



OPEN ACCESS

EDITED BY

Yanshan Dai,
Bristol Myers Squibb, United States

REVIEWED BY

Yu Wang,
University of Virginia, United States
Fei Shao,
Capital Medical University, China
Siqian Wang,
Columbia University, United States

*CORRESPONDENCE

Si Ri Gu Leng Sana
✉ sana820816@163.com

†These authors have contributed
equally to this work and share
first authorship

RECEIVED 03 October 2023

ACCEPTED 01 December 2023

PUBLISHED 19 December 2023

CITATION

Shen J, Wang Y, Deng X and Sana SRGL
(2023) Combining bioinformatics and
machine learning algorithms to identify
and analyze shared biomarkers and
pathways in COVID-19 convalescence
and diabetes mellitus.
Front. Endocrinol. 14:1306325.
doi: 10.3389/fendo.2023.1306325

COPYRIGHT

© 2023 Shen, Wang, Deng and Sana. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Combining bioinformatics and machine learning algorithms to identify and analyze shared biomarkers and pathways in COVID-19 convalescence and diabetes mellitus

Jinru Shen^{1†}, Yaolou Wang^{1†}, Xijin Deng²
and Si Ri Gu Leng Sana^{3*}

¹The First Clinical Medical School, Harbin Medical University, Harbin, China, ²Department of Anaesthesiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, ³Department of Anaesthesiology, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Background: Most patients who had coronavirus disease 2019 (COVID-19) fully recovered, but many others experienced acute sequelae or persistent symptoms. It is possible that acute COVID-19 recovery is just the beginning of a chronic condition. Even after COVID-19 recovery, it may lead to the exacerbation of hyperglycemia process or a new onset of diabetes mellitus (DM). In this study, we used a combination of bioinformatics and machine learning algorithms to investigate shared pathways and biomarkers in DM and COVID-19 convalescence.

Methods: Gene transcriptome datasets of COVID-19 convalescence and diabetes mellitus from Gene Expression Omnibus (GEO) were integrated using bioinformatics methods and differentially expressed genes (DEGs) were found using the R programme. These genes were also subjected to Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to find potential pathways. The hub DEGs genes were then identified by combining protein-protein interaction (PPI) networks and machine learning algorithms. And transcription factors (TFs) and miRNAs were predicted for DM after COVID-19 convalescence. In addition, the inflammatory and immune status of diabetes after COVID-19 convalescence was assessed by single-sample gene set enrichment analysis (ssGSEA).

Results: In this study, we developed genetic diagnostic models for 6 core DEGs between type 1 DM (T1DM) and COVID-19 convalescence and 2 core DEGs between type 2 DM (T2DM) and COVID-19 convalescence and demonstrated statistically significant differences ($p < 0.05$) and diagnostic validity in the validation set. Analysis of immune cell infiltration suggests that a variety of immune cells may be involved in the development of DM after COVID-19 convalescence.

Conclusion: We identified a genetic diagnostic model for COVID-19 convalescence and DM containing 8 core DEGs and constructed a nomogram for the diagnosis of COVID-19 convalescence DM.

KEYWORDS

COVID-19 convalescence, diabetes mellitus (DM), differentially expressed genes (DEGs), gene ontology (GO), protein-protein interaction (PPI), hub gene, machine learning

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for coronavirus disease 2019 (COVID-19) worldwide pandemic, can cause mild to severe respiratory disease and non-respiratory symptoms (1). This virus can be rapidly transmitted, infection is associated with a relatively high mortality rate, and the virus can often evade the host's immune response (2–4). Patient prognosis and long-term complications have become an increasingly important issue. In particular, an estimated 87.4% of patients who recovered from COVID-19 had at least 1 persistent symptom, especially in the neurological and respiratory systems (5).

Diabetes mellitus (DM) is a group of metabolic disorders characterized by chronic hyperglycemia that have multiple etiologies, all of which manifest as defects in insulin secretion and/or utilization. As of 2010, the global prevalence of DM in adults (20 to 79-years-old) was 6.4%, corresponding to 285 million cases. This prevalence was predicted to increase to 7.7% by 2030 (corresponding to 439 million adults), with the number of affected adults increasing by 69% in developing countries and by 20% in developed countries (6). DM is often divided into four categories according to its cause: Type 1 DM (T1DM), Type 2 DM (T2DM), gestational DM (GDM), and other types of DM (7). The destruction of pancreatic beta cells leads to the development of T1DM, and this type of diabetes often leads to an absolute deficiency of insulin. T2DM is characterized by insulin resistance, and decreased insulin secretion and decreased function of pancreatic beta cells may be the initiating factor in most cases. Although short-term hyperglycemia has no serious effects on the body, long-term hyperglycemia can lead to chronic changes, such as microvascular complications (e.g., diabetic nephropathy, diabetic retinopathy, and neuropathy) and devastating macrovascular complications, such as cardiovascular diseases, that can have irreversible and even fatal effects (8).

Many previous studies showed significant increases in the prevalence, severity, and mortality of COVID-19 in patients with DM compared to non-diabetic patients, suggesting an association of COVID-19 severity with poor glycemic control (9, 10). Other studies suggested that COVID-19 may predispose infected

individuals to hyperglycemia and promote the development of DM (11, 12).

SARS-CoV-2 binds to angiotensin-converting enzyme 2 (ACE2), and this protein is expressed in the lungs and many other organs, including the pancreas (13). This suggests that new-onset hyperglycemia and DM in patients with COVID-19 may be due to a direct attack of SARS-CoV-2 on islet β -cells in the pancreas. Therefore, it is crucial to identify the common biomolecules and pathways that are altered in patients with DM and patients undergoing convalescence following COVID-19. These shared biomolecules may have potential as biomarkers or therapeutic targets.

In this study, we used bioinformatics and machine learning algorithms to identify differentially expressed genes (DEGs) and predict altered molecular regulatory networks in patients undergoing convalescence from COVID-19, patients with T1DM, and patients with T2DM. Our findings may provide a basis for development of new measures that could be used for disease prevention and treatment in these patients.

Materials and methods

Acquisition of chip data

Data from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo>) were used to determine similarities of gene expression in patients undergoing COVID-19 convalescence (1, 3, and 6 months after hospital discharge), patients with T1DM, and patients with T2DM, with the search restricted to humans (Table 1). The GSE227116 dataset was generated by RNA sequencing (RNA-seq) of whole blood, and contains data from 75 samples: 65 patients after COVID-19 convalescence and 10 healthy donors. In particular, the study group was the population who had already recovered from COVID-19 infection. This sample was selected for follow-up analysis to investigate the long-term alterations after COVID-19 convalescence. For analysis of T1DM, three microarray datasets of peripheral blood mononuclear cells

TABLE 1 Training and validation datasets used for analysis.

Clinical status	Accession No.	Source	Platform	Cases : Controls
COVID-19 convalescence	GSE227116	Whole blood	GPL16791	65:10
Training				
T1DM	GSE193273	PBMCs	GPL20844	20:20
T1DM	GSE29142	PBMCs	GPL13507	9:10
T1DM	GSE55098	PBMCs	GPL570	12:10
T2DM	GSE156993	PBMCs	GPL570	12:6
T2DM	GSE163980	PBMCs	GPL20115	5:5
T2DM	GSE9006	PBMCs	GPL96	24:12
Validation				
T1DM	GSE33440	PBMCs	GPL6947	16:6
T2DM	GSE41762	pancreatic islets	GPL6244	20:57
COVID-19	GSE166253	PBMC	GPL20795	6:10

PBMCs, peripheral blood mononuclear cells; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

(PBMCs; GSE193273, GSE29142, and GSE55098) were selected (Table 1). GSE193273 contains data from 20 T1DM patients and 20 healthy controls; GSE29142 contains data from 9 T1DM patients and 10 healthy controls; and GSE55098 contains data from 12 T1DM patients and 10 healthy controls (Table 1). Similarly, for analysis of T2DM, three microarray datasets of PBMCs (GSE163980, GSE156993, and GSE9006) were selected; these data consist of gene expression data from 29 T2DM patients and 35 healthy controls.

Data processing and differential expression analysis

For the dataset of COVID-19 convalescent patients (GSE227116), the limma package in R software (version 4.3.0) was used to identify changes in gene expression (fold change ≥ 1.2 , $|\log_2(FC)| \geq 0.263$) (14, 15). All *P*-values were adjusted using the Benjamini-Hochberg correction, and the false discovery rate (FDR) threshold for DEGs was 0.05. Due to factors such as theoretical approximations, methodological difficulties, limitations in the sensitivity and resolving power of experimental instruments, instability of the surrounding environment, limitations in the observer's ability to discriminate between senses, and variability in technical proficiency, there will always be a deviation between the measurement results and the true value of the measurement. The problem of measurement error is equally present in this study. For analysis of the T1DM and T2DM datasets, the ComBat package was first used to process the gene expression data to eliminate batch effects (16). Then, the differentially expressed genes were obtained using the limma package in R software, as described above, and a heat map was generated (17).

