

## Research Article

# Comprehensive Network Analysis Reveals the Targets and Potential Multitarget Drugs of Type 2 Diabetes Mellitus

Wan Zhou <sup>1</sup>, Qiang Liu <sup>2</sup>, Wei Wang,<sup>1</sup> Xiao-Jing Yuan,<sup>3</sup> Chun-Chun Xiao,<sup>1</sup> and Shan-Dong Ye <sup>1</sup>

<sup>1</sup>Laboratory for Diabetes, Department of Endocrinology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230001, China

<sup>2</sup>Institute on Aging and Brain Disorders, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230026, China

<sup>3</sup>The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China

Correspondence should be addressed to Shan-Dong Ye; [yeshandong123@163.com](mailto:yeshandong123@163.com)

Received 1 April 2022; Revised 29 May 2022; Accepted 6 July 2022; Published 28 July 2022

Academic Editor: Swayam Srivastava

Copyright © 2022 Wan Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Type 2 diabetes mellitus (T2DM) is a metabolic disease with increasing prevalence and mortality year by year. The purpose of this study was to explore new therapeutic targets and candidate drugs for multitargets by single-cell RNA expression profile analysis, network pharmacology, and molecular docking. Single-cell RNA expression profiling of islet  $\beta$  cell samples between T2DM patients and nondiabetic controls was conducted to identify important subpopulations and the marker genes. The potential therapeutic targets of T2DM were identified by the overlap analysis of insulin-related genes and diabetes-related genes, the construction of protein-protein interaction network, and the molecular complex detection (MCODE) algorithm. The network distance method was employed to determine the potential drugs of the target. Molecular docking and molecular dynamic simulations were carried out using AutoDock Vina and Gromacs2019, respectively. Eleven cell clusters were identified by single-cell RNA sequencing (scRNA-seq) data, and three of them (C2, C8, and C10) showed significant differences between T2DM samples and normal samples. Eight genes from differential cell clusters were found from differential cell clusters to be associated with insulin activity and T2DM. The MCODE algorithm built six key subnetworks, with five of them correlating with inflammatory pathways and immune cell infiltration. Importantly, CCR5 was a gene within the key subnetworks and was differentially expressed between normal samples and T2DM samples, with the highest area under the ROC curve (AUC) of 82.5% for the diagnosis model. A total of 49 CCR5-related genes were screened, and DB05494 was identified as the most potential drug with the shortest distance to CCR5-related genes. Molecular docking illustrated that DB05494 stably bound with CCR5 (-8.0 kcal/mol) through multiple hydrogen bonds (LYS26, TYR37, TYR89, CYS178, and GLN280) and hydrophobic bonds (TRP86, PHE112, ILE198, TRP248, and TYR251). This study identified CCR5 as a potential therapeutic target and screened DB05494 as a potential drug for T2DM treatment.

## 1. Introduction

Diabetes is a metabolic syndrome that affects more than 536 million people worldwide. The prevalence of diabetes is consistently growing and is expected to rise to 780 million by 2045 [1]. As a result, the high incidence and mortality dra-

matically increased, placing a huge burden on the health care system [2]. The main categories of diabetes include type 1 diabetes (T1DM), type 2 diabetes mellitus (T2DM), gestational diabetes, and other less common types, such as monogenic diabetes syndrome, pancreatic exocrine diseases, and diabetes caused by drugs or chemicals [3]. Among them,

T2DM is the most common type of diabetes, accounting for 90% of all diabetic patients [4]. T2DM is characterized by hyperglycemia and is caused by insulin resistance, reduced or insufficient insulin secretion, and/or improper secretion of glucagon hormones [5]. Diabetes that is not effectively controlled for a long time usually causes various serious complications such as acute exacerbation of lethal diabetes pyruvate and hyperosmolar coma, in addition to severe chronic and organic lesions, such as cardiovascular disease, diabetic neuropathy, chronic kidney disease, and retinopathy. Metformin is the first-line drug for the treatment of T2DM by inhibiting liver gluconeogenesis and increasing peripheral insulin sensitivity. Additionally, sulfonylureas, thiazolidinediones, glucose-like peptide-1 (GLP-1) analogs, and dipeptidyl peptidase-4 (DPP4) inhibitors are widely used in clinical therapy [6]. However, these hypoglycemic drugs have some limitations such as long treatment time, large side effects, and high price, and there is a great demand for effective, low-toxic or nontoxic, and reasonably priced drugs for treating diabetes. Therefore, early screening and the development of efficient, safe, and sustainable drugs are still the main objectives of T2DM management [7]. For this purpose, a precise understanding of the molecular mechanism of T2DM and the key genes that play critical roles in T2DM development is a basis for further exploiting potential therapeutic drugs for T2DM.

Recent technologies and methodologies such as single-cell RNA sequencing (scRNA-seq) and molecular docking create the possibility for efficiently exploring potential drug targets. The dominant paradigm in drug discovery is the concept of designing maximally selective ligands to act on individual drug targets [8]. The emergence of functional genomics, proteomics, chemical informatics, and other system-based scientific methods has raised expectations for drug discovery and development of new targets [9]. In recent years, web-based drug discovery has been considered a promising method for cost-effective drug development [10]. Network pharmacology analyses consider drug responses in the context of cellular or phenotypic networks [11]. Some studies have developed drugs with certain therapeutic effects on specific diseases through the application of network pharmacology, but these studies are not comprehensive enough as there was a lack of taking the influence of T2DM heterogeneity into consideration [12].

Molecular docking is a widely used technique for large-scale virtual screening of the interactions between small-molecule ligands and their target proteins [13]. Docking can identify new compounds with therapeutic significance, predict ligand-target interactions at the molecular level, or depict the structure-activity relationship, without knowing the chemical structure of other target regulators in advance [14]. Recently, *in silico* approaches which compute the dynamic motion of atoms and molecules with respect to time were widely used and had made some progress in the screening of the target of finding drugs for diseases and the mechanism exploration for medicine [15]; it relies on defining the interface of complexes based on the surface geometry complementarity and amino acid pairwise affinities of the

3D structure of the unbound molecules without prior inclusion of any experimental information [16].

In this study, we explored potential targets and therapeutic agents for T2DM based on scRNA-seq data and transcriptome data of islet  $\beta$  cells of T2DM samples and normal samples, the analysis of integrated network pharmacology, and molecular dynamic simulation of drug-protein interactions.

## 2. Materials and Methods

### 2.1. The Source of T2DM Samples and Insulin-Related Genes.

The GSE182923 dataset with expression profiles of T2DM samples and the GSE137766 dataset with scRNA-seq data were downloaded from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) database. The GSE182923 dataset contained ten normal samples and eight T2DM samples. The GSE137766 dataset contained four islet  $\beta$  cell samples, where two of them had T2DM features (named as T2DM samples in the following) and the other two of them were nondiabetic samples (named as normal samples in the following). DM-related genes were obtained from DisGeNET (<http://www.disgenet.org/>) and Comparative Toxicogenomics Database (<http://ctdbase.org/>). The intersected DM-related genes from the two databases were used as a gene list of DM-related genes in the current study. Insulin-related genes were obtained from the ChEMBL (<http://www.ebi.ac.uk/chembl/>) database.

### 2.2. Quality Control of scRNA-seq Data.

Seurat R package was used to exclude low-quality cells [17]. First, the percentage of mitochondria and rRNA was calculated using the PercentageFeatureSet function, and high-quality cells were screened according to the following quality control criteria: for each cell, the total gene counts should be between 100 and 7500, the mitochondrial gene content was less than 35%, and the number of unique molecular identifiers (UMIs) was more than 1000 (Supplemental Figure S1A-C).

