerhans islet cells (Supplementary File 2). Probe IDs were converted to gene names by BioMart's online tool. One gene can produce multiple proteins due to splicing. Therefore, since some genes were identified by more than one protein ID, the resulting number of outputs from BioMart was 2663 proteins (Supplementary File 3).

Protein Analysis Through Evolutionary Relationships (PANTHER)

With the aim of promote high-throughput analysis, the PANTHER (Protein ANalysis THrough Evolutionary Relationships) classification system was created to categorize proteins (and their genes). An extensive, annotated "library" of gene family phylogenetic trees is the basis of PANTHER. From 500 stable gene IDs from BioMart, 336 genes were accepted in the PANTHER calcification system. Among them, 284 proteins were classified based on biological processes of PANTHE software (Figure 4). The results showed that most of the top expressed genes in pancreatic islets were categorized as "Translational proteins", "Metabolism enzymes", and "Protein modifying enzymes" respectively. The list of all protein groups was prepared in supplementary file 4.

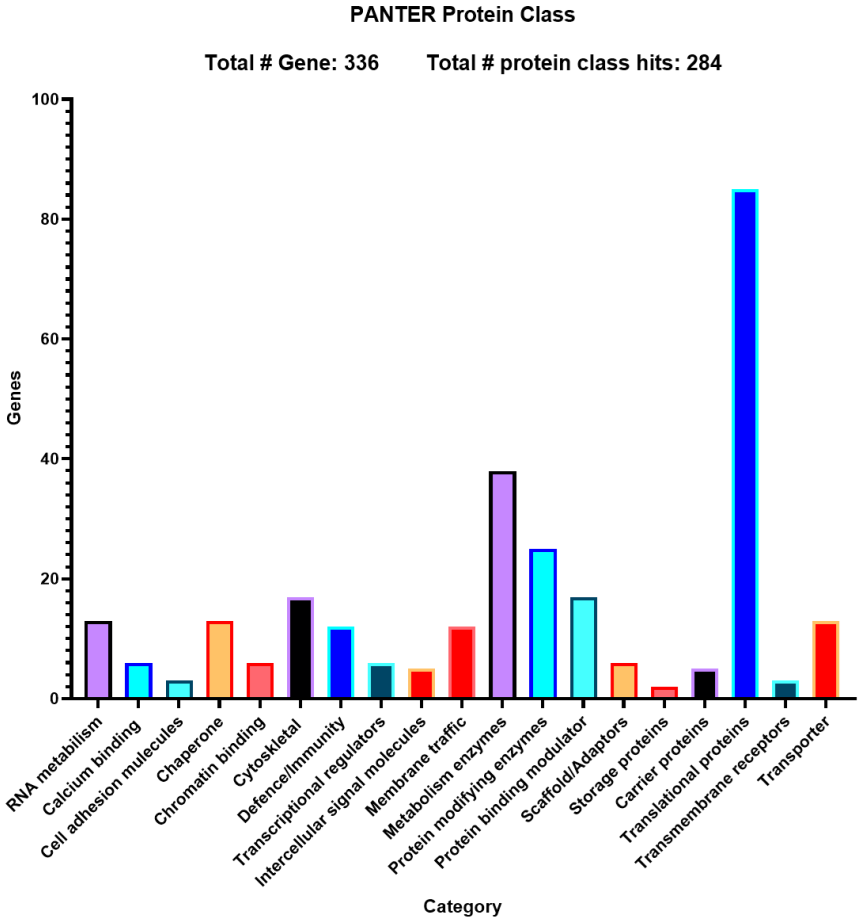


Figure 4. Protein classification of top 500 expressed genes in PANTHER online database. The most top expressed genes in pancreatic islets were classified as "Translational proteins", "Metabolism enzymes", and "Protein modifying enzymes" respectively. In contrast, fewer genes are in the "Storage proteins" category.

[DOI: 10.52547/jhgg.67]

Protein Atlas

Visions on the expression and spatiotemporal supply of proteins encoded by 13041 genes (65% of the human protein-coding genes) are prepared in high-resolution by the human protein atlas' subcellular segment. Immunofluorescence (ICC-IF) and confocal imaging have been used for each gene to examine the protein's subcellular localization. Following image processing, the protein's subcellular localization was divided into one or more of 35 distinct organelles and fine subcellular structures. The section also contains annotations for genes that differ in protein expression levels and/or subcellular localization in a single cell. The Sub-Cellular Section provides a database for in-depth research into specific genes and proteins of interest, as well as for methodical examination of proteomes in a wider context. The results revealed that among 284 selected proteins, two proteins have subcellular localization in the "Actin" category, twenty proteins for "Golgi", seventeen proteins for "Mitochondria", twenty proteins for "Plasma", thirdly one protein for "Vesicles, Peroxisomes, Endosomes, Lysosomes, and Lipid", and sixty four proteins for "Endoplasmic", thirty proteins for "Nucleoli", one hundred and thirty proteins for "Aggresome, Cytosol, Cytoplasmic", forty eight proteins for "Nucleoplasm, Nuclear" (supplementary file 5).

