Cell J
Cell Journal (Yakhteh)
Homepage: https://www.celljournal.org/
Original Article
Vol 26, No 1, January 2024, Pages: 51-61

Candidate Biomarkers for Targeting in Type 1 Diabetes; A Bioinformatic Analysis of Pancreatic Cell Surface Antigens

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

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

Abstract

Objective: Type 1 diabetes (T1Ds) is an autoimmune disease in which the immune system invades and destroys insulin-producing cells. Nevertheless, at the time of diagnosis, about 30-40% of pancreatic beta cells are healthy and capable of producing insulin. Bi-specific antibodies, chimeric antigen receptor regulatory T cells (CAR-Treg cells), and labeled antibodies could be a new emerging option for the treatment or diagnosis of type I diabetic patients. The aim of the study is to choose appropriate cell surface antigens in the pancreas tissue for generating an antibody for type I diabetic patients.

Materials and Methods: In this bioinformatics study, we extracted pancreas-specific proteins from two large databases; the Human Protein Atlas (HPA) and Genotype-Tissue Expression (GTEx) Portal. Pancreatic-enriched genes were chosen and narrowed down by Protter software for the investigation of accessible extracellular domains. The immunohistochemistry (IHC) data of the protein atlas database were used to evaluate the protein expression of selected antigens. We explored the function of candidate antigens by using the GeneCards database to evaluate the potential dysfunction or activation/hyperactivation of antigens after antibody binding.

Results: The results showed 429 genes are highly expressed in the pancreas tissue. Also, eighteen genes encoded plasma membrane proteins that have high expression in the microarray (GEO) dataset. Our results introduced four structural proteins, including NPHS1, KIRREL2, GP2, and CUZD1, among all seventeen candidate proteins.

Conclusion: The presented antigens can potentially be used to produce specific pancreatic antibodies that guide CAR-Treg, bi-specific, or labeling molecules to the pancreas for treatment, detection, or other molecular targeted therapy scopes for type I diabetes.

Keywords: Bioinformatics, Cell Surface Antigens, Molecular Targeted Therapies, Pancreatic Islets, Type 1 Diabetes

Citation: **Dabiri H, Habibi-Anbouhi M, Ziaei Y, Moghadasi Z, Sadeghizadeh M, Hajizadeh-Saffar E. Candidate biomarkers for targeting in type 1 diabetes; a bioinformatic analysis of pancreatic cell surface antigens: a prospective study. Cell J. 2024; 26(1): 51-61. doi: 10.22074/CELLJ.2023.1996297.1262**

Introduction

While Type 1 diabetes (T1D) is a mostly T-cell-mediated autoimmune disease, it is defined as a destruction of pancreatic beta-cells by autoreactive immune cells. This destruction event results in lifelong exogenous insulin dependency. The T1D development is complicated by immunological regulations and responses, that are accompanied by the key role of cellular immunity. More precisely, the destruction of pancreatic beta cells is caused by the invasion of multiple cells, including clusters of differentiation 4 positive [$CD4^+$] and $CD8^+$ T lymphocytes, B lymphocytes, natural killer cells, dendritic cells, and other immune cells, which leads to the incidence of T1D (1). While immunosuppressive

Received: 26/March/2023, Revised: 27/August/2023, Accepted: 24/October/2023
***Corresponding Addresses: P.O.Box: 66492-595, National Cell Bank of Iran, Pasteur Institute of Iran, Tehran, Iran**
P.O. Box: 14115-175, Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran
P.O.Box: 16635-174, Department of Regenerative Medicine and Advanced Therapy Medicinal Product Technology Development Center (ATMP-TDC), Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran
Emails: habibi_m@pasteur.ac.ir, sadeghma@modares.ac.ir, en.hajizadeh@royan-rc.ac.ir

Royan Institute
Cell Journal (Yakhteh)

medications are used to treat autoimmune diseases, they are not only partially efficient (2), but also They cause a general immune response attenuation, increasing susceptibility to infections and malignancies (3). In most cases, in type I diabetic patients, the balance between regulatory T cell functions or numbers is upset. Around a third of pancreatic beta cells are intact and capable of producing insulin at the time of diagnosis, therefore it would be a notable strategy to induce a T-cell tolerance for specific beta-cell antigens (Ag) to prevent further disease development that is a useful strategy for identification of individuals at risk of diabetes (4). It seems that the Bispecific antibodies and chimeric antigen receptor regulatory T cells (CAR-Treg cells) are two new implements to reach this goal.

Bispecific antibodies (bsAb) significantly have two antigen-binding sites. They are made up of two separate Fragment antigen-binding (Fab) arms or two different antibodies (Abs) joined by a common Fragment crystallizable (Fc) region (5). In autoimmune conditions, bsAbs can be utilized to: i. Deactivate multiple cytokines or receptors at the same time, ii. Drive cell-cell contacts between different immune cell populations, and iii. Cause receptor co-localization on the cell surface (5, 6). Blinatumomab, a CD19- and CD3-bispecific recombinant antibody, was the first bsAb to be used in clinical trials for the treatment of non-Hodgkin's B cell lymphoma. A Blinatumomab causes B cells to engage with cytotoxic T cells. As a result of the interaction between two cells by bsAb, B lymphoma cells are effectively eliminated, protective T lymphocytes are expanded, and the majority of patients have a longer life expectancy (5, 7). In the same way, pancreatic cell surface antigens can be effective for producing and using bsAb for the type I diabetic remission.

The immune system has a remarkable capacity to scan tissues and recognize and clear aberrant and malignant cells. In contrast, many immune system components, such as regulatory T cells, can locally modulate the aberrant immune response against the body's own cells. Adoptive immunotherapies utilize and improve the strength of our immune system against cancerous cells and aberrant immune responses. Among the various techniques, chimeric antigen receptors, or CARs, have recently been of large interest. Recently, many studies have been developed for using polyclonal Tregs in cell treatment approaches. They are now being studied in transplant recipients and patients with autoimmune disorders, especially in T1D (8). This method has some limitations, including a few numbers of Treg cells in the peripheral blood in comparison to other common T cells such as effector or cytotoxic ones. The use of antigen-specific modified Tregs has acceptable patient outcomes, because of their higher ability to generate a response and their requirement to fewer cells due to their local activity.

Animal model studies have been revealed that antigen-specific Tregs are functionally superior to polyclonal Tregs. Currently, in a few investigations genetic engineering has been used for expressing CAR in Tregs, as it has been used for hematological malignancies. Multiple studies have used various methods to confer antigen specificity to Tregs, with a focus on CAR-Treg cells. According to these investigations, CAR-Treg cells are preferable for type I diabetic patients. The local activity of CAR-Tregs in the pancreas provides a significant response and needs fewer cells (9-11). Hence, the pancreatic cell surface antigens can be used for the production of pancreas-specific CAR-Treg cells following the generation of appropriate antibodies (scFvs).

