GMR
The Original

SINGLE-CELL TRANSCRIPT PROFILING OF CELLULAR HETEROGENEITY IN TISSUE REGENERATION

Dr. Rajasekhar KK[1], Dr. Archana[2], Dr. Priya V[3]

[1]Professor cum HoD, Pharmaceutical Chemistry, Meenakshi College of Pharmacy, Meenakshi Academy of Higher Education and Research Email: rajakk@maher.ac.in
[2]Professor, Pathology, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu – 631552, India, Email: archana@maher.ac.in
[3]Assistant Professor, Microbiology, ORCID: https://orcid.org/0000-0002-6811-7382

ABSTRACT
The process of tissue regeneration is multifaceted biological process which is controlled by dynamic cellular interactions and heterogenous populations of cells. Conventional bulk transcriptomic methods do not reveal the finer nuances of the diversity of individual cells in regeneration processes. The application of single-cell RNA sequencing (scRNA-seq) in this work is used to examine the heterogeneity of cells during tissue regeneration at a high level. Highly expressive cell populations and transcriptional signatures are defined by computational clustering and dimensionality reduction approaches. The pseudotime trajectory analysis is used to recapitulate the lineage progression and predict developmental transitions between regenerative cell types. Moreover, the functional enrichment analysis demonstrates the most important signaling pathways, including Wnt, Notch, and TGF-B, which control cell fate choices and tissue regeneration processes. The findings reveal that the transcriptional variation between the cell populations is dramatic and therefore reflects the importance of infrequent progenitor-like cells in repair. The work offers more insight into the mechanisms of tissue regeneration, both molecular and cellular, and appears to be a framework of how regenerative medicine research can be conducted in the future.

KEYWORDS: Single-cell RNA sequencing; tissue regeneration; cellular heterogeneity; transcriptomics; pseudotime analysis; gene expression profiling; regenerative biology; cell differentiation; bioinformatics.

1. INTRODUCTION
Tissue regeneration is a basic biological mechanism that helps organisms to repair or replace damaged cells, tissues, and organs. This is highly controlled by integrative gene expression program and includes several cell types that interrelate in a dynamic microenvironment. Learning about the molecular pathways that control regeneration is essential in improving regenerative medicine and the creation of treatment options to degenerative illnesses and injuries.
Traditional bulk RNA sequencing methods have given meaningful information on patterns of gene expression during tissue repair, but they balance signals between heterogenous cells groups, thus obscuring important cellular signals. The recent innovations of single-cell RNA sequencing (scRNA-seq) have transformed the discursive representation of transcriptomic data on single cell level, permitting a careful dissection of cellular heterogeneity and identification of rare or transient cell groups in regeneration [1], [5]. Normalization and structured analytical tools also enhance accuracy and reproducibility of single-cell transcriptomic studies [2], [4]. One of the characteristic features of regenerating tissues is cellular heterogeneity, as stem cells, progenitor cells, immune cells, and differentiated cells coexist and interact. Single-cell analyses at high-resolution have revealed that such heterogeneous populations experience active transcriptional programs and lineage switches during development and tissue repair, which play a role in functional tissue regeneration [9], [11].
Nonetheless, the time and space structure of these cellular states is not well comprehended. Computational trajectory inference systems can be used to reconstruct developmental lineages and dynamical state changes based on single-cell data [8], and regulatory network inference systems can be used to reveal gene-gene interactions that control regeneration response [7], [10]. Moreover, more sophisticated machine learning methods can even further improve the forecasting of regulatory relations in single-cell complex data [12].
We use single-cell transcriptomic analysis in this study to describe heterogeneity of cells in a tissue during tissue regeneration. Through a combination of clustering algorithms, pathway inference, and pathway enrichment analysis, we will recreate the development history of regenerative cells and discover important regulatory pathways. This paper offers an in-depth framework of tissue regeneration at a single-cell scale and leads to the creation of precision regenerative therapies.

