Article

scGENA: A Single-Cell Gene Coexpression Network Analysis Framework for Clustering Cell Types and Revealing Biological Mechanisms

Yousif A. Algabri [ID], Lingyu Li [ID] and Zhi-Ping Liu *[ID]

Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan 250061, China; algabri@mail.sdu.edu.cn (Y.A.A.); lingyuli@mail.sdu.edu.cn (L.L.)
* Correspondence: zpliu@sdu.edu.cn

Abstract: Single-cell RNA-sequencing (scRNA-seq) is a recent high-throughput technique that can measure gene expression, reveal cell heterogeneity, rare and complex cell populations, and discover cell types and their relationships. The analysis of scRNA-seq data is challenging because of transcripts sparsity, replication noise, and outlier cell populations. A gene coexpression network (GCN) analysis effectively deciphers phenotypic differences in specific states by describing gene–gene pairwise relationships. The underlying gene modules with different coexpression patterns partially bridge the gap between genotype and phenotype. This study presents a new framework called scGENA (single-cell gene coexpression network analysis) for GCN analysis based on scRNA-seq data. Although there are several methods for scRNA-seq data analysis, we aim to build an integrative pipeline for several purposes that cover primary data preprocessing, including data exploration, quality control, normalization, imputation, and dimensionality reduction of clustering as downstream of GCN analysis. To demonstrate this integrated workflow, an scRNA-seq dataset of the human diabetic pancreas with 1600 cells and 39,851 genes was implemented to perform all these processes in practice. As a result, scGENA is demonstrated to uncover interesting gene modules behind complex diseases, which reveal biological mechanisms. scGENA provides a state-of-the-art method for gene coexpression analysis for scRNA-seq data.

Keywords: scRNA-seq; gene coexpression network analysis; cell heterogeneity; gene modules; biological mechanisms; human diabetic pancreas

1. Introduction

The vast majority of cells in a single organism share the same genome, although gene expression differs between tissues and cell types. One of the long-standing issues in biology and medicine is relating genotypes to phenotypes. The transcriptomic analysis is an effective way to address some of these issues [1]. One conventional method of transcriptomic analysis is an mRNA abundance measurement at the tissue or cell level and averaging it over hundreds of millions of cells in bulk RNA-seq data [2]. The bulk RNA-seq techniques have been successfully used in many studies, contributing to our understanding of gene expression [3,4]. In spite of that, the downside of bulk RNA-seq is that cell-specific mRNA abundance cannot be revealed, and biologically important gene expression regulation in individual cells may go undetected [5]. Yet, our current understanding of cell types and their dynamic alterations within biological systems is severely lacking [6]. With ongoing efforts to tackle this shortcoming, single-cell RNA sequencing (scRNA-seq) was introduced in 2009 by Tang et al. [7]. Since then, single-cell technique has been underdevelopment till it became easily accessible and was named the "method of the year" in 2014 by Nature Method [8].

The growth of scRNA-seq technology has many advantages over other existing methods, such as being a powerful technique to thoroughly characterize cellular disruption

within tissues since it assesses the gene expression in individual cells [9]. Furthermore, the scRNA-seq trajectory inference technique (which means a pseudotime analysis that arranges cells along a pathway based on expression pattern similarities) can provide a detailed understanding of dynamic cell differentiation [10]. Moreover, scRNA-seq datasets are primarily utilized to identify cell types or to discover new biomarkers. Gene expression in a single cell is considered stochastic, so the gene expression values and their interactions in different cells vary enormously [11,12]. However, it is significant to develop a comprehensive understanding of these interactions between components and coordination in the gene expression of an organism [13].

Gene coexpression networks (GCNs) have proven particularly useful in identifying relationships and annotating functions of uncharacterized genes [14–16]. GCNs are commonly developed to identify phenotype-specific biomarkers that contain genes with functional associations based on coexpression relationships [17–19].

Although there is a rapid increase in available tools to analyze scRNA-seq data, no systematic pipeline comprehensively analyzes scRNA-seq, including constructing a gene coexpression network analysis. There are some accessible packages for GCNs, such as weighted gene coexpression network analysis (WGCNA), coexpression of RNA-seq data (Coseq), and coexpression modules identification tool (CEMiTool), petal, CoXpress, and coexpressed biological processes (CoP) [20]. However, they were initially designed to analyze microarray and bulk RNA-seq datasets [21,22]. Furthermore, these GCNs packages cannot be directly used for analyzing single cells because of the sparse data in the scRNA-seq data. To the best of our knowledge, there is no complete systematic pipeline including all the above analysis procedures.

Therefore, this study aims to illustrate a single-cell gene coexpression network analysis (scGENA) framework in a systematic pipeline to analyze scRNA-seq data. scGENA aims to implement a complete R software package using scRNA-seq data with a step-by-step guide for the entire analysis, including data preprocessing, differential gene expression, data imputation, construction of gene coexpression networks, and investigating key gene modules that enrich critical functions in diverse cell types.