### 2.3. Dimensionality Reduction and Cell Annotation for scRNA-seq Data.

The scRNA-seq data was normalized by the log-normalization method in Seurat R package. Then, the FindVariableFeatures function was performed to filter highly variable genes through the analysis of variance (Supplemental Figure S1D). The data was scaled to mean 0 and variance 1, and principal component analysis (PCA) was used for dimension reduction and data visualization. Then, the cells were clustered to different subgroups (clusters) by FindNeighbors and FindClusters functions under conditions of  $\text{dim} = 30$  and  $\text{resolution} = 0.1$ . *T*-distributed stochastic neighbor embedding (*t*-SNE) was utilized to visualize the clustered cells. For screening the marker genes of cell clusters, the FindAllMarkers function was performed to identify the top 5 marker genes of each cluster with  $\log_2(\text{fold change (FC)}) = 0.5$ ,  $\text{Minpct} = 0.35$ , and  $P < 0.05$ .

### 2.4. Enrichment Analysis of Functional Pathway.

ClusterProfiler R package [18] was utilized to annotate Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. ClusterProfiler allows automation of the process of

biological-term classification and the enrichment analysis of gene clusters, which supports a universal interface for gene functional annotation from a variety of sources and therefore can be applied in diverse scenarios.

**2.5. Screening of Important Cell Clusters and Key Genes.** The difference between the nondiabetic control group and the T2DM group was analyzed using the Fisher test with the parameter of  $FC > 4$  or  $FC < 0.25$  and  $P < 0.05$ . The intersection of DM-related and insulin activity-related genes was shown by the Venn diagram. The STRING (<https://string-db.org/>) database was introduced to conduct protein-protein interaction (PPI) analysis within genes under a confidence score  $> 0.4$ . The visualization of the PPI network among the genes was outputted by using Cytoscape software [19]. The genes in PPI with few nodes were removed.

**2.6. Identification of Hub Gene in Subnetwork.** The molecular complex detection (MCODE) algorithm, which can detect closely connected proteins or dense regions of the protein interaction network, was employed to screen important subnetworks contributing to insulin activity and the development of T2DM. KEGG enrichment analysis was performed to identify key subnetworks [20]. Next, Pearson correlation analysis was used to evaluate the relation between the genes within key subnetworks, and immune cell infiltration was calculated by CIBERSORT [21], and hub genes significantly associated with immune infiltration were identified. The receiver operating characteristic (ROC) curve analysis was implemented using pROC package to estimate the performance of the hub genes [22]. The Wilcoxon test was conducted to test the expression of the hub genes between T2DM samples and normal samples. The hub gene with a significant difference between the two groups was considered the final hub gene.

**2.7. Prediction of CCR5-Related Gene Set and Potential Targeted Drugs.** To screen CCR5-related genes in the GSE182923 dataset, `rcorr` function of `Hmisc` package (<https://cran.r-project.org/web/packages/Hmisc/index.html>) was used. The genes with a correlation coefficient  $> 0.95$  and  $P < 0.001$  were considered potential target genes for the treatment of T2DM. Based on the drug target pairs obtained from DrugBank (<https://go.drugbank.com/>) and the key PPI network of STRING (threshold score was 400), the proximity between drugs and T2DM was calculated according to the formula in the published study as follows [23]:

$$d(S, T) = \frac{1}{|T|} \sum_{t \in T} \min_{s \in S} (d(s, t) + \omega). \quad (1)$$

$S$  represents a gene set related to T2DM.  $T$  is a gene set of drug targets.  $d(s, t)$  is the shortest distance between  $s$  nodes and  $t$  nodes, where  $s$  is a T2DM-related gene and  $t$  is a drug target gene.  $\omega$  indicates the weight of a target gene.  $\omega = -\ln(D + 1)$  when the target drug gene is within the T2DM-related gene set; otherwise,  $\omega = 0$ .

The significance of relatedness between a drug and hub gene was evaluated using a reference distance distribution

corresponding to the drug. Specifically, a group of proteins ( $R$ ) matching the number of drug targets were randomly selected from the network, and the distance  $d(S, R)$  between these proteins and genes related to hub genes was calculated. The reference distribution was obtained by repeating the randomization 1000 times. The mean  $\mu_{d(S,R)}$  and standard deviation  $\sigma_{d(S,R)}$  of the reference distribution were used to calculate the  $z$ -score by converting the observed distance to a normalized distance, i.e., proximity value:

$$z(S, T) = \frac{d(S, T) - \mu_{d(S,R)}}{\sigma_{d(S,R)}}. \quad (2)$$

**2.8. Molecular Docking.** CCR5 (PDB ID: 4MBS) was identified as a target to perform molecular docking using AutoDock Vina software using the semiflexible docking algorithm [24]. AutoDockTools 1.5.6 was used to remove excessive protein chains, water, and solvent molecules and add polar hydrogen and charge, etc. The coordinates of the grid box in each direction of  $XYZ$  during molecular docking were 150 Å, 108 Å, and 22.5 Å, respectively. The length in each direction of  $XYZ$  was 20 Å. The Lamarckian algorithm [25] was used to recognize the optimal binding mode of ligand molecules with the standards of exhaustiveness = 8, the output conformation number  $\leq 10$ , and the energy range  $\leq 3$  kcal/mol [26, 27]. The outputted plot was processed by Pymol.

**2.9. Molecular Dynamic Simulations.** The binding stability of the receptor-ligand complex was evaluated by 100 ns molecular dynamic simulations through Gromacs2019 software package [28]. The CHARMM36 force field was used, and the `str` file of ligands was obtained from the CGenFF program. The system was built in a dodecahedral box, with 0.154 M sodium and chloride ions [29] being added to neutralize the charge. The steepest descent algorithm (no more than 5000 steps) was used to minimize the energy of the solvation system. The LINCS algorithm was applied to constrain covalent bond length, and the electrostatic interaction was calculated by the PME algorithm. Then, 100 ps NVT and NPT simulations were carried out under constant temperature (300 K) and pressure (1 bar) to balance atoms of compounds in an initial coordinate. Finally, 100 ns molecular dynamic simulations were run through Prudent MD with a time step of 2 fs. The root mean square deviation (RMSD) of the ligands was computed at different times (ns), and the outputted plot was processed by Pymol [30].

**2.10. Statistical Analysis.** R software (v4.0.1) was used to perform all the statistical analyses and R packages. The parameters were defaulted if not specifically indicated. The Wilcoxon test was conducted to test the difference between the two groups.  $P < 0.05$  was considered significant.

### 3. Results

**3.1. Cell Clusters Were Identified in the T2DM scRNA-seq Data.** Through preprocessing scRNA-seq data of four islet  $\beta$  cell samples in the GSE137766 dataset, 1205 cells and