Gene ontology and pathway enrichment analysis

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used to analyze gene-related functions. The GO database classifies gene function as cellular component (CC), molecular function (MF), and biological process (BP) (18), and the KEGG database provides information about the related pathways. The ClusterProfiler package in R software was used for subsequent analysis and identification of information about gene function and potential pathways (19). The conditional filtering used a *P*-value cutoff of 0.05, the ggplot2 package in R software was used for visualizing BP in the GO enrichment analysis, and the graph package in R was used for visualizing the KEGG pathway.

Protein-protein interaction network analysis

The filtered DEGs were uploaded to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (<https://string-db.org/>) (20). The minimum required interaction score was set at 0.4 (medium confidence), and the disconnected nodes were hidden to construct a protein-protein interaction (PPI) network. Cytoscape software was used for network display, layout, and query, to integrate the biological networks and molecular information (such as gene expression and genotype) in a visual environment, and to link these networks with functional annotation databases (21). The resulting PPI networks were imported into Cytoscape software version 3.9.1, and the DEGs were ranked and filtered using the Matthews correlation coefficient (MCC) algorithm.

Transcription factors and miRNAs

To explore the potential impact of core DEGs on subsequent molecular regulatory mechanisms, predicted transcription factors (TFs) and miRNAs from databases were obtained from the networkanalyst (<https://www.networkanalyst.ca/>). These data were from JASPAR, an open source database of TF binding sites in the form of position frequency matrices (PFMs) and TF flexible models (TFFMs) that record DNA binding preference information of TFs in six major species (22). This database identified topologically credible TFs. The TF-gene relationships were then imported into Cytoscape software version 3.9.1 to construct a visual regulatory network.

The miRTarBase database is a specialized collection of microRNA-mRNA targeting relationships (MTI, MicroRNA-Target Interactions), and all of its data were experimentally validated (23). This database was used to obtain DEG-associated miRNAs. miRNAs were retained if they interacted with more than two DEGs for construction of a visual miRNA-gene interaction network.

Construction and validation of a diagnostic DEG signature

Two machine learning algorithms were used to screen for shared changes in biomarkers in the two pair-wise comparisons (COVID-19 convalescence + T2DM, COVID-19 + T1DM). Elastic Net Regressions is a regularization algorithm that combines the features of Lasso regression and ridge regression with the advantages of sparsity and variable selection. When multiple features are related, Lasso regression may only randomly choose one of them, ridge regression can choose all of the features. By combining these two regularization methods using elastic net regressions, we are able to bring together the strengths of both methods (24). The shrinkage regularization parameter λ , which controls the complexity of the model, was determined by 10-fold cross-validation of the partial likelihood deviance and the attendant '1 standard error rule'. The elastic net used 10-fold cross-validation and fitted a linear model using a penalty score ($\alpha = 0.9$). The elastic net algorithm for variable reduction and selection utilized the glmnet package in R (25); the independent variables were the normalized expression matrix of DEGs, and the dependent variables were the presence or absence of disease in the sample using a 10-fold cross validation. Support vector machine-recursive feature elimination (SVM-RFE) was used to identify the optimal hyperplane that partitioned the training dataset and maximized the geometric interval (26). This calculation used a 5-fold cross validation within the e1071 package in R (27). The top-ranked DEGs were selected from PPI network identification and the intersection of the two machine learning algorithms. Then, column line plots and receiver operating characteristic (ROC) analysis with area under the curve (AUC) values were used to assess the model. The dataset was selected from an extensive search in the GEO database. The selection criteria for the DM validation set were as follows: (a) The datasets are all of the type of expression

profiling by array. (b) Both contain patients with diagnosed diabetes and controls from healthy people. External validation datasets in GEO were used for model validation: GSE33440 (16 T1DM and 6 healthy individuals) and GSE41762 (20 T2DM and 57 healthy individuals). The COVID-19 convalescence validation dataset GSE166253 (6 convalescent patients, and 10 healthy controls) was obtained from an extensive search in GEO.

Evaluation and correlation of immune cell infiltration

Single-sample gene set enrichment analysis (ssGSEA), a method commonly used for analysis of immune cell infiltration, was used to estimate the relative enrichment of a gene set in each sample by comparing the gene expression data of a sample with a specific immune cell gene set and to estimate the relative abundance of different immune cell types in each sample (28). The immune cell marker genes were from **Supplementary Table S1** in the study of Bindea et al., which provided information on 24 immune cells (29). The immune infiltration status of each sample was then obtained using ssGSEA for two hub DEGs (*CD3G* and *YES1*) in the T1DM and COVID-19 convalescence datasets, and two hub DEGs (*PTRF* and *EHD1*) in the T2DM and COVID-19 convalescence datasets.

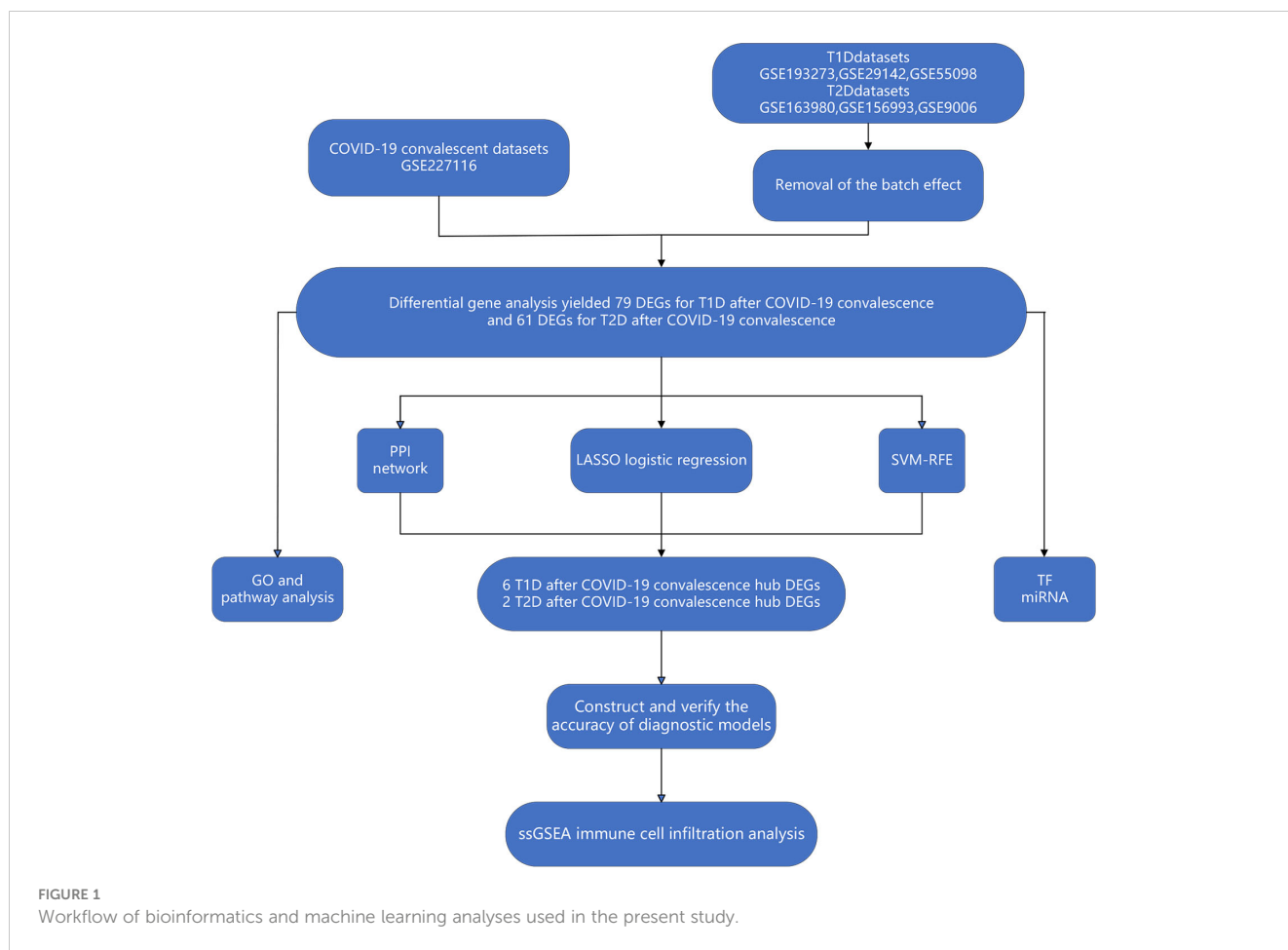
Results

Removal of batch effects

Different datasets can have statistically significant differences in the expression of the same genes in patients with the same disease due to differences in experimental reagents, operators, processing time, laboratory equipment, and other factors (30). We therefore used the ComBat function in the sva package to remove these batch effects and achieve a convergence in the distribution of expression values for the different datasets (**Figure 1**). Principal component analysis (PCA) of the sample distributions in the three T1DM datasets and the three T2DM datasets before and after the elimination of the batch effect showed that this method was successful (**Figure 2A**).