2. Materials and Methods

2.1. Overview of scGENA

scGENA is a systematic pipeline for single-cell data analysis and contains five phases, as illustrated in Figure 1. *Phase 1* set up and preprocesses the scRNA-seq dataset to filter low-dimensional and noisy single-cell expression genes. *Phase 2* performs a differentially expressed (DE) genes analysis to determine which genes are expressed significantly different in different conditions. These genes can reveal biological information about the processes that are influenced by the conditions of interest. *Phase 3* applies the SAVER imputation method to estimate and replace dropout values in each gene cross cell's actual missing expression level, reducing technical differences while preserving biological variability across cells [23]. *Phase 4* constructs a coexpression network analysis to shed light on the transcriptional regulatory mechanisms underpinning numerous biological processes [24]. *Phase 5* performs further analyses, including a functional enrichment analysis, differential coexpression network analysis, and overlapping genes identification across different cell-types to better interpret the biological insights. scGENA is fully implemented in R and available at GitHub (https://github.com/zpliulab/scGENA) (accessed on 1 September 2020). The following subheadings describe the details of each phase of scGENA.
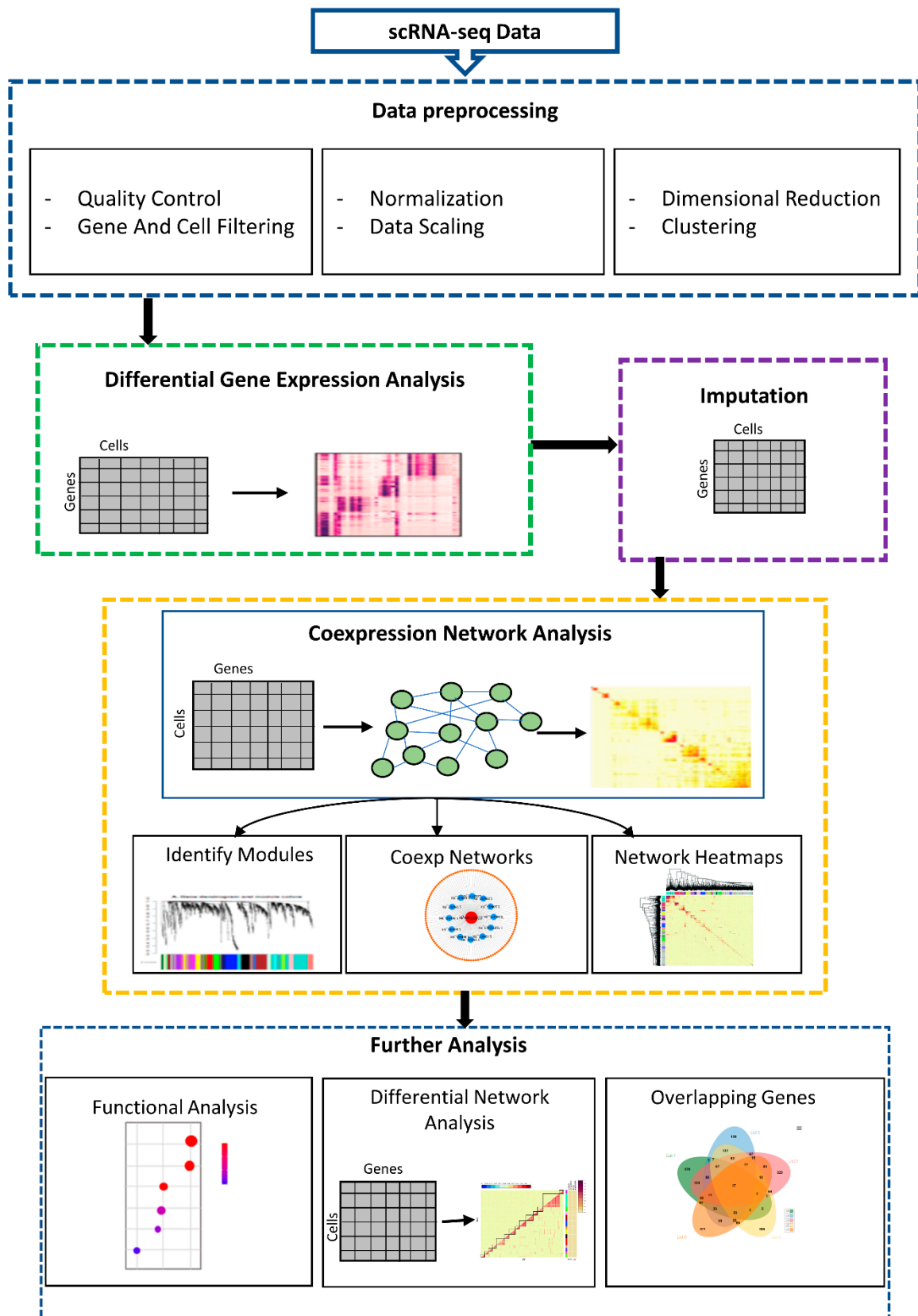
Figure 1. Flowchart of scGENA pipeline.

2.2. Preprocessing of scRNA-seq

In the first phase, scGENA sets up and preprocesses the scRNA-seq dataset. The single-cell data input is a matrix composed of genes as rows and cells as columns, containing the counts of gene expression in every cell.

This study implemented a human pancreas with a nondiabetic and type 2 diabetic disease dataset as case studies for illustration purposes. The dataset contained 1600 cells aggregated from 12 nondiabetic and 6 T2D organ donors of cell types of α-, β-, δ-, and PP-cells using the Fluidigm C1 cell-capturing process and 39,851 features (the human whole-genome sequencing including genes and ncRNA transcripts), as summarized in Table 1. We downloaded the data from the NCBI Gene Expression Omnibus (GEO) repository with accession ID GSE81608. After the data preparation, we used the Seurat method (https://satijalab.org/seurat/index.html (accessed on 1 September 2020)) for quality control, normalization, data exploration, and visualization of the preprocessing steps [25].

Table 1. Summary of single-cell data information in the proof-of-concept study.

GEO No.	Type of Cells	Cells	Features	Organism	Protocol	Ref.
GSE81608	α-	886	39,851	Homo sapiens	SMARTer	Xin et al., 2016 [26]
	β-	472				
	δ-	49				
	PP	85				

2.3. Differential Expression (DE) Analysis

Differential expression (DE) analysis is one of the most common tasks for scRNA-seq data. Although there are well-established techniques for such research in bulk RNA-seq data, tools for scRNA-seq data are still in the early stages [27]. We employed the R/Bioconductor MAST (model-based analysis of single-cell transcriptomics) package for this analysis based on empirical tests, which build two-part generalized linear models designed explicitly for bimodal and zero-inflated single-cell gene expression data [28,29]. MAST accounts for dropout events using a hurdle model while modeling variations in gene expression based on condition and technical variables. The hurdle model improves differential gene expression by summarizing differences between two groups with pairs of regression coefficients [28]. In this process, the highly differentiated genes were evaluated across the resulting four cell types and divided into clusters; the adaptive threshold was a cut-off value selected based on the gene's median expression value (in this case, we limited ourselves to genes that expressed in at least 0.25 within the cells). It also determined a single cluster's positive and negative expressed markers. We then selected the DE genes in each cell type to construct the gene coexpression networks and for further downstream analysis. In total, 2169 genes were selected as differentially expressed and presented in the Supplementary Materials Table S1.

2.4. Data Imputation

Due to the low transcript abundances in single cells, current scRNA sequencing technologies may fail to detect some gene expression. This can result in missing expressed gene values, known as a dropout event [30]. Dropouts can potentially cause significant bias in gene–gene correlations and other downstream analyses [31]. Recently, imputation methods have been developed to estimate actual expression levels directly. Here, we used the SAVER method in this pipeline because it imputed original zero values to actual values [32]. SAVER estimates expression levels by borrowing information across genes and applying a Bayesian technique. The main reason to use SAVER imputations is that it uses gene-to-gene relationships to impute the values of each gene expression level in each cell. It reduces technical variation while maintaining biological variability between cells [23].

2.5. Gene Coexpression Networks (GCNs) Analysis

The novel ensemble scGENA framework for constructing single-cell-based GCNs is based on combining multiple methods to establish a systematic R-based pipeline. Previous tools and packages for building GCNs were designed to analyze bulk RNA-seq and microarray data. Therefore, because single-cell data are intrinsically noisy and sparse, traditional measures (such as Pearson, Spearman, or cosine correlation) cannot be used effectively [33]. In contrast, mutual information (MI) has significant advantages over other measures because it may capture complex nonlinear and nonmonotonic interactions and represent the dynamics of groups or pairs of genes [34,35]. MI approaches are, therefore, often the preferred method for such a network inference analysis. We employed the *minet* package to discretize and compute the distance for the mutual information matrix as follows:

$$MIM_{ij} = I\left(X_i\,;X_j\right) \tag{1}$$

where i, j is the MI between X_i and X_j, and $X_i \,\epsilon\, \chi$, $i = 1,\ldots,n$, is a discrete variable representing the ith gene's expression level. Additionally, the empirical estimator is selected to estimate the amount of information shared by any pair of genes due to its ability to decrease the bias without affecting variance [36].

3. Results and Discussion

In this paper, we used the transcriptomic data of human pancreas cells from [26] to build an integrated systematic pipeline for a complete analysis of scRNA-seq data, including data exploration, quality control, normalization, dimensional reduction, a cell clustering, differential genes analysis, a gene coexpression network analysis, and a further downstream analysis.