Discussion

Early research has determined that type 1 diabetes mellitus (T1DM) is a condition brought on by an autoimmune attack on the pancreatic b-cells that produce insulin. Three key pillars of T1DM research are genetics, the environment, and the immune pathogenesis of the disease. Fang et al. attempted to comprehend how the gene expression profile changed during the pathogenesis of T1DM. In order to find microarray studies of T1DM with samples taken at or before the onset of T1DM, we conducted a thorough search in the Gene Expression Omnibus (GEO) database (17).

Their results of an integrated analysis of various GEO datasets and a comparison of the gene expression levels in T1DM samples collected at the time the islet autoantibodies first appeared, one year prior to the onset of T1DM, and at the time of the onset of T1DM revealed that CD274, which encodes PD-L1, was up-regulated in the samples taken at the time of the newly onset T1DM. In the control samples, CD274 expression was steady, but it gradually increased after the emergence of autoantibodies and before the development of T1DM. These findings suggest a link between CD274 upregulation in T1DM and disease etiology. Researchers may be able to develop techniques to stop the destruction of pancreatic b-cells by using PD-L1 to protect the pancreatic islets against autoimmune attack (18).

The bulk of the gene loci linked to type 2 diabetes affect how pancreatic islets function. Keller et al. exposed a genetically diverse mouse population to a western diet high in fat and sucrose. They then achieved genome-wide association of diabetes-related phenotypes to assess the relevance of islet gene expression causes diabetes. Keller et al. measured the total mRNA in the islets and found 18,820 expressions QTL. To find possible underlying developer genes at loci that influence the quantity of many transcripts, we used mediation analysis (19).

These include three genes having minimal associations with diabetes-associated features in humans, as well as two genes related to diabetes (PDX1 and HNF4A). In order to identify regulatory locus for elements enriched with mRNAs specific to alpha and delta cells, they classified transcripts into gene modules. A heterogeneous pattern of gene expression within the beta cell population is suggested by the fact that no one module was enriched for beta cell specific transcripts. The strongest correlation was found between a module rich in mRNAs linked to metabolism of specific amino acids and physiological characteristics that signal insulin resistance. Even though the animals employed in this examination were not clearly diabetic, by investigation the gene expression of islet under nutritional pressure, Keller et al. were nevertheless clear correlated difference in group of genes that are associated to physiological aspects linked to diabetes. The study by Keller et al. reveals how the mouse can support notional associations identified by human genome-wide association mapping and gives a typical degree of agreement between mouse and human populations' loci connected to diabetes (20).

Genome-wide association studies (GWASs) recognized several elements associated with type 2 diabetes (T2D). To understanding the causal molecular mechanisms, Alonso et al. have created the translational islet genotype of human resource, collecting >500 islet genomic databases. Alonso et al. known >1 million islets, 53 of which associated with T2D (21).

Efforts to find the subcellular localization of target proteins in the cell is important in different aspects. First, special methods must be used to plan the delivery of the potential drugs to the target area. For example, if the target protein is located in the nucleus or in the cytoplasm, the potential drug molecule must transfer between one membrane (plasma membrane) or two membranes (plasma membrane and nuclear envelope)(22). Therefore, the type of potential drug delivery tool will be different in these cases. In addition, pH conditions and viscosity of the subcellular location are important for targeting a protein. In such a way that the cytoplasm has more viscosity than the mitochondrial matrix. Finally, plasma membrane proteins can act as surface biomarkers for the binding of specific molecules such as antibodies (12). In addition, proteins involved in cell translation can be used as targets for the design of cell death inhibitors in cancer (24-25). All of these mentioned points can be the applications of the exact investigation of proteins with high expression levels in pancreatic islet cells.

Conclusion

In conclusion, we presented 500 top-ranked expressed genes in human pancreatic islets using microarray data analysis for the first time. We also categorized these high expressed genes into separate supplementary files that can be used for pancreatic research as well as potential drug design for type I diabetes or pancreatic cancers.

Acknowledgments

Dr. Saeed Ansari have helped us in the microarray data analysis, which we appreciate greatly.

[DOI: 10.52547/jhgg.67]

Conflicts of interest

One of the authors of this article is a member of the committee board of the Journal of Human Genetics and Genomics.

Funding /support

This project was financially supported by Iran National Science Foundation (INSF, Grant No. 97011360).