Similar DEGs in the different datasets

We analyzed 75 samples from the RNA-seq dataset (GSE227116) of controls and subjects with COVID-19 convalescence using the limma package in R software. There were 3436 DEGs, with 1724 up-regulated genes and 1712 down-regulated genes (**Figure 2B**). After removing batch effects, we used the same method for analysis of the three T1DM datasets and the three T2DM datasets. There were 81 samples in the T1DM datasets (41 cases and 40 controls), and the analysis identified 802 DEGs (**Supplementary Figure 1A**), with 432 up-regulated genes and 370 down-regulated genes (**Figure 2C**). There were 64 samples in the T2DM datasets (29 cases and 35 controls), and the analysis identified 782 DEGs (**Supplementary Figure 1B**), with 406 up-regulated genes and 376 genes downregulated genes (**Figure 2D**).



We used Venn diagrams to compare the DEGs in the COVID-19 convalescence, T1DM, and T2DM datasets (Figure 2E). T1DM and T2DM contained 38 co-up-regulated DEGs and 21 co-down-regulated DEGs (Supplementary Table 1). The COVID-19 convalescence and T1DM data had 79 of the same DEGs (32 up-regulated genes and 47 down-regulated genes). The COVID-19 convalescence and T2DM data had 61 of the same DEGs (38 up-regulated genes and 23 down-regulated genes).

KEEG enrichment analysis

We performed GO and KEGG enrichment analysis of the DEGs to identify the main biological processes and pathways in T2DM, T1DM, and COVID-19 convalescence. The upregulated genes in T1DM and COVID-19 convalescence were mainly clustered in “hemoglobin metabolic process”, “interleukin-18 production”, “iron ion homeostasis”, and “positive regulation of T cell differentiation”; the downregulated genes were mainly involved in “regulation of neuron projection development”, “steroid metabolic process”, “activation of immune response”, and “other biological processes” (Figure 3A, Supplementary Table 2). KEGG enrichment analysis demonstrated enrichment in “human T-cell leukemia virus 1 infection (HTLV-1)” and “Cholinergic synapse” pathways (Figure 4A).

The upregulated genes in T2DM and COVID-19 convalescence included “Rho protein signal transduction”, “neutral lipid metabolic

process”, and “regulation of cellular ketone metabolic process”; the downregulated genes included “regulation of nervous system development”, “response to alcohol”, and “regulation of GTPase activity” (Figure 3B, Supplementary Table 2). KEGG enrichment analysis demonstrated enrichment in “Tight junction”, “Adherens junction”, “Rap1 signaling pathway”, and “Cell adhesion molecules” (Figure 4B).

Construction of the PPI network

We then separately entered the 79 DEGs from comparison of the T1DM and COVID-19 convalescence datasets and the 61 DEGs from comparison of the T2DM and COVID-19 convalescence datasets into the STRING database to determine their relationships. The average node degree is the average value of the interaction of proteins in the network. It is used to measure the strength of the interaction relationship between proteins. For the first comparison (T1DM + COVID-19 convalescence), the PPI network contained 77 points, 21 edges, and the average node degree was 0.545 (Figure 4A). For the second comparison (T2DM + COVID-19 convalescence), the PPI network contained 61 nodes, 18 edges, and the average node degree was 0.59 (Figure 4B).

We then performed PPI network analysis using Cytoscape software with the cytoHubba plugin. Comparison of the T1DM and COVID-19 convalescence datasets indicated the 11 major DEGs were *CD3G*, *CAMK4*, *PIK3R1*, *YES1*, *CD69*, *ALAS2*, *STMN1*, *MYO1C*, *NCR3*,

TLN1, *PRKACB* (Figure 5A) sorting by degree value. Comparison of the T2DM and COVID-19 convalescence datasets indicated the 8 major DEGs were *CDH1*, *ALAS2*, *KLF4*, *ITGA6*, *DYSF*, *PTRF*, *EHD1*, and *FSTL1* (Figure 5B) sorting by degree value. These are considered to be genes with a high degree of gene-gene interaction and they have a strong association with DEGs. These results suggest that future studies that focus on these DEGs may provide new therapeutic strategies for disease prevention or treatment in these patients.

Determination of regulatory signatures

We then used the JASPAR database to analysis the relationships of TFs in the different datasets (Figures 6A, B). Cytoscape identified 58 shared TFs in a comparison of the T1DM and COVID-19 convalescence datasets, and 38 shared TFs in a comparison of the T2DM and COVID-19 convalescence datasets. Analysis of gene-miRNA relationships using the miRTarBase database (Figures 6C,

D) identified 17 shared miRNAs in a comparison of the T1DM and COVID-19 convalescence datasets, and 6 shared miRNAs in a comparison of the T2DM and COVID-19 convalescence datasets. Each of these miRNAs was associated with two or more DEGs.

Construction of a prognostic model

We also used elastic net regression to analyze the DEGs from the two comparisons. The penalty factor λ for the comparison of T1DM and COVID-19 convalescence was 0.01 ($\log(\lambda) = -4.54$) and the regression identified 30 genes (Figure 7A). The penalty factor λ for the comparison of T2DM and COVID-19 convalescence was 0.06 ($\log(\lambda) = -2.81$), and the regression identified 17 genes (Figure 7B). The input of SVM algorithm was 79 and 61 DEGs that were up-regulated or down-regulated at the same time in COVID-19 convalescence and T1/2DM, and other non-essential genes were not included. The results from the SVM algorithm (Figure 7C) showed that the highest accuracy

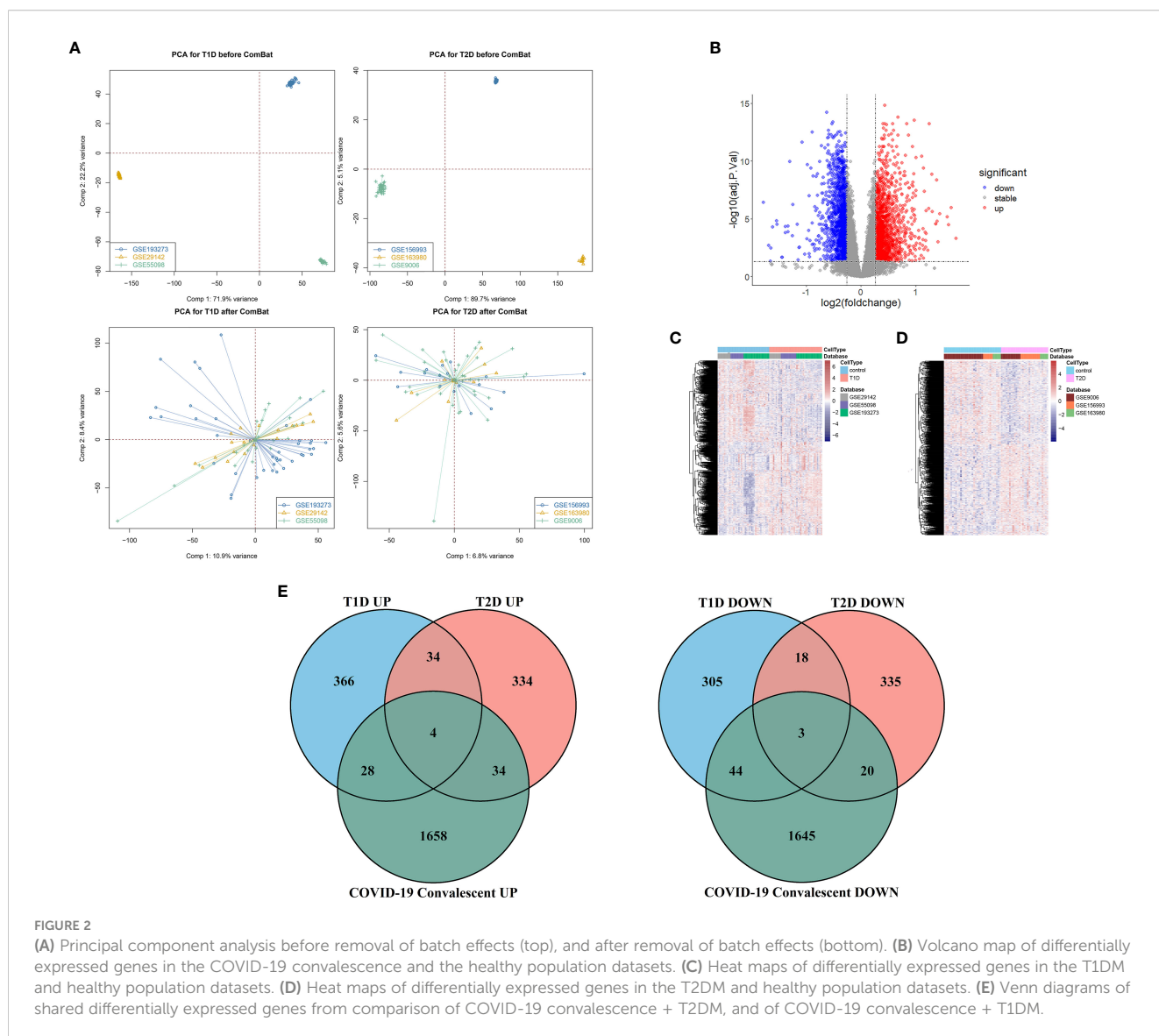


FIGURE 2 (A) Principal component analysis before removal of batch effects (top), and after removal of batch effects (bottom). (B) Volcano map of differentially expressed genes in the COVID-19 convalescence and the healthy population datasets. (C) Heat maps of differentially expressed genes in the T1DM and healthy population datasets. (D) Heat maps of differentially expressed genes in the T2DM and healthy population datasets. (E) Venn diagrams of shared differentially expressed genes from comparison of COVID-19 convalescence + T2DM, and of COVID-19 convalescence + T1DM.