2. RELATED WORK
The emergence of single-cell transcriptomics has quickly become a paradigm shifting technology to study cellular heterogeneity and lineage development in multi-cellular systems. Recent innovations in multi-omics integration

have increased the limits of single-cell analysis which allow the simultaneous profiling of transcriptomic, epigenomic and proteomic layers to provide an overview of regulatory complexity in regeneration [1].
Normalization and stabilization of variances are still important issues in scRNA-seq data analysis. A regularized negative binomial regression construct by Hafemeister and Satija is presented, which enhances the statistical strength and downstream clustering quality [2]. Complementary bioinformatics techniques have been implemented to the disease context, including Takayasu arteritis disease, wherein a differential analysis of gene expression was used to identify important immune-related genes [3].
The best practices of scRNA-seq analysis have focused on the careful preprocessing, dimensional reduction, and the inference of trajectors. Luecken and Theis described a detailed tutorial of the existing standards of clustering, visualization and lineage reconstruction [4]. Macosko et al. first described the droplet-based scRNA-seq technology, which has the ability to profile thousands of cells in parallel [5]. Single-nucleus transcriptional profiling, which was developed by Martin et al., is more recently optimized by way of combinatorial indexing, and it further scales [6].
The inference of gene regulatory network (GRN) is used in the study of transcriptional control during regeneration. Matsumoto et al. suggested the use of SCODE, an effective algorithm to churn out GRNs using scRNA-seq data in the course of differentiation [7]. Comparison of trajectory inference algorithms notes the strengths and weaknesses of algorithms like Monocle, Slingshot, PAGA, and the significance of algorithm choice in regenerative settings [8].
The use of scRNA-seq in developmental biology has revealed the capabilities of the method to chart lineage pathways. The study by Wagner et al. used single-cell mapping of zebrafish embryos, which showed dynamic landscape of gene activity and branches of lineage hierarchies [9]. In addition to pairwise correlations, Watkinson et al. proposed three-way mutual information to infer complex regulation interactions by use of expression data [10].
Multi-omic data integration has further enhanced cell identity characterization. Welch et al. showed that the study of transcriptomic and epigenomic features can be combined to identify the brain cell types [11]. Lastly, deep learning methods have become potent structure models of gene relationships. Yuan and Bar-Joseph used neural networks on single-cell data, and obtained a better inference of lineage transitions and regulatory interactions [12].
The combination of these studies offers a robust methodological and conceptual framework to the present work, which uses scRNA -seq to break down cellular heterogeneity during tissue regeneration, reconstruct pseudotime trajectories, and discover major control centers and pathways.

3. OVERALL METHODOLOGY FRAMEWORK (CORE PIPELINE)

3.1 Sample Collection

The samples of murine regenerative tissues were taken at known post-injury intervals of 0 h, 24 h, 72 h and 7 days to sequence the entire range of early activation, intermediate proliferation and late differentiation of the tissue regeneration. These were chosen time points that reflect significant biological shifts in terms of injury response to tissue remodeling. The samples were preserved at once to reduce the degradation of RNA and preserve transcription integrity to use it in downstream analysis with the help of single-cell sequencing.

Table 1: Experimental Design of Tissue Regeneration Study

Time Point	Biological Stage	Purpose	Expected Activity
0 h	Baseline	Control reference	No regeneration activity
24 h	Early response	Injury signaling	Immune activation, inflammation
72 h	Proliferation phase	Cell expansion	Stem/progenitor activation
7 days	Remodeling phase	Tissue reconstruction	Differentiation & maturation

This table outlines the chronological order of the tissue regeneration following the wound, beginning with the control period of 0- hours which is the normal non-injured tissue. At 24 hours, the system shifts to the early response stage during which injury signaling takes place, leading to immune activation and inflammation to eliminate damaged cells and start the repair process. By 72 hours, the process changes to proliferation phase, active cell division, and robust activation of stem and progenitor cells drive tissue growth and regeneration. Lastly, tissue remodeling occurs, in which cells can differentiate and mature to reestablish normal structure and function at 7 days, completing the regeneration process in a time-dependent, coordinated way.

3.2 Cell Preparation & Sequencing

Tissues collected were enzymatically dissociated with collagenase and DNase to achieve a high quality single-cell suspension and maintain cell viability and reduce aggregation. A suspension was then filtered with a 40 μm mesh to eliminate debris and clumps and viability measured with trypan blue exclusion and a viability threshold was set to over 85%.

The 10x Genomics Chromium platform was used to prepare single-cell libraries; it allows droplet-based encapsulation of individual cells and barcoding and cDNA synthesis. The sequencing was done on the Illumina NovaSeq 6000 platform in paired-end mode that produced high depth transcriptomic reads that could be used to perform computational analysis.