The appropriate beta-cell mass (BCM) is a valuable indication of the pancreatic proper function. Beta-cells are destroyed in the T1Ds, leading to a significant drop of BCM. Due to peripheral insulin resistance and increasing insulin demand, the BCM decreases slowly and steadily in people with type 2 diabetes. Surface antigens of beta cells can be recruited to evaluate the BCM as well as pancreatic proper function in a non-invasive manner. Some antigens have been studied in the human brain and pancreas, [11C] dihydrotetrabenazine (2-hydroxy-3-isobutyl-9-[11C] methoxy-10-methoxy-1,2,3,4,6,7-hexahydro-11bH-benzo[a]quinolizine), commonly known as [11C] DTBZ, was used to investigate the BCM with a positron emission tomography (PET) in a non-invasive in vivo method. In another study, researchers showed that the specific antibodies were bound to beta cells in normal mice but not to mice administered with streptozotocin (STZ) which induced BCM loss and diabetes (12). Consequently, the appropriate beta cell surface antigens accompanied by their specific antibodies can be used for monitoring BCM changes in diabetic patients.

As mentioned above, the most preliminary step for the production of bi-specific antibodies, CAR-Treg cells, or labeled antibodies for T1Ds is choosing the appropriate cell surface antigen in the pancreas tissue. The selection of surface antigens is sensitive from different viewpoints. First, the cell surface antigen should have an acceptable expression level in pancreatic cells. This is an influential event for the potential binding of antibodies or CAR-Treg cell receptors to the antigen. The antigen frequency can be surveyed in two levels of RNA expression and protein expression. Second, according to the length and three-dimensional structure of the polypeptide chain in the plasma membrane, the antigen should have enough accessible extracellular domains. Accessible domains are important for the potential binding activity of antibodies or CAR-Treg cells. Third, antigen enrichment in the pancreas in comparison to other body tissues is another aspect that should be considered in this regard. This feature is important in order to make antigens, the least off-target effect for final products. According to the binding of antibodies or CAR-Treg cells to other tissues, off-target events decrease pancreas accumulation of the final products leading to a decrease their effectiveness. Moreover, gathering final products in other tissues causes unpleasant side effects as well. Fourth, the function of the surface antigen is vital for the pancreatic cells in terms of inactivation or activation/hyperactivation after antibody

binding.

In the present study, we investigated important features of pancreatic surface antigens from the genes preferentially expressed in the pancreas tissues. To reach the best candidate antigens, we categorized them into two distinct groups according to their functions: "Structural proteins" that do not alter their activity after antibody binding, and "Functional proteins" which can be divided into two subgroups, water channels and signaling proteins. Water channels may have less impact on cell behavior upon antibody binding. In contrast, signaling proteins, which may bring about substantial changes in normal cell function due to the gain of function or loss of functions.

Materials and Methods

Data mining

In this bioinformatics study, to find the genes that are preferentially expressed in the pancreas tissue, candidate genes were extracted from two different sources, the Human Protein Atlas (HPA, Karolinska Institutet, 2023, Sweden) and the Genotype-Tissue Expression (GTEx) project (Broad Institute, 2021, USA) (13). The protein atlas database has an RNA dataset that was utilized to group genes based on their expression in different tissues. The clustering of 19019 genes which are expressed in different tissues led to obtain 87 expression groups. The groups were manually annotated to explain functional and specificity properties. The HPA is divided into 10 parts, each of which focuses on a different component of the genome-wide investigation of human proteins.

The protein distribution throughout all major tissues and organs in the human body is shown in the Tissue section of Protein Atlas tool) https//:www.proteinatlas.org/humanproteome/tissue). Pancreas high-expressed genes as well as genes that are only detected in the pancreas tissue were extracted from the Protein Atlas database.

The GTEx project is a long-term program aimed at preparing a comprehensive free source for investigating gene expression in body tissues as well as their gene regulations. Nearly 1000 people had samples taken from 54 non-diseased tissue locations, largely for molecular omics such as Whole Genome and Exome Sequencing (WGS and WES), and RNA-Seq data. The GTEx Biobank has also the remainder of the samples. The GTEx Portal makes data available such as gene expression, QTLs, and histology pictures. The 100 top genes expressed in the pancreas were got from GTEx and added to the genes which were extracted from the protein atlas database. The names of candidate genes were converted to ensemble ID by BioMart (https://www.ensembl.org/biomart) and prepared in the Ext. data format. The files were used as input materials for the consequent analysis (14).

Calculating tissue-specific gene enrichment using "TissueEnrich" tool

The TissueEnrich is an R-based package that finds gene enrichment in the body tissues from a set of input genes. Using TissueEnrich, tissue-specific genes were identified through the HPA methodology to analyze RNA-Seq data from the Human Protein Atlas, GTEx, and ENCODE projects (15). To evaluate if tissue-specific genes were enriched among the input genes, the hypergeometric test was performed. In addition, the TissueEnrich package also was used to identify enrichment through user-supplied expression datasets, which subsequently led to computing tissue-specific gene enrichment parameters such as exact P values.

The TissueEnrich package uses RNA Sequencing results from the HPA, GTEx, and ENCODE database to identify gene enrichments in tissues. This package only employed tissues with two biological replicates to have strong estimating. The following datasets were employed in this tool: i. RNA Sequencing results from human tissues (n=35) in the HPA Database, ii. RNA sequencing results from human tissues (n=29) in the GTEx Database, and iii. RNA sequencing results from mouse tissues (n=17) in the Mice ENCODE Dataset (16).

Tissue-specific genes definition in HPA were categorized as follows:

Genes having an expression level greater than one [Transcript Per Million (TPM) or Fragments per kilobase of transcript per million mapped fragments (FPKM)] as well as minimum five-fold expression levels in a specific tissue in comparison to all body tissues were referred to as "TissueEnriched". Genes having an expression level of more than 1 (TPM or FPKM) that also had at minimum five-fold expression levels in comparison to all body tissues were considered to as "Group Enriched". Genes having an expression level over 1 (TPM or FPKM) and minimum five-fold expression levels in a specific tissue in comparison with the average levels in all body tissues but were non-TissueEnriched or non-Group Enriched was referred as a "TissueElevated". To employ the TissueEnrich package, the R programming language (R v4.1.1 for Windows) and RStudio-Integrated Development Environment (IDE)- (RStudio 1.2.5042) were installed. The package was acquired from the Bioconductor version, Release 3.14, and added to the R program. Four hundred twenty-nine input genes were prepared as an Ext. data format and located in the program work directory (setwd). The command codes were written and run in the RStudio environment. Just the Tissue Enrich function of the software was used for a gene expression analysis. The resulting data were evaluated according to the suggested pipeline.

Evaluation RNA expression level of transmembrane proteins by using microarray analysis from the GEO dataset

The Output of the Tissue Enrich tool was analyzed to select the genes that encode plasma transmembrane proteins which can be targeted as pancreatic cell surface antigens. For this aim, the information on total human

transmembrane proteins was extracted from the HPA (Karolinska Institutet, 2023, Sweden). The correspondence between the total transmembrane and output proteins of Tissue Enrich was achieved using the Barc software (Bioinformatics and Research Computing, Massachusetts Institute of Technology, 2023, USA). To confirm the expression of plasma membrane genes in beta and acinar cells, microarray data of 3 samples were investigated again. For this aim, microarray raw data of acinar and beta cells, adipose tissue, heart, small intestine, diaphragm tissues, bone marrow, and spleen were downloaded from the GEO dataset (accession No. GSE24207). Data were analyzed by using "affy", "Affymetrix HG-U133A" and "annotate" packages from Bioconductor version Release 3.14 in the R programming language (R v4.1.1 for Windows).