(90.1%) was achieved at gene number 73 in the T1DM and COVID-19 convalescence datasets. The mean accuracy with standard deviation was $86.49 \pm 4.72\%$ (Supplementary Table 3, Supplementary Figure 2). This represents that after ranking the features and selecting the most important features for classification, when the number of genes is 73, its accuracy in high-dimensional space is the highest, which can effectively separate different categories in the datasets. The highest accuracy (85.5%) was at gene number 59 in the T2DM and COVID-19 convalescence datasets (Figure 7D). The mean accuracy with standard deviation was $80.21 \pm 4.66\%$ (Supplementary Table 3, Supplementary Figure 3). We then identified the intersection of genes screened by the two machine learning algorithms with the top-ranked genes of the MCC algorithm in the PPI network (Figure 7E).

They are considered to be the most prominent hub DEGs affecting DM and COVID-19 convalescence. The results indicated six hub DEGs for comparison of T1DM and COVID-19 convalescence (*CD3G*, *YES1*, *ALAS2*, *MYO1C*, *NCR3*, *PRKACB*) and two hub DEGs for comparison of T2DM and COVID-19 convalescence (*PTRF*, *EHD1*).

We used the rms package to construct column lineage maps of the signature genes from the two comparisons (Figures 8A, B), and used ROC curves to assess the performance of the prediction model. They are all based on combinations of training datasets. The AUC was 0.916 for prediction of T1DM based on COVID-19 convalescence data (Figure 8C), and the AUC was 0.759 for prediction of T2DM based on COVID-19 convalescence data (Figure 8D). We then used two other datasets (GSE33440 for T1DM and GSE41762 for T2DM) from

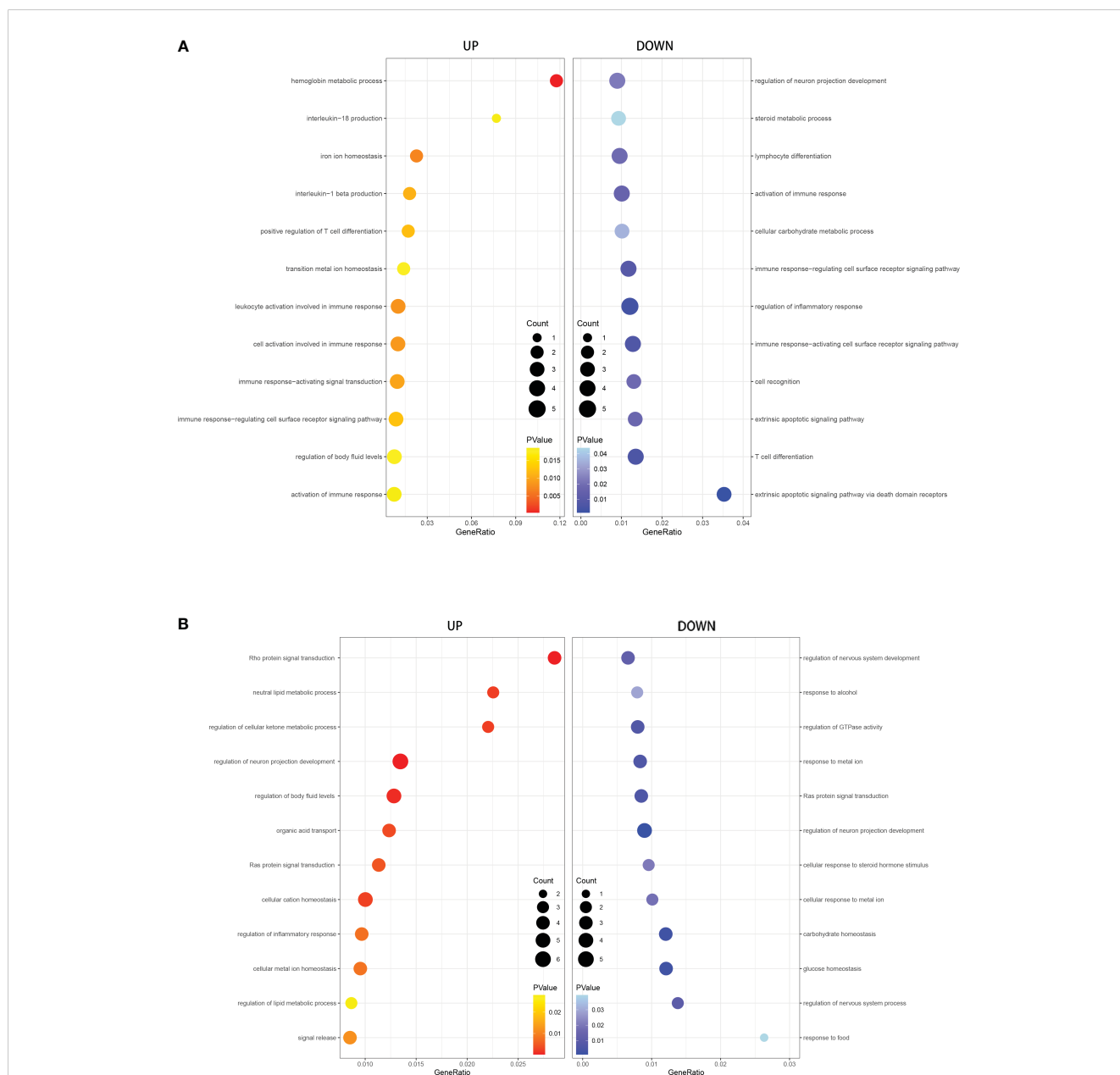


FIGURE 3 GO analysis of DEGs in T1DM and COVID-19 convalescence (A) and in T2DM and COVID-19 convalescence (B), showing genes enriched in biological processes that were upregulated (left, red) and down-regulated (right, blue).

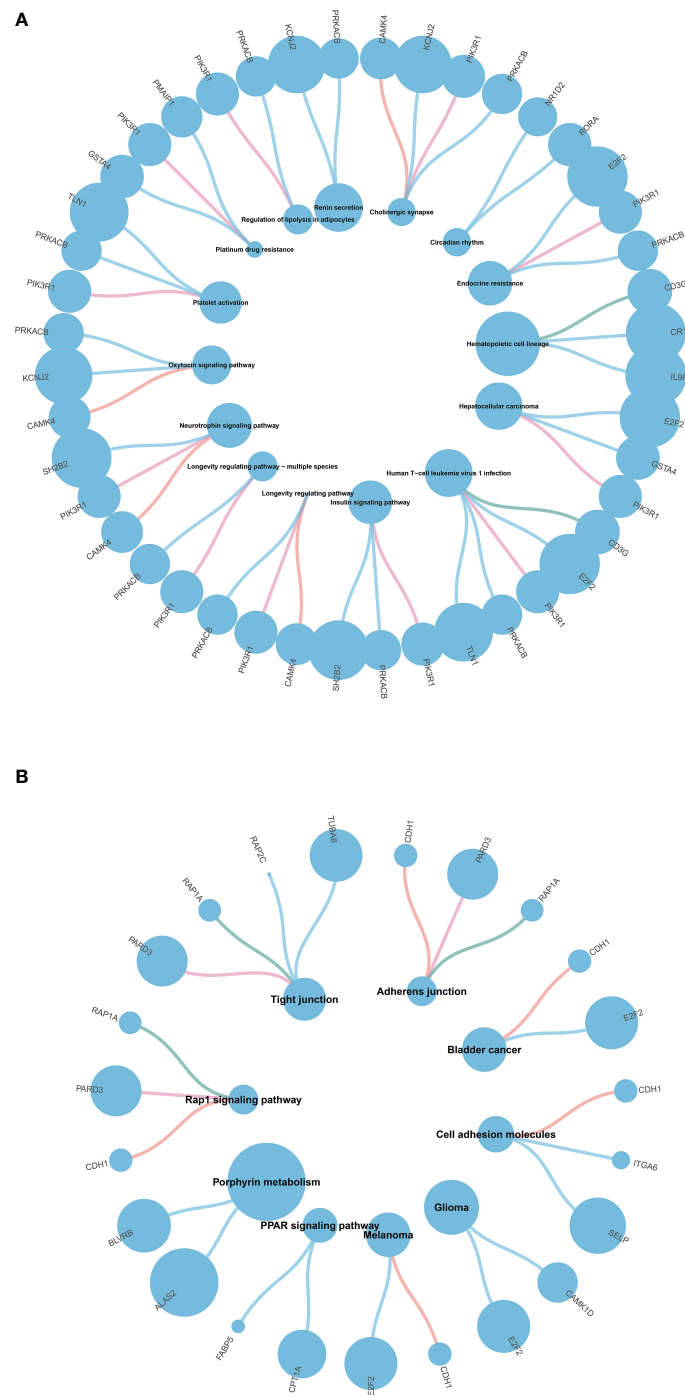


FIGURE 4 (A) KEGG pathway enrichment analysis of T1DM and COVID-19 convalescence (A) and T2DM and COVID-19 convalescence (B).

the GEO database to construct ROC curves and validate the model. The AUC value in the validation analysis was 0.854 for prediction of T1DM data (Figure 8E) and 0.734 for prediction of T2DM data (Figure 8F). Similarly, we constructed predictive models with 6 hub DEGs for T1DM and 2 hub DEGs for T2DM in the COVID-19 convalescence validation set, both with an AUC value of 1.0 (Supplementary Figure 4).