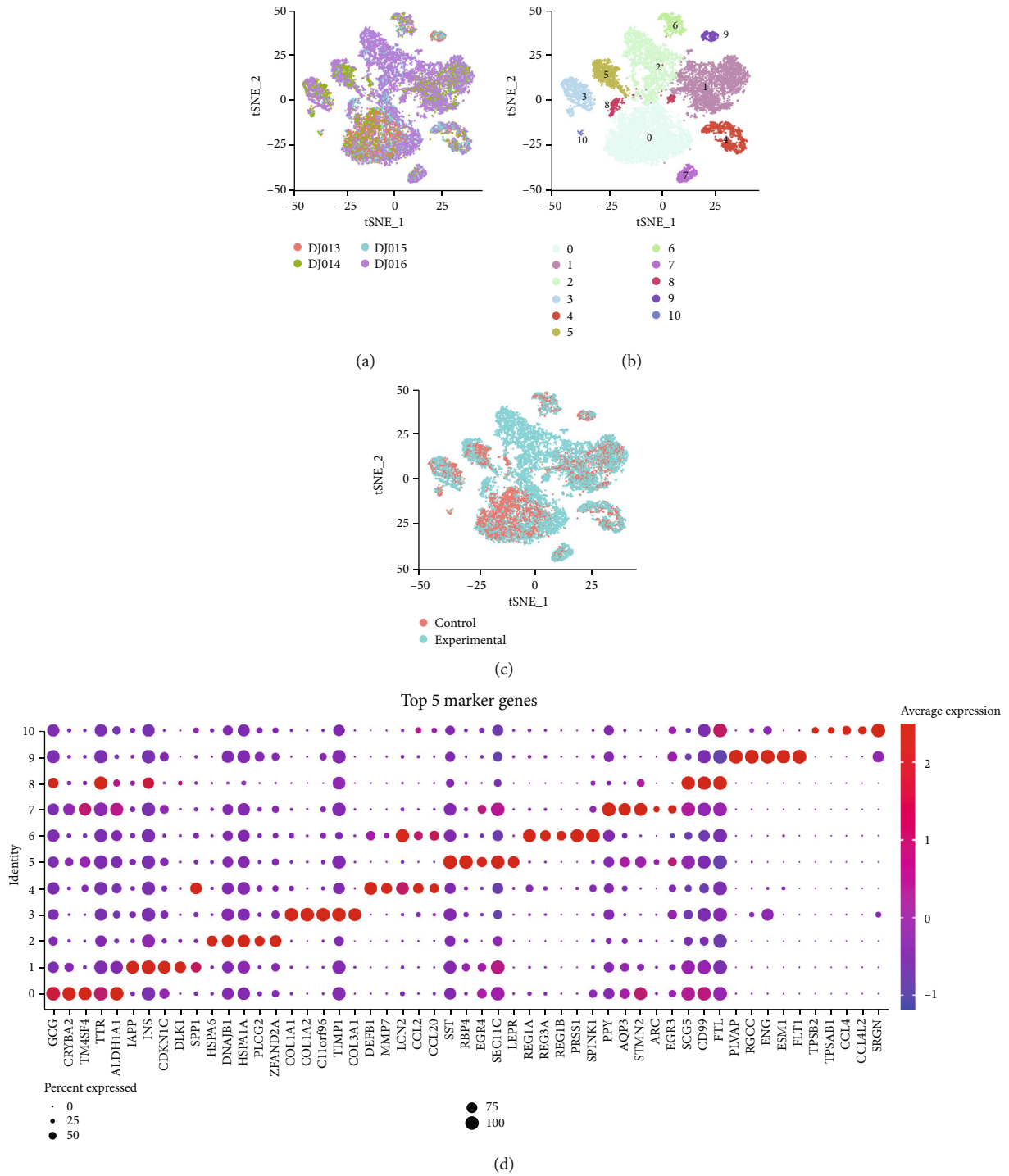


FIGURE 1: Continued.

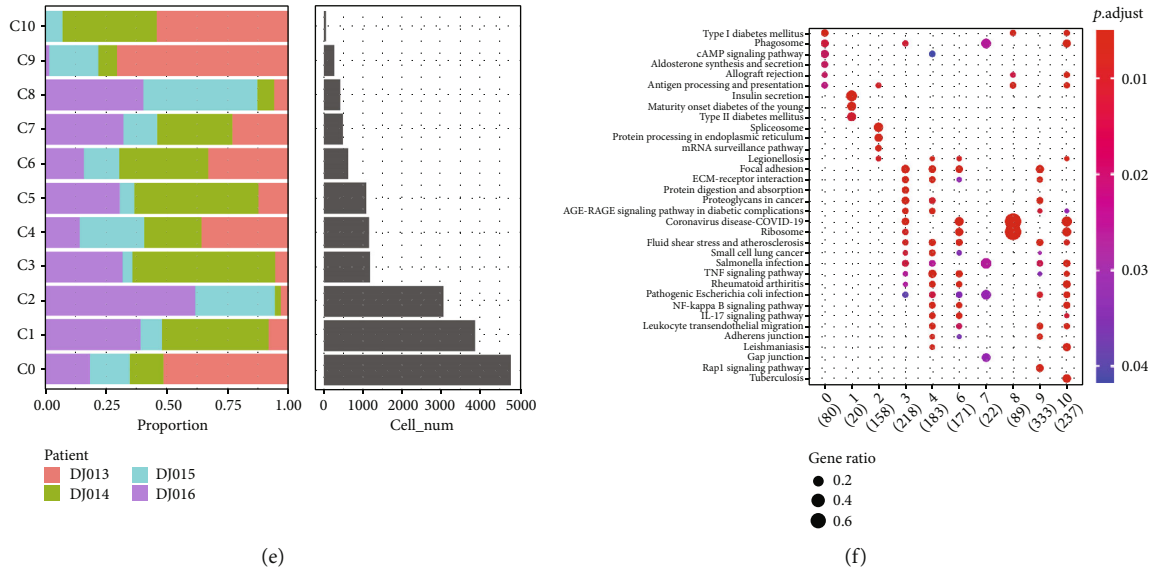


FIGURE 1: scRNA-seq identified 11 cell clusters in T2DM samples. (a) The  $t$ -SNE diagram of the cells in each islet  $\beta$  cell sample. (b) 11 subgroups of scRNA-seq cluster analysis. (c) The distribution of cells from T2DM samples and normal controls in the subpopulation. (d) The bubble plots show the top five marker genes of each subpopulation. (e) The proportion and cell number of 11 subpopulations in 4 samples. (f) The KEGG analysis of cell enrichment pathway was annotated according to the marker genes of each cluster.

17106 cells were finally included for further analysis (Supplemental Figure S1, see Sections 2.2 and 2.3). Next, the  $t$ -SNE algorithm clustered the cells in four islet  $\beta$  cell samples into 11 independent subgroups or cell clusters (Figures 1(a)–1(c)). The marker genes of the top five subpopulations were visualized by the bubble chart (Figure 1(d)). C0, C1, and C2 consisted the majority of cell populations, where T2DM samples were significantly accumulated in C2 (Figure 1(e)). In addition, C0, C3, C4, C5, C6, C9, and C10 accounted for a large proportion of normal samples (Figure 1(e)). KEGG enrichment analysis on the marker genes of each cluster revealed that DM-related pathways such as T1DM, T2DM, and insulin secretion were noticeably enriched, where T1DM was enriched in C0, C8, and C10 and T2DM and insulin secretion were both enriched in C1 (Figure 1(f)).

**3.2. Identification of Key Cell Clusters and Important Marker Genes.** Analysis of differences in 11 subsets of differences of islet  $\beta$  cell samples between T2DM patients and normal controls showed that there were three subsets within our threshold range, namely, C2, C8, and C10 (Table 1). Among the 93 genes, 8 belonged to the marker genes of the three key cell clusters, namely, METAP2, PLAU, BRD4, NR3C1, HSP90AA1, CTSS, CTSB, and ADRB2 (Figures 2(a) and 2(b)). The expression of each gene in C2, C8, and C10 was analyzed. The results showed that METAP2, BRD4, and NR3C1 were high-expressed in C2 and C10, that PLAU, CTSS, and ADRB2 were high-expressed in C10, and that HSP90AA1 and CTSB were high-expressed in C2, C8, and C10 (Figure 2(c)). After removing the genes with few nodes in PPI, 88 genes remained and were selected to construct a PPI network, with the above eight genes located at the core of the network (Figure 2(d)).

TABLE 1: Differences of 11 subgroups in islet  $\beta$  cell samples between T2DM patients and nondiabetic controls.