3.1. Data Preprocessing

We first explored the data library size and the distributions of genes within different cells for 1600 samples of human donors α-, β-, δ- and PP cells from nondiabetic and T2D organ in the preprocessing phase, as shown in Figure 2A,B. This step is necessary because it enables researchers to understand the dataset comprehensively. Next, we employed Seurat for data quality control (QC) and visualization. Seurat aims to detect and evaluate heterogeneous sources of single-cell transcriptomic measurements and dataset integrations [25]. The QC was performed to exclude cells with <1500 or >10,000 expressed genes and with >15% of unique molecular identifiers (UMIs), as well as the contaminated cells. Therefore, the number of remaining single cells was 1472, with 29,043 and variable genes of 2000 selected for the downstream analysis by calculating a group of genes in the dataset with a significant level of cell-to-cell variation (using FindVariableFeatures, as a Seurat function).

The quality control, variables, and nonvariable genes of single-cell data are displayed in Figure 2C,D. The data are then logarithm-normalized by employing the LogNormalize technique, which is regarded as a global normalization method that divides the gene counts for a single cell and then multiplies by the scale factor. The result was then transformed with the natural log using log(x + 1) to account for zero counts. The normalization process is essential for uncovering a dataset's underlying biological heterogeneity. The normalization approaches are also important to prevent noise and bias and are necessary for dimensionality reduction [37].
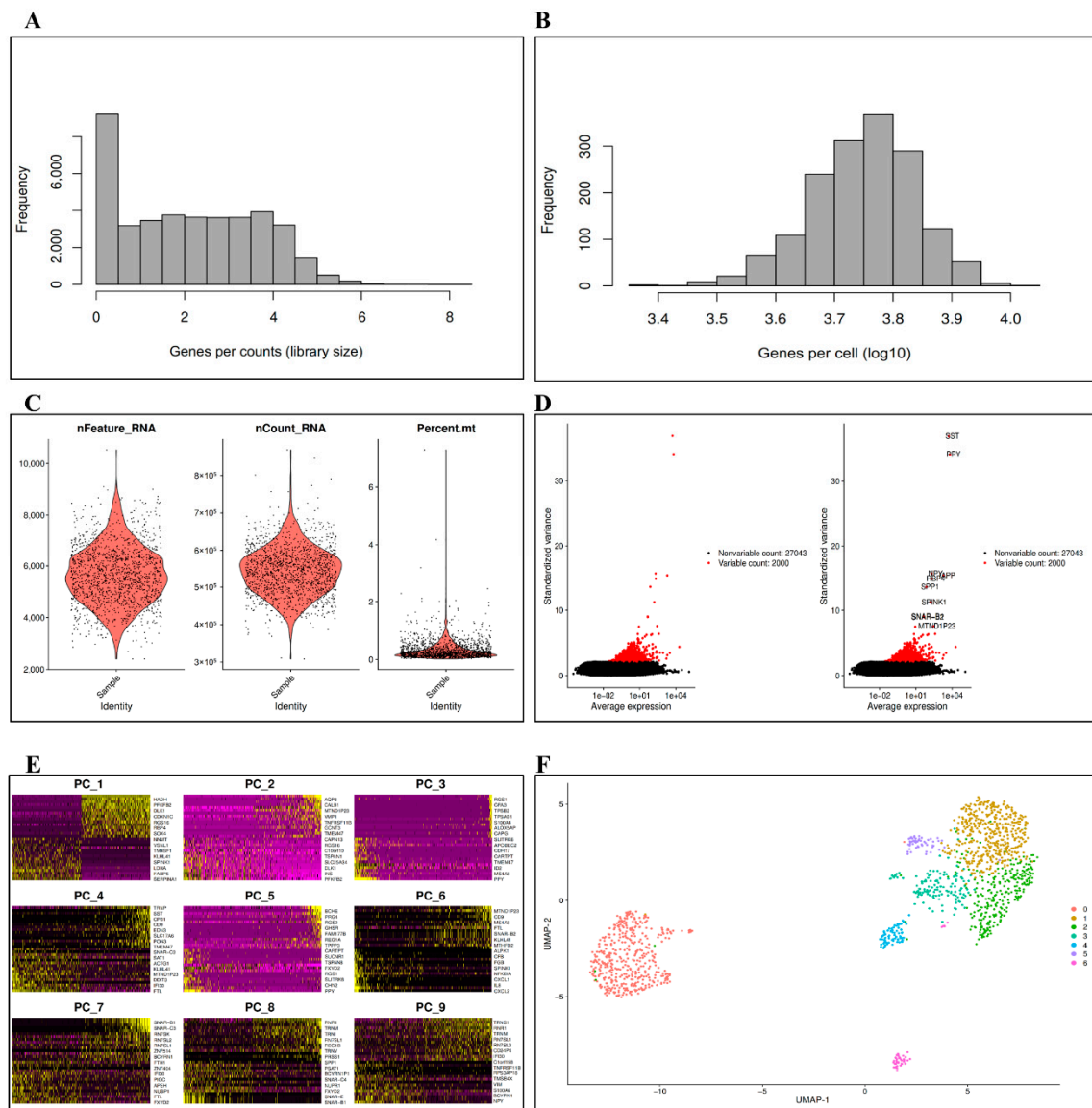
Figure 2. Preprocessing of scRNA-seq data. (**A**) Measurement of library size for genes per counts; (**B**) Genes distributions among cells; (**C**) Quality control; (**D**) Variable genes in the dataset; (**E**) Dimensionality reduction heatmap PCA; (**F**) UMAP clustering of the dataset.

These upstream analysis procedures, such as quality control filtering and normalization, can significantly affect clustering and trajectory inference. Following the completion of the preprocessing procedures, the subsequent analytical phases, which included dimensionality reduction, clustering, and trajectory inference, focused on discovering patterns in the data that gave biological insights. Dimensionality reduction reduces the dataset to a more compact and potentially interpretable representation that enables researchers to capture the key biological axes of variation and enhances clustering and trajectory inference performance [38]. Figure 2E illustrates the dimensionality reduction plotted by a principal component analysis (PCA) heatmap and gene clustering across cells by UMAP (uniform manifold approximation and projection). The PCA, which provides a linear combination of genes that best reflects the variation in the data, is the most often used dimensionality reduction approach for scRNA-seq analysis. The PCA's ability to lessen data dimensionality while identifying the dimensions with the most variance makes it a useful dimensionality reduction method prior to clustering [39]. Figure 2F visualizes the gene clusters in the dataset.

