

## Research Article

# Identification of the Hub Genes Involved in Stem Cell Treatment for Intervertebral Disc Degeneration: A Conjoint Analysis of Single-Cell and Machine Learning

Jianfeng Chen <sup>1</sup>, Fuwei Qi,<sup>2</sup> Guanshen Li,<sup>1</sup> Qiaosong Deng,<sup>1</sup> Chenlin Zhang,<sup>3</sup> Xiaojun Li,<sup>1</sup> and Yafeng Zhang <sup>1</sup>

<sup>1</sup>Department of Spine, Wuxi Hospital Affiliated to Nanjing University of Chinese Medicine, Wuxi 214000, China

<sup>2</sup>Department of Anesthesiology, The First People's Hospital of Taicang, Taicang Affiliated Hospital of Soochow University, Jiangsu Province, 215413 Suzhou City, China

<sup>3</sup>Department of Orthopaedics, Qidong Hospital of Chinese Medicine, Nantong, Jiangsu 226200, China

Correspondence should be addressed to Jianfeng Chen; wxzy009@njucm.edu.cn and Yafeng Zhang; orthozhang@sina.com

Received 5 August 2022; Revised 8 October 2022; Accepted 17 October 2022; Published 24 January 2023

Academic Editor: Fanglin Guan

Copyright © 2023 Jianfeng Chen 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.

Intervertebral disc degeneration (IDD), which is distinguished by a variety of pathologic alterations, is the major cause of low back pain (LBP). Nonetheless, preventative measures or therapies that may delay IDD are scarcely available. In this study, we sought to identify new diagnostic biological markers for IDD. In this first-of-a-kind study combining machine learning, stem cell treatment samples and single-cell sequencing data were collected. Differentially expressed genes (DEGs) were detected from the treatment group and clusters. To filter potential markers, support vector machine analysis and LASSO were performed. LPTM5 was found to be the hub gene for IDD. In addition, the results of single-cell sequencing demonstrated the critical function of stem cells in IDD. Finally, we found that aging is significantly associated with the rate of stem cells. In general, our results may offer fresh insights that may be used in the investigation of innovative markers for diagnosing IDD. The critical genes identified by the machine learning algorithm could provide new perspectives on IDD.

## 1. Introduction

Intervertebral disc degeneration (IDD) leads to low back pain (LBP), which places a significant burden on individuals and society alike [1]. Chronic low back pain is a major feature of spinal degenerative diseases, the prevalence of which has substantially increased with the burgeoning aging population. Spinal degenerative disease negatively impacts patients' quality of life and intensifies the financial load on their families and society as a whole [2]. In addition to lumbar disc herniation and discogenic low back pain, degenerative disc diseases (DDD) result from intervertebral disc degeneration [3]. Currently, most patients obtain pain relief from rest or other forms of conservative therapy, in addition to receiving steroids, local anesthetic, and other blocking

medications [4]. When other treatment options fail to alleviate symptoms and enhance the quality of life, clinicians often turn to surgical procedures as the last resort. Even though surgery can alleviate pain, it is not capable of replacing the lost nucleus pulposus (NP) cells or reversing the pathologic condition of the intervertebral disc (IVD). Furthermore, surgery is associated with the risk of intraoperative and postoperative complications.

Studies in both preclinical and clinical settings have tested a variety of advanced and novel treatments for degenerative IVD, including the use of mesenchymal stem cells (MSCs), gene transfection, biologics, and growth factors [5]. The capacity of cell-based therapy, both autologous and allogeneic, to target a wide variety of pathways that contribute to IVDD is one reason it is becoming an increasingly

