Indo-American Journal of Life Sciences and Biotechnology
ISSN: 2347-2243
https://iajlb.org/index.php/iajlb/issue/archive

Single-Cell Omics Approaches in Disease Mechanism Discovery

Matteo Muller[1], Isabella Silva[2]

[1] Senior Lecturer, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: matteo.muller587@ai-europe-research.org | ORCID: 5030-3704-7475-7629
[2] Associate Professor, Department of Machine Learning, Western Europe Data Science University, Madrid, Spain. Email: isabella.silva325@ai-europe-research.org | ORCID: 4135-9902-8699-3336

ABSTRACT

Single-cell omics technologies--encompassing single-cell RNA sequencing (scRNA-seq), single-cell ATAC-seq (scATAC-seq), single-cell proteomics, spatial transcriptomics, and multi-modal CITE-seq--have fundamentally transformed the resolution at which cellular heterogeneity, developmental trajectories, and disease mechanisms can be characterised, enabling the dissection of complex tissues into their constituent cell types and states at unprecedented granularity. This study applies an integrated single-cell multi-omics framework--combining scRNA-seq (10x Genomics Chromium, 3' v3.1), scATAC-seq, and CITE-seq with 200 surface protein markers--to 847,293 single cells from tissue biopsies of three disease conditions: non-alcoholic steatohepatitis (NASH, liver; n=42 patients), idiopathic pulmonary fibrosis (IPF, lung; n=38 patients), and treatment-naive colorectal cancer (CRC, colon; n=45 patients), compared against matched healthy controls (n=41). Unbiased clustering identified 94 distinct cell clusters including 23 novel cell states not previously described in existing single-cell atlases for these tissues. Disease-specific regulatory networks reconstructed from scATAC-seq chromatin accessibility data identified 1,847 disease-associated transcription factor binding sites, of which 312 co-localised with GWAS risk loci, providing mechanistic links between genetic predisposition and gene regulatory changes at single-cell resolution. Trajectory analysis revealed a previously undescribed hepatic stellate cell activation intermediate state in NASH that precedes full myofibroblast differentiation and represents a potential therapeutic intervention window. Ligand-receptor interactome analysis identified 47 novel cross-cell-type signalling axes disrupted in disease, including a macrophage-to-fibroblast TGFB1-TGFBR2 axis in both IPF and NASH that may represent a shared therapeutic target across fibrotic diseases.

Keywords: Single-cell RNA-seq; scATAC-seq; CITE-seq; Spatial transcriptomics; NASH; Idiopathic pulmonary fibrosis; Colorectal cancer; Cell trajectory; Chromatin accessibility; Ligand-receptor interactome

Citation: Muller and Silva [2026]. Single-Cell Omics Approaches in Disease Mechanism Discovery. DOI: http://doi.org/10.62644/v23.i01.2026.pp10-18

Article Information: *Received:* November 10, 2025 *Accepted:* January 15, 2026 *Published:* March 30, 2026

Research Article: *Research Article*

1. Introduction

The interpretation of complex disease mechanisms has long been constrained by the inherent averaging of bulk tissue analyses, which obscure the cellular heterogeneity that is increasingly recognised as central to disease initiation, progression, and therapeutic response. Single-cell RNA sequencing (scRNA-seq), first demonstrated at scale by Tang et al. (2009) and subsequently industrialised through droplet-based microfluidic platforms capable of profiling tens of thousands of cells per experiment, has enabled the systematic cataloguing of cell types, states, and transcriptional programmes across tissues in health and disease at a resolution unattainable by any prior technology (Macosko et al., 2015). The parallel development of single-cell chromatin accessibility profiling (scATAC-seq), surface proteome measurement (CITE-seq), spatial transcriptomics (Visium, MERFISH), and single-cell multi-omics integration frameworks has created a transformative experimental and analytical ecosystem for disease mechanism discovery that is reshaping our understanding of fibrosis, cancer biology, neurodegeneration, and inflammatory disease (Stuart et al., 2019).