3.2. DEs and Imputation

The field of biomedical research has entered the omics era with the emergence of high-throughput technology, and research tools focused on bulk RNA-seq data no longer meet this development's objectives. However, dealing with such a massive volume of data presents significant difficulties in extracting and analyzing data. Sequencing data analysis frequently yields a list of genes that are differentially expressed. However, it is challenging for many researchers to connect the vast number of differential genes or proteins to a biological event to be examined. We need to group genes with similar expressions and associate them with their biological phenotypes using DE tools to study the functional enrichment analysis for these genes. In our pipeline, we used Seurat's MAST method to perform a differential gene expression analysis based on a generalized linear model [28]. We selected the differentially expressed genes based on the criteria previously mentioned (see Section 2.3), which necessitate a gene being identified at the cut-off threshold in either of the two cell groups and being differentially expressed (on average) by some proportion between the two different cell groups, see Supplementary Materials Table S1. In a study [40], MAST performed as the best single-cell DE testing technique, outperforming bulk and single-cell approaches in a small-scale comparison on a benchmark dataset [41,42]. Figure 3A presents the DE genes based on the number of genes detected in each cell, while Figure 3B shows how to classify cells depending on the presence of mitochondrial genes. Out of these DE genes, we selected the top 10 from each group for further analysis based on the logarithm fold-change threshold of the average gene expressions. These genes are visualized in Figure 3C,D. We then clustered all the DE genes for each cell type using UMAP and heatmap plots, as depicted in Figure 3E,F.

An imputation method was performed to impute the missing gene expression values in the count matrix to reduce the effect of noise and dropout events. In this pipeline, we used the SAVER method, which showed good performance for imputing most of the missing values. Figure 4 compares the difference before and after imputation for the selected 2000 genes from the DE step across 50 samples. It dramatically imputed gene expression compared to original and differentially expressed genes. As can be noticed in Figure 4, the level of nonzero expression values changed from 28% to 94.6% after imputation. Therefore, we used these imputed data for our downstream analysis.

3.3. Gene Coexpression Networks Analysis

GCN is an approach for inferring gene module function and gene-disease interactions from genome-wide gene expression. This method builds networks of genes with a tendency to coactivate across a group of samples and then interrogates and analyzes this network [43]. GCNs can be employed for various purposes, including candidate disease gene selection, functional gene identification, and gene regulation discovery.

Since WGCNA was initially designed to analyze bulk RNA data [44], its performance on single-cell data is limited because of the inherent sparsity of scRNA-seq data [21,22]. To resolve this, our pipeline has a function that aggregates the transcriptionally similar cells into a pseudobulk cell type before running WGCNA in our framework. Figure 5A shows the pseudotime aggregation for the four cell types of this dataset. However, because of the MI symmetry characteristic, it relies on the pseudotime input to infer the GCNs [45]. Then, we utilized a signed consensus network based on the WGCNA algorithm for a particular cell type (see [46,47]), computing component-wise values for topological overlap in the dataset. Biweighted midcorrelations (defined in WGCNA as *bicor*) were computed for each pair of genes, followed by a signed similarity matrix. The similarity between genes in the signed network showed the sign of the connection of their expression patterns.
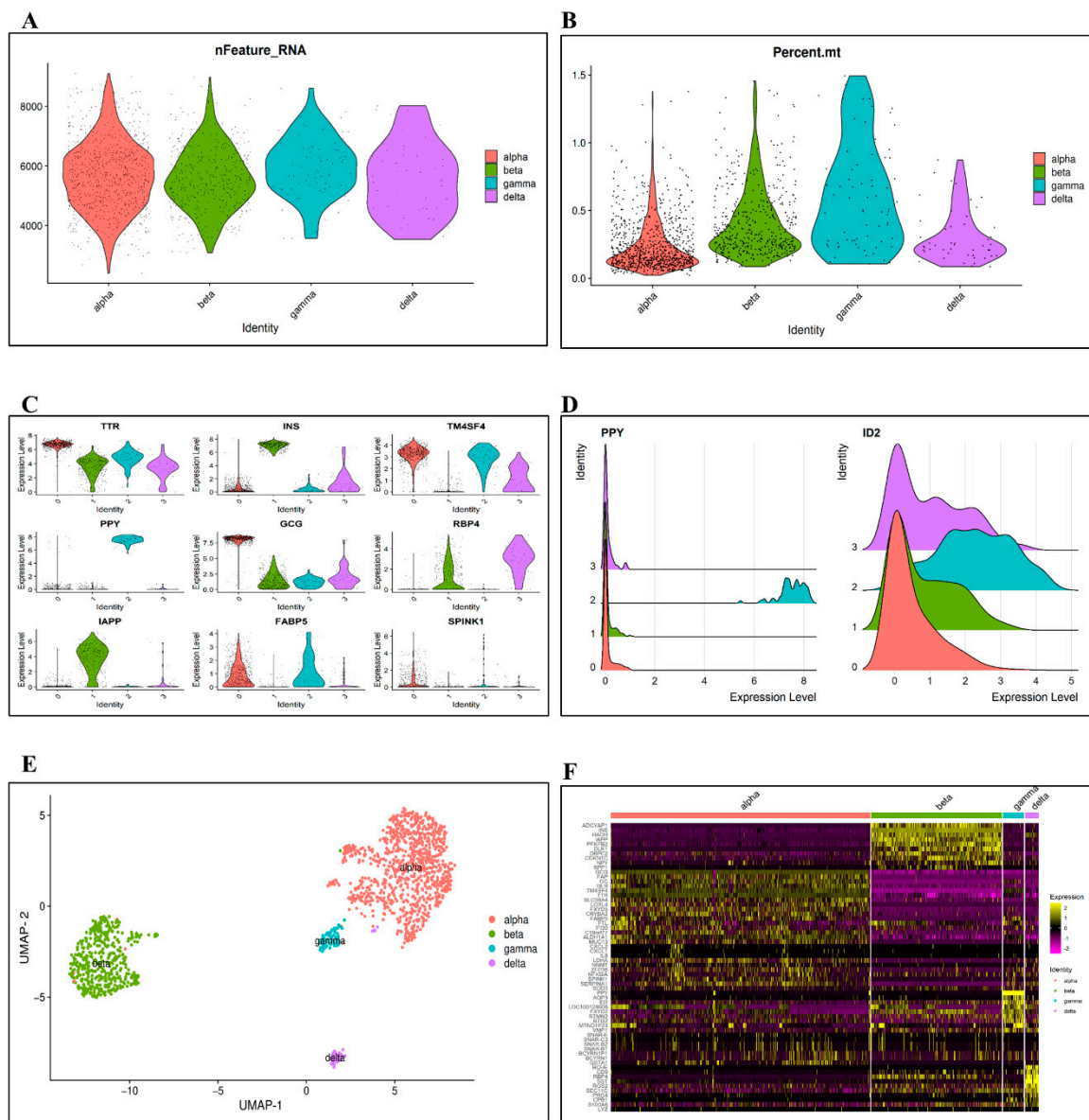
Figure 3. Differential expression analysis in scGENA. (**A**) Gene clustering and distribution in all data cell-types. (**B**) Gene distribution based on the percentage of mitochondria. (**C,D**) Highly differentiated genes. (**E**) Cell-types clustering using UMAP. (**F**) Heatmap showing the level of gene expression in all cell-types.

On an exponential scale, the signed similarity matrix was then boosted to power β, varying the cell types to accentuate strong correlations and lessen the emphasis on weak correlations. The selection of the power β or soft thresholding, as called in the WGCNA package, was based on the number of data samples (see [46]). The resulted adjacency matrix was then turned into a topological overlap matrix. Figure 5B,C visualize the hierarchal clustering tree (dendrogram) and gene coexpression networks of β-cells data. We have selected only 25 genes for each module to visualize a clear network. Therefore, modules were formed by utilizing module-cutting criteria such as a minimum module size of 100 genes, with a deepSplit score of 4, and a correlation threshold (mergCutHeight) of 0.2 that can be used to merge modules. Modules having a correlation larger than 0.8 were combined. These parameters were selected to construct the block-wise networks based on the samples' data sets. As a result, four coexpression modules were significantly correlated in β-cells; therefore, these modules' genes were used for the functional enrichment analysis.