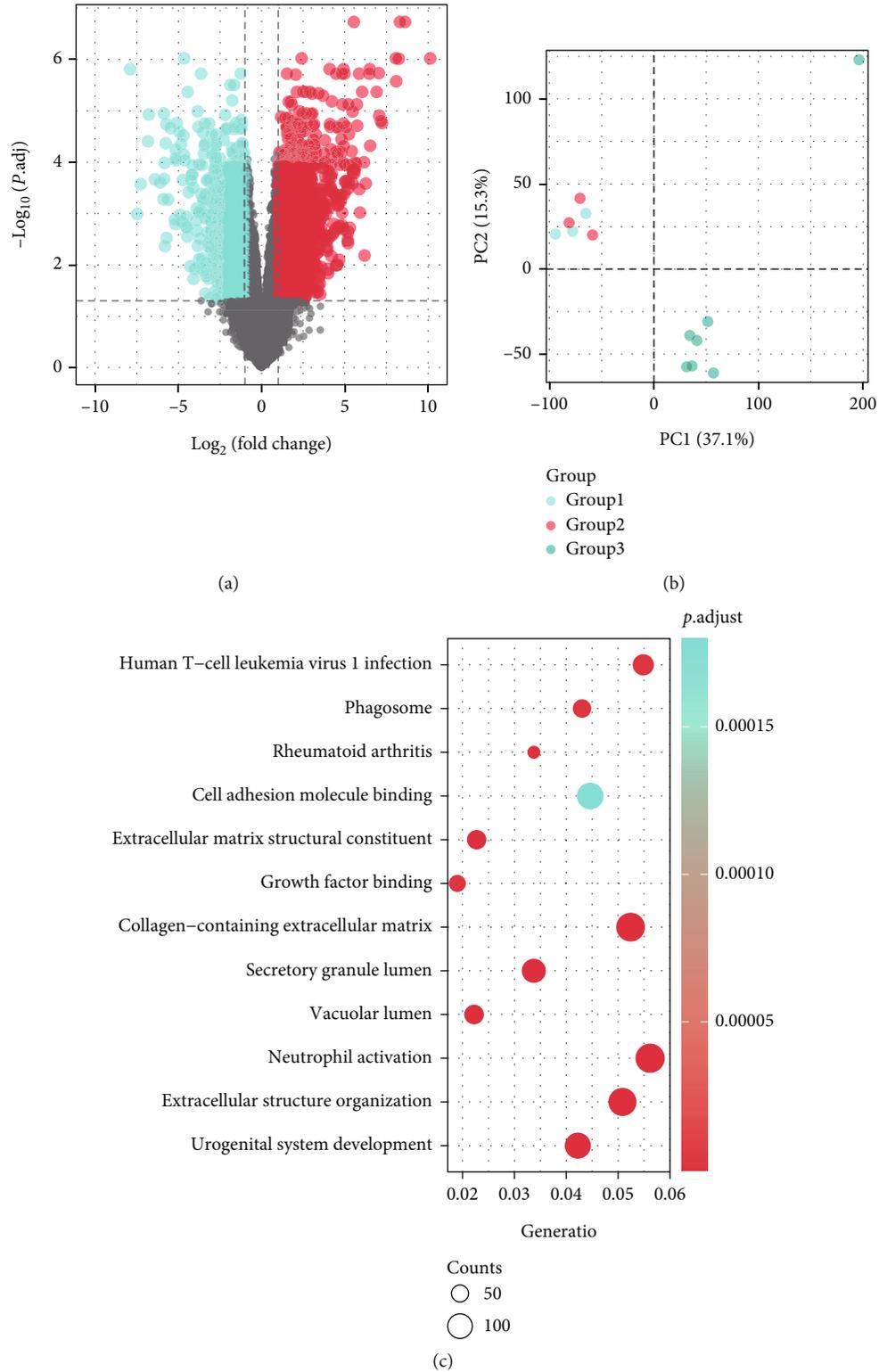


FIGURE 1: Continued.

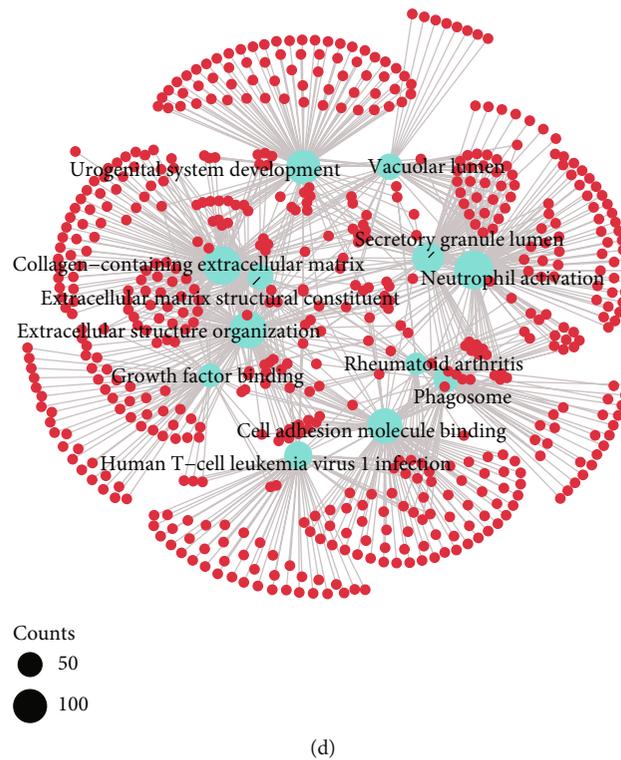


FIGURE 1: (a) Volcano plot: the fold change and  $P$  adjusted values were used in the construction of the volcano plot. (b) Principal component analysis of the samples. (c) The pop plot of enrichment analysis of differentially expressed genes. (d) Network of enrichment analysis.

popular treatment option. Nonetheless, many obstacles need to be overcome to achieve a satisfactory survival rate in transplanted cells and attain control of cell viability and regulation of differentiation [6].

Currently, MSCs are widely regarded as the most desirable and viable cell source for IVD regeneration, regardless of whether the process takes place *in vitro* or *in vivo* [7]. Most recently, numerous preclinical investigations and clinical trials have shown the potential utility of MSCs in the treatment of degenerated discs [8, 9]. Research has shown that MSCs can develop into an NP-like phenotype, which may optimize hydration in the IVD, elevate disc height, and enhance matrix formation of glycosaminoglycans (GAGs) [10]. The coculturing of MSCs with NP cells (NPCs) has also been shown to contribute to bidirectional communication between the MSCs and NPCs. Consequently, transplanted cells can alter NPC functions via the release of bioactive substances, such as anabolic growth factors, in addition to undergoing differentiation and the *de novo* production of the extracellular matrix (ECM) [11]. Moreover, MSCs have anticatabolic and anti-inflammatory characteristics, both of which have the potential to lower cytokine levels; this leads to modulation of the inflammatory niche, resulting in a native NPCs' phenotype that is healthy and nondegenerative [12].

Stem cells have garnered a great deal of interest from researchers and clinicians alike. The fast-paced advancement of stem cell technologies has enabled their adoption in a

variety of fields, including hematology, circulation, and orthopedics, among others [13]. Extensive research on IVD and IDD, including that into their underlying mechanism, has led us to believe that a combination of stem cell technology and therapeutic interventions for IDD can sustain the normal physiological structures and functions of the IVD, in addition to reversing the IDD cascade [14]. The benefit of IDD therapy and subsequent recovery may be substantially enhanced by the organic combination of IVD and stem cell technologies; nevertheless, this approach is still contentious [15]. In this study, we investigate the potential of stem cell treatment for IVD based on microarray and single-cell sequencing, with the goal of offering a new perspective on the treatment of IDD.

## 2. Material and Methods

**2.1. Data Collection and Processing.** The microarray data on stem cell therapy for IVD was acquired from the Gene Expression Omnibus (GEO) database in GSE17782 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17782>). Single-cell RNA sequencing data of purified wild-type rat intervertebral disc (GSE154884) were acquired from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>).

**2.2. Differentially Expressed Gene Identification.** To standardize and convert the raw data into expression values, we employed the packages available on the Bioconductor