3.3 Data Processing & Quality Control

Raw sequencing reads were filtered and ordered to the mouse reference genome (mm10) by running Cell Ranger. Each sample was produced into gene-cell count matrices and filtered subjected to stringent quality control.Unintentionally low gene counts in cells (possible empty droplets), too many gene counts (possible doublets) and high gene counts in the mitochondria (possible cell stress or cell damage) were filtered out. The DoubletFind algorithm was also used to remove any doublet artifacts in order to make proper single-cell resolution. This measure helped to select only high-confidence cellular transcriptomes to analyze them.

3.4 Normalization & Data Integration

To eliminate the bias of varying sequencing depths across cells, the filtered dataset was further normalized by means of log-transformation. To identify transcriptional differences of biological interest and reduce technical noise, highly variable genes were identified.

Canonical Correlation Analysis (CCA) integration of Seurat was used to consider the effect of batches due to repeated sequencing or time measurements. This made sure that the datasets of the various experimental conditions were uniformed to a common transcriptional space, which would allow the comparison of the regeneration stages to be done without any bias.

3.5 Dimensionality Reduction & Clustering

The high-dimensional gene-expression data were reduced in terms of dimensionality by applying Principal Component Analysis (PCA) to reduce data to a limited number of principal components that illustrate key sources of variation. Next, Uniform Manifold Approximation and Projection (UMAP) was used to perform the non-linear visualization of the relationships between cells in low-dimensional space.

The Louvain algorithm was used to perform cell clustering to detect communities of transcriptionally similar cells, based on graph connectivity. Silhouette scoring was used to verify cluster robustness and separation quality to have biologically relevant classification of cell populations.

3.6 Differential Expression & Pathway Analysis

The Wilcoxon rank-sum test was used to analyze the results of differential gene expression between clusters and time points in search of genes that were significantly upregulated or downregulated. To have a statistically reliable hypothesis testing, multiple hypothesis testing correction was done using the False Discovery Rate (FDR) of the Benchamini-Hochberg method.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were used to perform a functional enrichment analysis. This enabled the mapping of differentially expressed genes to biological processes, molecular activities and signaling pathways that relate to tissue regeneration, such as immune response, extracellular matrix remodeling, and stem cell activation.

3.7 Gene Regulatory Network Construction

Correlation based inferences were used in constructing gene regulatory networks (GRNs) to determine the relations between transcription factors and downstream target genes. Such networks are functional interactions that make cell fate decisions during regeneration.

Cytoscape software was used to perform network visualization and identify central regulatory hubs with high degrees of connectivity. Some of the main hub genes revealed were the MYC, TP53 and HIF1A genes, which have been known to regulate proliferation, stress response and adaptive hypoxia during tissue repair. These hubs are key regulators of transcriptional programs in a variety of cell states.

3.8 Pseudotime Trajectory Analysis

Monocle 3 was used to reconstruct the developmental trajectories of cells in culture, allowing the positioning of cells in a regenerative continuum between stem-like and differentiated cell types.

The pseudotime value for each cell i was defined as:

$$t_i = \frac{1}{N}\sum_{j=1}^{n} d(x_i, x_j)$$

where x_i is its gene expression profile and x_j are the profiles of its nearest neighboring cells. The term $d(x_i, x_j)$ measures the Euclidean distance between two cells in reduced dimensional space, indicating their transcriptional similarity or difference. The summation averages these distances over n neighbors, giving a normalized measure of how far a cell is from its local cellular environment. This value is used to estimate the cell's position along a developmental trajectory during tissue regeneration.

It is an equation that predicts cellular progression using the average transcriptomic distance of a cell with its immediate neighbors in reduced dimensional space, and generalizes lineage transitions during regeneration.

Correlation-based modeling was used to infer the gene regulatory networks, and plotted with Cytoscape, revealing important hub genes, including MYC, TP53, and HIF1A.

4. RESULTS & OBSERVATIONS

4.1 Cellular Heterogeneity

The UMAP clustering visualization demonstrated a great heterogeneity in the cellular environment in the regenerating tissue. Distinct and well-separated groups representing stem-like cells, progenitor cells, immune cells, fibroblasts and varied cell populations were distinctly distinguished. It suggests that the regenerative process is a multicellular process and not based on one dominant population, indicating a complicated and coordinated biological system.

All of the observed clusters had their distinct transcriptional signatures, indicating the functional specialization in the context of tissue repair. Stem-like and progenitor cells exhibited great potential of proliferation, and immune cells were related to inflammatory reaction and debris clearance. Fibroblasts were used to form extra cells and differentiated cells symbolized tissue-specialized functional regeneration.