Evaluation of accessible extracellular domains

To evaluate the accessible extracellular domains of the plasma membrane proteins, PROTTER (http://wlab.ethz.ch/protter/#), an interactive protein feature visualization software (ETH University, 2023, Switzerland), was used (17).

Evaluation of gene expression at the protein level

To confirm the protein expression of the candidate genes, besides the RNA expression, the pancreatic human proteome of the Protein Atlas database (https://www.proteinatlas.org/ENSG00000138161 and https://www.proteinatlas.org/ENSG00000164756-SLC30A8/tissue/pancreas#img - available from v21.0.proteinatlas.org) was used (18, 19). In this database, a protein profiling has been done through the immunohistochemistry (IHC) method and the final images have been deposited. To obtain the human proteome, protein expression data were extracted from 44 normal human tissue types. All primary pictures of the normal tissues IHC profile, as well as scientific annotations of protein expression stages, have been provided. The databank has been included 15323 genes (76%) that antibodies were accessible. The IHC process has been done according to the standard protocol of the Protein Atlas database. Briefly, Paraffin slides have been dried at room temperature overnight (20).

The slides have been deparaffinized in xylene (Cat No. 22-050-283, Thermo Fisher Scientific, Germany) and graded ethanol (Cat No. 64-17-5, Sigma, USA) to distilled water before immunostaining. Endogenous peroxidase has been blocked with 0.3% H_2O_2 plus 95% ethanol. The Heat-Induced Epitope Retrieval (HIER) has been performed by heating the slides immersed in the antigen retrieval buffer (Cat No. ab93684, Abcam, UK) for 4 minutes at 125°C. When the boiling step has been done, the immunohistochemical staining program has been run by Autostainer 480® instrument (AUTOSTAINER 480S-2D 220-240., ThermoFisher Scientific, Runcorn, UK). The slides have been cleaned in the washing buffer (Cat No. ab206977, Abcam, UK) and incubated with Ultra V Block (TA-060-UB, ThermoFisher Scientific, UK) for 5 minutes. After that, the slides were washed in the washing buffer (2X) Or a washing solution (Cat No. ab252273, Abcam, UK), then incubated with primary antibody (Thermo Fisher Scientific, Germany) for 30 minutes, and consequently rinsed in the wash buffer (3X) (Cat No. ab252273, Abcam, UK). Then, the slides were incubated with labeled HRP antibodies (Cat No. ab102890, Abcam, UK) for 30 minutes and cleaned in the wash buffer (2X) solution (Cat No. ab252273, Abcam, UK). Moreover, slides have been incubated in 3,3′-Diaminobenzidine (DAB) solution for 5 minutes and bleached in hematoxylin for 7.5 minutes. In the next step, slides were washed in the lithium carbonate water (Cat No. ab235613, Abcam, UK) and diluted in 1:5 in the saturated PBS solution for 1 minutes. Finally, the slides have been dehydrated in graded ethanol and followed by coverslips mounted. All reagents are applied at a volume of 300 µl per slide and antibody concentrations have been optimized based on their manufacture data sheets (20). The DAB labeled antibodies have been used to stain tissue microarrays, which have been then counterstained with hematoxylin (Cat No. ab220365, Abcam, UK). With the exception of the endometrium, skin, soft tissue and stomach, which are each represented by samples from six people, each tissue type is represented by samples from three individuals. Additional tissues have been stained for chosen proteins, including mouse brain, human samples of different tissues, including lactating breast, eye, thymus, and extended adrenal gland, skin, and brain. Immunohistochemical staining slides from tissue microarrays have been digitized and uploaded to the HPA online portal for analysis and presentation. All tissue samples have been obtained from the Department of Clinical Pathology, Uppsala University Hospital, Uppsala, Sweden, as part of the sample collection managed by the Uppsala Biobank and have been collected and handled in compliance with Swedish laws and regulations. All tissue samples have been anonymized in compliance with the Uppsala Ethical Review Board's permission and advisory report. Finally, selected genes were searched by gene names and IHC images and appendix data were extracted from the pancreas tissue section.

Investigation of antigen functions using the GeneCards database

Some cell surface antigens become inactivated or activated/hyperactivated after targeting. Hence, natural functions or homeostasis of the target cells can be altered after binding the antibodies to their surface antigens. For investigating the functions of selected antigens, the GeneCards database was employed. Using the GeneCards database, the candidate proteins were divided into two groups according to their functional or structural roles.

Statistical analysis

To better discover the tissue enrichment of total extracted genes, the analysis of the genes was carried out using the TissueEnrich package. According to the standard protocol, One-way ANOVA was employed to

measure variances between 35 separate tissue groups for every single gene. According to the TissueEnrich tutorials, the threshold of fold changes can be altered as the grade of tissue specificity of certain genes. The threshold was set as 4-fold changes by the TissueEnrich tool. In other words, genes from pancreas tissue-enhanced groups which reach at least 4-fold expression level in comparison with the average levels in all other tissues, were considered as pancreas tissue-specific genes (21).

For investigation of the significance of gene expression levels with the mean expression level of all genes, the ANOVA approach was recruited as the most valuable model for microarray data analysis. The standard protocol of Benjamini and Hochberg was employed to measure the gene differential expression (22, 23). The ANOVA model for microarray data was determined in two stages. The first stage is the normalization formula as follows:

yijgr=μ+Ai+Dj+ADij+rijgr

yijgr: logarithm of signal intensity for the ith array

jth dye, gth gene and rth measurement: the overall average expression level

A: The effect of the array at the measured intensity

D: The effect of the dye

AD: The effect of interaction between dye and array

* The first stage aims to calculate the μ term from the measured intensities.

In the second stage, the effects of factors on the gene expression were modeled in terms of the normalization phase as follows:

rijgr=G+Vgi+DGj+AGi+€ijr

G: The average intensity for a particular gene

AGi: The effect of an array on this gene

DGj: The effect of dye on this gene

€ijr: The residual value

The major interest in the microarray data analysis was the variety-by-gene (VG) label which represented the expression level variability among samples for a specific gene. The VGij was introduced as a ''call parameter'' which in the main cases, is a value of the state corresponding to the sample for j dye and i array.

Results

Four hundred twenty-nine genes are highly expressed in the pancreas tissue

To evaluate the pancreatic-specific antigens, proteins that are preferentially expressed in the pancreas tissue were detected. For this aim, a data mining method was carried out for the RNA sequencing expression results from multiple comprehensive sources. Based on the protein atlas data source, transcriptome research indicated that the pancreas expresses 69% of all human proteins (n=13887/20090), and also 321 of that genes showed an increase expression level in the pancreas in comparison with other body tissues. One of the input gene sets for analyzing pancreatic antigens is the group of 321 genes that are elevated in the pancreas that were extracted from the HPA database. Moreover, 8 genes were detected only in the pancreas which were added to the previous 321 gene set. The top 100 genes expressed from 305 pancreas samples were chosen from the GTEx source. Totally, 429 genes were prepared as genes that were highly expressed in the pancreas tissue for the following analysis (Supplementary File 1, See Online Information at www.celljournal.org).