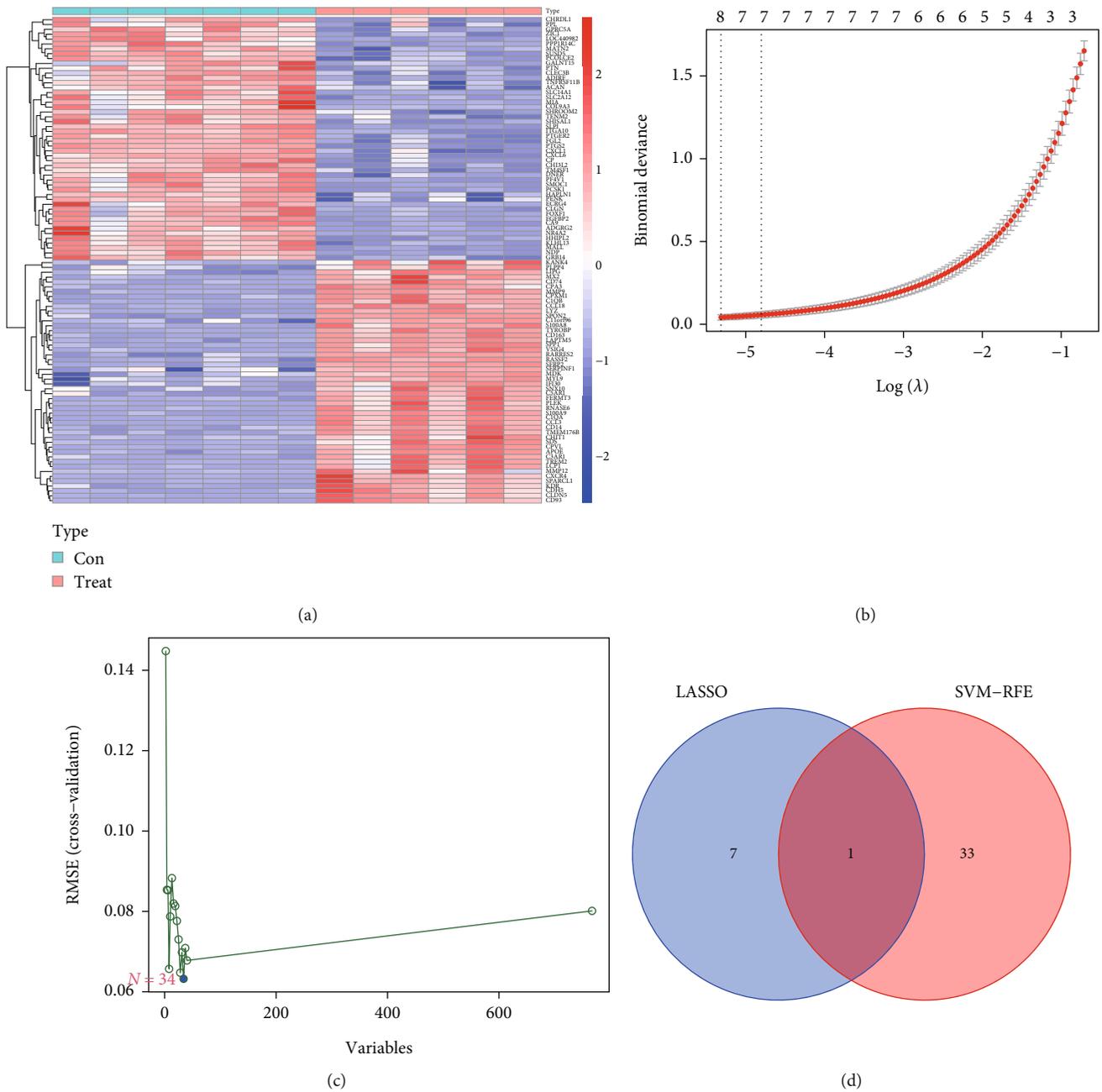


FIGURE 2: (a) Heatmap of top differentially expressed genes. (b) Number of genes identified using the least absolute shrinkage and selection operator (LASSO) regression method. (c) Number of genes detected using the support vector machine (SVM) regression method. (d) A Venn plot of overlapping genes obtained from LASSO and SVM.

website (<http://www.bioconductor.org/>). To detect the differentially expressed genes (DEGs) between treatment and control samples, the significance analysis included in the empirical Bayes methods of the “limma” package was utilized [16, 17]. The threshold values of  $|\logFC| > 1$  and  $P$  value  $< 0.05$  were employed for selecting significant DEGs.

**2.3. Analysis of Functions and Pathway Enrichment.** Annotation of the functions and pathways of the overlapping DEGs was achieved by employing clusterProfiler (version 3.14.3)

for Gene Ontology (GO) terms and performing Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, with a backdrop of the most recent GO annotations (version 3.1.0) and KEGG pathway annotations, respectively [18, 19]. The corresponding bar graph and network graph were then generated (<http://www.xiantao.love/>).

**2.4. Validation of Hub Genes Based on Machine Learning Models.** The least absolute shrinkage and selection operator (LASSO) model is a regressive analytical arithmetic method