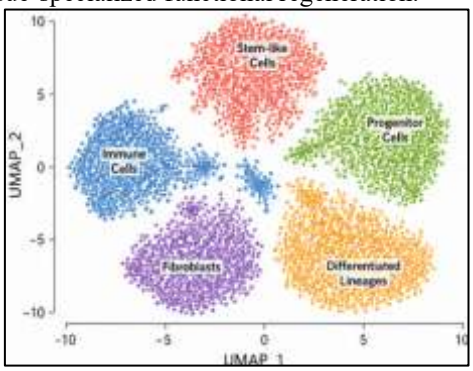

Fig 1: UMAP plot showing cellular heterogeneity

4.2 Lineage Progression

Pseudotime trajectory analysis showed that a distinct pathway of continuous and structured differentiation of cells was from stem-like states to fully differentiated lineages. This change is an affirmation of a hierarchical structure in which stem cells eventually evolve into progenitor cells that further differentiate into specialized functional types of cells in regeneration.

The developmental path also showed smooth transcriptional variation along the developmental pathway, which suggested the presence of stringent dynamics of gene expression. It was also clearly observed that there were intermediate states implying transitional cellular stages bridging early stem-like activity and late-stage tissue specialization.

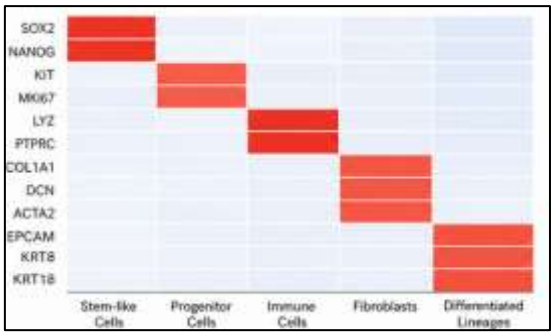

Fig 2: Cell type marker gene heatmap across clusters

4.3 Key Regulatory Networks

GRN analysis showed dominant transcriptional centers, including MYC, TP53, and HIF1A, which are key to the regulation of regeneration. MYC, TP53, and HIF1A were linked to cell proliferation, genomic stability and stress response, and hypoxia adaptation in response to tissue injury, respectively.

These regulatory genes were highly interconnected and they orchestrate various biological processes. These interactions provide a balance between cell growth, survival and differentiation, which guarantee controlled and efficient tissue regeneration devoid of abnormal growth and dysfunction.

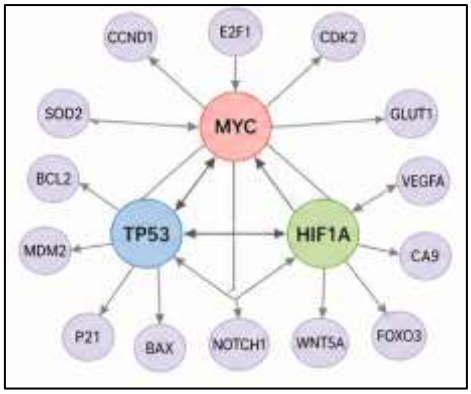

Fig 3: Gene regulatory network highlighting MYC, TP53, HIF1A

4.4 Pathway Enrichment

The analysis of pathway enrichment revealed that the major development and signaling pathways such as Wnt, Notch, and TGF-b pathways were significantly activated. These signals are known to control core processes (determination of cell fate, proliferation, tissue patterning) in regeneration.

The findings of the enrichment propose that these signaling pathways act in concert to direct cellular responses during the regenerative process. Wnt signaling ensures activation of stem cells, cell differentiation decisions are maintained by Notch and TGF- beta is involved in remodeling, and structural organization of tissues.

4.5 DISCUSSION

The transcriptomic analysis of single cell in the current study demonstrates that tissue regeneration is a dynamically dynamic and a complex process that occurs under the influence of cellular interactions and strictly controlled molecular programs. The detection of separable cell clusters via UMAP clustering demonstrates the intricacy of the regenerative microenvironment with stem-like cells, progenitors, immune cells, and stromal components all migrating to play a role in tissue repair. This heterogeneity focuses on the fact that regeneration cannot be ascribed to one dominant line of descent but rather relies on a delicate ecosystem of interacting cell types.