Sixty-eight genes have a high preferential expression in the pancreas in comparison with other tissues

For a better analysis of this 429 genes group, the TissueEnrich package in the R programming language was employed. We found that 68 genes (69 of 429) have a high preferential expression level in the pancreas tissue. Genes, that showed an expression level of more than 1 (TPM or FPKM), also had a minimum five-fold expression levels in the pancreas tissue in comparison to all body tissues. Among the group of 429 extracted genes, 24 genes had a moderate to low preferential expression level in some other body tissues (Fig.1). The results showed that all 68 genes had a high expression level in the pancreas in comparison to other body tissues, while some of them had different expression levels in some tissues (Fig.2).

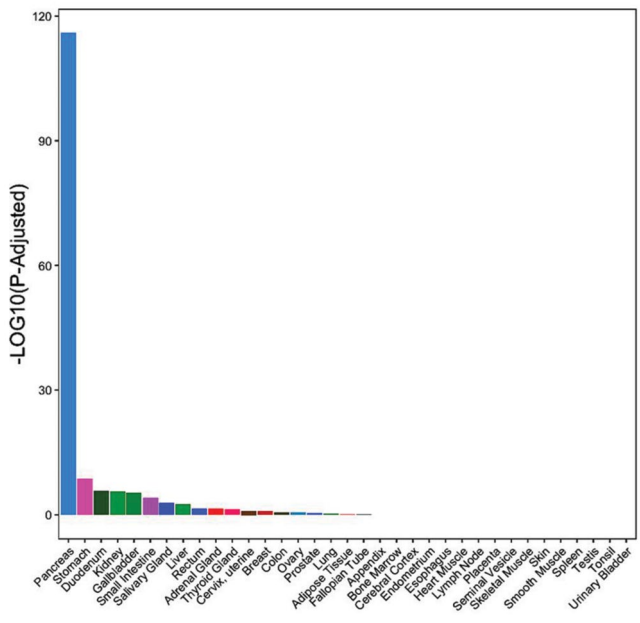

Fig.1: The tissue-specific gene enrichment plot of pancreatic chosen genes. The graph shows the enrichment expression of 429 genes that were evaluated in different 35 tissues.

Eighteen genes encoded plasma membrane proteins that have enough expression level microarray data analysis based on GEO database

The output of the TissueEnrich tool, a group of 68 genes, was analyzed to select the genes that encoded plasma membrane proteins and could be considered as a pancreatic cell surface antigen. The results showed that eighteen genes were encoded proteins that were single or multiple-pass trans plasma-membrane as well as anchored membrane proteins (Table 1). To confirm the expression of these genes in the islet and acinar cells, microarray data (accession No. GSE24207) of 3 samples were analyzed separately for multiple tissues including islet cells of the pancreas, acinar cells of the pancreas, adipose tissue, heart, small intestine, diaphragm tissues, bone marrow, and spleen. The results revealed that 8 genes have high RNA expression levels in islets in comparison with the insulin gene, as a positive control gene of pancreatic beta cells. Also, the results showed that 8 genes have a good expression level in the pancreatic islet cells. Interestingly, 5 genes that exhibit a strong RNA expression level in the pancreatic islet cells also have high RNA expression levels in the pancreatic acini cells in comparison to other tissues (Fig.3).

Except for *SLC30A8*, all seventeen candidate genes encode plasma membrane proteins with enough accessible extracellular domains

The appropriate cell surface antigen should have enough accessible extracellular domains for the production of specific antibodies. Evaluating these extracellular domains of the plasma membrane proteins, the Protter software, an interactive protein feature visualization, was employed. The results showed that most of the plasma membrane proteins had accessible extracellular domains for targeting antibodies except for SLC30A8. The SLC30A8 protein just had intracellular and transmembrane domains without sufficient extracellular regions and had to be omitted from the candidate gene list (Fig.4, Supplementary File 2, See Online Information at www.celljournal.org).

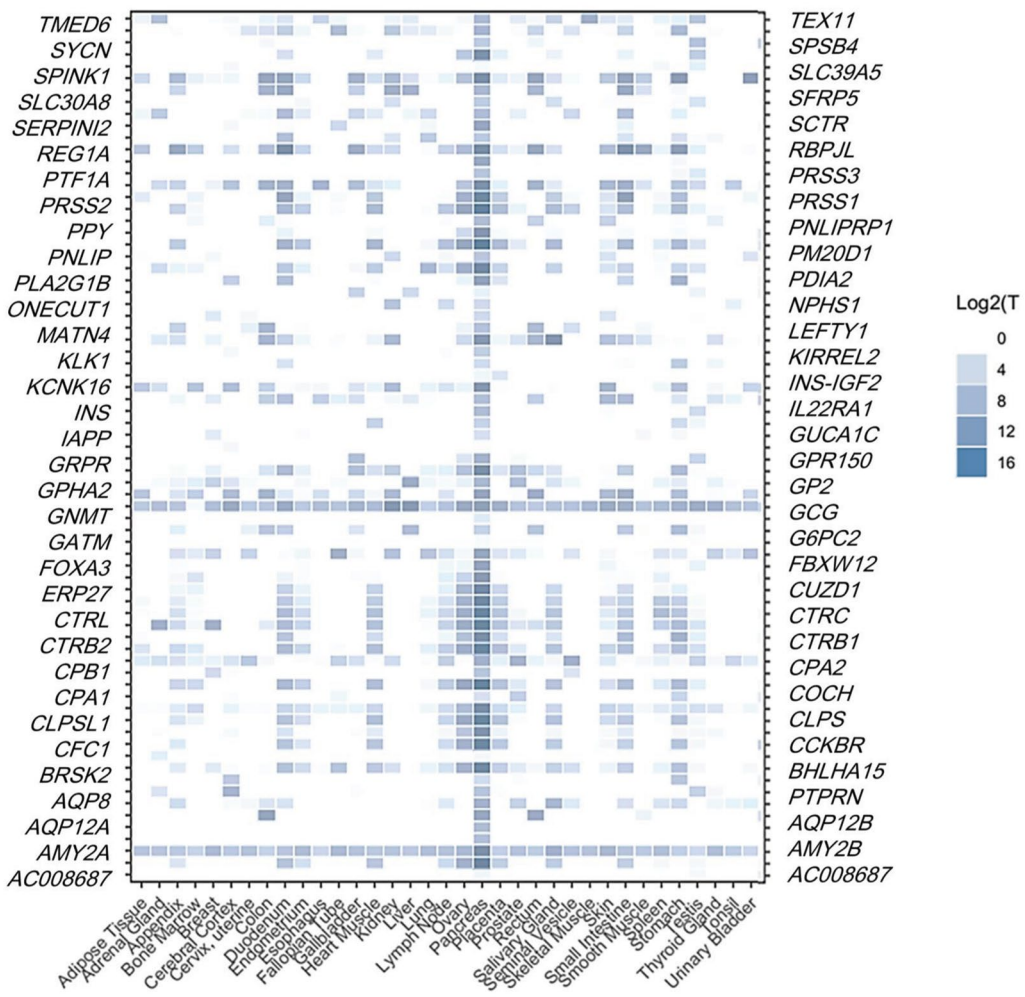

Fig.2: The Heatmap plot of pancreas enriched expression genes. The plot showed all 68 genes have a high preferential expression level in the pancreas. Some genes, such as insulin (*INS*), were only detected in the pancreas, and some others, such as Glycine Amidinotransferase (*GATM*), had a basic expression level in other body tissues.

