



Exploring Sex Differences in Type 2 Diabetes via a Male-Dominant Beta-Cell Cluster from Single-Cell Pancreatic Sequencing of Public Datasets

Nguyen Van Anh^{1,2,*}, Hyo-Wook Gil^{3,*}, Samel Park³, Seongho Ryu^{1,2,4}

¹Soonchunhyang Institute of Medi-bio Science (SIMS), Soonchunhyang University; ²Departments of Integrated Biomedical Science, Soonchunhyang University College of Medicine; ³Department of Internal Medicine, Soonchunhyang University Cheonan Hospital; ⁴Department of Pathology, Soonchunhyang University College of Medicine, Cheonan, Korea

Background: Type 2 diabetes is a complex metabolic disorder characterized by insulin resistance and progressive beta-cell dysfunction. Although sex differences in type 2 diabetes prevalence, progression, and complications have been reported, the molecular mechanisms underlying these differences remain largely unknown. We aimed to utilize single-cell RNA sequencing to identify a beta-cell cluster that is more prevalent in males than in females and exhibits distinct gene expression patterns, gene set enrichment profiles, and cell-cell communication compared to other clusters.

Methods: FASTQ files from four public datasets were preprocessed, aligned to the human genome (GRCh38), and integrated into a high-quality matrix to mitigate batch effects. We focused on beta-cells from type 2 diabetes patients, performed trajectory inference to identify clusters, and conducted differential gene expression and gene set enrichment analyses. These findings were validated using bulk RNA-seq datasets. Additionally, cell-cell communication analysis was performed to identify ligand-receptor interactions, followed by a sensitivity analysis to assess sex-specific differences.

Results: We identified a male-dominant beta-cell cluster (adjusted P value = 4.2×10^{-6}) that displayed unique gene expression patterns and downregulation of pathways associated with protein metabolism and insulin synthesis. Differentially expressed genes (e.g., interleukin 24 [IL24], regulator of G protein signaling like 1 [RGSL1]) were confirmed through bulk analysis. Moreover, the cluster demonstrated distinct communication patterns with other cell types, underscoring sex-specific differences.

Conclusion: We have identified a male-dominant beta-cell cluster characterized by distinct gene expression, signaling pathways, and cell interactions. These findings provide insights into the pathophysiology of type 2 diabetes and may inform the development of more effective, sex-specific therapeutic strategies in the future.

Keywords: Diabetes mellitus, type 2; Single-cell gene expression analysis; Beta-cell cluster

Received: 27 December 2024, Revised: 25 January 2025,
Accepted: 18 February 2025

Corresponding authors: Seongho Ryu
Department of Integrated Biomedical Science, Soonchunhyang University,
25 Bongjeong-ro, Dongnam-gu, Cheonan 31151, Korea
Tel: +82-41-413-5020, E-mail: ryu@sch.ac.kr

Samel Park
Department of Internal Medicine, Soonchunhyang University Cheonan Hospital,
31 Suncheonhyang 6-gil, Dongnam-gu, Cheonan 31151, Korea
Tel: +82-41-413-5006, E-mail: samelpark17@schmc.ac.kr

*These authors contributed equally to this work.

Copyright © 2025 Korean Endocrine Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Type 2 diabetes is a complex metabolic disorder characterized by insulin resistance and progressive beta-cell dysfunction [1]. The complications associated with type 2 diabetes significantly impact disease prognosis and are influenced by multiple factors, with pancreatic islet cell dysfunction playing a critical role in their development and progression [2]. Patients with diabetes mellitus are two to three times more likely to die from any cause [3]. Furthermore, diabetes mellitus is the second-largest contributor to the decline in global health-adjusted life expectancy [4].

Although sex differences in type 2 diabetes prevalence, progression, and complications have been documented, the molecular mechanisms underlying these differences remain largely unexplored. Clinical studies indicate that women present with a higher burden of risk factors at the time of type 2 diabetes diagnosis, including excess weight gain and hypertension [5,6]. Furthermore, female type 2 diabetes patients face a higher relative risk of cardiovascular complications and mortality compared to their male counterparts [7]. These observations suggest potential disparities in endocrine cell function between males and females with type 2 diabetes. Elucidating these differences may facilitate the development of novel, sex-specific treatments for type 2 diabetes.

Single-cell RNA sequencing of pancreatic islets is a powerful tool for studying the pathology of type 2 diabetes. This technology enables researchers to analyze gene expression profiles of individual cells within the islets of Langerhans, including insulin-producing beta-cells and other cell types [8]. By examining these cells at the single-cell level, specific gene expression changes that contribute to beta-cell dysfunction in type 2 diabetes can be identified [9].

In our study, we integrated four public datasets to mitigate batch effects, focusing solely on cells derived from type 2 diabetes patients. Next, we performed trajectory inference analysis to identify dominant beta-cell clusters. We then investigated these clusters by comparing differential gene expression and signaling pathway activities between each cluster and the others. The genes identified in these comparisons were subsequently validated using an additional bulk dataset. Given that beta-cells comprise the majority of the Langerhans islets (60%) [10], we further examined how each cluster interacts with other cell types. Finally, we bolstered our findings on cell-cell communication by conducting a sensitivity analysis of differential gene and signaling pathway expression between beta-cells in males and females.

METHODS

Gene expression matrix construction

Sequence Read Archive (SRA) files were obtained from four public datasets (GSE81547, GSE81608, E-MTAB-5061, GSE86469) and converted to FASTQ files using the sra-toolkit library [8,9,11-13]. The quality of the FASTQ reads was assessed using the fastqc and multiqc libraries [14,15]. Quality control was then performed using the fastp library with a Phred score threshold of 20 [16,17]. The resulting output files were mapped to the human reference genome version 38 (GRCh38.111) to construct gene expression matrices using the STAR library [18,19]. Finally, four annotated data objects (anndata objects) were created from these matrices using the anndata library [20].