Correlations of hub DEGs and immune cell infiltration

We analyzed the relationship of immune cell infiltration with two hub DEGs (*CD3G* and *YES1*) in the T1DM and COVID-19 convalescence data, and with two other hub DEGs (*PTRF* and *EHD1*) in the T2DM and COVID-19 convalescence data. The

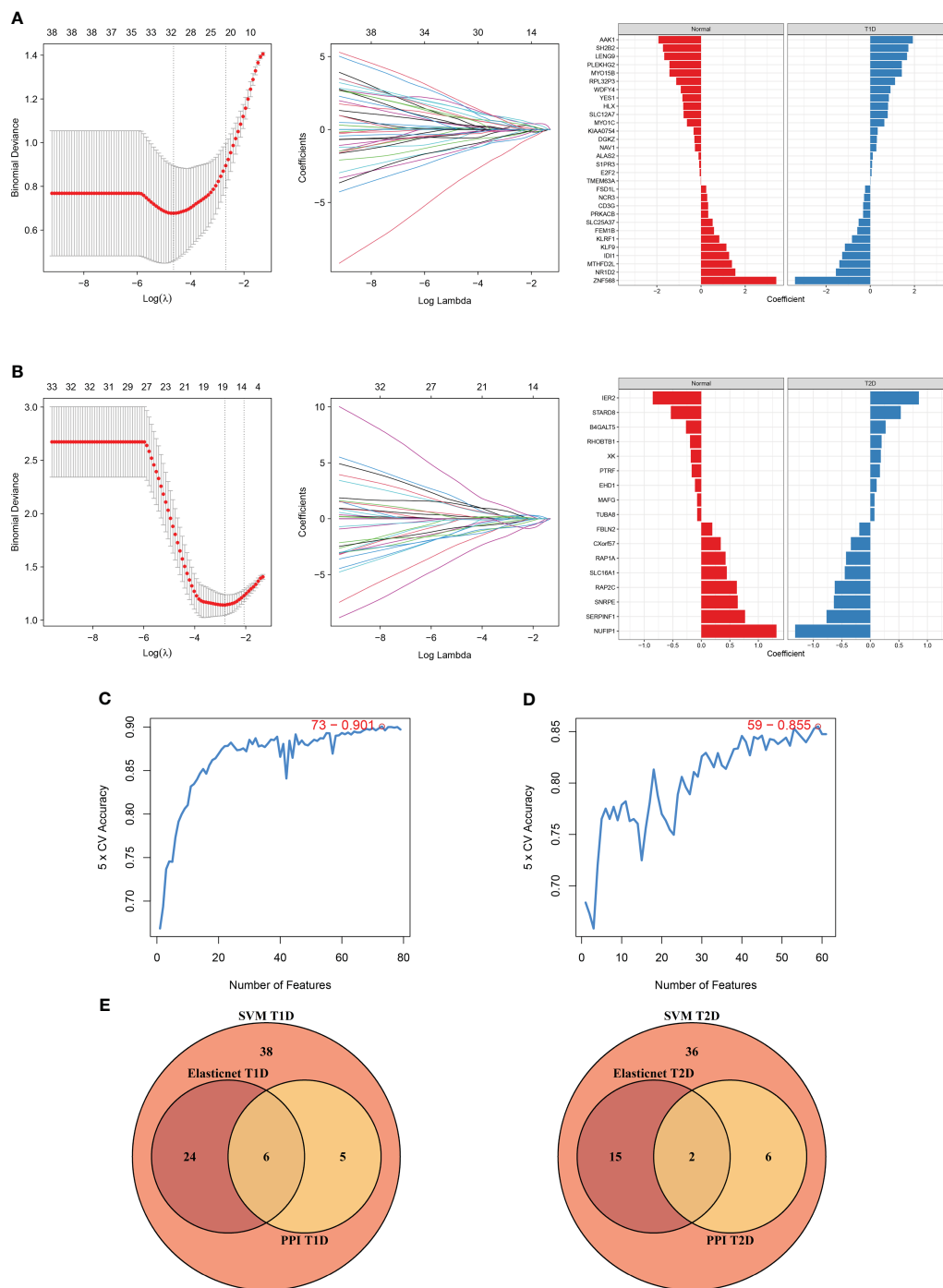


FIGURE 7 Diagram of the machine algorithm obtained by inputting overlapping DEGs. **(A)** Elastic Net Regression screening for shared diagnostic markers in T1DM and COVID-19 convalescence (left), identification of different genes by color (middle), and coefficient values of the resulting genes(right). **(B)** Elastic Net Regression screening for shared diagnostic markers in T2DM and COVID-19 convalescence (left), identification of different genes by color (right), and coefficient values of the resulting genes(right). **(C)** SVM screening of biomarkers with highest accuracy for T1DM and COVID-19 convalescence. **(D)** SVM screening of biomarkers with highest accuracy for T2DM and COVID-19 convalescence. **(E)** Venn diagrams hand over hub DEGs common to LASSO algorithm, SVM algorithm and PPI network MCC algorithm.

Down-regulated genes in T2DM and COVID-19 convalescence were enriched in the regulation of nervous system development. The arcuate nucleus (ARC) of the hypothalamus integrates insulin signals and primary sensory information about circulating nutrients (e.g., glucose) to coordinate the neuroendocrine system and

maintain glucose homeostasis (51). Thus, imbalances in the nervous system may disrupt glucose metabolism. The results of our KEGG analysis showed that alterations of tight junctions were a key pathway alteration in T2DM and COVID-19 convalescence. The tight junctions of the intestinal mucosa have an important role

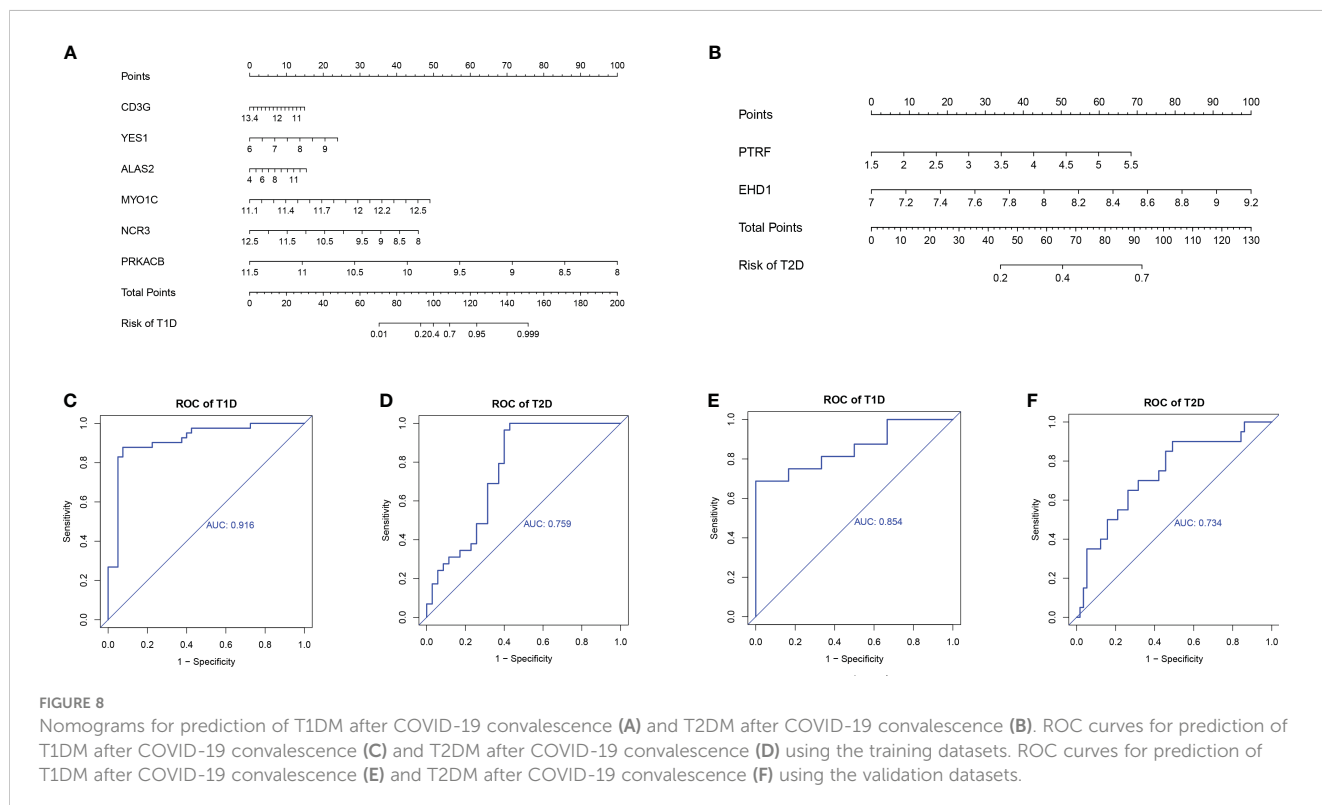
in maintaining the permeability and integrity of the intestinal mucosa. Dysfunction of the intestinal mucosal barrier is closely related to the development of diabetes, and some studies suggested that maintaining the ecological balance in the intestine may be a novel approach to overcome insulin resistance (52, 53).