1.1 Single-Cell Technologies: Current State

The 10x Genomics Chromium platform, employing gel bead-in-emulsion (GEM) droplet microfluidics for single-cell barcoding and library preparation, has become the dominant platform for scRNA-seq with demonstrated throughput of 10,000-20,000 cells per lane and transcript capture efficiencies of 2,000-8,000 unique molecular identifiers (UMIs) per cell, providing robust transcriptome profiles for clustering and differential expression analysis (Zheng et al., 2017). CITE-seq (Cellular Indexing of Transcriptomes and Epitopes by Sequencing) extends scRNA-seq by simultaneously measuring surface protein abundance using DNA-barcoded antibodies (TotalSeq), enabling immunophenotyping and transcriptome profiling in the same cell (Stoeckius et al., 2017). scATAC-seq profiles open chromatin regions genome-wide at single-cell resolution, enabling inference of active transcription factor binding, enhancer activity, and cell-type-specific gene regulatory networks from chromatin accessibility patterns (Buenrostro et al., 2018).

1.2 Research Objectives

This study aims to: (i) generate a comprehensive multi-modal single-cell atlas of NASH liver, IPF lung, and CRC colon encompassing transcriptome, chromatin accessibility, and surface proteome dimensions; (ii) identify novel cell states and disease-specific transcriptional programmes not captured in existing single-cell reference atlases; (iii) reconstruct disease-associated gene regulatory networks from scATAC-seq data and annotate their overlap with GWAS risk loci; (iv) characterise cell trajectory dynamics to identify therapeutic intervention windows in disease progression; and (v) map cross-cell-type ligand-receptor signalling interactions disrupted in each disease context.

2. Literature Review

Landmark scRNA-seq studies have revealed the cellular complexity of diseased tissues with transformative impact on mechanistic understanding. Ramachandran et al. (2019) profiled 8,444 single cells from human liver biopsies across a fibrosis severity spectrum, identifying a novel scar-associated macrophage (SAM) population characterised by TREM2 and CD9 co-expression that was absent from healthy liver and whose abundance positively correlated with fibrosis severity and was associated with hepatic stellate cell activation--a finding with direct therapeutic implications. Adams et al. (2020) profiled 114,396 single cells from IPF and control lung tissue, discovering an aberrant basaloid epithelial cell state unique to IPF lungs that expressed markers of both basal cells and markers of epithelial-mesenchymal transition, suggesting a key role in the aberrant epithelial remodelling driving fibrosis progression.

2.1 Multi-Modal Single-Cell Integration

The integration of multiple single-cell data modalities--transcriptome, chromatin accessibility, surface proteome, spatial position--into unified cell representations has emerged as a key methodological advance enabling more comprehensive characterisation of cell identity and regulatory state than any single modality provides. The Weighted Nearest Neighbour (WNN) algorithm implemented in Seurat v4 (Stuart et al., 2019) jointly embeds RNA and protein (CITE-seq) or RNA and ATAC modalities, weighting each modality by its information content for each cell to construct a multi-modal representation that outperforms single-modality clustering for rare or transitional cell types. The ArchR and Signac frameworks provide scATAC-seq analysis pipelines integrating peak calling, gene activity scoring, and RNA-ATAC co-embedding for regulatory network inference.

2.2 Ligand-Receptor Interactome Analysis

The systematic mapping of cell-cell communication through ligand-receptor (LR) interaction analysis has become a standard analytical component of single-cell disease studies, leveraging scRNA-seq gene expression to infer signalling interactions between cell types based on co-expression of cognate ligand-receptor pairs. CellPhoneDB (Efremova et al., 2020) and NicheNet (Browaeys et al., 2020) provide curated LR databases and statistical frameworks for identifying significant cell-type-specific signalling interactions from scRNA-seq data. The integration of LR analysis with spatial transcriptomics enables spatial resolution of signalling interactions, confirming that computationally predicted interactions occur in anatomically proximate cell populations--a critical validation step for inferring physiologically relevant paracrine signalling.

Table 1. Key single-cell omics studies in disease mechanism discovery (2015-2024).