Quality control and normalization of the anndata objects

Quality control was performed on the four anndata objects following the guidelines [21]. Low-quality cells were filtered using the numpy, scanpy, seaborn, and scipy libraries [22-25]. Three covariates were used for cell quality control: the number of counts per cell, the number of genes expressed per cell, and the fraction of counts from mitochondrial genes per cell. The thresholds for these covariates were determined using the median absolute deviations (MAD); the cutoffs for the number of counts per cell and the number of genes expressed per cell were set at 5 MADs, while the threshold for the mitochondrial gene fraction was set at 3 MADs [21]. Correction for ambient RNA was performed using the anndata2ri, rpy2, and SoupX libraries [26-28]. This step involved first reducing the data dimensions using principal component analysis, followed by clustering the cells on a k-nearest neighbors algorithm (KNN) graph with the Leiden algorithm, and then applying SoupX [21]. Doublet detection was performed using the scDblFinder library [29]. After eliminating all doublets and low-quality cells, the objects were normalized using Scran's pooling-based size factor estimation method via the Scran library [30]. The datasets were first dimensionally reduced using principal component analysis and clustered using the Leiden algorithm; the resulting clusters were then used to compute size factors, which were subsequently applied to normalize the datasets.

Data integration

The four datasets were merged into a single final dataset, and highly variable genes were selected using the scanpy library [23]. The dataset was then annotated with cell types, and batch effects were removed through batch integration using the scVI

and scANVI libraries [31–33]. The expression matrix is represented as log-normalized counts. Finally, the dataset was filtered to retain only cells derived from type 2 diabetes patients.

Trajectory inference analysis

This step was performed using the Seurat and Monocle libraries [34,35]. First, a Seurat object was created from a type 2 diabetes beta anndata object that contained only type 2 diabetes beta-cells. This object was then converted into a CellDataSet (CDS) object compatible with the Monocle library. The CDS object was processed following the library's tutorial and dimension reduction was performed using the Discriminative dimensionality reduction via learning a tree (DDRTree) algorithm to elucidate cellular trajectories. Differences in the proportions of clusters between sexes were examined using Fisher's exact row-wise test. The threshold for the adjusted *P* value was 0.05.

Differential gene expression

Differential gene expression analysis was conducted using the scanpy library [23]. Data were expressed as log₂ fold change, and the unpaired *t* test was used for statistical comparisons. Thresholds were set at an adjusted *P* value of 0.05 and a log₂ fold change of 0.5.

Bulk analysis

To validate the differential gene expression results between cluster 1 and the remaining clusters, we employed a bulk dataset (accession number GSE86468) and pyDESeq2 to obtain differential expression data [12,36]. Only bulk samples derived from type 2 diabetes patients were included. The Wald test was used for statistical analysis, with an adjusted *P* value threshold of 0.05.

Gene set enrichment analysis

Gene set enrichment analysis was performed using the gseapy library to compare one cluster against the remaining clusters, using the BO: Biological Processes 2023 dataset [37–39]. Gene lists with corresponding scores from the differential gene expression analysis were utilized, and a false discovery rate (FDR) cutoff of 0.25 was applied [40].

Cell-cell communication

Interactions between each cluster and other cell types were examined using the CellChat library, which analyzed both the number and strength of interactions and identified the ligand-receptor pairs involved in these communications [41].

RESULTS

Trajectory inference of beta-cells to identify the dominant cluster in the type 2 diabetes group

We examined the distribution of cell types within the type 2 diabetes group (Fig. 1A). Clustering beta-cells based on trajectory analysis revealed differences in cluster proportions between sexes (Fig. 1B, C). Specifically, cluster 1 represented 52.67% of beta-cells in males, compared to 33.52% in females (Fig. 1D). Although cluster 4 was the second-largest cluster in both sexes, its proportion was higher in females (31.27%) than in males (21.33%). Clusters 2 and 3 exhibited similar patterns, with proportions around 8% in males and 15% in females. The smallest cluster in females, cluster 5, accounted for approximately 10% (Fig. 1D). A Fisher's exact row-wise test was performed for each cluster to compare sexes, revealing that cluster 1 is dominant in both groups and has the smallest adjusted *P* value (4.2×10^{-6}) (Fig. 1D).

Differences in gene and signaling pathway expression between cluster 1 and the remaining clusters

Next, we compared gene expression and gene set enrichment profiles between cluster 1 and the other clusters. As shown in Fig. 2A, cluster 1 displayed distinct gene expression patterns, with vesicle-associated membrane protein 4 (VAMP4), cell division cycle 7 (CDC7), and polymeric immunoglobulin receptor (PIGR) being downregulated relative to their expression in other clusters. This downregulation may be associated with reduced activity in protein transportation and protein metabolism processes in cluster 1 (Fig. 2B). Additionally, signaling pathways involved in protein metabolism and transportation—such as the endosomal transport pathway, transcription by RNA polymerase II pathway, and protein localization pathway, which may relate to insulin production—were downregulated in cluster 1 compared with the other clusters (Fig. 2B, G, H, I). In contrast, when comparing cluster 2 to the remaining clusters, signaling pathways related to apoptotic processes—including positive regulation of apoptotic process, regulation of apoptotic process, and positive regulation of programmed cell death—were upregulated (Fig. 2C). Gene set enrichment analyses for clusters 3 and 4 relative to the other clusters revealed similar trends regarding protein metabolism and transportation pathways. Specifically, pathways such as peptidyl-serine phosphorylation, peptidyl-serine modification, phosphorylation, and endosomal transport were upregulated in both clusters (Fig. 2D, E). Notably, in cluster 4, pathways related to ribosome synthesis (e.g., ri-