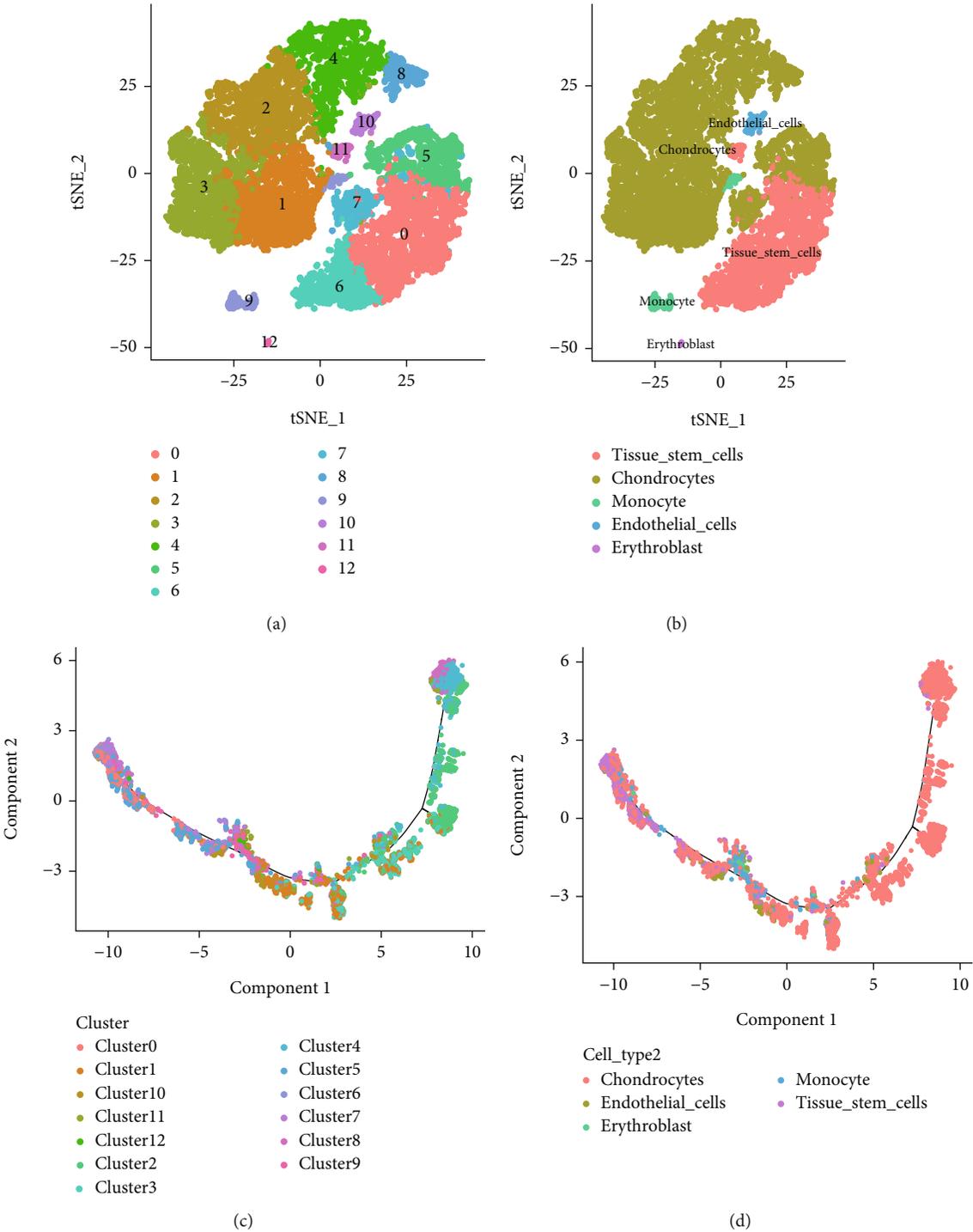
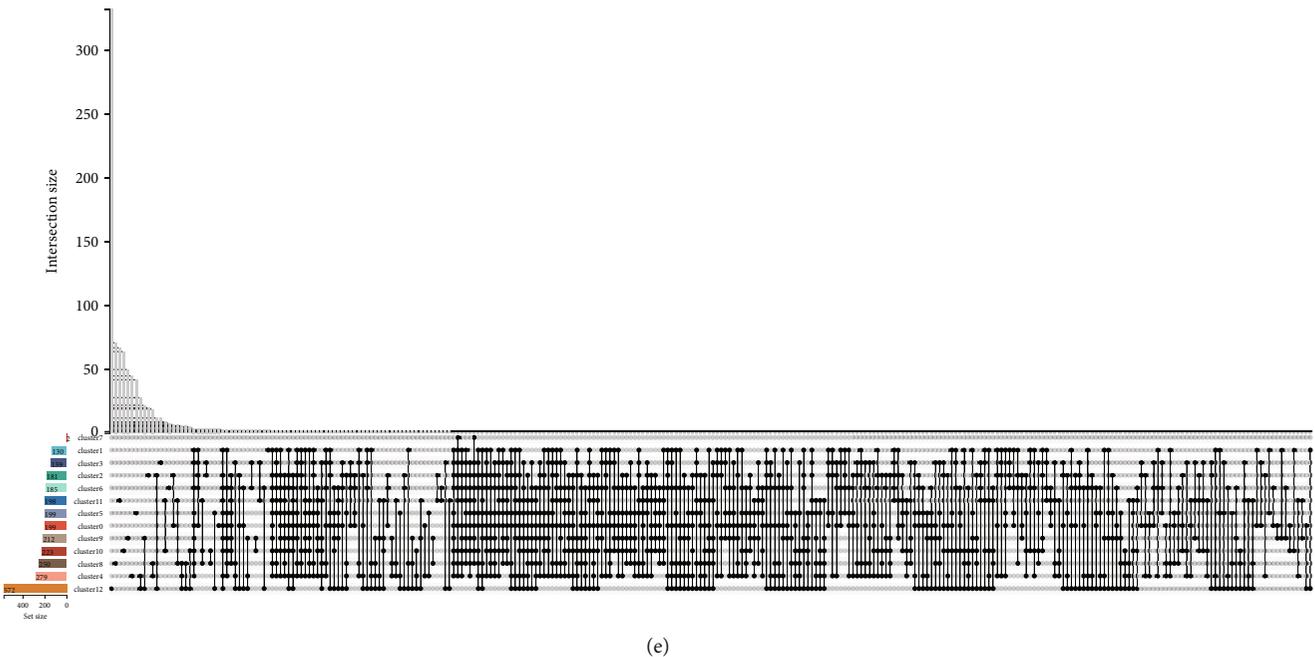


FIGURE 3: Continued.



(e)

FIGURE 3: (a, b) Using  $t$ -distributed stochastic neighbor embedding, a dimensionality reduction algorithm, cells were grouped into 18 kinds, with each hue representing a different annotated phenotype for each cluster. (c, d) Timeline of cells and related cell types. (e) A Venn plot of overlap DEGs in each clusters.

that employs regularization to improve the accuracy of the predictive process. The “glmnet” package in R was employed to carry out the LASSO analysis to identify the genes linked to the ability to discriminate between treatment and normal samples [20]. Support vector machine (SVM) is a kind of monitored machine learning technology that is utilized widely for classification and regression analysis. To prevent the problem of overfitting, recursive feature elimination was applied for screening genes from the metadata cohort. SVM recursive feature elimination (SVM-RFE) was applied to screen candidate features to determine which gene set had the best discriminative capacity.

**2.5. Single-Cell RNA Processing.** To examine the transcription heterogeneity of cells inside the IVD milieu, the “Seurat” R package was utilized for integration, preprocessing, elimination of batch effect, and reduction of the nonlinear dimension of the datasets [21]. Cells were clustered with the help of the FindCluster tool. To obtain marker genes within clusters, we employed the FindMarkers tool. Automatic annotation of the cells was done with the SingleR program. We conducted pseudotime trajectory analysis using the default configuration parameters of the Monocle 2 program [22].

**2.6. Construction of Enrichment Analysis and PPI Network.** Enrichment and protein-protein interaction (PPI) analyses were both performed on the DEGs after they were imported into the Metascape database (<http://metascape.org>). To identify network components at tight junctions, the Molecular Complex Detection algorithm 10 was applied, and all interactions from the STRING database that had a combined

score of more than 0.187 were employed. Next, the Cytoscape program was applied to display the PPI network for DEGs in each cluster.

### 3. Results

**3.1. Determination of DEGs in Stem Cell Treatment.** Within the scope of this study, GSE17782 was investigated using a retrospective approach. The “Limma” package was used to investigate the DEGs within the metadata. Initially, 2944 DEGs were identified, as shown in Figure 1(a). A two-dimensional principal component analysis cluster diagram was used to illustrate the gene expression matrix associated with the dataset. The vast majority of samples exhibited considerable clustering, suggesting that the sample source was reliable (Figure 1(b)).