The results from the PPI networks and screening by two machine learning algorithms indicated that *PTRF* and *EHD1* were the most important hub DEGs. Since the T2D validation set (GSE51762) was derived from human islets and the ratio of patients to healthy controls was somewhat different from T2DM training, this may be an important difference factor for the relatively low AUC values. However, the AUC value of the ROC curve was still greater than 0.7 and the difference with the AUC value of the training model was less than 0.1, so we believed that it still had certain accuracy. Polymerase I and transcription release factor (*PTRF*), also known as cavin-1, is associated with caveolae (“pits”) in the plasma membrane, and functions directly in the formation and secretion of cell-derived exosomes. Most studies of *PTRF* have focused on generalized lipodystrophy (GL), and mutations in this gene are highly associated with type 4 CGL. GL is a heterogeneous congenital disease (CGL) or acquired disease (AGL) characterized by loss of adipose tissue and increased insulin resistance, and an increased predisposition to metabolic complications, such as DM, hypertriglyceridemia, and hepatic steatosis (54). Mice with *PTRF* knockout have elevated triglycerides, decreased adipose tissue mass, glucose intolerance, and hyperinsulinism (55). Although there is no definitive evidence of a mechanistic relationship of *PTRF* and T2DM, we hypothesize that downregulation of *PTRF* after COVID-19 convalescence may lead to symptoms of AGL, thus increasing the risk of T2DM. *EHD1* (EH Domain Containing 1) is a protein coding

gene. Diseases associated with *EHD1* include plasmacytoid cystic tumor of the pancreas and cerebral hypoplasia, neuropathy, ichthyosis, and palmoplantar keratosis syndrome. Its related pathways include Angiopoietin-like protein 8 regulatory pathway and Response to elevated platelet cytosolic Ca²⁺. Insulin stimulates the translocation of glucose transporter 4 (GLUT4) from a perinuclear location to the plasma membrane (56). *EHD1* controls the normal perinuclear localization of GLUT4-containing membranes and facilitates retrograde transport of GLUT4 vesicles from early endosomes to recycling endosomes or perinuclear compartments (57, 58). This suggests that *EHD1* deficiency disrupts the insulin-regulated GLUT4 cycle in cultured adipocytes.

Our analysis of *EHD1* led to the identification of eleven TFs (*HOXA5*, *PPARG*, *STAT3*, *KLF5*, *NFKB1*, *RELA*, *MAX*, *USF1*, *USF2*, *SREBF1*, *NFATC2*) and two miRNAs (hsa-mir-34a-5p, hsa-mir-26b-5p). The hsa-mir-34a-5p had been found to be highly correlated with the occurrence of T2DM caused by mixed heavy metals (59). For T2DM, hsa-mir-26b-5p was significantly down-regulated after metformin treatment (60). These substances might be suggestive of subsequent studies of diabetes after COVID-19 convalescence. We believe that the reason why the number of hub genes found in T2DM is significantly less than that in T1DM is due to the high correlation between T1DM and genetics. Studies have shown that genetic defects are the basis of T1DM. T2DM is mostly perennial onset, which is related to acquired factors and may have relatively little influence on genes. Unfortunately, we found no TFs or miRNAs related to the *PTRF* gene, possibly due to the incompleteness of the databases.

We identified T-cell and CD8 T-cell infiltration in T1DM and COVID-19 convalescence, and cytotoxic T cell infiltration in T2DM



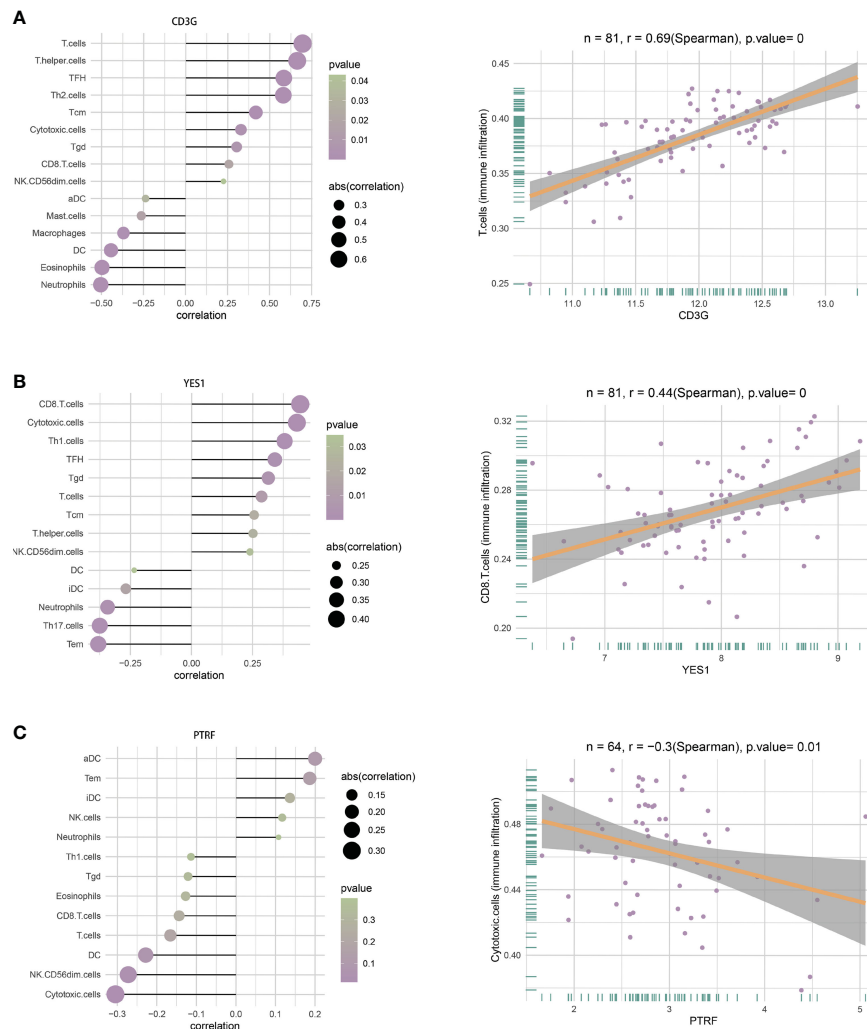


FIGURE 9 Correlations between CD3G and infiltration of different immune cells (left), and between CD3G and T cells (right). **(B)** Correlation between YES1 and infiltration of different immune cells (left), and between YES1 and CD8 T cells (right). **(C)** Correlation between PTRF and the infiltration of different immune cells (left), and between PTRF and cytotoxic T cells (right).

and COVID-19 convalescence were not sufficiently correlated. T1DM is an autoimmune disease in which T cells attack and destroy insulin-producing beta cells in the pancreatic islets. Effector T cells respond to pancreatic beta cell-derived peptides presented by HLA class I and II molecules, and this leads to death of beta cells and insulin deficiency (61, 62). Previous research showed that CD8 T-cell-mediated autoimmune diseases are caused by disruption of auto-reactive CD8 T-cell self-tolerance mechanisms, and that an increase in the number of auto-reactive CD8 T cells drives the transition from autoimmune progenitor cells to autoimmune mediators (63). A higher percentage of cytotoxic T cells can also occur in T2DM (64). Even so, the analysis of the characteristics of immune cells in T2DM showed that there was little correlation between T2DM and immune cells. It must be mentioned that there are certain limitations in our study. The number of public COVID-19 convalescence data sets is limited, and

we did not find open PBMC data sets in the GEO database. The data sets used to find differential genes and subsequent analysis are from whole blood, which is relatively inferior in the control of variables. This cannot exclude that other components of the blood would have an effect on the exploration of its pathway. On the other hand, due to the incompleteness of the TF and miRNA databases, we did not retrieve the PTRF hub DEG that was present in T2DM and COVID-19 convalescence. In addition, we did not examine the relationships of multiple risk factors affecting diabetes, the extent of glycemic control, the presence or absence of complications, and survival rate with different molecular targets.