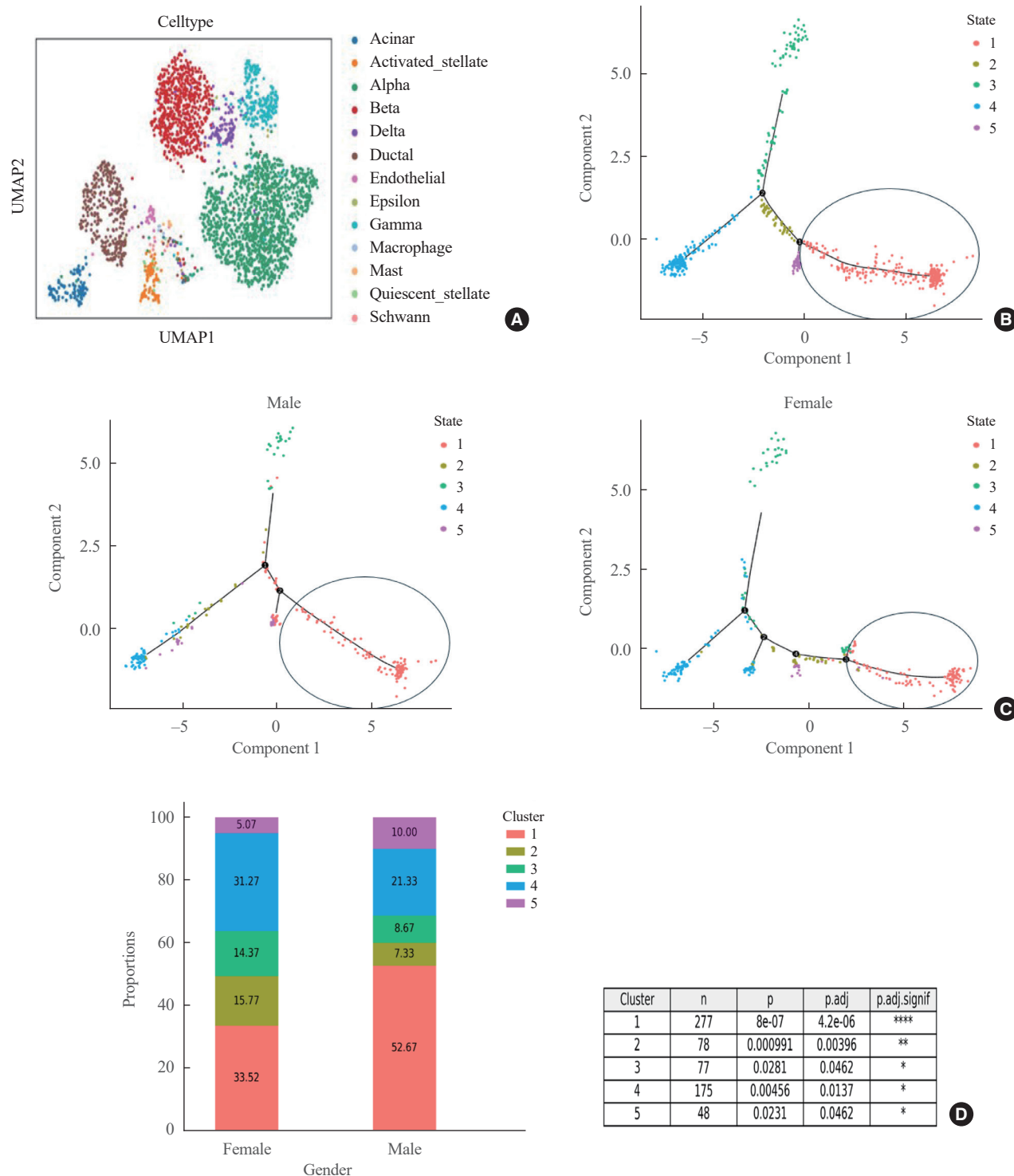


Fig. 1. (A) Uniform Manifold Approximation and Projection (UMAP) representing all cell types in the type 2 diabetes group created using the UMAP method of the scanpy library. (B) Trajectory inference of beta-cells in both sexes created using the Monocle library. (C) Trajectory inference of beta-cells in males (left) and females (right) created using the Monocle library. These plots reveal different beta-cell clusters based on cell differentiation processes. (D) Stacked bar graph representing the proportion of each cluster by sex (left) and the results of Fisher's exact row-wise test for differences in cluster proportions between sexes (right). Two graphs show the most dominant beta-cell cluster in the dataset (cluster 1).