Cell_name	P.val	Adjust.P.val	fc
C0	$8.20E - 140$	$4.51E - 139$	0.41759
C1	0.002008	0.002209289	0.892653
C2	0	$1.46E - 14$	22.39432
C3	$1.10E - 24$	0	0.533703
C4	$1.40E - 11$	$2.73E - 24$	0.660653
C5	$1.95E - 20$	$1.72E - 11$	0.554438
C6	$1.24E - 24$	$3.06E - 20$	0.422659
C7	0.061583	$2.73E - 24$	0.842438
C8	$1.96E - 60$	0.061582664	6.80658
C9	$2.12E - 22$	$7.20E - 60$	0.274513
C10	$1.06E - 14$	$3.89E - 22$	0.076347

**3.3. Regulatory Relationship of Six Key Subnetworks and Their Effects on Infiltrating Immune Cells.** The MCODE algorithm found out six key subnetworks from the PPI network (Figure 3(a)). KEGG enrichment analysis screened the KEGG pathways significantly related to the 88 previously screened genes, among which the genes in MCODE 1-5 were involved in the regulation of these KEGG pathways. Specifically, MCODE 1 was enriched in many signaling pathways, such as apoptosis, Kaposi sarcoma-associated herpesvirus infection, platinum drug resistance, estrogen signaling pathway, and endocrine resistance. MCODE 2 was related to the IL-17 signaling pathway and Alzheimer disease. All the genes of MCODE 3 were enriched in complement and coagulation cascades. The pathways regulated by MCODE 4 include renin secretion, tuberculosis, regulation of lipolysis

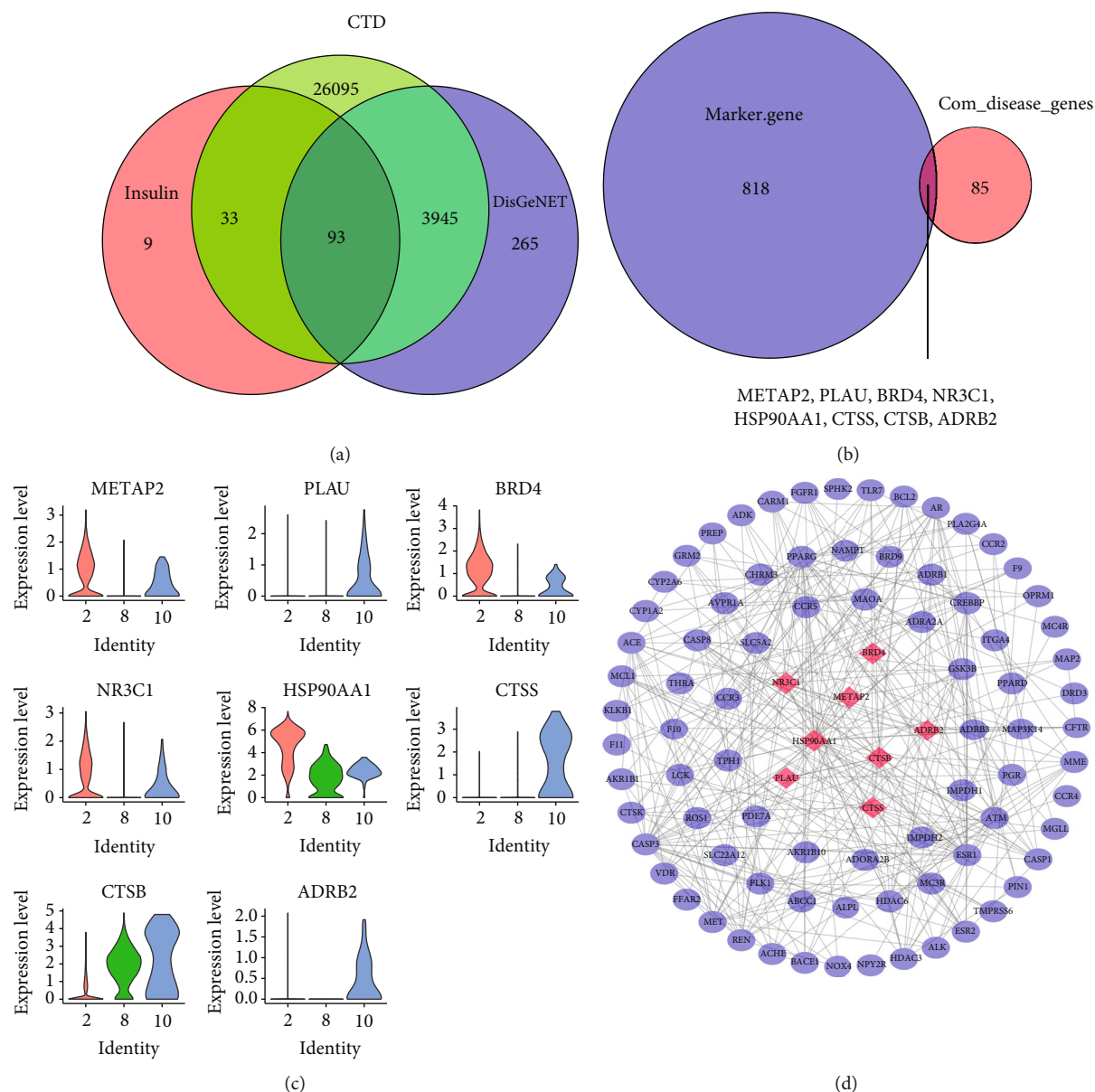


FIGURE 2: Analysis of important genes related to diabetes mellitus and insulin activity. (a) The Venn diagram shows the intersection of diabetes-related genes and ChEMBL-derived insulin activity-related genes in DisGeNET and Comparative Toxicogenomics Database. (b) Venn diagram displaying the intersection of C2, C8, and C10 marker genes with 93 genes. (c) The expression level of METAP2, PLAU, BRD4, NR3C1, HSP90AA1, CTSS, CTSB, and ADRB2 in C2, C8, and C10. (d) PPI network based on 88 genes. Red rhombus indicates the eight intersected genes associated with both T2DM and insulin activity.

in adipocytes, and neuroactive ligand-receptor interaction. The gene in MCODE 5 showed many common regulatory pathways with MCODE 1 and MCODE 4, and it also regulated legionellosis alone (Figure 3(b)). Pearson correlation analysis demonstrated that MCODE 1 and MCODE 5 were significantly correlated with memory B cells, M1 macrophages, M2 macrophages, and resting dendritic cells (Figure 3(c)), where BCL2, CARM1, CASP8, HDAC3, HDAC6, MCL1, CASP1, CCR5, and CTSBT2D had significant correlations with all four cell types.

**3.4. Identification of Genes with the Most Diagnostic Significance for T2DM.** Among the genes in MCODE 1-5,

the expression of 10 genes (BCL2, CARM1, CASP8, HDAC3, HDAC6, NR3C1, BACE1, CASP1, CCR5, and CTSB) can be detected in the GSE182923 dataset. Evaluating the diagnostic performance of each of the 10 genes showed that the diagnostic efficiency of HDAC3, HDAC6, BACE1, CASP1, CCR5, and CTSB was higher than 70% and that the area under the ROC curve (AUC) of CCR5 was the highest of 82.5% (Figure 4(a)). Among the 10 genes, only CCR5 was differentially expressed between normal samples and T2DM samples ( $P < 0.05$ , Figure 4(b)). Moreover, we extracted CCR5-related genes, chemicals, and the corresponding pathways (Supplemental Figure S2). The result showed that CASP and CCR protein families closely