In summary, we used bioinformatics methods with machine learning algorithms to identify specific shared hub DEGs, potential TFs, and altered pathways that occur in DM and after COVID-19 convalescence. We also constructed and validated a diagnostic model of DM after COVID-19 convalescence. Our

results provide a new point of reference for subsequent studies and also provide a basis for a new approach that could be used for prevention and management of new onset DM after COVID-19 convalescence.

Conclusions

Our study examined biomolecules and pathways that were related to the development of new-onset DM after COVID-19 convalescence by analysis of three PBMC datasets for T1DM, three PBMC datasets for T2DM, and one whole blood dataset for COVID-19 convalescence. We also used separate datasets for model validation. The results demonstrated multiple similarities of DM and COVID-19 convalescence in terms of DEGs, TFs, miRNAs, and pathways. The results from two machine learning algorithms showed that six core DEGs were shared by T1DM and COVID-19 convalescence, and that two core DEGs were shared by T2DM and COVID-19 convalescence. We therefore consider these genes as reliable indicators of DM after COVID-19 convalescence. Our finding of the importance of these several hub DEGs suggests new directions for subsequent research, and that these molecules have potential use as therapeutic targets for patients who develop new-onset DM after COVID-19 convalescence.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

JS: Data curation, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing. YW: Data curation, Validation, Visualization, Writing – original draft. XD: Supervision, Writing – review & editing. SS: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported grants from the Beijing Xinyue Foundation (2022IIT037) and The First Affiliated Hospital of Harbin Medical University Foundation (2023M17) that were awarded to SS.

Acknowledgments

We would like to acknowledge the GEO (GSE227116, GSE193273, GSE29142, GSE55098, GSE156993, GSE163980, GSE9006, GSE33440, GSE41762 and GSE166253) network for providing data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2023.1306325/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

(A) Volcano map of differentially expressed genes in the T1DM and the healthy population datasets. (B) Volcano map of differentially expressed genes in the T2DM and the healthy population datasets.

SUPPLEMENTARY FIGURE 2

(A) curve of T1DM fold1 cross-validation (B) curve of T1DM fold2 cross-validation (C) curve of T1DM fold3 cross-validation (D) curve of T1DM fold4 cross-validation (E) curve of T1DM fold5 cross-validation.

SUPPLEMENTARY FIGURE 3

(A) curve of T2DM fold1 cross-validation (B) curve of T2DM fold2 cross-validation (C) curve of T2DM fold3 cross-validation (D) curve of T2DM fold4 cross-validation (E) curve of T2DM fold5 cross-validation.

SUPPLEMENTARY FIGURE 4

Nomograms for prediction of T1DM after COVID-19 convalescence (A) and T2DM after COVID-19 convalescence (B) on COVID-19 convalescence validation dataset. (C) ROC curves of 6 hub DEGs in the COVID-19 convalescence validation dataset. (D) ROC curves of 2 hub DEGs in the COVID-19 convalescence validation dataset.

SUPPLEMENTARY FIGURE 5

Correlation between EHD1 and the infiltration of different immune cells (left), and between EHD1 and immune infiltration (right).

References

- Rai P, Kumar BK, Deekshit VK, Karunasagar I, Karunasagar I. Detection technologies and recent developments in the diagnosis of covid-19 infection. *Appl Microbiol Biotechnol* (2021) 105(2):441–55. doi: 10.1007/s00253-020-11061-5
- Long MJC, Aye Y. Science's response to covid-19. *ChemMedChem* (2021) 16(15):2288–314. doi: 10.1002/cmdc.202100079
- Zhang L, Li Q, Liang Z, Li T, Liu S, Cui Q, et al. The significant immune escape of pseudotyped sars-cov-2 variant omicron. *Emerg Microbes Infect* (2022) 11(1):1–5. doi: 10.1080/22221751.2021.2017757
- Tuekprakhon A, Notalai R, Dijokaite-Guraliuc A, Zhou D, Ginn HM, Selvaraj M, et al. Antibody escape of sars-cov-2 omicron ba.4 and ba.5 from vaccine and ba.1 serum. *Cell* (2022) 185(14):2422–33 e13. doi: 10.1016/j.cell.2022.06.005
- Carfi A, Bernabei R, Landi F, Gemelli Against C-P-ACSG. Persistent symptoms in patients after acute covid-19. *JAMA* (2020) 324(6):603–5. doi: 10.1001/jama.2020.12603
- Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* (2010) 87(1):4–14. doi: 10.1016/j.diabres.2009.10.007
- Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a who consultation. *Diabetes Med* (1998) 15(7):539–53. doi: 10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S
- Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol* (2020) 16(7):377–90. doi: 10.1038/s41581-020-0278-5
- Singh AK, Khunti K. Covid-19 and diabetes. *Annu Rev Med* (2022) 73(1545-326X (Electronic)):129–47. doi: 10.1146/annurev-med-042220-011857
- Li R, Shen M, Yang Q, Fairley CK, Chai Z, McIntyre R, et al. Global diabetes prevalence in covid-19 patients and contribution to covid-19-related severity and mortality: A systematic review and meta-analysis. *Diabetes Care* (2023) 46(4):890–7. doi: 10.2337/dc22-1943
- The Lancet Diabetes E. Covid-19 and diabetes: A co-conspiracy? *Lancet Diabetes Endocrinol* (2020) 8(10):801. doi: 10.1016/S2213-8587(20)30315-6
- Lim S, Bae JH, Kwon HS, Nauck MA. Covid-19 and diabetes mellitus: from pathophysiology to clinical management. *Nat Rev Endocrinol* (2021) 17(1):11–30. doi: 10.1038/s41574-020-00435-4
- Memon B, Abdelalim EM. Ace2 function in the pancreatic islet: implications for relationship between sars-cov-2 and diabetes. *Acta Physiol (Oxf)* (2021) 233(4):e13733. doi: 10.1111/apha.13733
- Chen H, Peng L, Wang Z, He Y, Tang S, Zhang X. Exploration of cross-talk and pyroptosis-related gene signatures and molecular mechanisms between periodontitis and diabetes mellitus via peripheral blood mononuclear cell microarray data analysis. *Cytokine* (2022) 159:156014. doi: 10.1016/j.cyto.2022.156014
- Li X, Liao M, Guan J, Zhou L, Shen R, Long M, et al. Identification of key genes and pathways in peripheral blood mononuclear cells of type 1 diabetes mellitus by integrated bioinformatics analysis. *Diabetes Metab J* (2022) 46(3):451–63. doi: 10.4093/dmj.2021.0018
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
- Ritchie ME, Phipson B, Wu D, Hu YF, Law CW, Shi W, et al. Powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A.* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102
- Yu GC, Wang LG, Han YY, He QY. Clusterprofiler: an R package for comparing biological themes among gene clusters. *Omic* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* (2021) 49(D1):D605–D12. doi: 10.1093/nar/gkaa1074
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303
- Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. Jasp 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* (2018) 46(D1):D260–D6. doi: 10.1093/nar/gkx1126
- Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. Mirtarbase 2020: updates to the experimentally validated microrna-target interaction database. *Nucleic Acids Res* (2020) 48(D1):D148–D54. doi: 10.1093/nar/gkz896
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
- Engelbrecht S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenet* (2019) 11(1):123. doi: 10.1186/s13148-019-0730-1
- Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (Svm) learning in cancer genomics. *Cancer Genomics Proteomics* (2018) 15(1):41–51. doi: 10.21873/cgp.20063
- Qiu JL, Peng BG, Tang YQ, Qian YB, Guo P, Li MF, et al. CpG methylation signature predicts recurrence in early-stage hepatocellular carcinoma: results from a multicenter study. *J Clin Oncol* (2017) 35(7):734–+. doi: 10.1200/Jco.2016.68.2153
- Finotello F, Trajanoski Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol Immunother* (2018) 67(7):1031–40. doi: 10.1007/s00262-018-2150-z
- Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenaus AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* (2013) 39(4):782–95. doi: 10.1016/j.immuni.2013.10.003
- Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* (2010) 11(10):733–9. doi: 10.1038/nrg2825
- Sane AH, Mekonnen MS, Tsegaw MG, Zewde WC, Mesfin EG, Beyene HA, et al. New onset of diabetes mellitus and associated factors among covid-19 patients in covid-19 care centers, addis ababa, Ethiopia 2022. *J Diabetes Res* (2022) 2022(2314-6753 (Electronic)):9652940. doi: 10.1155/2022/9652940
- Toldi G, Vasarhelyi B, Kaposi A, Meszaros G, Panczel P, Hosszufalusi N, et al. Lymphocyte activation in type 1 diabetes mellitus: the increased significance of kv1.3 potassium channels. *Immunol Lett* (2010) 133(1):35–41. doi: 10.1016/j.jimlet.2010.06.009
- Talukder MR, Woodman R, Pham H, Wilson K, Gessain A, Kaldor J, et al. High human T-cell leukemia virus type 1c proviral loads are associated with diabetes and chronic kidney disease: results of a cross-sectional community survey in central Australia. *Clin Infect Dis* (2023) 76(3):e820–e6. doi: 10.1093/cid/ciac614
- Love JA, Yi E, Smith TG. Autonomic pathways regulating pancreatic exocrine secretion. *Auton Neurosci* (2007) 133(1):19–34. doi: 10.1016/j.autneu.2006.10.001
- Rowe JH, Delmonte OM, Keles S, Stadinski BD, Dobbs AK, Henderson LA, et al. Patients with cd3g mutations reveal a role for human cd3gamma in T(Reg) diversity and suppressive function. *Blood* (2018) 131(21):2335–44. doi: 10.1182/blood-2018-02-835561
- Wang J, Gu X, Cao L, Ouyang Y, Qi X, Wang Z, et al. A novel prognostic biomarker cd3g that correlates with the tumor microenvironment in cervical cancer. *Front Oncol* (2022) 12(2234-943X:979226(2234-943X (Print))). doi: 10.3389/fonc.2022.979226
- Chen Z, Wang M, De Wilde RL, Feng R, Su M, Torres-de la Roche LA, et al. A machine learning model to predict the triple negative breast cancer immune subtype. *Front Immunol* (2021) 12(1664-3224:749459(1664-3224 (Electronic))). doi: 10.3389/fimmu.2021.749459
- Li N, Li L, Wu M, Li Y, Yang J, Wu Y, et al. Integrated bioinformatics and validation reveal potential biomarkers associated with progression of primary sjogren's syndrome. *Front Immunol* (2021) 12(1664-3224:697157(1664-3224 (Electronic))). doi: 10.3389/fimmu.2021.697157
- Gokturk B, Keles S, Kirac M, Artac H, Tokgoz H, Guner SN, et al. Cd3g gene defects in familial autoimmune thyroiditis. *Scand J Immunol* (2014) 80(5):354–61. doi: 10.1111/sji.12200
- Garmendia I, Pajares MJ, Hermida-Prado F, Ajona D, Bertolo C, Sainz C, et al. Yes1 drives lung cancer growth and progression and predicts sensitivity to dasatinib. *Am J Respir Crit Care Med* (2019) 200(7):888–99. doi: 10.1164/rccm.201807-1292OC
- Mao L, Yuan W, Cai K, Lai C, Huang C, Xu Y, et al. Epha2-yes1-ANXA2 pathway promotes gastric cancer progression and metastasis. *Oncogene* (2021) 40(20):3610–23. doi: 10.1038/s41388-021-01786-6
- Zhou Y, Wang C, Ding J, Chen Y, Sun Y, Cheng Z. Mir-133a targets yes1 to reduce cisplatin resistance in ovarian cancer by regulating cell autophagy. *Cancer Cell Int* (2022) 22(1):15. doi: 10.1186/s12935-021-02412-x
- Fujihara M, Shien T, Shien K, Suzawa K, Takeda T, Zhu Y, et al. Yes1 as a therapeutic target for her2-positive breast cancer after trastuzumab and trastuzumab-entansine (T-dm1) resistance development. *Int J Mol Sci* (2021) 22(23):12809. doi: 10.3390/ijms222312809
- Redin E, Garrido-Martin EM, Valencia K, Redrado M, Solorzano JL, Carias R, et al. Yes1 is a druggable oncogenic target in sclc. *J Thorac Oncol* (2022) 17(12):1387–403. doi: 10.1016/j.jtho.2022.08.002
- Yoder SM, Dineen SL, Wang Z, Thurmond DC. Yes, a src family kinase, is a proximal glucose-specific activator of cell division cycle control protein 42 (Cdc42) in pancreatic islet beta cells. *J Biol Chem* (2014) 289(16):11476–87. doi: 10.1074/jbc.M114.559328
- Cao S, Zhang S. Forkhead-box C1 attenuates high glucose-induced trophoblast cell injury during gestational diabetes mellitus via activating adenosine monophosphate-activated protein kinase through regulating fibroblast growth factor 19. *Bioengineered* (2022) 13(1):1174–84. doi: 10.1080/21655979.2021.2018094