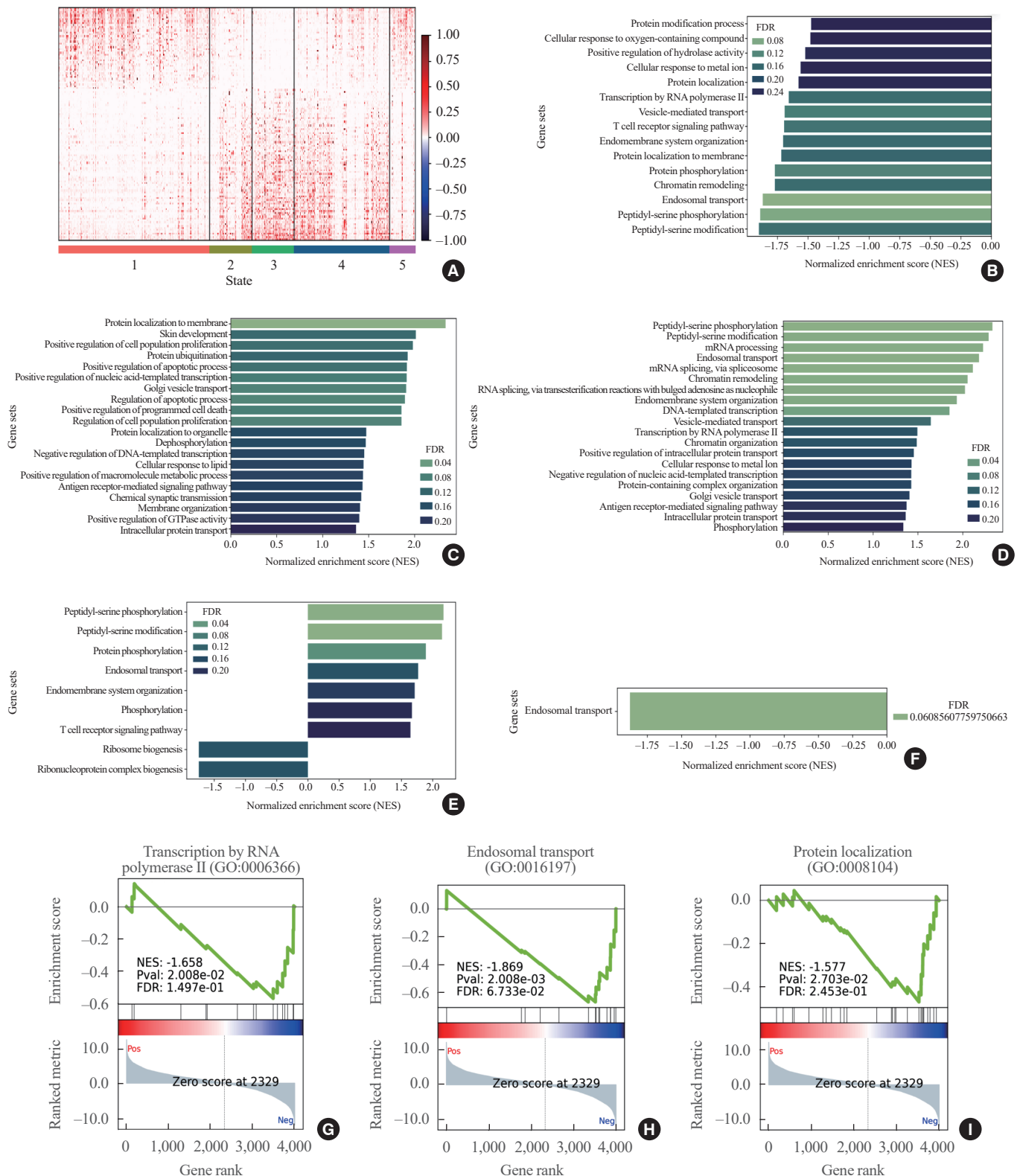


Fig. 2. (A) Heatmap representing the differences in gene expression between cluster 1 and other clusters created using the heatmap method of the scanpy library. This heatmap shows the differences in gene expression which relates to differences in biological processes between cluster 1 and the remaining clusters. (Continued to the next page)

Fig. 2. Continued. (B) Differences in gene set enrichment between cluster 1 and the remaining clusters performed by the gseapy library. The signaling pathways related to protein metabolism and protein transportation are downregulated in cluster 1 compared to those in other clusters. (C) Gene set enrichment for cluster 2 compared to the remaining clusters conducted using the gseapy library. (D) Gene set enrichment for cluster 3 compared to the remaining clusters performed by the gseapy library. (E) Gene set enrichment for cluster 4 compared to the remaining clusters conducted by the gseapy library. (F) Gene set enrichment for cluster 5 compared to the remaining clusters performed using the gseapy library. The cutoff for the false discovery rate (FDR) is 0.25. Statistical details of the differences in three pathways between cluster 1 and the remaining clusters: endosomal transport pathway (H), transcription by RNA polymerase II (G), and protein localization (I). All three gene set enrichment analysis (GSEA) plots were drawn using the gseapy library. All three pathways are downregulated in cluster 1 compared to other clusters. GTPase, guanosine triphosphate.

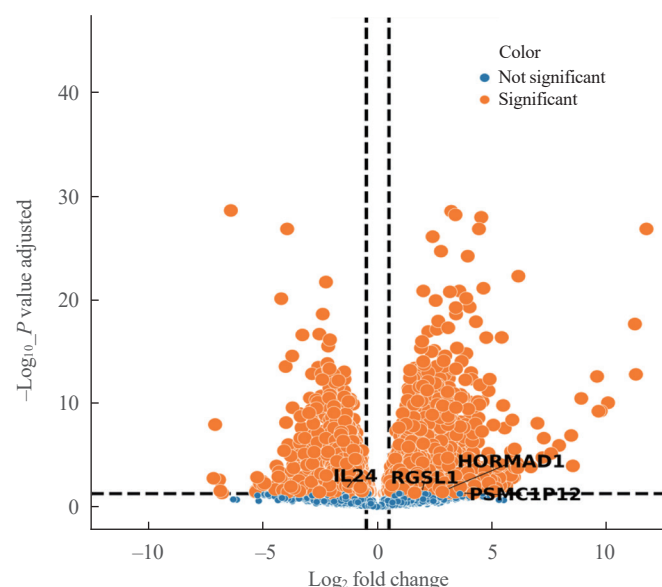


Fig. 3. Volcano plot representing the differences in gene expression between bulk samples in male and female patients performed using pyDeseq2. The cutoff for the adjusted P value and $\log_2$ fold change were 0.05 and 0.5, respectively. IL24, interleukin 24; RGSL1, regulator of G protein signaling like 1; HORMAD1, HORMA domain containing 1; PSMC1P12, proteasome 26S subunit, ATPase 1 pseudogene 12.

bosome biogenesis and ribonucleoprotein complex biogenesis) were downregulated, while the T-cell receptor signaling pathway is upregulated compared to other clusters (Fig. 2E). Finally, in cluster 5, the endosomal transport pathway was downregulated relative to the other clusters (Fig. 2F).