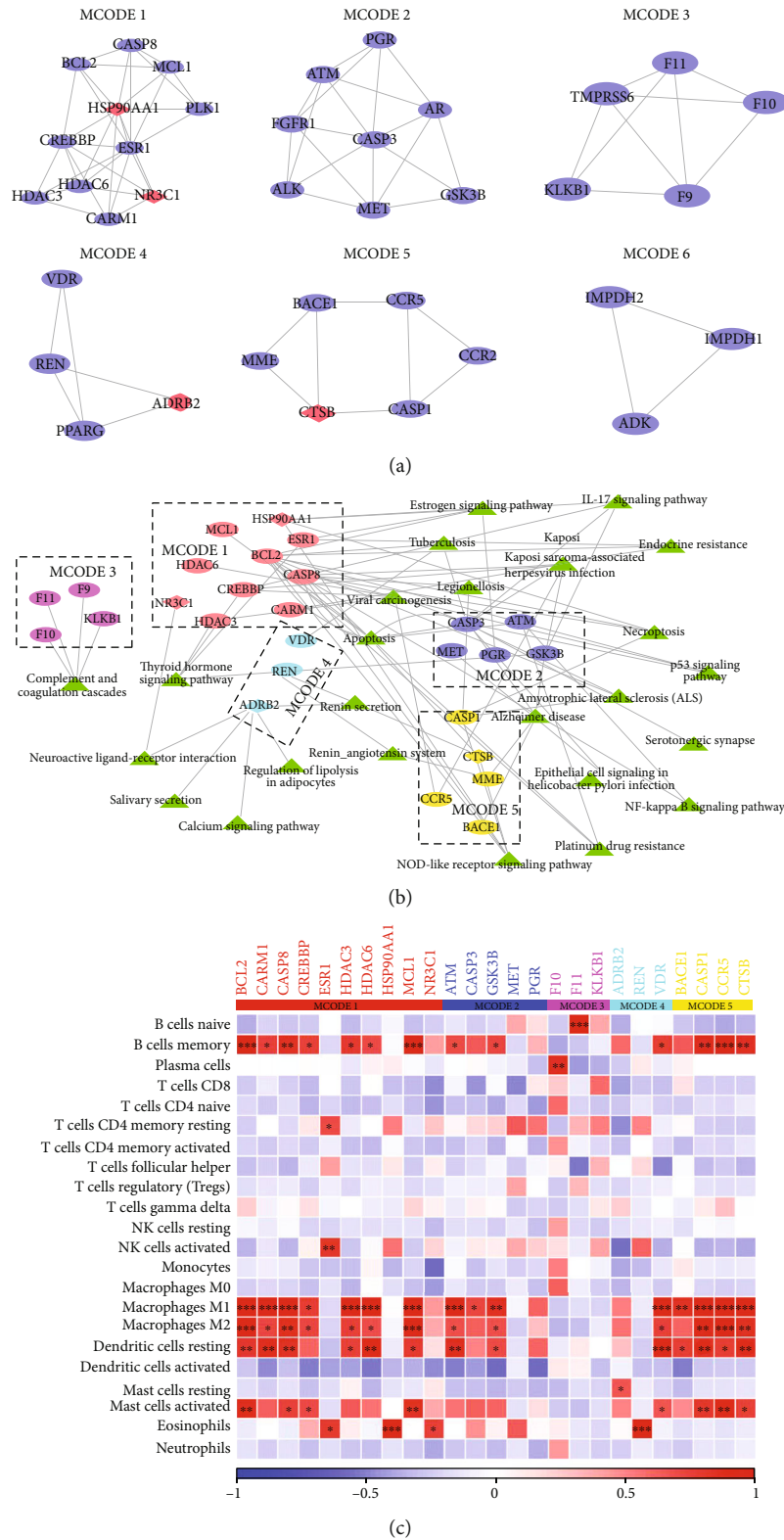
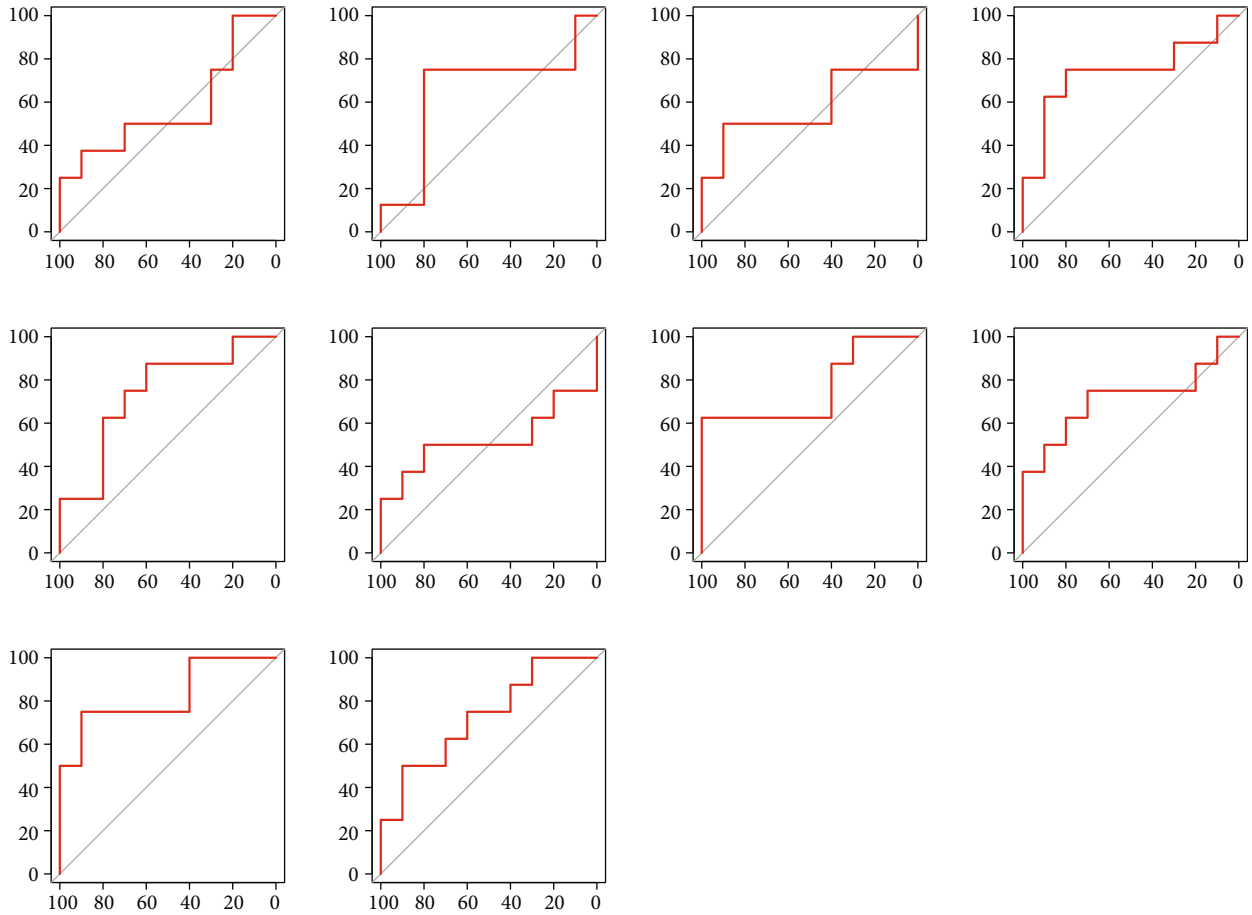
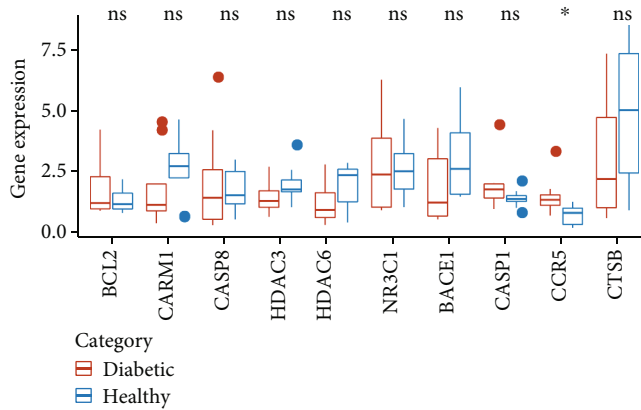


FIGURE 3: Regulatory relationship of 6 key clusters and their effects on infiltrating immune cells. (a) The key subnetworks in the PPI network were screened based on the MCODE algorithm. (b) KEGG pathways of five key subnetworks (MCODE 1-5). The triangle green represents the pathway, the circle is the disease-related gene, and the diamond is the marker gene of the differential subgroup, and the genes in the same red dotted line belong to the same MCODE. (c) Pearson correlation analysis between each gene within the five key subnetworks and 22 immune cells. Red indicates positive correlation, and blue indicates negative correlation. \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$ .



(a)



(b)

FIGURE 4: Identification of genes with the most diagnostic significance for T2DM. (a) The ROC curve of diabetes was diagnosed by the expression of BCL2, CARM1, CASP8, HDAC3, HDAC6, NR3C1, BACE1, CASP1, CCR5, and CTSBT2D. (b) The differential expression of 10 genes between healthy samples and T2DM samples was analyzed. The Wilcoxon test was conducted. ns: no significance. \* $P < 0.05$ .

interacted with CCR5 and that two pathways (viral carcinogenesis and Kaposi sarcoma-associated herpesvirus infection) were enriched.