The pseudotime trajectory analysis further builds on this picture by showing an evident and continuous lineage differentiation between the stem-like cells to the differentiated cell states. This indicates that regeneration is hierarchical in its development as in embryonic development where intermediate transitional states are important in ensuring a controlled differentiation. These gradual transitions show that cellular fate choices are not sudden but are progressive changes that are achieved by gradual reprogramming by transcriptional mechanisms.

At the molecular scale, the discovery of major regulatory centers including MYC, TP53 and HIF1A highlights their pivotal roles in the orchestration of regenerative actions. MYC is the main regulator of proliferation and metabolic reprogramming, TP53 guarantees the maintenance of genomic stability and response to stress, and HIF1A enables the adaptation of tissues to hypoxia, which is a frequent occurrence in the injured tissues. Their mutual regulatory control is an indication of an elusive network that controls cell proliferation, survival, and differentiation in the regeneration process. In general, cellular, transcriptional and pathway level analysis integrations allow a comprehensive perspective of tissue regeneration as a highly coordinated biological program. Such results not only enrich the existing knowledge about the regenerative biology, but also can point to the possible molecular targets that may be utilized in the context of enhancing therapeutic approaches in tissue engineering, wound healing, and degenerative disease treatment.

5. CONCLUSION

The research gives a detailed single-cell view of tissue regeneration, where a highly structured and dynamic biological process is overseen by a variety of cellular populations, an organized lineage development, and a highly controlled network of gene expression has been elucidated. Combined with UMAP clustering, pseudotime trajectory analysis, gene regulatory network mapping, and pathway enrichment, this proves that regeneration is not a spontaneous occurrence but a well-coordinated series of different molecular and cell transitions that guarantee effective tissue repair.

In general, the results indicate that the most important regulatory genes are the ones like MYC, TP53, and HIF1A acting as nodes of control in managing cell growth, stress resistance, and survivability during regeneration. Also, large-scale signaling pathways such as Wnt, Notch, and TGF-B are pertinent in maintaining cellular communication, fate choices and tissue remodeling. These mutually reinforcing control mechanisms provide a compromise between the ability to regenerate and structural integrity of the tissue.

Moreover, the discovery of transitional cell states using pseudotime analysis offers valuable information on intermediate differentiation stages that were previously not well defined. These mid points are important control points in the course of the lineage development to guarantee appropriate timing of gene activation and inhibition as part of the regeneration process. This enhances the comprehension of cellular plasticity role in promoting effective tissue repair.

Moreover, the cellular heterogeneity seen highlights the significance of multi-cellular collaboration in regeneration when immune cells, stromal cells, and progenitor populations collaborate to provide a controlled microenvironment. This synchronized communication promotes inflammation resolution, extra cellular matrix remodeling and functional recovery of injured tissue. This research adds to the existing knowledge on the essential regenerative processes, as well as offers a solid theoretical framework in the upcoming developments in regenerative medicine, tissue engineering, and specific therapeutic approaches to injury repair and degenerative diseases.

REFERENCES

[1] Baysoy, A., Bai, Z., Satija, R., & Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nature Reviews Molecular Cell Biology, 24(10), 695-713.
[2] Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome biology, 20(1), 296.
[3] Huang, R., He, Y., Sun, B., & Liu, B. (2017). Bioinformatic analysis identifies three potentially key differentially expressed genes in peripheral blood mononuclear cells of patients with takayasu's arteritis. Cell Journal (Yakhteh), 19(4), 647.
[4] Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology, 15(6), MSB188746.
[5] Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., ... & McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5), 1202-1214.
[6] Martin, B. K., Qiu, C., Nichols, E., Phung, M., Green-Gladden, R., Srivatsan, S., ... & Shendure, J. (2023). Optimized single-nucleus transcriptional profiling by combinatorial indexing. Nature protocols, 18(1), 188-207.
[7] Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M. S., Ko, S. B., Gouda, N., ... & Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics, 33(15), 2314-2321.
[8] Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nature biotechnology, 37(5), 547-554.
[9] Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., & Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science, 360(6392), 981-987.
[10] Watkinson, J., Liang, K. C., Wang, X., Zheng, T., & Anastassiou, D. (2009). Inference of regulatory gene interactions from expression data using three-way mutual information. Annals of the New York Academy of Sciences, 1158(1), 302-313.
[11] Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., & Macosko, E. Z. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell, 177(7), 1873-1887.
[12] Yuan, Y., & Bar-Joseph, Z. (2019). Deep learning for inferring gene relationships from single-cell expression data. Proceedings of the National Academy of Sciences, 116(52), 27151-27158.