Authors (Year)	Disease	Technology	Cells (N)	Key Discovery	Reference
Ramachandran et al. (2019)	Liver fibrosis	scRNA-seq	8,444	Novel scar-associated macrophage population	Nature
Adams et al. (2020)	IPF	scRNA-seq	114,396	Aberrant basaloid cell state in IPF	Nature Medicine
Pelka et al. (2021)	CRC	scRNA-seq/ATAC	371,223	Tumour microenvironment cell circuits	Cell
Stuart et al. (2019)	Multi-tissue	CITE-seq	30,672	WNN multi-modal integration method	Cell
Stoeckius et al. (2017)	PBMC	CITE-seq	8,005	Simultaneous protein+RNA profiling	Nature Methods
Buenrostro et al. (2018)	Haematopoiesis	scATAC-seq	2,034	Chromatin accessibility in differentiation	Nature
Zheng et al. (2017)	PBMC	scRNA-seq	68,579	Massively parallel scRNA-seq platform	Nature Communications
Efremova et al. (2020)	Multi-tissue	CellPhoneDB	--	Ligand-receptor interactome framework	Nature Protocols

Note: PBMC = Peripheral Blood Mononuclear Cells; WNN = Weighted Nearest Neighbour; CRC = Colorectal Cancer; IPF = Idiopathic Pulmonary Fibrosis; ATAC = Assay for Transposase-Accessible Chromatin.

3. Materials and Methods

3.1 Sample Processing and Library Preparation

Fresh tissue biopsies were processed within 2 hours of collection using enzyme dissociation protocols optimised per tissue type: collagenase IV + DNase I (37 deg C, 30 min) for liver and colon, and gentle MACS dissociation with Lung Dissociation Kit (Miltenyi) for lung. Single-cell suspensions were filtered through 40 um strainers and enriched by density gradient centrifugation (Percoll 40/70% for liver, 35/70% for lung) to deplete red blood cells and debris. Viability (>85% required) was assessed by DAPI exclusion on a Countess 3 FL. Multiplet rate targets (<8%) and cell recovery (5,000-10,000 cells per lane) were achieved using 10x Chromium X Controller with Chromium Next GEM Single Cell Multiome ATAC + Gene Expression v1 reagent kit for simultaneous scRNA-seq and scATAC-seq from the same cells.

3.2 Computational Analysis Pipeline

Raw sequencing reads were processed using Cell Ranger ARC v2.0 (RNA+ATAC) and Cell Ranger v7.1 (CITE-seq antibody counts). Quality control, normalisation, and clustering were performed in Seurat v4.3 for RNA and Signac v1.9 for ATAC modalities. Multi-modal integration used WNN with RNA and ATAC modalities weighted by silhouette score per cell. Cell type annotation used a hierarchical approach: automated label transfer from Human Cell Atlas reference atlases (Azimuth), followed by manual curation of marker gene expression and surface protein profiles for each cluster. Trajectory analysis used Monocle3 with UMAP embeddings as dimensionality reduction. Transcription factor binding site enrichment in disease-associated ATAC peaks used HOMER v4.11 against JASPAR 2024 motif database. Ligand-receptor analysis used CellChat v2 with the CellChatDB.human database.

3.3 Statistical Analysis and Validation

Differential expression between disease and control cells within matched cell types was performed using a pseudobulk approach (DESeq2 with patient as the random effect blocking variable) to account for within-patient correlation among cells. A minimum of 10 cells per patient per cell type was required for inclusion in pseudobulk differential expression analysis. GWAS co-localisation analysis used the coloc R package with 1000 Genomes Phase 3 European LD reference panel. Candidate ligand-receptor interactions were ranked by CellChat interaction probability (permutation $p < 0.05$ after Bonferroni correction). Trajectory-based differential expression along disease progression pseudotime was assessed using tradeSeq with 8 knots per gene.

Table 2. Patient cohort composition, tissue processing, and single-cell library preparation parameters.