**3.5. Identification of Potential Drugs Targeting CCR5.** To identify potential drugs targeting CCR5, 49 genes significantly related to CCR5 in GSE182923 were screened. Then,

the proximal drugs were collected by the application of web-based proximity analysis, and the drugs without a significant correlation with CCR5 were excluded. By mapping CCR-related genes to the network, the proximity of drugs to CCR5-related genes was calculated. The density map showed that the distance distribution between drugs and CCR5-related genes showed partial overlap but was smaller



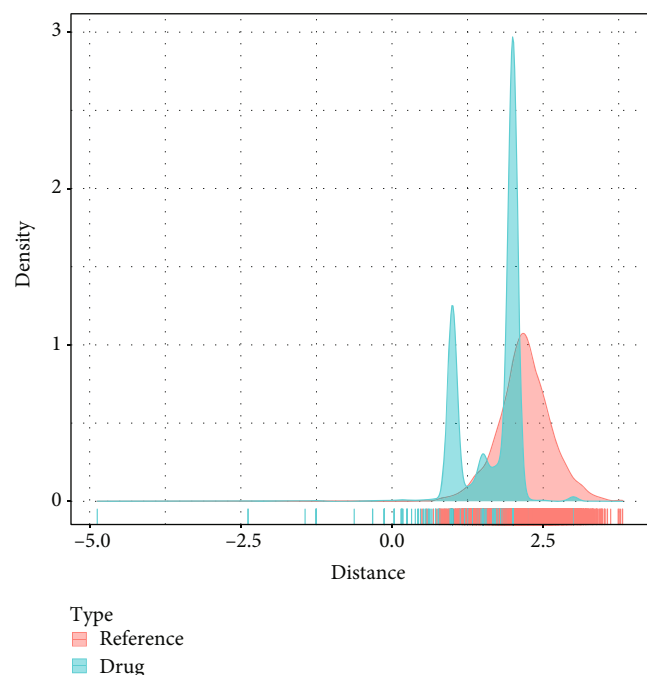


FIGURE 5: Proximity between drugs and CCR5. Density plot showing the distance distribution of all drugs to PD-related genes (blue) and the reference data (red).

than that of reference background drugs and CCR-related genes (Figure 5). We selected drugs with global FDR < 0.001 and obtained 145 drug candidates. The first five drugs with the smallest distance were DB05494, DB00629, DB01018, DB02594, and DB00484. The docking score of each drug with CCR was calculated, and we detected that the docking score between DB05494 and CCR5 was determined to be  $-8.0$  kcal/mol (Table 2). DB05494 produces hydrogen bonds with LYS26, TYR37, TYR89, CYS178, and GLN280 of receptor proteins and hydrophobically interacts with TRP86, PHE112, ILE198, TRP248, and TYR251 of CCR5 (Figures 6(a) and 6(b)). In the 100 ns molecular dynamic simulation, DB05494 binding to CCR5 showed light microwave motion from 0 to 10 ns. When compared to the simple protein, the simulation from 10 to 100 ns showed a highly stable ligand RMSD (Figure 6(c)). These results indicated that DB05494 could stably conjugate with CCR5 and exert potential biological activity. Moreover, docking the ligand of 4MBS with CCR5 showed a significantly low root-mean-square distance (RMSD) of  $0.4550 \text{ \AA}$  that was far below  $2 \text{ \AA}$  (Supplemental Figure S3), supporting the reliability of our docking method.

#### 4. Discussion

T2DM is a common chronic illness characterized by insulin resistance and eventually by decreased insulin secretion by pancreatic beta cells, leading to chronic hyperglycemia and associated long-term disease complications. The rising burden of T2DM remains one of the biggest health challenges today [31]. So far, T2DM is still incurable, and many drugs do not always fully control T2DM unless they are used or

combined used in the early stages of disease progression [32]. However, this strategy may be affected by issues associated with the multipharmacological methods, such as some side effects, toxicity, and unwanted drug-drug interactions. An alternative strategy is to use a single drug to selectively modulate different targets, which may have the potential to improve the balance between efficacy and safety compared to single-target drugs [33].

In this study, we designed a systematic computing framework and used network analysis to identify potential therapeutic drugs and drug targets of T2DM. Firstly, we collected insulin-related genes and T2DM-related genes from two databases and developed a PPI network based on the genes coexisting in the three databases. From the PPI network, we found 6 key MCODEs. The genes in MCODE, individually or jointly, regulated a wide range of biological pathways, either individually or jointly, including apoptosis, platinum drug resistance, estrogen signaling pathway, endocrine resistance, complement and coagulation cascades, renin secretion, and regulation of lipolysis in adipocytes. The relationship between most of these pathways and the development of diabetes have been reported in previous studies. It has been shown that the estrogen signaling pathway regulates rapid changes in systemic metabolism, fat distribution, and insulin action associated with diabetes [34]. Diabetes is caused by metabolic disorders of the endocrine system [35]. Previous studies have demonstrated that androgen receptor-mediated endocrine resistance was associated with diabetes [36]. A direct positive correlation between blood glucose control and renin has been reported. Hyperglycemia increased succinate concentration and succinate receptor activation in the kidney, leading to renin release [37]. Regulation of lipolysis in adipocytes is closely related to insulin resistance. Inhibition of adipocyte lipolysis could be a promising treatment strategy for T2DM related to insulin resistance and obesity prevention [38]. Therefore, the genes in MCODE are likely to be involved in the regulation of these pathways to affect the progression of diabetes.

Among the MCODE genes screened above, we found that BCL2, CARM1, CASP8, HDAC3, HDAC6, MCL1, NR3C1, BACE1, CASP1, CCR5, and CTSBT2D were expressed in another independent dataset (GSE182923) with T2DM samples. Only the expression of CCR5 was significantly different from that in normal samples, with the highest AUC of 82.5%. CCR5 is a kind of chemokine receptor (CCR), and it is known to contribute to metabolic disorders such as T2DM, obesity, atherosclerosis, and human immunodeficiency virus (HIV) infection-induced metabolic changes [39]. Some studies have reported the association of CCR5 polymorphism with clinical outcomes of T1DM, T2DM, and diabetic complications. There was a correlation between CCR5-del32 mutation and the clinical course of T1DM, which was the pathogenic factor of T1DM [40]. [41] reported that the expression of CCR5 in T2DM patients was increased by measuring the peripheral blood mononuclear cells, and CCR5 can be considered an indicator of atherosclerosis in diabetic people by promoting insulin resistance and aggravating pancreatic beta cell dysfunction during the development of diabetes. Muntinghe et al. [42]

TABLE 2: Docking results of CCR5 with five compounds.

Compound	Score (kcal/mol)	Hydrogen bond	Hydrophobic bond
DB05494	-8.0	LYS26, TYR37, TYR89, CYS178, GLN280	TRP86, PHE112, ILE198, TRP248, TYR251
DB00629	-6.5	TYR37, GLU283	TYR108, TRP86, TYR89
DB01018	-7.1	TYR37, GLU283	TRP86, TYR89, TYR108
DB02594	-5.8	TYR37	TRP86, TYR108
DB00484	-6.9	—	TRP86, TYR89, TYR108, THR167