Verification through bulk analysis

To validate the significant differences in gene expression between cluster 1 and the remaining clusters, bulk analysis was performed using an additional bulk dataset. The $\log_2$ fold change cutoff was set at 0.5. Insulin receptor related receptor (*INSRR*) expression was found to be downregulated in males compared to its expression in females (adjusted $P=0.041$). In contrast, *VAMP4*, *CDC7*, and *PIGR* expression levels are upregulated in

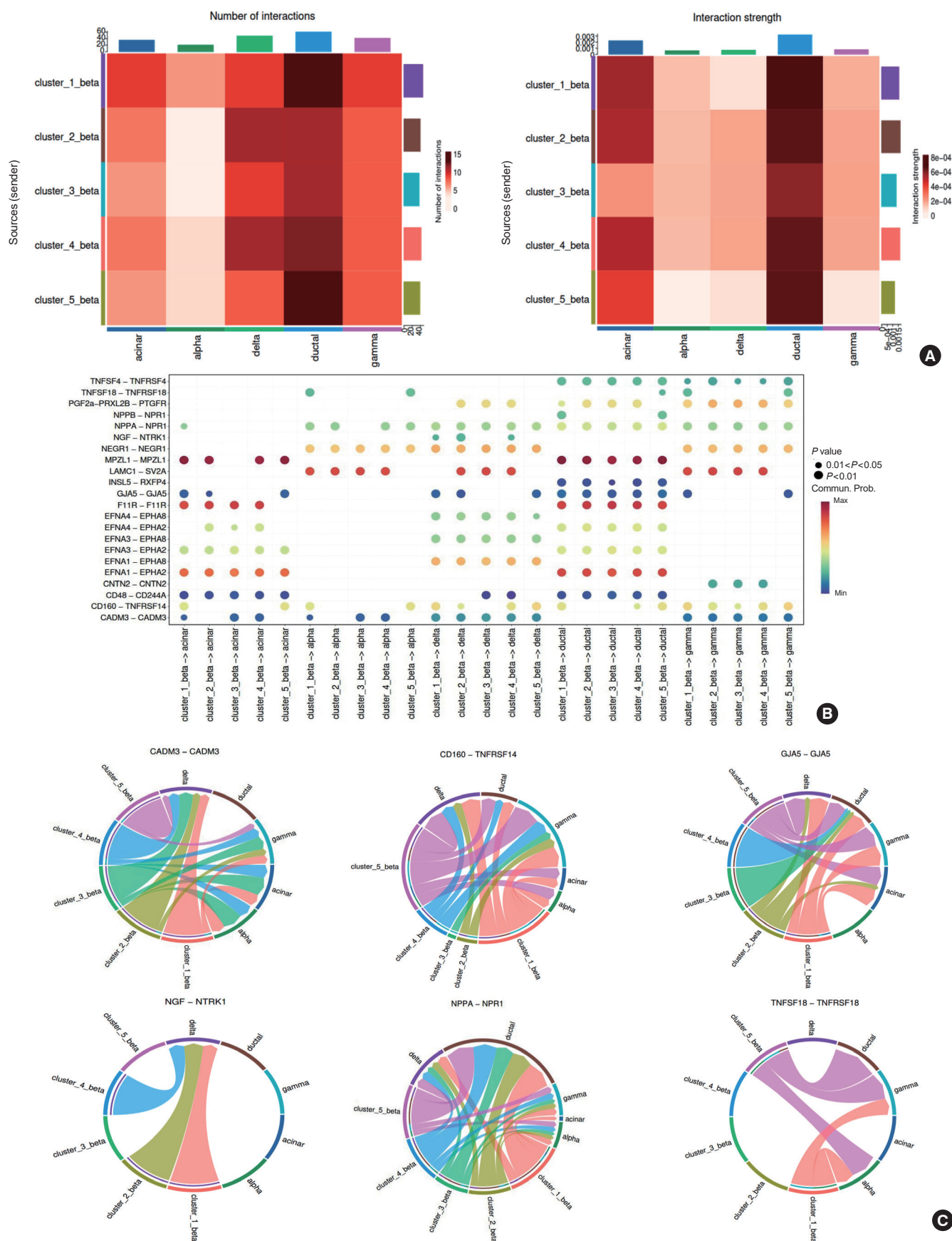
males compared to those in females (adjusted $P=0.0004$, 0.015, and 0.044, respectively) (Fig. 3).

Disparities in cell-cell communication between each cluster and other cell types

We next examined the interactions between each cluster and other cell types. In terms of interaction count, all clusters exhibited the highest number of interactions with ductal cells (Fig. 4A), while the fewest interactions occurred with alpha-cells. Notably, cluster 1 displayed the largest number of interactions with most cell types, except for delta cells (Fig. 4A). Regarding interaction strength, communication between each cluster and ductal cells was the strongest, whereas interactions with alpha-cells were the weakest across all clusters. Specifically, cluster 1 demonstrated the strongest interaction with ductal cells among all clusters (Fig. 4A). Analysis of ligand-receptor pairs revealed that, in addition to common pairs, the natriuretic peptide A (NPPA)-natriuretic peptide receptor 1 (NPR1) pair was uniquely employed by cluster 1 to communicate with acinar cells (Fig. 4B, C). Furthermore, clusters 1, 2, and 4 utilized the nerve growth factor (NGF)-neurotrophic receptor tyrosine kinase 1 (NTRK1) pair for interaction with delta cells. In contrast, most clusters used the cell adhesion molecule 3 (CADM3)-CADM3 pair for communication with other cell types, except for ductal cells (Fig. 4B, C). Only clusters 1 and 5 employed the TNF superfamily member 18 (TNFSF18)-TNF receptor superfamily member 18 (TNFRSF18) pair to communicate with alpha cells and gamma cells. The CD160-TNF receptor superfamily member 14 (TNFRSF14) pair was used exclusively by clusters 1 and 5 for interactions with acinar and alpha cells (Fig. 4B, C). Lastly, clusters 1 and 5 used the gap junction protein alpha 5 (GJA5)-GJA5 pair to interact with nearly all cell types, although neither cluster used this pair to communicate with alpha-cells (Fig. 4B, C).