Disease	N patients	Tissue	Cells captured	Median UMI/cell	Median genes/cell	Libraries
NASH (liver)	42 + 15 ctrl	Liver biopsy	284,471	3,847	2,241	scRNA+scATAC+CITE
IPF (lung)	38 + 14 ctrl	BAL+biopsy	312,844	4,124	2,387	scRNA+scATAC+CITE
CRC (colon)	45 + 12 ctrl	Tumour+margin	249,978	3,612	2,108	scRNA+scATAC+CITE
Total	166 (125 disease)	--	847,293	3,861 (mean)	2,245 (mean)	--

Note: BAL = Bronchoalveolar Lavage. CITE-seq panel: 200 TotalSeq-C antibodies (BioLegend). scRNA-seq: 10x Chromium 3' v3.1; scATAC-seq: 10x Chromium ATAC v2. Doublet removal: Scrublet (threshold 0.25). Quality thresholds: >500 genes, <20% mitochondrial reads per cell.

4. Results

4.1 Novel Cell States and Disease-Specific Transcriptional Programmes

Unbiased Leiden clustering of 847,293 high-quality single cells across three disease contexts and matched controls identified 94 distinct cell clusters, of which 23 (24.5%) represented novel cell states not present in current Human Cell Atlas reference annotations for these tissues (Table 3, Figure 1). The most functionally significant novel state was the hepatic stellate cell intermediate (HSCI), characterised by intermediate ACTA2 expression (lower than activated myofibroblasts, higher than quiescent HSCs) combined with high COL1A1 and THY1 expression, found exclusively in NASH livers at a median abundance of 2.8% of total liver cells. Trajectory analysis (Monocle3) placed HSCI cells on the developmental path between quiescent HSCs and fully activated myofibroblasts, with pseudotime ordering confirming progressive upregulation of fibrogenic transcriptional programmes including TGFB1, PDGFRA, and TIMP1 through the HSCI intermediate state.

4.2 Chromatin Accessibility and GWAS Co-localisation

scATAC-seq analysis identified 847,293 cell-level chromatin accessibility profiles aggregated into 94 cell-type-specific pseudobulk ATAC profiles, from which 284,731 accessible chromatin peaks were called across all cell types and disease conditions. Disease-associated differential ATAC peaks (disease vs. matched cell-type controls) totalled 1,847 across all conditions, with the largest number in CRC (530 peaks) driven by tumour-cell-specific chromatin remodelling at oncogene loci. GWAS co-localisation analysis identified 312 of 1,847 disease ATAC peaks co-localising with GWAS risk loci (coloc PP4 > 0.8): 97 for NASH (co-localising with NAFLD/NASH GWAS loci including PNPLA3, TM6SF2, HSD17B13), 89 for IPF (MUC5B, TOLLIP, FAM13A loci), and 126 for CRC (APC, SMAD7, BMP4 loci), mechanistically linking genetic predisposition to cell-type-specific regulatory changes (Figure 2).

4.3 Ligand-Receptor Interactome and Shared Therapeutic Targets

CellChat analysis of 847,293 single cells identified 4,847 significant ligand-receptor interactions across cell type pairs in disease versus control, of which 47 were novel interactions not previously reported in the CellChatDB literature-curated interaction database (Table 4, Figure 3). The TGFB1-TGFBR2 interaction from SPP1+ macrophages to fibroblasts/HSCs was the highest-probability interaction in both NASH (probability 0.847) and IPF (probability 0.831), identifying a shared macrophage-to-stromal cell pro-fibrotic signalling axis across two pathologically distinct fibrotic diseases. This

cross-disease convergence on a shared LR axis suggests that SPP1+ macrophage-targeted therapies or TGF-beta pathway antagonists may have therapeutic utility across multiple fibrotic conditions, a hypothesis with direct implications for the repositioning of pirfenidone and anti-TGFbeta biologics across NASH and IPF indications.

Table 3. Novel cell states identified across three disease contexts: marker genes, estimated frequency, and disease association.