Figure 5D depicts the gene networks for β-cells module genes using a heatmap. The heatmap of each row and column correspond to a single gene and the relationship between genes. Module colors and gene dendrograms are also plotted on the left and top sides of the heatmap. For the construction of gene coexpression networks for the other three cell types and their identified modules, see Supplementary Material Figures S1–S3.
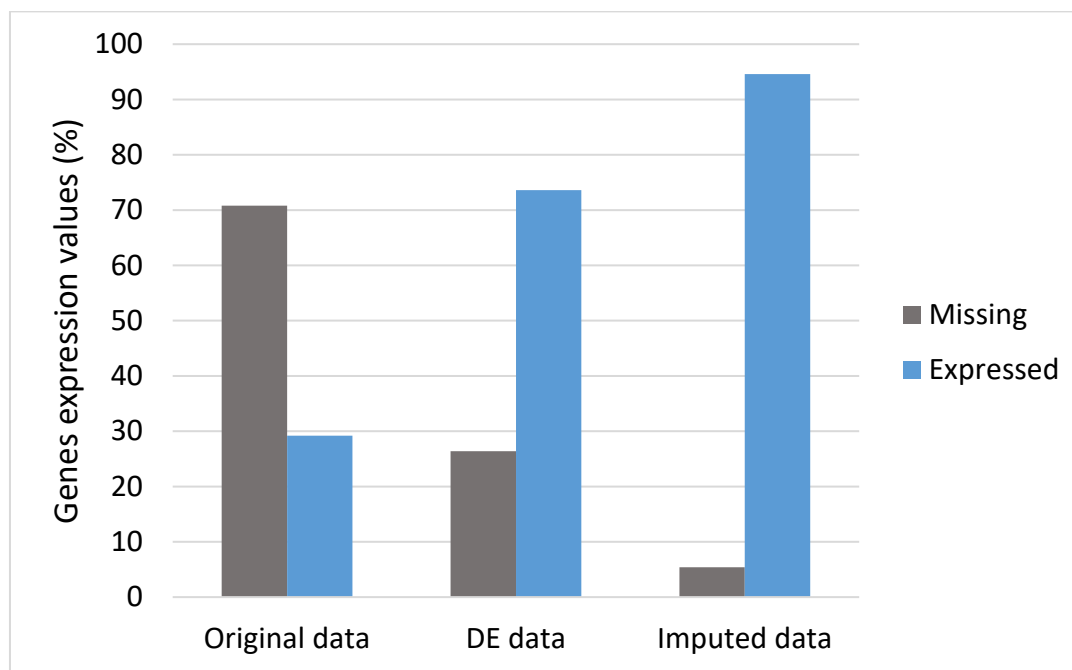

Figure 4. Comparing the data before and after performing imputation.

3.4. Further Analysis

In this substep, we performed a functional enrichment analysis, differential coexpression analysis, and the identification of overlapping genes across the cell types for co-expression network genes in each module to investigate the biological insights. A functional enrichment analysis is specifically used to identify the gene sets linked with a biological process or molecular function in order to interpret the underlying physiological insights and reveal the dysfunctional mechanisms. Therefore, we carried out a Gene Ontology (GO) terms enrichment using the R package ClusterProfiler, which included the following three categories: biological process (BP), cellular component (CC) and molecular function (MF). In this analysis, we demonstrated the top ten terms within each category that were considerably enriched significantly in M1, as illustrated in Figure 6A. A subset of each enriched term was chosen and shown as a network plot, with terms with similarity > 0.3 linked by edges with the best p-values from each cluster to further understand the relations among the terms. All network modules' GO terms listed are presented in Supplementary Materials Tables S2–S12.

The essential GO terms in the three categories were, respectively, protein folding (GO:0006457) for BP, the cell-substrate junction (GO:0030055) for CC, and unfolded protein binding (GO:0051082) and protein folding chaperone (GO:0044183) for MF. The study of [48] emphasizes that high inflammatory cytokines might induce the accumulation of unfolded or misfolded proteins, i.e., endoplasmic reticulum stress in diabetic pancreatic islets. In the CC category functional enrichment investigation, Zhang et al., 2022 [49] discovered that the upregulated differentially expressed genes concerning acute pancreatitis were connected with the cell–substrate junction. On the other hand, chaperones' adaptive unfolded protein response signaling maintained endoplasmic reticulum protein folding equilibrium in healthy beta cells, according to Yong et al., 2021 [50]. Figure 6A shows that several of our targeted genes included protein folding terms. We assumed that the BP

and MF protein-folding-related GO terms were relevant in pancreatic or diabetes disease. Furthermore, several of our identified genes were enriched for the BP and MF categories, with only a few enriched for CC. The enhanced GO terms may be crucial in revealing the progression of diabetes disease.
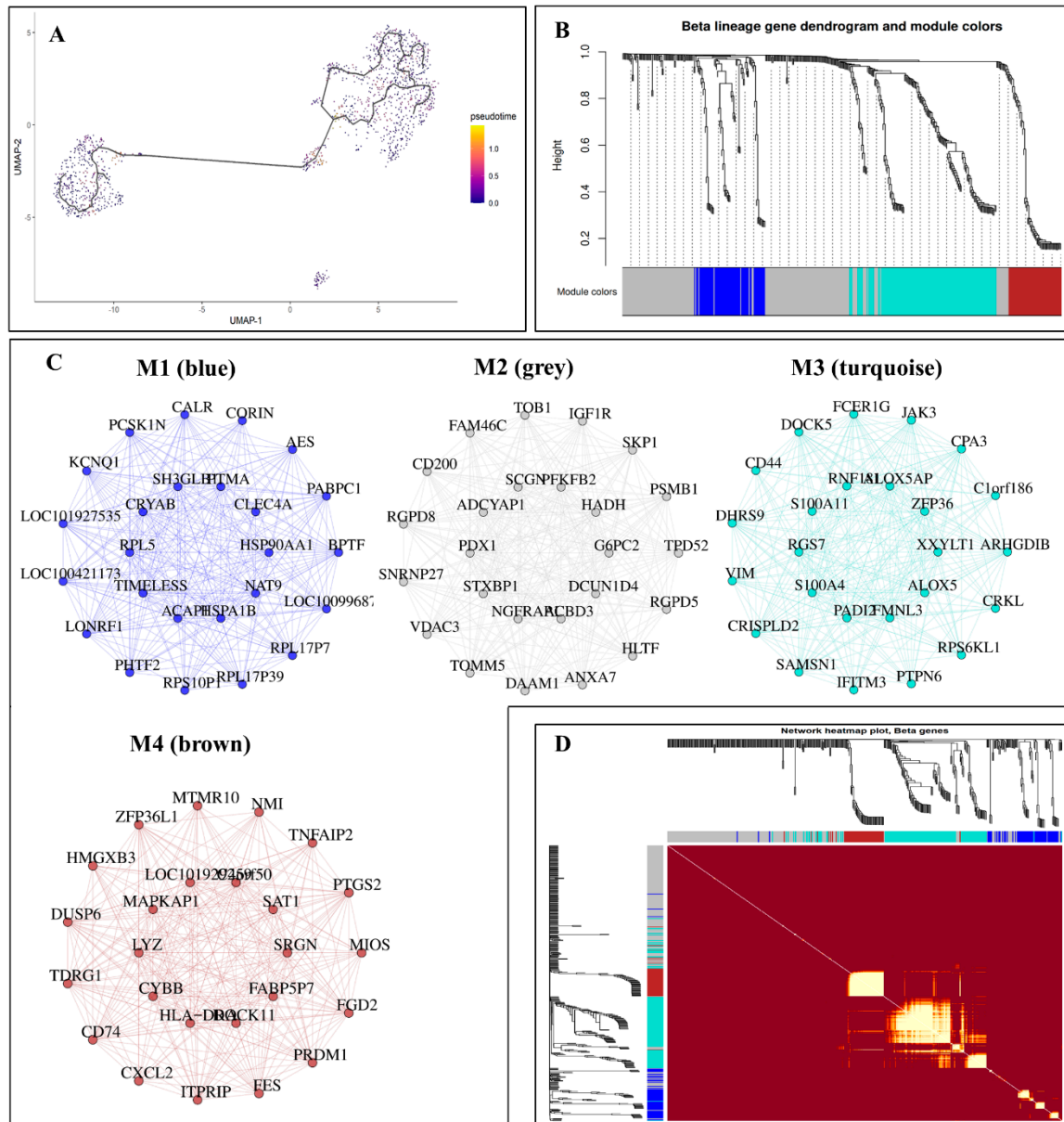