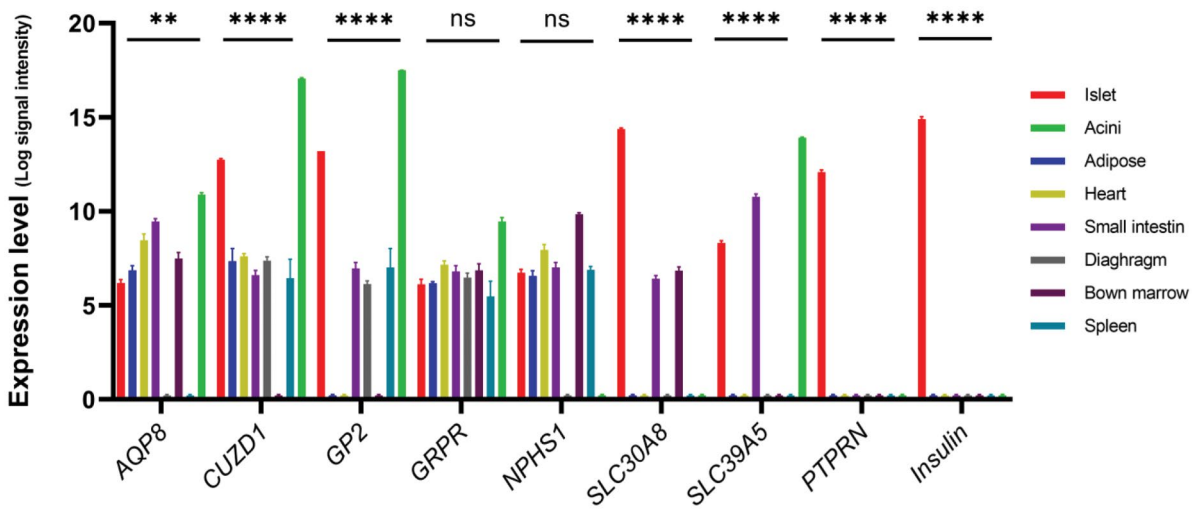

Fig.3: RNA expression level of plasma membrane proteins of the pancreatic islet cells and other tissues. Islets of the pancreas, acinar cells of the pancreas, adipose tissue, heart, small intestine, diaphragm tissues, bone marrow, and spleen tissues were evaluated by microarray data analysis (accession No. GSE24207). All 8 candidate genes have acceptable expression levels in the Pancreas beta cells. Five genes also have a good expression level in the Pancreas acini cells. The results showed that *AQP8, CUZD1, GP2, SLC30A8, SLC39A5,* and *PTPRN* genes were significantly expressed in the Pancreatic islet and acini cells. Also, the data showed that *GRPR* and *NPHS1* were not significantly expressed in the Pancreas islets and acini cells in compression with other tissues. According to the results of analysis, P Value of *AQP8, CUZD1, GP2, SLC30A8, SLC39A5*, and *PTPRN* genes were 3×10^{-3}, 5×10^{-5}, 7×10^{-5}, 3×10^{-5} and 2×10^{-5}, respectively. ****; P<0.0001, **; P<0.01, and ns; Not significant.

Table 1: Different functional and structural proteins among selected pancreatic surface antigens

Functional proteins		Structural proteins	
Gene name	**Function in the pancreas**	**Gene name**	**Function in the pancreas**
SLC39A5	Zinc transporters	*NPHS1*	Cell adhesion molecules
SLC30A8	Zinc-efflux transporter	*KIRREL2*	Cell adhesion molecules. The encoded protein localizes to adherent junctions in pancreatic beta cells and regulates insulin secretion.
SCTR	G protein-coupled receptor and belongs to the glucagon-VIP-secretin receptor	*GP2*	Integral membrane protein associates with the plasma membrane via glycosylphosphatidylinositol (GPI) linkage
KCNK16	Potassium channel proteins	*CUZD1*	Localized to zymogen granules, where it functions in trypsinogen activation (by similarity). May indirectly regulate cell motility, cell-cell and cell/extracellular matrix interactions
IL22RA1	Cytokine receptor		
GRPR	Receptor for gastrin-releasing peptide		
GPR150	Rhodopsin-like family of G-protein-coupled receptors		
CFC1	Member of the epidermal growth factor which are involved in signaling during embryonic development		
CCKBR	G-protein coupled receptor for gastrin and cholecystokinin (CCK)		
AQP8	Aquaporins facilitate the transport of water and small neutral solutes across cell membranes		
AQP12B	Aquaporin		
AQP12A	Aquaporin		
PTPRN	Protein tyrosine phosphatase receptor type N		

Seventeen candidate proteins are highly expressed in the pancreatic tissue and localized in the cytoplasm/ plasma membrane area

To validate the expression of the candidate antigens at the protein level, the IHC data of the pancreatic tissue from the HPA database were used. The results of normal pancreas tissue sections illustrated that five genes of these ten genes, including *GP2, CUZD1, AQP12A, AQP12B,* and *SCTR*, had a high protein expression level in the exocrine cells of the pancreas. In addition, three genes, including *CFC1, KIRREL2*, and *PTPRN*, showed a high expression level in the pancreatic endocrine cells. Based on the IHC data from the protein atlas database, three other genes, including *GPR150, AQP8*, and *NPHS1*, were highly or moderately expressed in both endocrine and exocrine cells at the protein level. Interestingly, IHC data revealed that all selected antigens were localized in the cytoplasm or plasma-membrane (Fig.5, Supplementary File 3, See Online Information at www.celljournal.org).

Four structural proteins were among all seventeen candidate proteins

After interaction with antibodies, several cell surface proteins become inactivated or activated/hyperactivated. As a result, the target cells homeostasis, and regular functions were disrupted. However, appropriate surface antigens should not have vital functions for natural cells. For the investigation of the functions of the candidate antigens, the GeneCards database was used. Finally, the antigens were categorized into two groups "Functional genes" and "Structural genes". Among all antigens, NPHS1, KIRREL2, GP2, and CUZD1 were structural proteins that will be less affected after an antibody binding.