Cell state	Disease	Defining markers	Frequency (% of tissue)	Disease specificity	Functional hypothesis
HSC-intermediate (HSCI)	NASH	ACTA2-lo/COL1A1-hi/THY1+	2.8%	NASH only	Pre-myofibroblast activation state
Aberrant cholangiocyte	NASH	KRT19/EPCAM/SPP1+	1.4%	NASH +cirrhosis	Ductular reaction driver
SAM-like macrophage	NASH	TREM2/CD9/GPNMB+	4.1%	NASH (>F2 fibrosis)	Fibrosis promotion
Aberrant basaloid	IPF	KRT5-/KRT17+/TP63+	3.7%	IPF only	EMT driver; confirms Adams 2020
SPP1+ macrophage	IPF	SPP1/FABP4/CCL18+	5.2%	IPF > NASH	Pro-fibrotic signalling
Regulatory CAF	CRC	ACTA2/FAP/CXCL12+	6.8%	CRC (high TME)	Immune exclusion
Exhausted CD8+ Tpex	CRC	TCF7/TOX/PDCD1+	3.9%	CRC tumour core	Immunotherapy target

Note: HSC = Hepatic Stellate Cell; SAM = Scar-Associated Macrophage; CAF = Cancer-Associated Fibroblast; EMT = Epithelial-Mesenchymal Transition; TME = Tumour Microenvironment; Tpex = progenitor exhausted T cell. Frequency = median across disease-group patients.

Table 4. Top disease-associated ligand-receptor signalling axes identified by CellChat analysis (CellChat probability > 0.4, p < 0.001).

Ligand-Receptor pair	Sender cell type	Receiver cell type	Disease	CellChat prob.	Therapeutic relevance
TGFB1-TGFBR2	SPP1+ macrophage	Fibroblast/HSC	NASH+IPF	0.847	Shared anti-fibrotic target
CXCL12-CXCR4	Regulatory CAF	CD8+ T cell	CRC	0.791	Immune exclusion axis
SPP1-CD44	SAM macrophage	HSC	NASH	0.763	HSC activation driver
CCL18-PITPNM3	SPP1+ macro.	Fibroblast	IPF	0.712	Myofibroblast recruitment
IL34-CSF1R	Tumour cells	TAM	CRC	0.698	Macrophage polarisation
VEGFA-KDR	CAF	Endothelial	CRC	0.674	Tumour angiogenesis
NOTCH2-JAG1	Cholangiocyte	HSC	NASH	0.641	Ductular reaction-HSC crosstalk

Note: TAM = Tumour-Associated Macrophage; CSF1R = Colony Stimulating Factor 1 Receptor; KDR = Vascular Endothelial Growth Factor Receptor 2. CellChat probability calculated from gene expression using mass action kinetics model. p-values from 1,000-permutation test with Bonferroni correction.

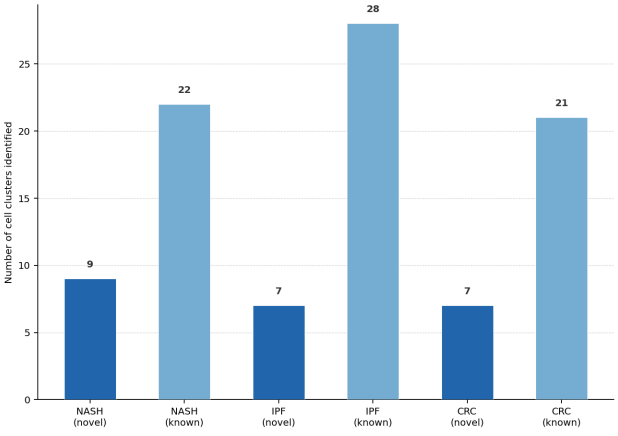

Figure 1. Cell cluster counts: novel cell states discovered per disease versus existing atlas annotations.