47. Sun GP, Kohno M, Guo P, Nagai Y, Miyata K, Fan YY, et al. Involvements of rho-kinase and tgf-beta pathways in aldosterone-induced renal injury. *J Am Soc Nephrol* (2006) 17(8):2193–201. doi: 10.1681/ASN.2005121375
48. Furukawa N, Ongusaha P, Jahng WJ, Araki K, Choi CS, Kim HJ, et al. Role of rho-kinase in regulation of insulin action and glucose homeostasis. *Cell Metab* (2005) 2(2):119–29. doi: 10.1016/j.cmet.2005.06.011
49. Lee DH, Shi J, Jeoung NH, Kim MS, Zabolotny JM, Lee SW, et al. Targeted disruption of rock1 causes insulin resistance in vivo. *J Biol Chem* (2009) 284(18):11776–80. doi: 10.1074/jbc.C900014200
50. Zhou H, Li YJ. Rho kinase inhibitors: potential treatments for diabetes and diabetic complications. *Curr Pharm Des* (2012) 18(20):2964–73. doi: 10.2174/138161212800672688
51. Jais A, Bruning JC. Arcuate nucleus-dependent regulation of metabolism-pathways to obesity and diabetes mellitus. *Endocr Rev* (2022) 43(2):314–28. doi: 10.1210/edrv/bnab025
52. Qin X, Dong H, Lu FE. Research progress of relationship between diabetes and intestinal epithelial tight junction barrier and intervention of berberine. *Zhongguo Zhong Yao Za Zhi* (2016) 41(11):1973–7. doi: 10.4268/cjcm20161101
53. Sato J, Kanazawa A, Watada H. Type 2 diabetes and bacteremia. *Ann Nutr Metab* (2017) 71 Suppl 1(1421-9697(1421-9697 (Electronic):17–22. doi: 10.1159/000479919
54. Patni N, Garg A. Congenital generalized lipodystrophies—new insights into metabolic dysfunction. *Nat Rev Endocrinol* (2015) 11(9):522–34. doi: 10.1038/nrendo.2015.123
55. Hayashi YK, Matsuda C, Ogawa M, Goto K, Tominaga K, Mitsuhashi S, et al. Human ptrf mutations cause secondary deficiency of caveolins resulting in muscular dystrophy with generalized lipodystrophy. *J Clin Invest* (2009) 119(9):2623–33. doi: 10.1172/JCI38660
56. Guilherme A, Soriano NA, Furciniti PS, Czech MP. Role of ehd1 and ehbp1 in perinuclear sorting and insulin-regulated glut4 recycling in 3t3-L1 adipocytes. *J Biol Chem* (2004) 279(38):40062–75. doi: 10.1074/jbc.M401918200
57. Matsui K, Emoto M, Fukuda N, Nomiya R, Yamada K, Tanizawa Y. Snare-binding protein synaptosomal-associated protein of 29 Kda (Snap29) regulates the intracellular sequestration of glucose transporter 4 (Glut4) vesicles in adipocytes. *J Diabetes Invest* (2022) 14(1):19–27. doi: 10.1111/jdi.13912
58. Ishiki M, Klip A. Minireview: recent developments in the regulation of glucose transporter-4 traffic: new signals, locations, and partners. *Endocrinology* (2005) 146(12):5071–8. doi: 10.1210/en.2005-0850
59. Nguyen HD. An evaluation of the effects of mixed heavy metals on prediabetes and type 2 diabetes: epidemiological and toxicogenomic analysis. *Environ Sci Pollut R* (2023) 30(34):82437–57. doi: 10.1007/s11356-023-28037-3
60. Demirsoy İH, Ertural DY, Balci Ş, Çinkır Ü, Sezer K, Tamer L, et al. Profiles of circulating mirnas following metformin treatment in patients with type 2 diabetes. *J Med Biochem* (2018) 37(4):499–506. doi: 10.2478/jomb-2018-0009
61. Bluestone JA, Buckner JH, Herold KC. Immunotherapy: building a bridge to a cure for type 1 diabetes. *Science* (2021) 373(6554):510–6. doi: 10.1126/science.abh1654
62. James EA, Mallone R, Kent SC, DiLorenzo TP. T-cell epitopes and neo-epitopes in type 1 diabetes: A comprehensive update and reappraisal. *Diabetes* (2020) 69(7):1311–35. doi: 10.2337/dbi19-0022
63. Gearthy SV, Dundar F, Zumbo P, Espinosa-Carrasco G, Shakiba M, Sanchez-Rivera FJ, et al. An autoimmune stem-like cd8 T cell population drives type 1 diabetes. *Nature* (2022) 602(7895):156–61. doi: 10.1038/s41586-021-04248-x
64. Menart-Houtermans B, Rutter R, Nowotny B, Rosenbauer J, Koliaki C, Kahl S, et al. Leukocyte profiles differ between type 1 and type 2 diabetes and are associated with metabolic phenotypes: results from the German diabetes study (Gds). *Diabetes Care* (2014) 37(8):2326–33. doi: 10.2337/dc14-0316