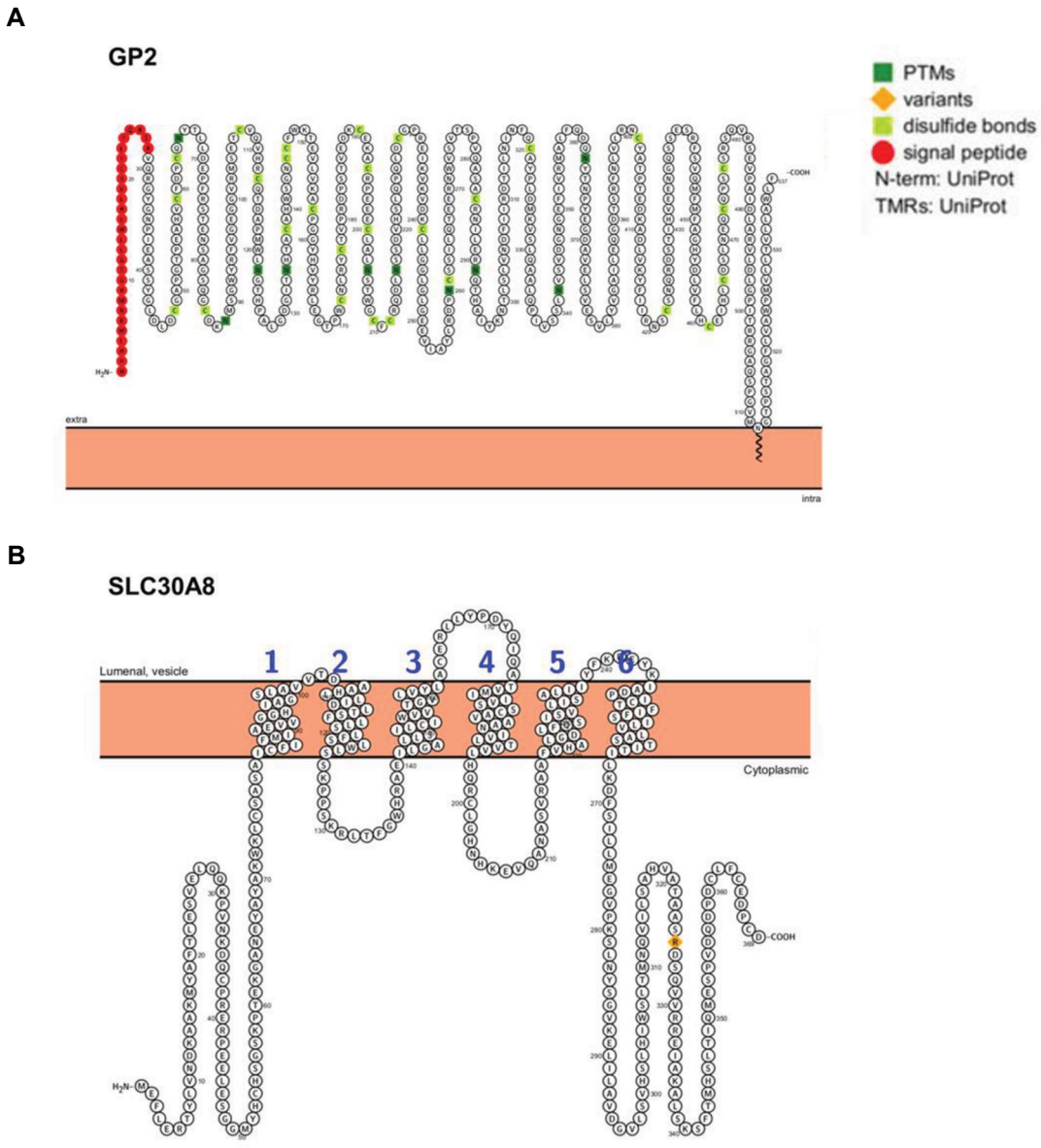

Fig.4: Two examples of the Protter outputs of candidate plasma membrane antigens. **A.** The results showed that glycoprotein 2 (GP2) had an extracellular domain of about 510 amino acids which binds to the plasma membrane by a glycosylphosphatidylinositol (GIP) anchor. **B.** In comparison, the SLC30A8 protein did have not an adequate extracellular domain (just about 15 amino acids) for being targeted by antibodies or immune Treg cells.

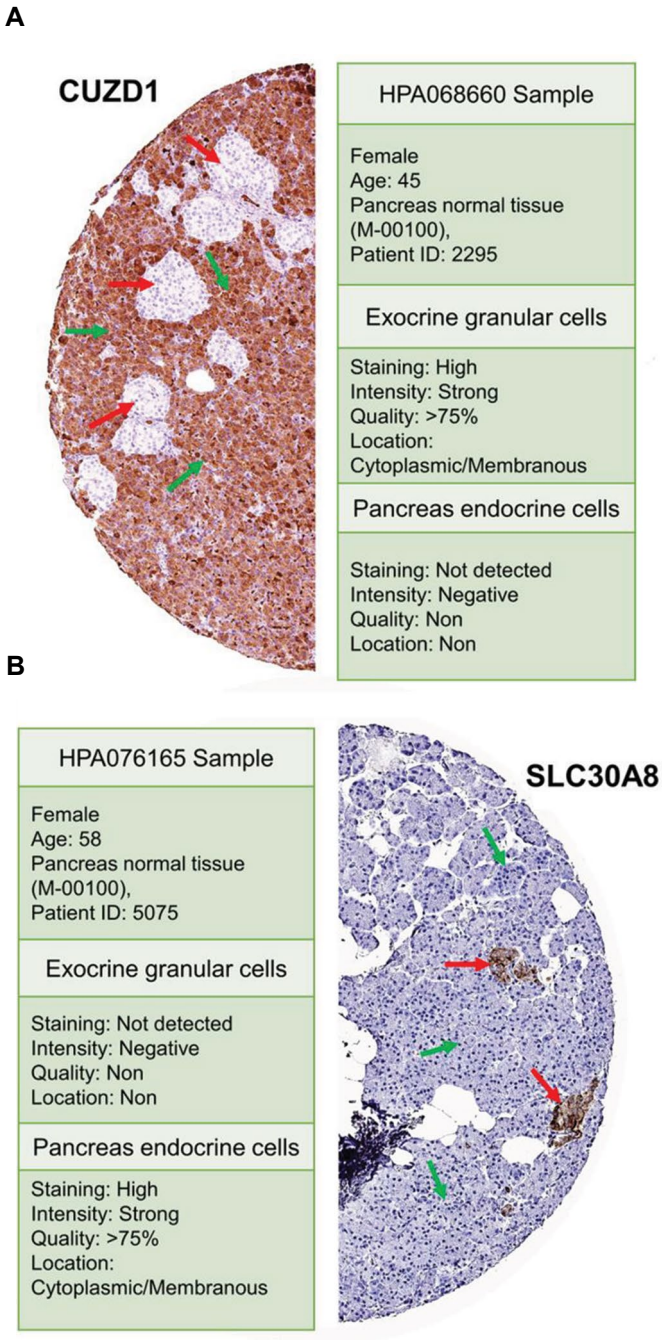

Fig.5: Two examples of immunohistochemistry (IHC) images and localized area for candidate antigens. **A.** IHC data showed that the CUZD1 protein had a high expression in the pancreas exocrine cells and displayed a cytoplasmic/membranous subcellular localization. **B.** IHC data showed that the SLC30A8 protein had a high expression in the pancreatic islet endocrine cells and showed a cytoplasmic/membranous subcellular localization. Red arrows indicate the islets of Langerhans, endocrine cells including insulin-producing beta cells, and green arrows reveal the exocrine cells, including acini.

Discussion

As mentioned earlier, in T1Ds, insulin-producing beta cells are destroyed by the patient's immune cells. Since in most cases, about one-third of beta cells are healthy at diagnosis, preventing immune system attack can lead to the disease avert through repairing pancreatic tissue. Some strategies have been proposed and used to suppress the immune system in T1Ds, but approaches that do it locally are preferred. Therefore, one of the main strategies is targeting pancreatic cells which become desirable recently and is of significant importance (24). In this study, to investigate comprehensively the total antigens of pancreatic tissues to achieve the appropriate surface antigen to target beta cells in T1Ds, bioinformatics systems, and comprehensive databases were employed. Overall, a suitable antigen should have three main features, as follows: i. Having high and preferential expression in pancreatic cells, ii. Having enough accessible extracellular domains, and iii. Not having any critical function that affects the natural behavior of the cells after being targeted. It can be said with almost certainty that there is no antigen in the body tissues that possesses all the desired features (25). So, the candidate antigens should be investigated more carefully in different aspects.