References

1. Da Silva Xavier G. The cells of the islets of Langerhans. Journal of clinical medicine. 2018;7(3):54.
2. Virtanen I, Banerjee M, Palgi J, Korsgren O, Lukinius A, Thornell L-E, et al. Blood vessels of human islets of Langerhans are surrounded by a double basement membrane. Diabetologia. 2008;51(7):1181-91.
3. Koh D-S, Cho J-H, Chen L. Paracrine interactions within islets of Langerhans. Journal of molecular neuroscience. 2012;48(2):429-40.
4. Hodson DJ, Tarasov AI, Gimeno Brias S, Mitchell RK, Johnston NR, Haghollahi S, et al. Incretin-modulated beta cell energetics in intact islets of Langerhans. Molecular endocrinology. 2014;28(6):860-71.
5. Yang Y, Kim JW, Park HS, Lee EY, Yoon KH. Pancreatic stellate cells in the islets as a novel target to preserve the pancreatic β-cell mass and function. Journal of Diabetes Investigation. 2020;11(2):268-80.
6. J. Ozougwu, K. Obimba, C. Belonwu and C. Unakalamba. The pathogenesis and pathophysiology of type 1 and type 2 diabetes mellitus. J Physiol Pathophysiol 2013 Vol. 4 Issue 4 Pages 46-57
7. R. A. Oram, E. K. Sims and C. Evans-Molina. Beta cells in type 1 diabetes: mass and function; sleeping or dead?. Diabetologia 2019 Vol. 62 Issue 4 Pages 567-577
8. C. Limbert, G. Päth, F. Jakob and J. Seufert. Beta-cell replacement and regeneration: Strategies of cell-based therapy for type 1 diabetes mellitus. Diabetes research and clinical practice 2008 Vol. 79 Issue 3 Pages 389-399.
9. J. A. Bluestone, J. H. Buckner, M. Fitch, S. E. Gitelman, S. Gupta, M. K. Hellerstein, et al. Type 1 diabetes immunotherapy using polyclonal regulatory T cells. Science translational medicine 2015 Vol. 7 Issue 315 Pages 315ra189-315ra189.
10. S. A. Beheshti, K. Shamsasenjan, M. Ahmadi and B. Abbasi. CAR Treg: A new approach in the treatment of autoimmune diseases. International Immunopharmacology 2022 Vol. 102 Pages 108409.
11. R. E. Kontermann and U. Brinkmann. Bispecific antibodies. Drug discovery today 2015 Vol. 20 Issue 7 Pages 838-847.
12. Y. Vandenbrouck, D. Christiany, F. Combes, V. Loux and V. Brun. Bioinformatics tools and workflow to select blood biomarkers for early cancer diagnosis: an application to pancreatic cancer. Proteomics 2019 Vol. 19 Issue 21-22 Pages 1800489.
13. T. S. Assmann, M. Recamonde-Mendoza, B. M. De Souza and D. Crispim. MicroRNA expression profiles and type 1 diabetes mellitus: systematic review and bioinformatic analysis. Endocrine connections 2017 Vol. 6 Issue 8 Pages 773-790.
14. Qi M. Transplantation of encapsulated pancreatic islets as a treatment for patients with type 1 diabetes mellitus. Advances in medicine. 2014;2014.
15. Carlsson PO, Espes D, Sedigh A, Rotem A, Zimerman B, Grinberg H, et al. Transplantation of macroencapsulated human islets within the bioartificial pancreas βAir to patients with type 1 diabetes mellitus. American Journal of Transplantation. 2018;18(7):1735-44.
16. Ludwig B, Reichel A, Steffen A, Zimerman B, Schally AV, Block NL, et al. Transplantation of human islets without immunosuppression. Proceedings of the National Academy of Sciences. 2013;110(47):19054-8.
17. Shapiro A, Pokrywczynska M, Ricordi C. Clinical pancreatic islet transplantation. Nature Reviews Endocrinology. 2017;13(5):268-77.
18. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic acids research. 2012;41(D1):D991-D5.
19. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart–biological queries made easy. BMC genomics. 2009;10(1):1-12.
20. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic acids research. 2019;47(D1):D419-D26.
21. Fang C, Huang Y, Pei Y, Zhang H-h, Chen X, Guo H, et al. Genome-wide gene expression profiling reveals that CD274 is up-regulated new-onset type 1 diabetes mellitus. Acta Diabetologica. 2017;54(8):757-67.
22. Keller MP, Gatti DM, Schueler KL, Rabaglia ME, Stapleton DS, Simecek P, et al. Genetic drivers of pancreatic islet function. Genetics. 2018;209(1):335-56.
23. Elzek MA, Christopher JA, Breckels LM, Lilley KS. Localization of Organelle Proteins by Isotope Tagging: Current status and potential applications in drug discovery research. Drug Discovery Today: Technologies. 2021.
24. Dahlgren D, Lennernäs H. Antibody-drug conjugates and targeted treatment strategies for hepatocellular carcinoma: a drug-delivery perspective. Molecules. 2020;25(12):2861.
25. Jia Y, Polunovsky V, Bitterman PB, Wagner CR. Cap-Dependent Translation Initiation Factor e IF4E: An Emerging Anticancer Drug Target. Medicinal research reviews. 2012;32(4):786-814.

[DOI: 10.52547/jhgg.67]