Differences in gene and signaling pathway expression between beta-cells in males and females

To reinforce our findings regarding ligand-receptor interactions,



(Continued to the next page)

Fig. 4. (A) Heatmaps representing the number of interactions (left) and the strength of interactions (right) between each cluster (cluster 1–cluster 5) and other cell types created using the `netVisual_heatmap` method of the CellChat library. Cluster 1 has the largest number of interactions with most cell types. (B) Dot plot of ligand-receptor pairs used by each cluster to communicate with other cell types. This plot was created using the `netVisual_bubble` method of the CellChat library. (C) Chord plots representing how each cluster uses six ligand-receptor pairs: cell adhesion molecule 3 (CADM3)-CADM3, CD160-TNF receptor superfamily member 14 (TNFRSF14), gap junction protein alpha 5 (GJA5)-GJA5, nerve growth factor (NGF)-neurotrophic receptor tyrosine kinase 1 (NTRK1), natriuretic peptide A (NPPA)-natriuretic peptide receptor 1 (NPR1), and TNF superfamily member 18 (TNFSF18)-TNF receptor superfamily member 18 (TNFRSF18) to communicate with other cell types. These plots were drawn using the `netVisual_chord_cell` method of the CellChat library.

a sensitivity analysis was conducted. We assessed differences in gene expression in beta-cells between males and females using cutoffs of 0.5 for log₂ fold change and 0.05 for the adjusted *P* value, as well as gene set enrichment analysis with an FDR cutoff of 0.25. The analysis revealed significant differences in gene and signaling pathway expression between beta-cells in males and females. Notably, *NGF* expression was downregulated in males relative to females (adjusted *P*=0.0017). In contrast, *TNFSF18* and *CADM3* displayed the opposite pattern, with higher expression in males (adjusted *P*=0.004 and 0.036, respectively) (Fig. 5A, B). Further gene set enrichment analysis showed that pathways related to protein transportation and metabolism, such as protein localization to the membrane, positive regulation of intracellular protein transport, and protein autophosphorylation, were downregulated in beta-cells in males compared to beta-cells in females (Fig. 5C-E). Notably, pathways associated with ribosome synthesis—specifically ribosome biogenesis and ribonucleoprotein complex biogenesis—were upregulated, while the T-cell receptor signaling pathway was downregulated in beta-cells in males relative to beta-cells in females (Fig. 5C, F).

DISCUSSION

Our research investigated the underlying causes of sex differences in patients with type 2 diabetes mellitus. Our results reveal significant differences in the pancreatic cellular patterns of a beta-cell cluster that extend beyond conventional blood sugar measurements. Specifically, we observed variations in the proportion of cluster 1 beta-cells, as well as differences in gene expression, signaling pathway activity, and cell-cell communication between this cluster and the others.

Our findings are consistent with previous studies that have identified sex differences in type 2 diabetes pathophysiology; however, our study provides a more detailed view at the single-cell level. Unlike prior research that primarily focused on blood sugar levels and clinical properties, we uncovered cellular patterns that offer deeper insights into sex-specific disease mecha-

nisms. We found that cluster 1 beta-cells exhibit downregulated signaling pathways related to protein synthesis and transportation. Moreover, these cells utilize the TNFSF18-TNFRSF18 pair to communicate with alpha and gamma-cells—a mechanism that plays a critical role in the progression of autoimmune diabetes [42]. The dominance of cluster 1 in males may be linked to the higher incidence of type 1 diabetes observed in males than in females [43,44]. Additionally, clusters 1, 2, and 4 use the NGF-NTRK1 pair to interact with delta cells; NGF-NTRK1 is pivotal for beta-cell survival, and NGF withdrawal triggers apoptosis by inhibiting phosphatidylinositol 3-kinase (PI3K), protein kinase B, and the Bad survival pathway while activating c-Jun kinase [45]. The downregulation of protein metabolism and transportation pathways in males, which can lead to endoplasmic reticulum (ER) stress [46], may ultimately result in cell death if prolonged [47-49]. This observation helps explain why female islets maintain greater beta-cell function during ER stress compared to male islets [50]. In addition to previously reported genes, we identified several genes encoding ligands—such as the NPPA-NPR1 and CD160-TNFRSF14 pairs—whose roles in type 2 diabetes pathology remain unexplored. These ligands may regulate biological processes that contribute to sex discrepancies in the development of type 2 diabetes.

The strength of our study lies in the discovery of a male-dominant pancreatic beta-cell cluster and its association with the sex dimorphism of diabetes. However, our study is limited by the relatively small number of beta-cells analyzed, which constrained the comprehensiveness of our analysis.

The sex-specific cellular patterns identified in our study underscore the need for sex-tailored treatments for type 2 diabetes. These findings have the potential to enhance treatment effectiveness and improve the prevention and management of complications, ultimately benefiting patients' quality of life.

Future work will further investigate the molecular mechanisms underlying interactions between this beta-cell cluster and other cell types to deepen our understanding of type 2 diabetes pathophysiology.