Figure 5. Coexpression analysis in scGENA. (**A**) Pseudotime trajectory of the cell types; (**B**) Dendrogram and modules colors clustering for β-cells; (**C**) Gene coexpression networks for the four modules in β-cells (M1, M2, M3, and M4); (**D**) Coexpression network heatmap.

Once the coexpression gene modules are defined, researchers can utilize these findings to perform several analyses, such as a differential coexpression network analysis. Consequently, we used MODA for the differential coexpression network analysis [51], which can analyze the different networks of each cell type of data. In order to reduce the number of modules as much as possible, the gene was sampled; in other words, only differential genes were selected as the input, as shown for the beta cells in Figure 6B. The other cell types' results are shown in the Supplementary Materials Figures S4–S6. Last but not least, we identified the consistent and overlapped genes across all four cell types using

the *intervene* R package [52] to plot a Venn diagram, as shown in Figure 6C. It is observed that beta–alpha cells share many genes compared to other cell types. Identifying these overlapping genes will assist the researcher in learning more about human islet cell biology and pathophysiology, particularly the alpha-cell-derived paracrine signals' role in normal beta-cell survival and function [53]. The experimental analysis in the study [53] on human and mouse islet cells indicated that the crucial variable was not necessarily a different species but the differing alpha–beta-cell ratio. The set point of glucose homeostasis in the body appears to be determined by paracrine interactions between alpha and beta cells. It is still unclear how the various islet designs found in different animals connect to their glycemic set points.
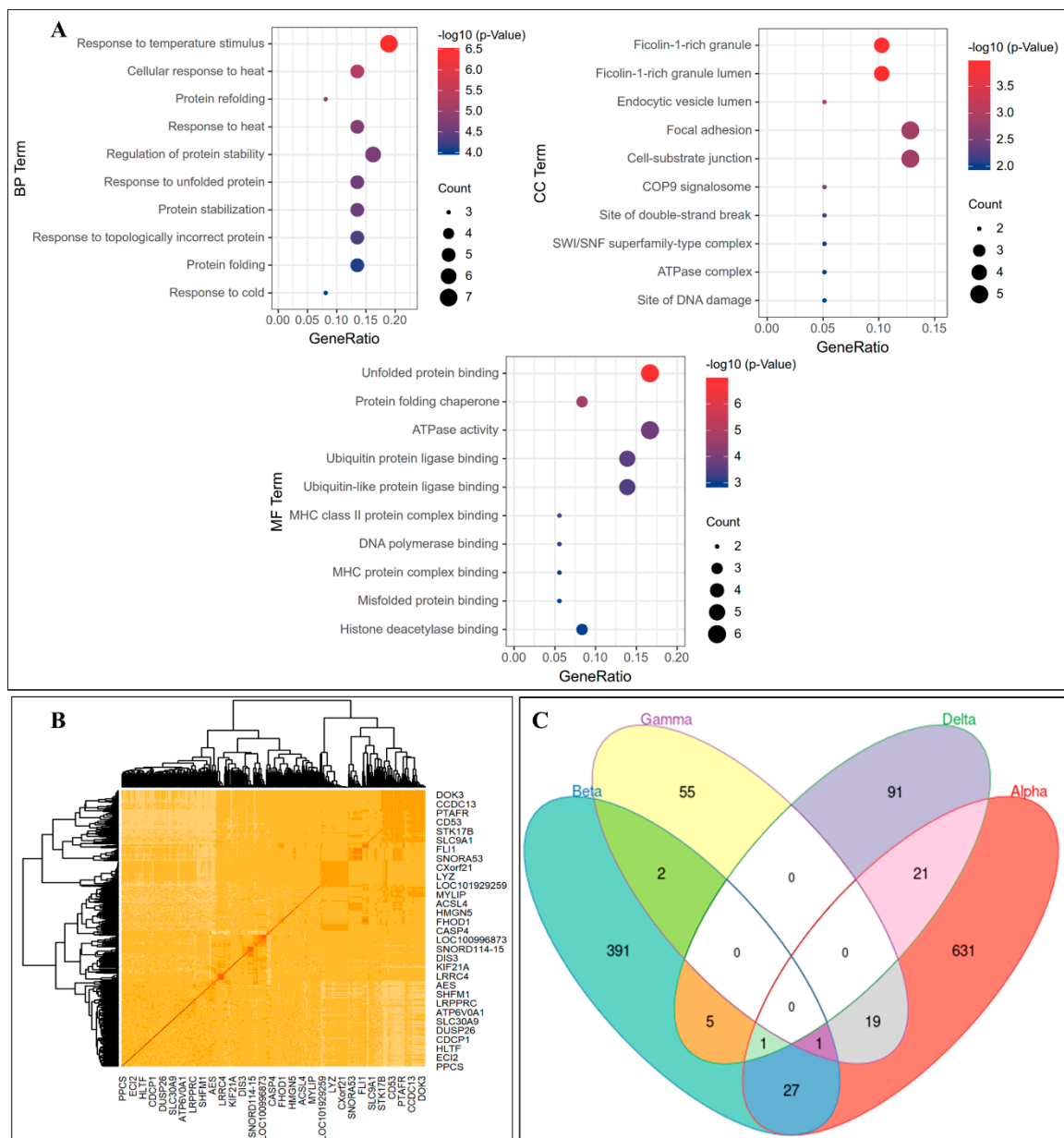
Figure 6. Further analyses in scGENA. (**A**) The significantly enriched GO terms for the β-cells for BP, CC, and MF. (**B**) Differential coexpression heatmap by MODA. (**C**) Genes overlapping among the four cell types by the Venn diagram.