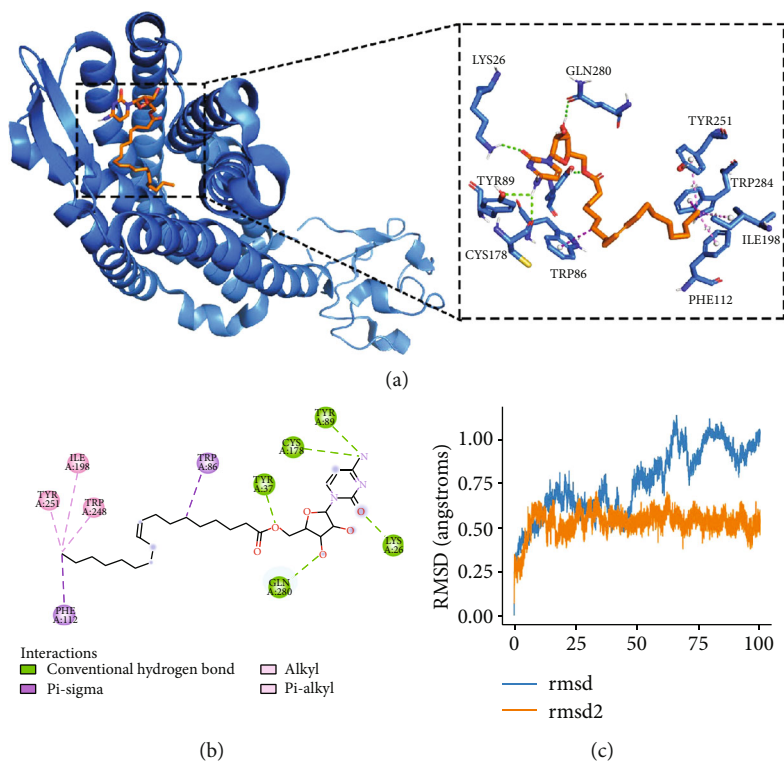


FIGURE 6: The combination mode of DB05494 and CCR5. (a) The three-dimensional diagram of the interaction between CCR5 and DB05494, the green dotted line represents the hydrogen bond, and the purple dotted line represents the hydrophobic interaction. (b) 2D diagram of the interaction between DB05494 and CCR5. (c) The RMSD of the simple protein (1-blue) and molecular complex (2-orange).

reported that the existence of CCR5-delta32 was associated with the high survival rate of patients with T2DM. These results are similar with that reported by Slominski et al. who showed that CCR5-delta32 polymorphism is related to celiac disease and autoimmune thyroiditis coincidence in patients with T1DM [43]. Moreover, the variation of CCR5 was associated with the susceptibility to nephropathy in patients with T2DM [44]. Considering the relationship between CCR5 and T2DM, we screened 49 genes associated with CCR5 in the dataset to explore and find multitarget drugs that simultaneously target these genes. By analyzing the network-based proximity between drugs and CCR-related genes, we obtained 145 drug candidates and calculated the binding energy of the first five genes with the smallest distance to CCR5. Finally, the DB05494 with the smallest score of docking with CCR5 was selected for simulated molecular docking. It has been observed that DB05494 produced hydrogen bonds with LYS26, TYR37, TYR89,

CYS178, and GLN280 of CCR5 and hydrophobically interacts with TRP86, PHE112, ILE198, TRP248, and TYR251 to form a stable conformation. This indicated that DB05494 may be a potential drug targeting CCR5.

There are limitations to this study. Firstly, T2DM samples in single-cell data are transited from nondiabetic islet  $\beta$  cells infected by adenovirus (packaging with T2DM-related transcriptional factors that can result in T2DM-like feature) [45]. The nonoriginal T2DM samples may affect the accuracy of the analysis. Nevertheless, Gribov et al. [17] demonstrated that T2DM-like islet  $\beta$  cells could be reversed to nondiabetic islet  $\beta$  cells, supporting the reliability of the cell model. Therefore, to some extent, the analysis based on T2DM-like samples was reasonable and reliable. Secondly, although DB05494 is a potential therapeutic candidate for treating T2DM, whether it is effective against T2DM has not been confirmed by in vitro or in vivo experiments. In addition, CCR5 is a well-known target for treating

HIV infection [46], and it is also a potential target for neuro-inflammatory diseases [47]. Whether targeting CCR5 in T2DM patients may induce other inflammatory responses remains unclear. On the one hand, expanded biological experiments are needed to demonstrate the exact role of CCR5 blockade and the efficiency of DB05494 in inhibiting CCR5 for T2DM treatment. On the other hand, the adverse effects of targeting CCR5 such as on the aspect of immune response should also be verified.

## 5. Conclusion

In conclusion, our work combined scRNA-seq data and transcriptome data of T2DM samples and identified CCR5 as a potential target for T2DM treatment based on PPI network analysis. Through employing molecular docking and simulative molecular dynamics, DB05494 was screened as a potential therapeutic drug for treating T2DM. Our study clarified the potential role of CCR5 in T2DM from comprehensive bioinformatics analysis and supported the previous findings that CCR5 variants were associated with the development of diabetes.

## Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Conflicts of Interest

No potential conflicts of interest relevant to this article were reported.

## Authors' Contributions

ZW designed the research, carried out the experiment, analyzed the data, and wrote the manuscript. YX and XC collected data. WW provided technical support. YS and LQ reviewed and edited the manuscript. YS is the guarantor of this work. All authors provided final approval of the version to be submitted.

## Acknowledgments

This work was supported by the Project of National Natural Science Foundation of China (Grant No. 81800713 to Wan Zhou and Grant No. 81971264 to Wei Wang), the project of Natural Science Foundation of Anhui Province (Grant No. 1808085QH292 to Wan Zhou), and the Fundamental Research Funds for the Central Universities (Grant No. WK9110000041 to Wan Zhou).

## Supplementary Materials

Supplemental Figure S1: preprocessing scRNA-seq data of four T2DM samples. Supplemental Figure S2: CCR5-acted compounds and biological networks. Supplemental Figure S3: molecular docking between the ligand of 4MBS and CCR5. (Supplementary Materials). (*Supplementary Materials*)

## References

- [1] H. Sun, P. Saeedi, S. Karuranga et al., "IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Research and Clinical Practice*, vol. 183, article 109119, p. 13, 2022.
- [2] A. American Diabetes, "2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020," *Diabetes Care*, vol. 43, Supplement 1, pp. S14–S31, 2020.
- [3] C. Baradez, J. Liska, C. Brulle-Wohlhueter, D. Pushkarna, M. Baxter, and J. Piette, "Brief digital solutions in behavior change interventions for type 2 diabetes mellitus: a literature review," *Diabetes Therapy*, vol. 13, no. 4, pp. 635–649, 2022.
- [4] M. S. Alqunai and F. F. Alrashid, "Bariatric surgery for the management of type 2 diabetes mellitus-current trends and challenges: a review article," *American Journal of Translational Research*, vol. 14, no. 2, pp. 1160–1171, 2022.
- [5] A. I. Adler, I. M. Stratton, H. A. W. Neil et al., "Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study," *BMJ*, vol. 321, no. 7258, pp. 412–419, 2000.
- [6] A. American Diabetes, "Standards of medical care in diabetes—2014," *Diabetes Care*, vol. 37, Supplement\_1, pp. S14–S80, 2014.
- [7] S. Chatterjee, K. Khunti, and M. J. Davies, "Type 2 diabetes," *Lancet*, vol. 389, Supplement 1, pp. 2239–5221, 2017.
- [8] A. L. Hopkins, "Network pharmacology: the next paradigm in drug discovery," *Nature Chemical Biology*, vol. 4, no. 11, pp. 682–690, 2008.
- [9] M. L. Billingsley, "Druggable targets and targeted drugs: enhancing the development of new therapeutics," *Pharmacology*, vol. 82, no. 4, pp. 239–244, 2008.
- [10] R. Zhang, X. Zhu, H. Bai, and K. Ning, "Network pharmacology databases for traditional chinese medicine: review and assessment," *Frontiers In Pharmacology*, vol. 10, p. 14, 2019.
- [11] B. Boezio, K. Audouze, P. Ducrot, and O. Taboureau, "Network-based approaches in pharmacology," *Molecular Informatics*, vol. 36, no. 10, p. 11, 2017.
- [12] P. Quan, K. Wang, S. Yan et al., "Integrated network analysis identifying potential novel drug candidates and targets for Parkinson's disease," *Scientific Reports*, vol. 11, no. 1, p. 9, 2021.
- [13] R. Bharti, V. K. Bhardwaj, C. B. Reddy, R. Purohit, and P. Das, "Benzosuberene-sulfone analogues synthesis from Cedrus deodara oil and their therapeutic evaluation by computational analysis to treat type 2 diabetes," *Bioorganic Chemistry*, vol. 112, article 104860, 2021.
- [14] S. Kumar, K. Sinha, R. Sharma, R. Purohit, and Y. Padwad, "Phloretin and phloridzin improve insulin sensitivity and enhance glucose uptake by subverting PPAR $\gamma$ /Cdk5 interaction in differentiated adipocytes," *Experimental Cell Research*, vol. 383, no. 1, p. 111480, 2019.
- [15] R. Singh, V. K. Bhardwaj, P. Das, and R. Purohit, "Identification of 11 $\beta$ -HSD1 inhibitors through enhanced sampling methods," *Chemical Communications*, vol. 58, no. 32, pp. 5005–5008, 2022.
- [16] V. Rajendran, C. Gopalakrishnan, and R. Sethumadhavan, "Pathological role of a point mutation (T315I) in BCR-ABL1 protein—a computational insight," *Biochemistry*, vol. 119, no. 1, pp. 918–925, 2018.