**3.2. KEGG Pathway Enrichment Analyses of Stem Cell Treated DEGs.** KEGG analysis was performed using clusterProfiler to annotate DEGs that overlapped with one another in their functions and pathways. The DEGs were found to be most enriched in human T-cell leukemia virus 1 infection, phagosome, and rheumatoid arthritis pathways (Figures 1(c) and 1(d)).

**3.3. Determination of Hub Genes.** To determine the key biomarkers, two different arithmetic methods were applied. Figure 2(a) displays the top DEGs that were found. The LASSO method led to the identification of eight variates as hub genes for stem cell therapy (Figure 2(b)). With the aid of the SVM-RFE algorithm, a subset of 34 features shared by the DEGs was discovered (Figure 2(c)). Lastly, the

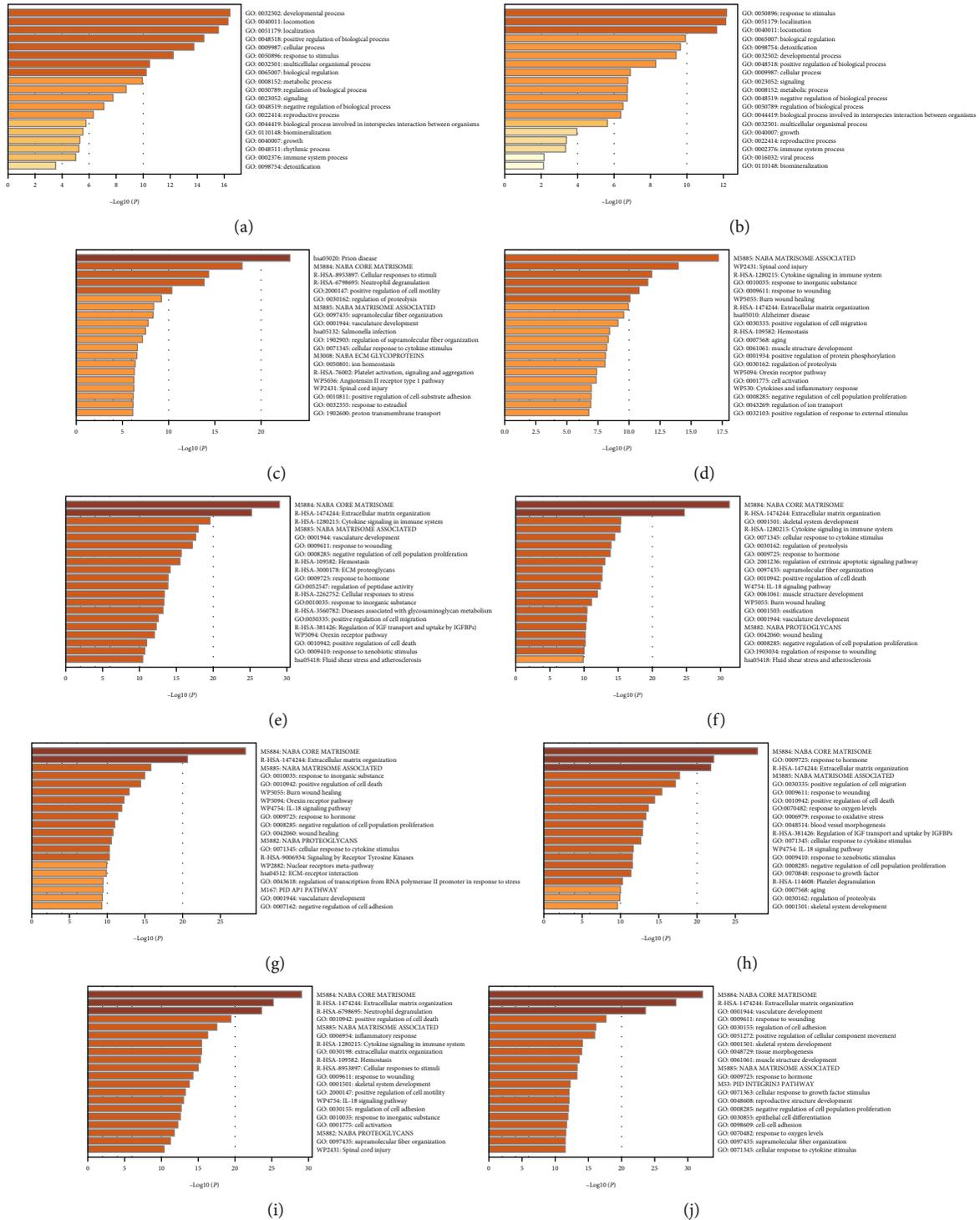


FIGURE 4: Continued.

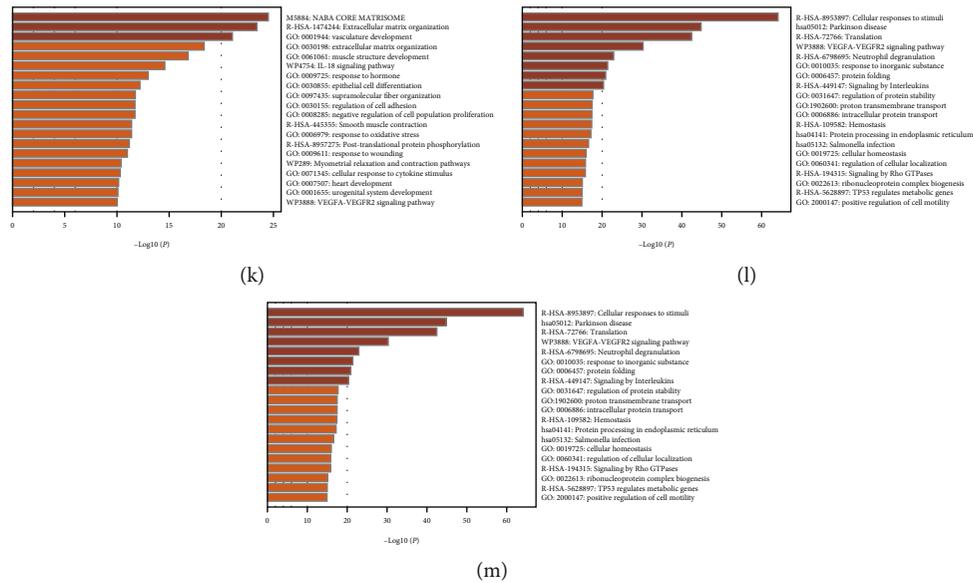


FIGURE 4: (a–m) Gene Ontology enrichment analysis of each cluster.

overlapping characteristic, LPTM5, was selected between the two arithmetic approaches, and this gene may be critical in stem cell treatment (Figure 2(d)).