In summary, our study reveals significant sex-specific differ-

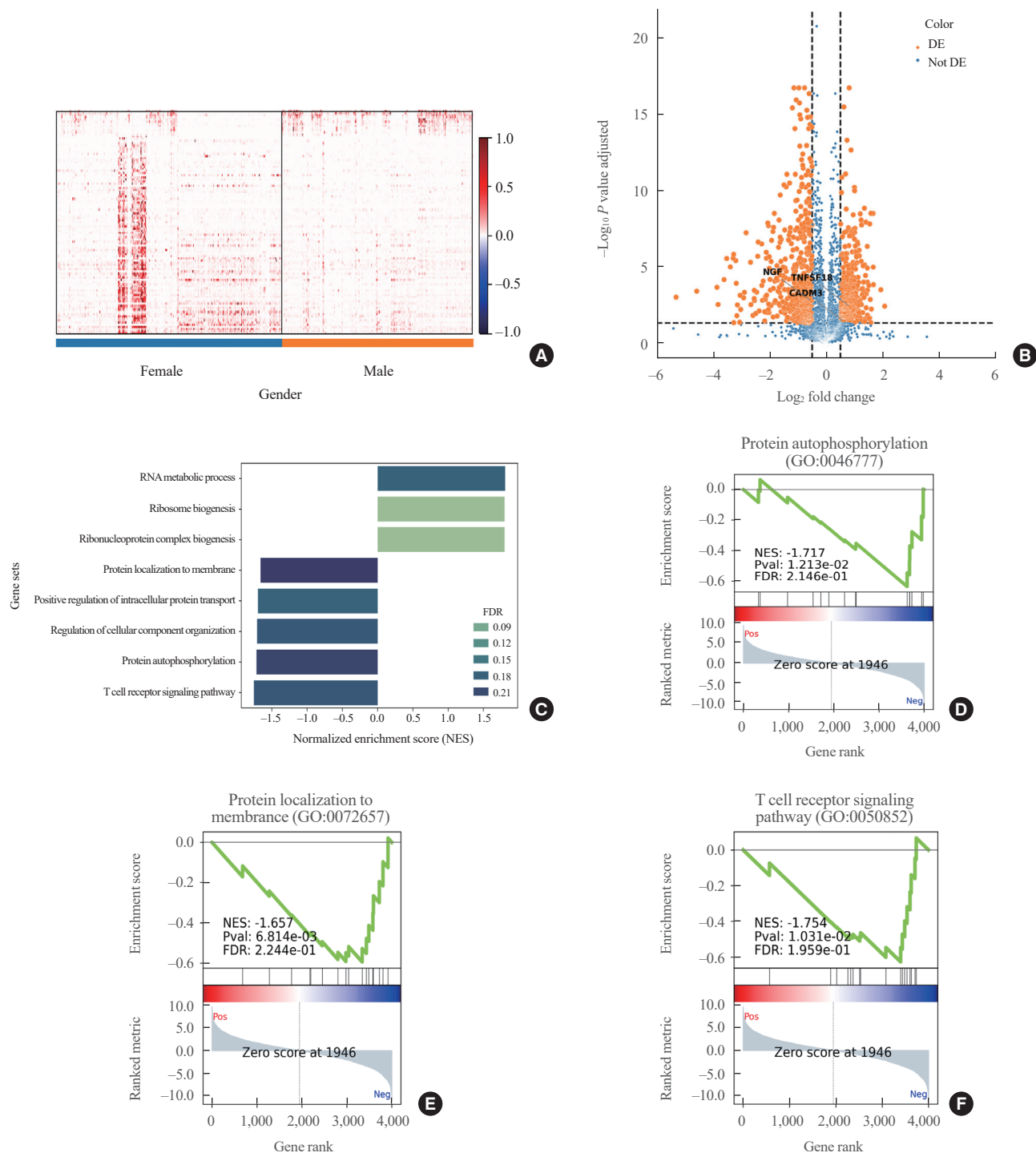


Fig. 5. (A) Heatmap of differential gene expression between beta-cells in males and females. (B) Volcano plot showing some genes, that are statistically different between cluster 1 and the remaining clusters, are downregulated or upregulated between beta-cells in males and females. The cutoff for the adjusted P value and $\log_2$ fold change are 0.05 and 0.5, respectively. The statistical test used was an unpaired t test. (C) Bar plot representing the differences in signaling pathways between beta-cells in males and females created using the gseapy library. The cutoff for false discovery rate (FDR) is 0.25. (D, E, F) Gene set enrichment analysis (GSEA) plots showing details of three signaling pathways related to three genes from the volcano plot drawn using the gseapy library. All three pathways are downregulated in males when compared to those of females. DE, differentially expressed; NES, normalized enrichment score.

ences in the pancreatic cellular patterns of a beta-cell cluster in type 2 diabetes, offering new insights beyond traditional blood sugar metrics and clinical observations. These findings highlight the importance of personalized medicine and lay the groundwork for future research into sex-specific mechanisms and treatments for type 2 diabetes.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

This research was supported by the Bio & Medical Technology Development Program (2019M3E5D3073092) and the Basic Science Research Program (NRF-2021R1A2C3012633 and MSIT, RS-2023-00219563) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT, and the Soonchunhyang University Research Fund.

AUTHOR CONTRIBUTIONS

Conception or design: S.R. Acquisition, analysis, or interpretation of data: N.V.A., H.W.G., S.P. Drafting the work or revising: H.W.G., S.P. Final approval of the manuscript: N.V.A., H.W.G., S.P., S.R.