Furthermore, suppose specific genes in scRNA-seq data consistently exhibit identical expression changes in biological processes or various tissues. In that case, we suspect these

genes are functionally associated and may be categorized as a module. We can utilize the findings of the gene modules to perform many different analysis tasks.

4. Conclusions

Advances in scRNA-seq technology have resulted in the generation of datasets with increasing size and complexity. As a result, an ecosystem of computational approaches has been developed to address the issues associated with evaluating big datasets. In this work, we proposed an integrative pipeline scGENA for a complete single-cell gene coexpression analysis based on scRNA-seq data. scGENA integrates numerous models to comprehensively perform several steps: preprocessing, dimensionality reduction, clustering, differential genes identification, imputation, and network construction analysis. Because scRNA-seq data are often sparse and noisy, it is challenging to build coexpression and differential coexpression networks. We showed how to use the scGENA framework to construct and analyze coexpression networks using scRNA-seq data in an integrative and reliable way. The results demonstrated that the scRNA-seq-based method was good and valuable for identifying cell types and revealing biological insights by analyzing gene coexpression patterns.

References

1. Hwang, B.; Lee, J.H.; Bang, D. Single-Cell RNA Sequencing Technologies and Bioinformatics Pipelines. *Exp. Mol. Med.* **2018**, *50*, 96. [CrossRef]
2. Stark, R.; Grzelak, M.; Hadfield, J. RNA Sequencing: The Teenage Years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [CrossRef] [PubMed]
3. Lister, R.; O'Malley, R.C.; Tonti-Filippini, J.; Gregory, B.D.; Berry, C.C.; Millar, A.H.; Ecker, J.R. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* **2008**, *133*, 523–536. [CrossRef] [PubMed]
4. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtukova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **2008**, *456*, 470–476. [CrossRef] [PubMed]
5. Byrne, A.; Beaudin, A.E.; Olsen, H.E.; Jain, M.; Cole, C.; Palmer, T.; DuBois, R.M.; Forsberg, E.C.; Akeson, M.; Vollmers, C. Nanopore Long-Read RNAseq Reveals Widespread Transcriptional Variation among the Surface Receptors of Individual B Cells. *Nat. Commun.* **2017**, *8*, 16027. [CrossRef]

6. Schumacher, A.; Rookmaaker, M.B.; Joles, J.A.; Kramann, R.; Nguyen, T.Q.; van Griensven, M.; LaPointe, V.L.S. Defining the Variety of Cell Types in Developing and Adult Human Kidneys by Single-Cell RNA Sequencing. *NPJ Regen. Med.* **2021**, *6*, 45. [CrossRef] [PubMed]

7. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A. MRNA-Seq Whole-Transcriptome Analysis of a Single Cell. *Nat. Methods* **2009**, *6*, 377–382. [CrossRef]

8. Eberwine, J.; Sul, J.-Y.; Bartfai, T.; Kim, J. The Promise of Single-Cell Sequencing. *Nat. Methods* **2014**, *11*, 25–27. [CrossRef] [PubMed]

9. Tirosh, I.; Suvà, M.L. Deciphering Human Tumor Biology by Single-Cell Expression Profiling. *Annu. Rev. Cancer Biol.* **2019**, *3*, 151–166. [CrossRef]

10. Saelens, W.; Cannoodt, R.; Todorov, H.; Saeys, Y. A Comparison of Single-Cell Trajectory Inference Methods. *Nat. Biotechnol.* **2019**, *37*, 547–554. [CrossRef]

11. Wang, T.; Nabavi, S. SigEMD: A Powerful Method for Differential Gene Expression Analysis in Single-Cell RNA Sequencing Data. *Methods* **2018**, *145*, 25–32. [CrossRef] [PubMed]

12. Elowitz, M.B.; Levine, A.J.; Siggia, E.D.; Swain, P.S. Stochastic Gene Expression in a Single Cell. *Science* **2002**, *297*, 1183–1186. [CrossRef] [PubMed]

13. Komili, S.; Silver, P.A. Coupling and Coordination in Gene Expression Processes: A Systems Biology View. *Nat. Rev. Genet.* **2008**, *9*, 38–48. [CrossRef]

14. Furlong, L.I. Human Diseases through the Lens of Network Biology. *Trends Genet.* **2013**, *29*, 150–159. [CrossRef] [PubMed]

15. Gysi, D.M.; de Fragoso, T.M.; Zebardast, F.; Bertoli, W.; Busskamp, V.; Almaas, E.; Nowick, K. Whole Transcriptomic Network Analysis Using Co-Expression Differential Network Analysis (CoDiNA). *PLoS ONE* **2020**, *15*, e0240523. [CrossRef]

16. Wang, J.; Xia, S.; Arand, B.; Zhu, H.; Machiraju, R.; Huang, K.; Ji, H.; Qian, J. Single-Cell Co-Expression Analysis Reveals Distinct Functional Modules, Co-Regulation Mechanisms and Clinical Outcomes. *PLoS Comput. Biol.* **2016**, *12*, e1004892. [CrossRef] [PubMed]

17. Chen, X.; Hu, L.; Wang, Y.; Sun, W.; Yang, C. Single Cell Gene Co-Expression Network Reveals FECH/CROT Signature as a Prognostic Marker. *Cells* **2019**, *8*, 698. [CrossRef]

18. Elo, L.L.; Järvenpää, H.; Orešič, M.; Lahesmaa, R.; Aittokallio, T. Systematic Construction of Gene Coexpression Networks with Applications to Human T Helper Cell Differentiation Process. *Bioinformatics* **2007**, *23*, 2096–2103. [CrossRef]

19. Reverter, A.; Chan, E.K.F. Combining Partial Correlation and an Information Theory Approach to the Reversed Engineering of Gene Co-Expression Networks. *Bioinformatics* **2008**, *24*, 2491–2497. [CrossRef]

20. Cheng, C.W.; Beech, D.J.; Wheatcroft, S.B. Advantages of CEMiTool for Gene Co-Expression Analysis of RNA-Seq Data. *Comput. Biol. Med.* **2020**, *125*, 103975. [CrossRef]

21. Rexach, J.E.; Polioudakis, D.; Yin, A.; Swarup, V.; Chang, T.S.; Nguyen, T.; Sarkar, A.; Chen, L.; Huang, J.; Lin, L.-C.; et al. Tau Pathology Drives Dementia Risk-Associated Gene Networks toward Chronic Inflammatory States and Immunosuppression. *Cell Rep.* **2020**, *33*, 108398. [CrossRef] [PubMed]

22. Zhang, B.; Gaiteri, C.; Bodea, L.-G.; Wang, Z.; McElwee, J.; Podtelezhnikov, A.A.; Zhang, C.; Xie, T.; Tran, L.; Dobrin, R.; et al. Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. *Cell* **2013**, *153*, 707–720. [CrossRef]

23. Huang, M.; Wang, J.; Torre, E.; Dueck, H.; Shaffer, S.; Bonasio, R.; Murray, J.I.; Raj, A.; Li, M.; Zhang, N.R. SAVER: Gene Expression Recovery for Single-Cell RNA Sequencing. *Nat. Methods* **2018**, *15*, 539–542. [CrossRef] [PubMed]