Muraro et al. (26) developed a platform that captured the transcriptomes of pancreatic cells obtained from dead organ donors by using FACS and the CEL-Seq2 protocol. They presented the top 10 expressed genes in islet cells, including alpha, beta, delta, and PP cells. Among all 40 genes that Muraro et al. (26) presented, 8 genes were contained within our list of 429 genes set. The other 32 genes they presented were either intracellular or expressed in other body tissues. Therefore, they were not appropriate biomarker as specific surface antigens. Other authors also used a single-cell RNA sequencing approach to assess the transcriptomes of 12,000 separate pancreatic cells from four different human pancreases. They represented 15 pancreas marker genes, five of which are included in our list of 429 genes set. The remaining 10 proteins (10/15) were intracellular or expressed in other body tissues. Hence, as before, they were not suitable as specific surface antigens (27).

The SLC30A8, also known as ZnT8, is a transporter protein efflux zinc. In the secretory granules of insulin, the encoded protein co-localizes with an insulin structure in beta cells. Having multi-pass transmembrane helices and not having any accessible extracellular domains make this antigen unsuitable for targeting by antibodies (27). The *NPHS1* gene encodes a protein for cell adhesion from the immunoglobulin family that plays a role in the kidney's glomerular filtration barrier. Based on the protein atlas database, the protein expression level of NPHS1 antigen was low in the pancreas but high in kidney tissues, particularly in the glomeruli cells (28). Due to the potential the antibodies of the renal off-target, it is less suitable. The *KIRREL2* gene also encodes a transmembrane protein which is one of the immunoglobulin superfamily of cell adhesion molecules (29). This protein controls the insulin secretion by binding to adherent junctions in the pancreatic beta cells. The protein expression level of this gene, based on the protein atlas database, for pancreatic exocrine glandular cells is high, although in the kidney tissue, particularly in the tubule cells, is low. Therefore, it seems to be a more suitable biomarker than *NPHS1* for the pancreatic tissue targeting. The *GP2* gene encodes an inherent plasma membrane protein that forms a

glycosylphosphatidylinositol (GPI) connection with the plasma membrane (30). According to our IHC finding of the protein atlas database, the GP2 has a high protein expression level in the pancreas tissue. It seems to be a suitable aim for acinar cells targeting in comparison with other antigens in the pancreas. The CUZD1 is a protein found in zymogen granules that aids in the activation of trypsinogen (31). Cell motility, cell-cell, and cell-extracellular matrix interactions may all be influenced indirectly. Based on pancreatic expression protein levels, it is another suitable potential target of exocrine cells in the pancreas.

Several studies by subjects similar to the present investigation were done yet. One of them is the research that was completed by Bausch-Fluck et al. (32). They employed a machine-learning method that find specific domains of proteins to make a prediction software based on the transmembrane topology that predicts the surface proteins (33). Their goal was only highlighting overexpressed surface proteins of cancer cells. They developed an RNA sequencing analyser software named Qsurface that extracts the data from the Cancer Genome Atlas (TCGA) database as inputs and at the end presents surface proteins. The limitation of their method was that the Qsurface just uses gene expression at RNA levels and doesn't regard protein expression since it is a more important attribute. Two other similar studies were done by Pont et al. (34) and Orentas et al. (35). In the first gene expression effort from microarray, databases were employed to find surface antigens for the Hematological Malignancies immunotherapy. In the second, an algorithm was written based on the tumour cell gene expression to capture the surface antigens of cancerous cells. Both investigations were about cancers, not normal tissues, and used just RNA expression, not proteins. In the present study, about some gene cases such as *AQP8, CUZD1, GP2* and *SLC39A5*, it seems to be a mismatched expression in protein and RNA levels. This matter can be due to many causes. The stability of the transcript, translation rate of the transcript, stability of translated polypeptide as well as the association of multi-subunit protein components are the main reasons for different RNA and protein expression of the same genes. Our results showed that *PTPRN* and *SLC30A*8 genes have RNA expression just in islet cells. Our present finding revealed that *AQP8, CUZD1, GP2*, and *SLC39A5* genes express in the RNA level in both islets and acini cells. It seems that the only transcript of these genes is functional protein in the cells, although their amount is not enough for an IHC assay. The affinity of a specific antibody which used in the IHC assay also is important to detect the low-level expression of the proteins in the cells. On the other hand, several studies confirmed the RNA expression of *AQP8, CUZD1, GP2*, and *SLC39A5* genes in both islets and acini cells (36).

There are some limitations to the pipeline introduced in the presented study. One of them is that the PTPRN protein plays a role in vesicle-mediated secretory processes and was categorized as a "functional" group antigen, hence it is "structural" group antigen after the fusion of vesicles to the plasma membrane. On the other hand, what makes this protein less suitable antigen, as a surface antigen, is that the most of its extracellular domains are cut and separated before the vesicle fuses with the plasma membrane. This cutting phenomenon can be found just in the literature but not in expression databases. Evaluating all features of selected antigens for future practical investigations, the literature should be considered besides the bioinformatics (37). Another limitation of this investigation is that the categorizing of the candidate antigens was not absolute .Hence, among functional proteins, there were antigens that were less affected when were banded, and conversely, antigens that were highly affected. For example, receptors involved in signaling pathways or G protein-like receptors (e.g., SCTR, GPR150, CCKBR), hormone or cytokine receptors (e.g., IL22RA1, GRPR, CFC1, and CCKBR), and specific ion transporters or channels (e.g., SLC39A5, SLC39A8, and KCNK16) are greatly affected if targeted by antibodies. In contrast, water channels (e.g., Aquaporin), since they are more abundant, may not cause major problems upon some of them being blocked by binding antibodies in their overall function of the cells (38). Hence, the proteins just categorized as two 'Functional' and 'Structural' groups that should be evaluated in more detail in individual cases. However, the candidate surface antigens presented in this study can be recruited to develop the potential monoclonal antibodies as well as recombinant antibody fragments such as scFvs for binding to the pancreatic beta cells in the strategies to overcome type I diabetes disease.

Conclusion

One effective way to treat the T1Ds is a local suppression of the immune system to prevent it from invading insulin-producing cells. The first and perhaps most important step in local inhibition is to find the proper antigens to target pancreatic tissue cells. Therefore, in this study, we evaluated the antigens of pancreatic tissue cells comprehensively. Based on our results, 18 genes encoded plasma membrane proteins on pancreatic cells, of which, NPHS1, KIRREL2, GP2, and CUZD1 antigens are structural and probably more suitable for targeting. In addition, water channel proteins (Aquaporins), despite being a functional channel protein, can be used as target antigens. However, the proposed antigens or any other antigen needs to be verified using experimental and laboratory work to prove that the antigen is suitable for targeting. In addition, the antibody type, affinity, and specificity are very important factors for developing an efficient treatment.

Acknowledgements

Dr. Javad Zahiri, Dr. Maryam Hasanlou, Dr. Ali Fasihi and Dr. Hadi Najafi have helped us in the microarray data analysis, which we appreciate greatly. The research reported in this publication was supported by the Elite

Researcher Grant Committee under award number [989951] from the National Institute for Medical Research Development (NIMAD), Tehran, Iran. This project was also financially supported by Iran National Science Foundation (INSF, Grant No. 97011360). There is no conflict of interest in this study.