- [17] A. Gribov, M. Sill, S. Lück et al., "SEURAT: visual analytics for the integrated analysis of microarray data," *BMC Medical Genomics*, vol. 3, p. 21, 2010.
- [18] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "ClusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS*, vol. 16, no. 5, pp. 284–287, 2012.
- [19] G. Su, J. H. Morris, B. Demchak, and G. D. Bader, "Biological network exploration with Cytoscape 3," *Current Protocols In Bioinformatics*, vol. 47, no. 8, pp. 13.1–1324, 2014.
- [20] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinformatics*, vol. 4, no. 1, p. 2, 2003.
- [21] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, and A. A. Alizadeh, "Profiling tumor infiltrating immune cells with CIBERSORT," *Methods in Molecular Biology*, vol. 1711, pp. 243–259, 2018.
- [22] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *Bioinformatics*, vol. 12, no. 1, pp. 1–8, 2011.
- [23] Y. Peng, M. Yuan, J. Xin, X. Liu, and J. Wang, "Screening novel drug candidates for Alzheimer's disease by an integrated network and transcriptome analysis," *Bioinformatics*, vol. 36, pp. 4626–4632, 2020.
- [24] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, pp. 455–461, 2010.
- [25] N. A. Murugan, S. Kumar, J. Jayakanthan, and V. Srivastava, "Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach," *Scientific Reports*, vol. 10, article 19125, 2020.
- [26] N. Sharma, M. Kashif, D. Fontinha et al., "Novel antiplasmodial compounds leveraged with multistage potency against the parasite plasmodium falciparum: in vitro and in vivo evaluations and pharmacokinetic studies," *Journal of Medicinal Chemistry*, vol. 64, no. 12, pp. 8666–8683, 2021.
- [27] K. Gong, W. Guo, K. Du, F. Wang, M. Li, and J. Guo, "Mechanism of huoluo xiaoling dan in the treatment of psoriasis based on network pharmacology and molecular docking," *Evidence-Based Complementary and Alternative Medicine*, vol. 2022, Article ID 7053613, 12 pages, 2022.
- [28] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, pp. 1701–1718, 2005.
- [29] Z. Elkarhat, H. Charoute, L. Elkhatabi, A. Barakat, and H. Rouba, "Potential inhibitors of SARS-cov-2 RNA dependent RNA polymerase protein: molecular docking, molecular dynamics simulations and MM-PBSA analyses," *Journal of Biomolecular Structure & Dynamics*, vol. 40, no. 1, pp. 361–374, 2022.
- [30] R. Chen and S. H. Chung, "Molecular dynamics simulations of Na(+) and leucine transport by LeuT," *Biochemical and Biophysical Research Communications*, vol. 464, no. 1, pp. 281–285, 2015.
- [31] M. J. Redondo, W. A. Hagopian, R. Oram et al., "The clinical consequences of heterogeneity within and between different diabetes types," *Diabetologia*, vol. 63, no. 10, pp. 2040–2048, 2020.
- [32] J. S. Skyler, G. L. Bakris, E. Bonifacio et al., "Differentiation of diabetes by pathophysiology, natural history, and prognosis," *Diabetes*, vol. 66, no. 2, pp. 241–255, 2017.
- [33] A. Artasensi, A. Pedretti, G. Vistoli, and L. Fumagalli, "Type 2 diabetes mellitus: a review of multi-target drugs," *Molecules*, vol. 25, no. 8, p. 1987, 2020.
- [34] F. Mauvais-Jarvis, C. L. May, J. P. Tiano, S. Liu, G. Kilic-Berkmen, and J. H. Kim, "The role of estrogens in pancreatic islet physiopathology," *Advances In Experimental Medicine and Biology*, vol. 1043, pp. 385–399, 2017.
- [35] H. Alrefai, H. Allababidi, S. Levy, and J. Levy, "The endocrine system in diabetes mellitus," *Endocrine*, vol. 18, no. 2, pp. 105–109, 2002.
- [36] S. Sakkiah, T. Wang, W. Zou et al., "Endocrine disrupting chemicals mediated through binding androgen receptor are associated with diabetes mellitus," *International Journal of Environmental Research And Public Health*, vol. 15, no. 1, p. 25, 2017.
- [37] T. P. Griffin, D. Wall, G. A. Browne, M. C. Denny, and P. M. O'Shea, "Associations between glycaemic control and activation of the renin-angiotensin-aldosterone system in participants with type 2 diabetes mellitus and hypertension," *Annals of Clinical Biochemistry*, vol. 55, no. 3, pp. 373–384, 2018.
- [38] P. Morigny, M. Houssier, E. Mouisel, and D. Langin, "Adipocyte lipolysis and insulin resistance," *Biochimie*, vol. 125, pp. 259–266, 2016.
- [39] S. Y. Chou, R. Ajoy, C. A. Changou, Y. T. Hsieh, Y. K. Wang, and B. Hoffer, "CCL5/RANTES contributes to hypothalamic insulin signaling for systemic insulin responsiveness through CCR5," *Scientific Reports*, vol. 6, article 37659, p. 13, 2016.
- [40] J. Dytfeld, P. Bogdański, D. P. Musialik, P. P. Jagodziński, W. Bryl, and A. Kujawa, "Expression of chemokine receptor CCR5 in patients with type 2 diabetes," *Polski Merkurusz Lekarski*, vol. 20, no. 116, pp. 195–198, 2006.
- [41] I. Kalev, K. Oselin, P. Pärlist et al., "CC-chemokine receptor CCR5-del32 mutation as a modifying pathogenetic factor in type I diabetes," *Journal of Diabetes and Its Complications*, vol. 17, no. 6, pp. 387–391, 2003.
- [42] F. L. Muntinghe, S. Gross, S. J. Bakker et al., "CCR5delta32 genotype is associated with outcome in type 2 diabetes mellitus," *Diabetes Research and Clinical Practice*, vol. 86, no. 2, pp. 140–145, 2009.
- [43] B. Slominski, U. Lawryniewicz, J. Myśliwska, M. Ryba-Stanisławska, M. Skrzypkowska, and A. Brandt, "CCR5-delta32 gene polymorphism is associated with retinopathy in patients with type 1 diabetes," *Molecular and Cellular Endocrinology*, vol. 439, no. 2016, pp. 256–260, 2017.
- [44] A. K. Yadav, V. Kumar, P. Dutta, A. Bhansali, and V. Jha, "Variations in CCR5, but not HFE, ELMO1, or SLC12A3, are associated with susceptibility to kidney disease in north Indian individuals with type 2 diabetes," *Journal of Diabetes*, vol. 6, no. 6, pp. 547–555, 2014.
- [45] J. Son, H. Ding, T. B. Farb et al., "BACH2 inhibition reverses  $\beta$  cell failure in type 2 diabetes models," *The Journal of Clinical Investigation*, vol. 131, no. 24, p. 15, 2021.
- [46] T. J. Wilkin and R. M. Gulick, "CCR5 antagonism in HIV infection: current concepts and future opportunities," *Annual Review of Medicine*, vol. 63, pp. 81–93, 2012.
- [47] G. Martin-Blondel, D. Brassat, J. Bauer, H. Lassmann, and R. S. Liblau, "CCR5 blockade for neuroinflammatory diseases – beyond control of HIV," *Nature Reviews Neurology*, vol. 12, no. 2, pp. 95–105, 2016.