24. Li, W.V.; Li, Y. ScLink: Inferring Sparse Gene Co-Expression Networks from Single-Cell Expression Data. *Genom. Proteom. Bioinform.* **2021**, *19*, 475–492. [CrossRef]

25. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M.; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* **2021**, *184*, 3573–3587.e29. [CrossRef] [PubMed]

26. Xin, Y.; Kim, J.; Okamoto, H.; Ni, M.; Wei, Y.; Adler, C.; Murphy, A.J.; Yancopoulos, G.D.; Lin, C.; Gromada, J. RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab.* **2016**, *24*, 608–615. [CrossRef] [PubMed]

27. Soneson, C.; Robinson, M.D. Bias, Robustness and Scalability in Single-Cell Differential Expression Analysis. *Nat. Methods* **2018**, *15*, 255–261. [CrossRef]

28. Finak, G.; McDavid, A.; Yajima, M.; Deng, J.; Gersuk, V.; Shalek, A.K.; Slichter, C.K.; Miller, H.W.; McElrath, M.J.; Prlic, M.; et al. MAST: A Flexible Statistical Framework for Assessing Transcriptional Changes and Characterizing Heterogeneity in Single-Cell RNA Sequencing Data. *Genome Biol.* **2015**, *16*, 278. [CrossRef] [PubMed]

29. McDavid, A.; Finak, G.; Yajima, M. MAST: Model-Based Analysis of Single Cell Transcriptomics. R Package Version 1.22.0. 2022. Available online: https://github.com/RGLab/MAST/ (accessed on 1 September 2020).

30. Li, X.; Liu, L.; Goodall, G.J.; Schreiber, A.; Xu, T.; Li, J.; Le, T.D. A Novel Single-Cell Based Method for Breast Cancer Prognosis. *PLoS Comput. Biol.* **2020**, *16*, e1008133. [CrossRef]

31. van Dijk, D.; Sharma, R.; Nainys, J.; Yim, K.; Kathail, P.; Carr, A.J.; Burdziak, C.; Moon, K.R.; Chaffer, C.L.; Pattabiraman, D.; et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **2018**, *174*, 716–729.e27. [CrossRef]

32. Zhang, L.; Zhang, S. Comparison of Computational Methods for Imputing Single-Cell RNA-Sequencing Data. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2017**, *17*, 376–389. [CrossRef]

33. Iacono, G.; Massoni-Badosa, R.; Heyn, H. Single-Cell Transcriptomics Unveils Gene Regulatory Network Plasticity. *Genome Biol.* **2019**, *20*, 110. [CrossRef]

34. Liu, Z.-P. Quantifying Gene Regulatory Relationships with Association Measures: A Comparative Study. *Front. Genet.* **2017**, *8*, 96. [CrossRef] [PubMed]

35. Mc Mahon, S.S.; Lenive, O.; Filippi, S.; Stumpf, M.P.H. Information Processing by Simple Molecular Motifs and Susceptibility to Noise. *J. R. Soc. Interface* **2015**, *12*, 20150597. [CrossRef] [PubMed]

36. Meyer, P.E.; Lafitte, F.; Bontempi, G. Minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. *BMC Bioinform.* **2008**, *9*, 461. [CrossRef] [PubMed]

37. Lytal, N.; Ran, D.; An, L. Normalization Methods on Single-Cell RNA-Seq Data: An Empirical Survey. *Front. Genet.* **2020**, *11*, 41. [CrossRef]

38. Kiselev, V.Y.; Andrews, T.S.; Hemberg, M. Challenges in Unsupervised Clustering of Single-Cell RNA-Seq Data. *Nat. Rev. Genet.* **2019**, *20*, 273–282. [CrossRef]

39. Abdi, H.; Williams, L.J. Principal Component Analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]

40. Duò, A.; Robinson, M.D.; Soneson, C. A Systematic Performance Evaluation of Clustering Methods for Single-Cell RNA-Seq Data. *F1000 Research* **2020**, *7*, 1141. [CrossRef]

41. Luecken, M.D.; Theis, F.J. Current Best Practices in Single-Cell RNA-Seq Analysis: A Tutorial. *Mol. Syst. Biol.* **2019**, *15*, e8746. [CrossRef]

42. Vieth, B.; Ziegenhain, C.; Parekh, S.; Enard, W.; Hellmann, I. PowsimR: Power Analysis for Bulk and Single Cell RNA-Seq Experiments. *Bioinformatics* **2017**, *33*, 3486–3488. [CrossRef] [PubMed]

43. van Dam, S.; Võsa, U.; van der Graaf, A.; Franke, L.; de Magalhães, J.P. Gene Co-Expression Analysis for Functional Classification and Gene–Disease Predictions. *Brief. Bioinform.* **2017**, *19*, 575–592. [CrossRef] [PubMed]

44. Langfelder, P.; Horvath, S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]

45. Zeng, Y.; Yan, X.; Liang, Z.; Zheng, R.; Li, M. MKG: A Mutual Information Based Method to Infer Single Cell Gene Regulatory Network. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 223–228.

46. WGCNA: R Package for Performing Weighted Gene Co-Expression Network Analysis. Available online: https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/ (accessed on 18 July 2022).

47. Morabito, S.; Miyoshi, E.; Michael, N.; Swarup, V. Integrative Genomics Approach Identifies Conserved Transcriptomic Networks in Alzheimer's Disease. *Hum. Mol. Genet.* **2020**, *29*, 2899–2919. [CrossRef]

48. Wang, M.; Kaufman, R.J. Protein Misfolding in the Endoplasmic Reticulum as a Conduit to Human Disease. *Nature* **2016**, *529*, 326–335. [CrossRef] [PubMed]

49. Zhang, S.; Liang, Z.; Xiang, X.; Liu, L.; Yang, H.; Tang, G. Identification and Validation of Hub Genes in Acute Pancreatitis and Hypertriglyceridemia. *Diabetes Metab. Syndr. Obes.* **2022**, *15*, 559–577. [CrossRef] [PubMed]

50. Yong, J.; Johnson, J.D.; Arvan, P.; Han, J.; Kaufman, R.J. Therapeutic Opportunities for Pancreatic β-Cell ER Stress in Diabetes Mellitus. *Nat. Rev. Endocrinol.* **2021**, *17*, 455–467. [CrossRef] [PubMed]

51. Li, D.; Brown, J.B.; Orsini, L.; Pan, Z.; Hu, G.; He, S. MODA: MODA: MOdule Differential Analysis for Weighted Gene Co-Expression Network. R Package Version 1.22.0. 2022. Available online: https://doi.org/10.48550/arXiv.1605.04739 (accessed on 1 September 2020). [CrossRef]

52. Khan, A.; Mathelier, A. Intervene: A Tool for Intersection and Visualization of Multiple Gene or Genomic Region Sets. *BMC Bioinform.* **2017**, *18*, 287. [CrossRef]

53. Moede, T.; Leibiger, I.B.; Berggren, P.-O. Alpha Cell Regulation of Beta Cell Function. *Diabetologia* **2020**, *63*, 2064–2075. [CrossRef] [PubMed]