Authors' Contributions

H.D., M.H.-A., V.Z., Z.M., M.S., E.H.-S.; Conceptualization, Methodology, Resources, Visualization, Supervision, and Funding acquisition. All authors read and approved the final manuscript.

References

1. Ilonen J, Lempainen J, Veijola R. The heterogeneous pathogenesis of type 1 diabetes mellitus. Nat Rev Endocrinol. 2019; 15(11): 635-650.
2. Paschou SA, Papadopoulou-Marketou N, Chrousos GP, Kanaka-Gantenbein C. On type 1 diabetes mellitus pathogenesis. Endocr Connect. 2018; 7(1): R38-R46.
3. Jwad SM, AL-Fatlawi HY. Types of diabetes and their effect on the immune system. J Adv Pharm Pract. 2022; 4(1): 21-30.
4. Perry JS, Hsieh CS. Development of T-cell tolerance utilizes both cell-autonomous and cooperative presentation of self-antigen. Immunol Rev. 2016; 271(1): 141-155.
5. Labrijn AF, Janmaat ML, Reichert JM, Parren PWHI. Bispecific antibodies: a mechanistic review of the pipeline. Nat Rev Drug Discov. 2019; 18(8): 585-608.
6. Kufer P, Lutterbüse R, Baeuerle PA. A revival of bispecific antibodies. Trends Biotechnol. 2004; 22(5): 238-244.
7. Topp MS, Stelljes M, Zugmaier G, Barnette P, Heffner LT Jr, Trippett T, et al. Blinatumomab retreatment after relapse in patients with relapsed/refractory B-precursor acute lymphoblastic leukemia. Leukemia. 2018; 32(2): 562-565.
8. Balcerek J, Shy BR, Putnam AL, Masiello LM, Lares A, Dekovic F, et al. Polyclonal regulatory T cell manufacturing under cGMP: a decade of experience. Front Immunol. 2021; 12: 744763.
9. Mohseni YR, Tung SL, Dudreuilh C, Lechler RI, Fruhwirth GO, Lombardi G. The future of regulatory T cell therapy: promises and challenges of implementing CAR technology. Front Immunol. 2020; 11: 1608.
10. Tarbell KV, Yamazaki S, Olson K, Toy P, Steinman RM. CD25+ CD4+ T cells, expanded with dendritic cells presenting a single autoantigenic peptide, suppress autoimmune diabetes. J Exp Med. 2004; 199(11): 1467-1477.
11. Tang Q, Henriksen KJ, Bi M, Finger EB, Szot G, Ye J, et al. In vitro-expanded antigen-specific regulatory T cells suppress autoimmune diabetes. J Exp Med. 2004; 199(11): 1455-1465.
12. Chopra A, Shan L, Eckelman WC, Leung K, Latterner M, Bryant SH, et al. Molecular imaging and contrast agent database (MICAD): evolution and progress. Mol Imaging Biol. 2012; 14(1): 4-13.
13. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science. 2020; 367(6482): eaay5947.
14. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic Acids Res. 2015; 43(Database issue): D662-D669.
15. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Sci Adv. 2021; 7(31): eabh2169.
16. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 2013; 41(Database issue): D56-D63.
17. Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics. 2014; 30(6): 884-886.
18. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science. 2017; 356(6340): eaal3321.
19. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Sci Adv. 2021; 7(31): eabh2169.
20. Kampf C, Olsson I, Ryberg U, Sjöstedt E, Pontén F. Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. J Vis Exp. 2012; (63): 3620.
21. Peng S, Zhou Y, Xiong L, Wang Q. Identification of novel targets and pathways to distinguish suicide dependent or independent on depression diagnosis. Sci Rep. 2023; 13(1): 2488.
22. Kuo L, Yu F, Zhao Y. Statistical methods for identifying differentially expressed genes in replicated microarray experiments: a review. Stat Adv Biomed Science. 2008: 341.
23. Cramer AO, van Ravenzwaaij D, Matzke D, Steingroever H, Wetzels R, Grasman RP, et al. Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. Psychon Bull Rev. 2016; 23(2): 640-647.
24. von Scholten BJ, Kreiner FF, Gough SCL, von Herrath M. Current and future therapies for type 1 diabetes. Diabetologia. 2021; 64(5): 1037-1048.
25. Haen SP, Löffler MW, Rammensee HG, Brossart P. Towards new horizons: characterization, classification and implications of the tumour antigenic repertoire. Nat Rev Clin Oncol. 2020; 17(10): 595-610.
26. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A single-cell transcriptome atlas of the human pancreas. Cell Syst. 2016; 3(4): 385-394. e3.
27. Kawasaki E. ZnT8 and type 1 diabetes. Endocr J. 2012; 59(7): 531-537.
28. Zhuo L, Huang L, Yang Z, Li G, Wang L. A comprehensive analysis of NPHS1 gene mutations in patients with sporadic focal segmental glomerulosclerosis. BMC Med Genet. 2019; 20(1): 111.
29. Korshunov A, Okonechnikov K, Stichel D, Schrimpf D, Delaidelli A, Tonn S, et al. Gene expression profiling of Group 3 medulloblastomas defines a clinically tractable stratification based on KIRREL2 expression. Acta Neuropathol. 2022; 144(2): 339-352.
30. Lin Y, Nakatochi M, Hosono Y, Ito H, Kamatani Y, Inoko A, et al. Genome-wide association meta-analysis identifies GP2 gene risk variants for pancreatic cancer. Nat Commun. 2020; 11(1): 3175.
31. Rygiel AM, Unger LS, Sörgel FL, Masson E, Matsumoto R, Ewers M, et al. Variants in the pancreatic CUB and zona pellucida-like domains 1 (CUZD1) gene in early-onset chronic pancreatitis - A possible new susceptibility gene. Pancreatology. 2022; 22(5): 564-571.
32. Bausch-Fluck D, Goldmann U, Müller S, van Oostrum M, Müller M, Schubert OT, et al. The in silico human surfaceome. Proc Natl Acad Sci USA. 2018; 115(46): E10988-E1099.
33. Hong Y, Park C, Kim N, Cho J, Moon SU, Kim J, et al. QSurface: fast identification of surface expression markers in cancers. BMC Syst Biol. 2018; 12 Suppl 2: 17.
34. Pont MJ, Honders MW, Kremer AN, van Kooten C, Out C, Hiemstra PS, et al. Microarray gene expression analysis to evaluate cell type specific expression of targets relevant for immunotherapy of hematological malignancies. PLoS One. 2016; 11(5): e0155165.
35. Orentas RJ, Yang JJ, Wen X, Wei JS, Mackall CL, Khan J. Identification of cell surface proteins as potential immunotherapy targets in 12 pediatric cancers. Front Oncol. 2012; 2: 194.
36. Magaki S, Hojat SA, Wei B, So A, Yong WH. An introduction to the performance of immunohistochemistry. Methods Mol Biol. 2019; 1897: 289-298.
37. Toledo PL, Vazquez DS, Gianotti AR, Abate MB, Wegbrod C, Torkko JM, et al. Condensation of the β-cell secretory granule luminal cargoes pro/insulin and ICA512 RESP18 homology domain. Protein Sci. 2023; 32(6): e4649.
38. da Silva IV, Soveral G. Aquaporins in immune cells and inflammation: new targets for drug development. Int J Mol Sci. 2021; 22(4): 1845.