**3.4. Single-Cell Sequencing Analysis.** After preprocessing the scRNA-seq data, strict quality control techniques outlined before were used to successfully isolate high-quality cell samples from the disc tissues. Next, to visualize the high-dimensional scRNA-seq data, we applied the t-SNE method to the leading 20 principal components and effectively categorized the cells into 13 different subcategories (Figure 3(a)), which were then annotated to recognize cell types utilizing the “SingleR” R package (Figure 3(b)). After examining the differentiation trajectories and pseudotemporal analysis using the “Monocle2” R package, based on the marker genes found for each cluster, we found it possible to establish a link between these cell clusters and their status (Figures 3(c) and 3(d)). The overlapping genes of DEGs ++ in each cluster are displayed in Figure 3(e).

**3.5. Gene Enrichment Analysis and PPI Analysis for DEGs in Each Cluster.** The Metascape database was used to perform functional enrichment analysis on genes that were differentially expressed and methylated, which allowed us to examine the biological activities of these genes. The results for each cluster are presented in Figures 4(a)–4(m). Finally, the enrichment network of cluster 7 is displayed in Figures 5(a) and 5(b). At the same time, the PPI network is depicted in Figure 5(c).

## 4. Discussion

IDD is a degenerative illness that can lead to other conditions, including low back pain and disc herniation. These illnesses have a significant negative impact on the patient’s quality of life and impose a significant financial load on society as a whole. Currently, standard therapies are available for

IDD in the clinical setting; however, these treatment modalities encounter a “bottleneck,” in that they are unable to reverse the onset and progression of IDD and can only alleviate patients’ pain. Treatments found on MSCs and mRNA are prospective targets for IDD therapy, and they are also of considerable use to researchers. The specific molecular mechanisms by which MSCs can be useful for the therapy of IDD are controversial, and clinical studies ought to evaluate their usefulness and safety before being implemented. The effective use of microarrays and bioinformatics has enabled a comprehensive understanding of the molecular processes underlying the onset and progression of illness, which in turn can help determine biological markers and treatment targets.

Although MSCs are currently acknowledged as a high-priority subject in the area of regenerative medicine, they may be difficult to use in the clinical setting owing to storage limitations and the possibility of cell senescence during *in vitro* growth. A recent research study demonstrated that MSCs are capable of secreting exosomes via paracrine signaling and that these exosomes exhibit similar features as those of MSCs. Exosomes generated from MSCs have been shown to exhibit excellent reparability, lower cytotoxicity, and lower immunogenicity. Given that the intervertebral disc is the largest avascular organ in the body, it derives nutrition by rather unique means. The metabolic and nutritional products are transported by blood vessels that are located outside the disc. Nutrients that are transported by capillaries and nutrient canals must first pass through dense hyaline cartilage endplate (CEP) to reach the disc matrix. Additionally, MSCs are created at a nanoscale, which makes them easily transportable and storable. As a result, their potential uses in therapeutic settings are rapidly expanding. Presently, both autologous and allogeneic bone marrow stromal cells (BMSCs) have shown outstanding therapeutic efficacy. Despite the fact that autologous BMSCs have greater homogeneity, cheaper treatment costs, broader applicability,