ORCID

Nguyen Van Anh <https://orcid.org/0009-0000-1487-8373>

Hyo-Wook Gil <https://orcid.org/0000-0003-2550-2739>

Samel Park <https://orcid.org/0000-0002-5717-0743>

Seongho Ryu <https://orcid.org/0000-0001-9697-7972>

REFERENCES

1. Lu X, Xie Q, Pan X, Zhang R, Zhang X, Peng G, et al. Type 2 diabetes mellitus in adults: pathogenesis, prevention and therapy. *Signal Transduct Target Ther* 2024;9:262.
2. Tomic D, Shaw JE, Magliano DJ. The burden and risks of emerging complications of diabetes mellitus. *Nat Rev Endocrinol* 2022;18:525-39.
3. Li S, Wang J, Zhang B, Li X, Liu Y. Diabetes mellitus and cause-specific mortality: a population-based study. *Diabetes Metab J* 2019;43:319-41.
4. Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. *Sci Rep* 2020;10:14790.
5. Tramunt B, Smati S, Grandgeorge N, Lenfant F, Arnal JF, Montagner A, et al. Sex differences in metabolic regulation and diabetes susceptibility. *Diabetologia* 2020;63:453-61.
6. Wright AK, Welsh P, Gill JM, Kontopantelis E, Emsley R, Buchan I, et al. Age-, sex- and ethnicity-related differences in body weight, blood pressure, HbA1c and lipid levels at the diagnosis of type 2 diabetes relative to people without diabetes. *Diabetologia* 2020;63:1542-53.
7. Hu G; DECODE Study Group. Gender difference in all-cause and cardiovascular mortality related to hyperglycaemia and newly-diagnosed diabetes. *Diabetologia* 2003;46:608-17.
8. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, Kim SK, et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* 2017;171:321-30.
9. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab* 2016;24:608-15.
10. Johnson LR. *Encyclopedia of gastroenterology*. Amsterdam: Elsevier; 2004. Chapter, Endocrine pancreas; p. 675-80.
11. Segerstolpe A, Palasantza A, Eliasson P, Andersson EM, Andreasson AC, Sun X, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab* 2016;24:593-607.
12. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, Sivakamasundari V, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res* 2017;27:208-22.
13. SRA Toolkit Development Team. (Sequence Read Archive Toolkit) a collection of tools and libraries for using data in the INSDC Sequence Read Archives [Internet]. SRA Toolkit Development Team; 2016 [cited 2025 Mar 12]. Available from: <https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software> and the SRA Toolkit Development Team.
14. Andrews S. FastQC: a quality control tool for high throughput sequence data [Internet]. Cambridge: Babraham Bioinformatics; 2010 [cited 2025 Mar 12]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
15. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32:3047-8.

16. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884-90.
17. Sheng Q, Vickers K, Zhao S, Wang J, Samuels DC, Koues O, et al. Multi-perspective quality control of Illumina RNA sequencing data analysis. *Brief Funct Genomics* 2017;16:194-204.
18. Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, et al. Ensembl 2024. *Nucleic Acids Res* 2024;52(D1):D891-9.
19. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21.
20. Virshup I, Rybakov S, Theis FJ, Angerer P, Wolf FA. anndata: access and store annotated data matrices. *J Open Source Softw* 2024;9:4371.
21. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, et al. Best practices for single-cell analysis across modalities. *Nat Rev Genet* 2023;24:550-72.
22. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature* 2020;585:357-62.
23. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018;19:15.
24. Waskom ML. Seaborn: statistical data visualization. *J Open Source Softw* 2021;6:3021.
25. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261-72.
26. Philipp A, Eraslan G, Virshup I, Gigante S. theislabs/anndata2ri v1.0.4 [Internet]. Geneva: Zenodo; 2020 [cited 2025 Mar 12]. Available from: <https://doi.org/10.5281/zenodo.3992373>.
27. rpy2 Contributors. rpy2: Python interface to R [Internet]. GitHub; 2021 [cited 2025 Mar 12]. Available from: <https://github.com/rpy2/rpy2>.
28. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* 2020;9:giaa151.
29. Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDbfFinder. *F1000Res* 2021;10:979.
30. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 2016;5:2122.
31. Virshup I, Bredikhin D, Heumos L, Palla G, Sturm G, Gayoso A, et al. The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat Biotechnol* 2023;41:604-6.
32. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018;15:1053-8.
33. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol* 2021;17:e9620.
34. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381-6.
35. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multi-modal and scalable single-cell analysis. *Nat Biotechnol* 2024;42:293-304.
36. Muzellec B, Telenczuk M, Cabeli V, Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* 2023;39:btad547.
37. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 2023;39:btac757.
38. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25-9.
39. Gene Ontology Consortium; Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. *Genetics* 2023;224:iyad031.
40. 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;102:15545-50.
41. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat Protoc* 2025;20:180-219.
42. You S, Poulton L, Cobbold S, Liu CP, Rosenzweig M, Ringler D, et al. Key role of the GITR/GITRLigand pathway in the development of murine autoimmune diabetes: a potential therapeutic target. *PLoS One* 2009;4:e7848.
43. Varlamov O, Bethea CL, Roberts CT Jr. Sex-specific differences in lipid and glucose metabolism. *Front Endocrinol (Lausanne)* 2015;5:241.
44. Mauvais-Jarvis F. Gender differences in glucose homeosta-

- sis and diabetes. *Physiol Behav* 2018;187:20-3.
45. Pierucci D, Cicconi S, Bonini P, Ferrelli F, Pastore D, Matteucci C, et al. NGF-withdrawal induces apoptosis in pancreatic beta cells in vitro. *Diabetologia* 2001;44:1281-95.
 46. Scheuner D, Song B, McEwen E, Liu C, Laybutt R, Gillespie P, et al. Translational control is required for the unfolded protein response and in vivo glucose homeostasis. *Mol Cell* 2001;7:1165-76.
 47. Luciani DS, Gwiazda KS, Yang TL, Kalynyak TB, Bychkivska Y, Frey MH, et al. Roles of IP3R and RyR Ca²⁺ channels in endoplasmic reticulum stress and beta-cell death. *Diabetes* 2009;58:422-32.
 48. Riggs AC, Bernal-Mizrachi E, Ohsugi M, Wasson J, Fatrai S, Welling C, et al. Mice conditionally lacking the Wolfram gene in pancreatic islet beta cells exhibit diabetes as a result of enhanced endoplasmic reticulum stress and apoptosis. *Diabetologia* 2005;48:2313-21.
 49. Li J, Lee B, Lee AS. Endoplasmic reticulum stress-induced apoptosis: multiple pathways and activation of p53-up-regulated modulator of apoptosis (PUMA) and NOXA by p53. *J Biol Chem* 2006;281:7260-70.
 50. Brownrigg GP, Xia YH, Chu CM, Wang S, Chao C, Zhang JA, et al. Sex differences in islet stress responses support female β cell resilience. *Mol Metab* 2023;69:101678.