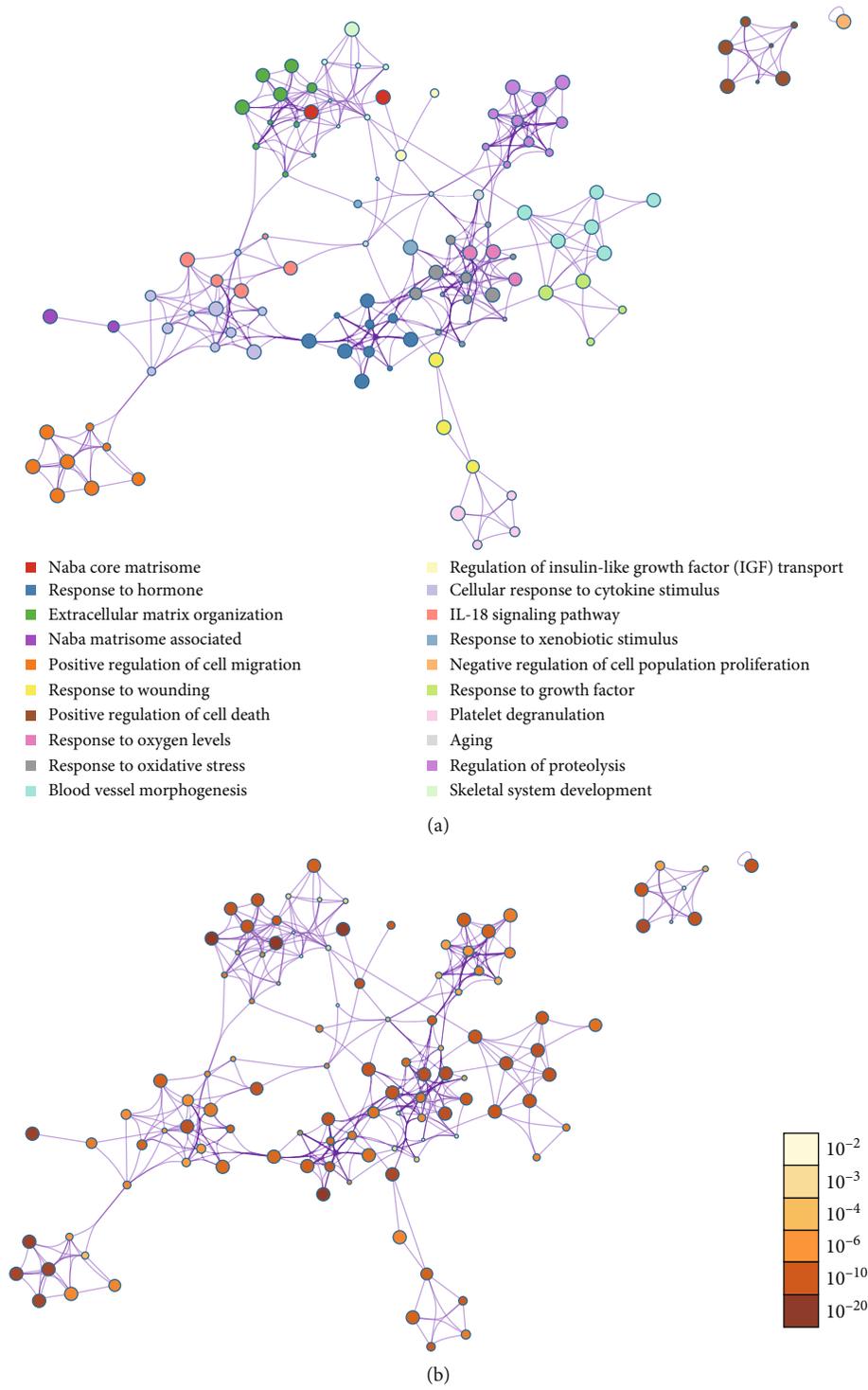


FIGURE 5: Continued.

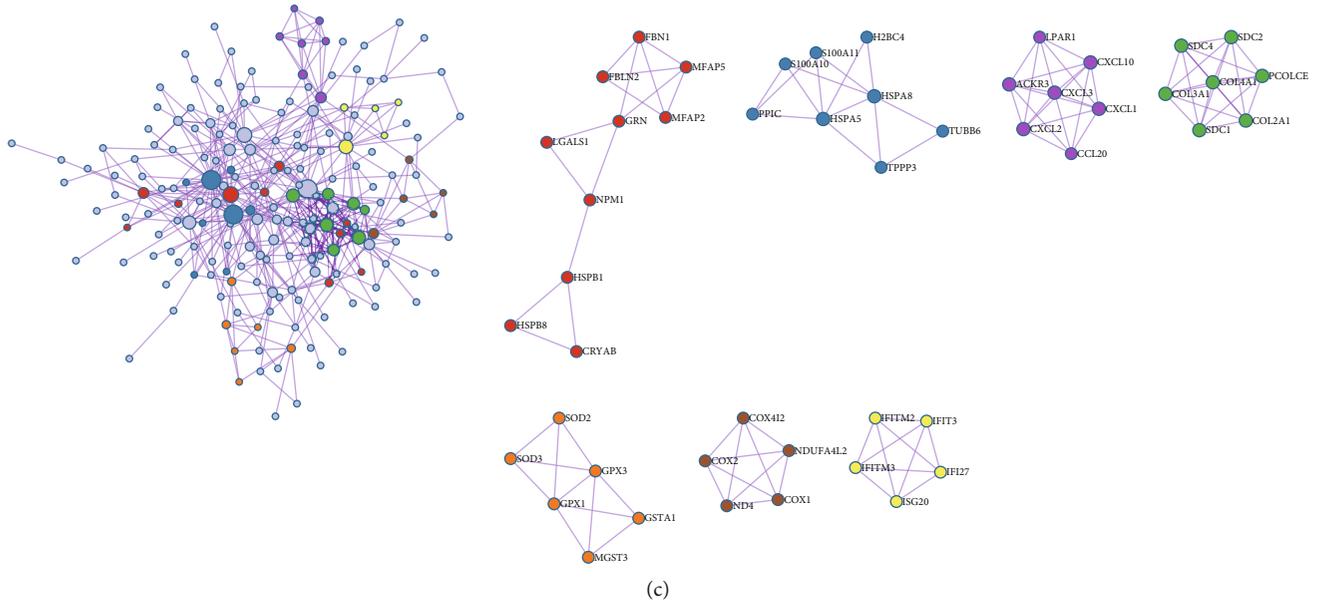


FIGURE 5: Network of enrichment analysis by (a) pathways and (b)  $P$  value; protein-protein interaction (PPI) network of differentially expressed genes.

and superior therapeutic potential, allogeneic BMSCs continue to be the cells of choice in stem cell treatment. However, the process of acquiring BMSCs involves invasive procedures; the sample size of clinical studies has been small, and the long-term effectiveness is uncertain; thus, more studies are required.

Chondrocytes undergo terminal differentiation under hypertrophic conditions, which occurs throughout the process of cartilage growth. The hypertrophy-like alterations in chondrocytes are generally accepted to be involved in osteoarthritis. Furthermore, hypertrophic degeneration has been revealed to occur in the NP throughout the course of DDD. As a result, we analyzed the cell fate to evaluate the differentiation statuses of chondrocytes in CEP tissues. In addition, with the use of machine learning, we discovered that LAPT5 is the hub gene for IDD. This was achieved by stem cell therapy. A recent research report discovered that RUNX2 transactivated the expression of the LAPT5 gene, which was in turn implicated in the translocation of RANKL. These results confirmed that RUNX2 could modulate osteoclast differentiation via lysosome-related genes that mediate RANKL transport in bone cells. The current findings pointed to the existence of a potential coupling mechanism between osteogenesis and osteoclastogenesis [23]. During the differentiation process of chondrocytes, homeostatic and stem cells were present in the beginning, but stem cells were not present at the end of this process. Through this phenomenon, we speculate that stem cell therapy will be an important means to delay the aging of the lumbar intervertebral disc. To examine the transcriptional heterogeneity of cells in IVD environment, we obtained 13 cell clusters and 4 cell subtypes using single-cell analysis based on GSE154884 dataset. We also use the interaction network of the Cytoscape program cluster 7.

However, our study is mainly based on bioinformatics and relies on public databases, so there are some limitations. First, the patient data is completely from the open database and has not been verified by clinical experiments. Second, the sample size is small, which may limit the statistical power. Therefore, more sample size studies are needed to validate our study in the future.

## 5. Conclusion

To conclude, we provided the first integrated assessment of scRNA-seq and machine learning in the field of IDD. In the diagnosis and treatment of IDD, having a full grasp of cell heterogeneity as well as stem cell treatment might help establish innovative insights and techniques.

## Data Availability

The data can be available via GEO database. More information can be accessed from the corresponding authors.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

Jianfeng Chen, Fuwei Qi, and Guanshen Li contributed equally to this work.

## Acknowledgments

This work was funded by the Guiding Project of Wuxi Science and Technology Bureau (NZ2019025) and Top Talent Support Program for Young and Middle-Aged People of Wuxi Health Committee (HB2020068).

## References

- [1] M. C. Battié, A. B. Joshi, and L. E. Gibbons, “Degenerative Disc Disease,” *Spine*, vol. 44, no. 21, pp. 1523–1529, 2019.
- [2] I. Han, A. E. Ropper, D. Konya et al., “Biological approaches to treating intervertebral disk degeneration: devising stem cell therapies,” *Cell Transplantation*, vol. 24, no. 11, pp. 2197–2208, 2015.
- [3] L. Zhang, Q. Chen, H. Wang, J. Yang, and S. Sheng, “Andrographolide mitigates IL-1 $\beta$ -induced human nucleus pulposus cells degeneration through the TLR4/MyD88/NF- $\kappa$ B signaling pathway,” *Molecular Medicine Reports*, vol. 18, no. 6, pp. 5427–5436, 2018.
- [4] Y. Chang, M. Yang, S. Ke, Y. Zhang, G. Xu, and Z. Li, “Effect of platelet-rich plasma on intervertebral disc degeneration *in vivo* and *in vitro*: a critical review,” *Oxidative Medicine and Cellular Longevity*, vol. 2020, Article ID 8893819, 10 pages, 2020.
- [5] N. Piazza, M. Dehghani, T. R. Gaborski, and K. Wuertz-Kozak, “Therapeutic potential of extracellular vesicles in degenerative diseases of the intervertebral disc,” *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 311, 2020.
- [6] D. Sakai and G. B. Andersson, “Stem cell therapy for intervertebral disc regeneration: obstacles and solutions,” *Nature Reviews Rheumatology*, vol. 11, no. 4, pp. 243–256, 2015.
- [7] G. Vadalà, L. Ambrosio, F. Russo, R. Papalia, and V. Denaro, “Interaction between mesenchymal stem cells and intervertebral disc microenvironment: from cell therapy to tissue engineering,” *Stem Cells International*, vol. 2019, Article ID 2376172, 15 pages, 2019.
- [8] L. Orozco, R. Soler, C. Morera, M. Alberca, A. Sánchez, and J. Garcia-Sancho, “Intervertebral disc repair by autologous mesenchymal bone marrow cells: a pilot study,” *Transplantation*, vol. 92, no. 7, pp. 822–828, 2011.
- [9] C. Zhang, S. E. Gullbrand, T. P. Schaer et al., “Combined hydrogel and mesenchymal stem cell therapy for moderate-severity disc degeneration in goats,” *Tissue Engineering Part A*, vol. 27, no. 1-2, pp. 117–128, 2021.
- [10] S. M. Richardson, G. Kalamegam, P. N. Pushparaj et al., “Mesenchymal stem cells in regenerative medicine: focus on articular cartilage and intervertebral disc regeneration,” *Methods*, vol. 99, pp. 69–80, 2016.
- [11] F. Han, Q. Yu, G. Chu et al., “Multifunctional nanofibrous scaffolds with angle-ply microstructure and co-delivery capacity promote partial repair and total replacement of intervertebral disc,” *Advanced Healthcare Materials*, vol. 11, no. 19, article 2200895, 2022.
- [12] M. J. Kibble, M. Domingos, J. A. Hoyland, and S. M. Richardson, “Importance of matrix cues on intervertebral disc development, degeneration, and regeneration,” *International Journal of Molecular Sciences*, vol. 23, no. 13, p. 6915, 2022.
- [13] N. Sugimoto and K. Eto, “Platelet production from induced pluripotent stem cells,” *Journal of Thrombosis and Haemostasis*, vol. 15, no. 9, pp. 1717–1727, 2017.
- [14] A. S. Croft, S. Illien-Jünger, S. Grad, J. Guerrero, S. Wangler, and B. Gantenbein, “The application of mesenchymal stromal cells and their homing capabilities to regenerate the intervertebral disc,” *International Journal of Molecular Sciences*, vol. 22, no. 7, p. 3519, 2021.
- [15] L. I. Silverman, G. Dulatova, T. Tandeski et al., “In vitro and in vivo evaluation of discogenic cells, an investigational cell therapy for disc degeneration,” *The Spine Journal*, vol. 20, no. 1, pp. 138–149, 2020.
- [16] B. Zhang, Q. Wu, R. Xu et al., “The promising novel biomarkers and candidate small molecule drugs in lower-grade glioma: evidence from bioinformatics analysis of high-throughput data,” *Journal of Cellular Biochemistry*, vol. 120, no. 9, pp. 15106–15118, 2019.
- [17] Q. Wu, B. Zhang, Z. H. Wang et al., “Integrated bioinformatics analysis reveals novel key biomarkers and potential candidate small molecule drugs in gastric cancer,” *Pathology Research and Practice*, vol. 215, no. 5, pp. 1038–1048, 2019.
- [18] S. Q. Ren, W. Wang, H. Y. Shen et al., “Development and validation of a clinical prognostic model based on immune-related genes expressed in clear cell renal cell carcinoma,” *Frontiers in Oncology*, vol. 10, p. 1496, 2020.
- [19] X. Q. Ge, Z. H. Wang, R. Jiang et al., “SCAMP4 is a novel prognostic marker and correlated with the tumor progression and immune infiltration in glioma,” *International Journal of Biochemistry & Cell Biology*, vol. 139, article 106054, 2021.
- [20] M. He, G. J. Wu, Z. H. Wang, K. Ren, Z. Yang, and Q. Xue, “Development and validation of a TRP-related gene signature for overall survival prediction in lung adenocarcinoma,” *Frontiers in Genetics*, vol. 13, article 13905650, 2022.
- [21] Y. Hao, S. Hao, E. Andersen-Nissen et al., “Integrated analysis of multimodal single-cell data,” *Cell*, vol. 184, no. 13, pp. 3573–3587.e29, 2021.
- [22] Z. C. Lu, Z. H. Wang, Z. H. Song et al., “Single-cell sequencing of brain tissues reveal the central nervous system’s susceptibility to SARS-CoV-2 and the drug,” *Frontiers in Pharmacology*, vol. 13, article 971017, 2022.
- [23] Y. M. Geng, C. X. Liu, W. Y. Lu et al., “LAPTM5 is transactivated by RUNX2 and involved in RANKL trafficking in osteoblastic cells,” *Molecular Medicine Reports*, vol. 20, no. 5, pp. 4193–4201, 2019.