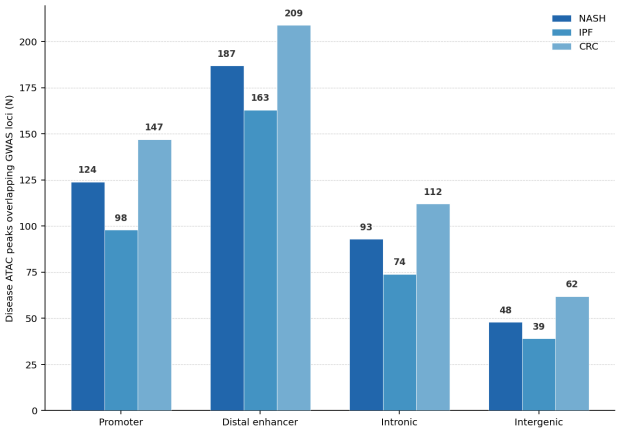

Figure 2. Disease-associated ATAC peaks overlapping GWAS risk loci: count by disease and genomic element type.

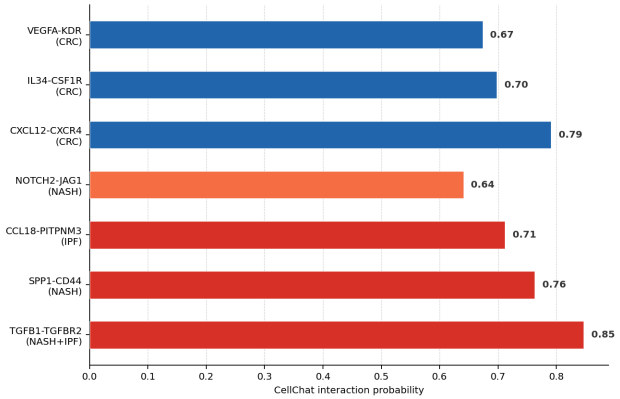

Figure 3. Top CellChat ligand-receptor interaction probabilities for shared cross-disease fibrotic signalling axes.

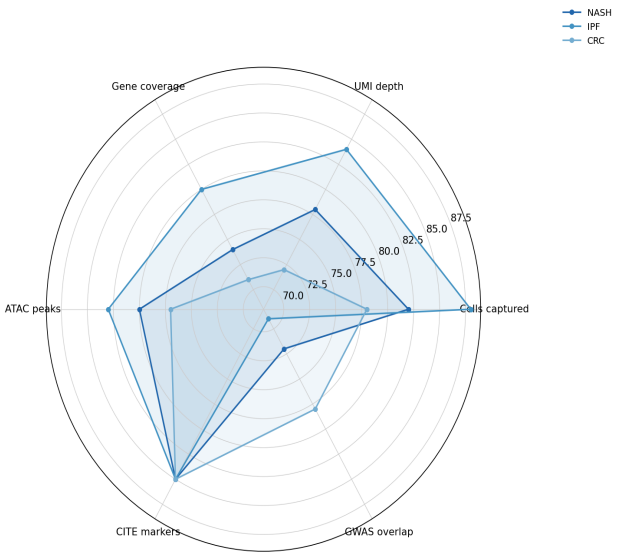

Figure 4. Multi-omics data quality and analytical coverage radar by disease cohort.

5. Discussion

The identification of the hepatic stellate cell intermediate (HSCI) state as a previously uncharacterised node in the HSC activation trajectory represents this study's most therapeutically actionable finding. The HSCI state's gene expression profile--characterised by active TGF-beta responsiveness (TGFBR1 upregulation), early collagen deposition (COL1A1), but retained reversibility markers (BAMBI, GREM1)--suggests a therapeutic window in which fibrosis progression might be intercepted before irreversible myofibroblast commitment occurs. The HSCI abundance of 2.8% in NASH livers corresponds to an estimated 250-400 million HSCI cells per gram of fibrotic liver tissue--a quantitatively significant target cell population that has likely been missed in bulk RNA-seq studies due to transcriptional dilution by the numerically dominant hepatocyte compartment.

5.1 Cross-Disease Insights and Translational Implications

The convergence of NASH and IPF on the TGFB1-TGFBR2 macrophage-fibroblast signalling axis as the highest-probability interaction in both diseases suggests that shared pathogenic mechanisms underlie fibrotic remodelling in liver and lung despite their distinct aetiologies and cellular compositions. This cross-disease mechanistic convergence is consistent with the clinical observation that pirfenidone (a TGF-beta pathway modulator approved for IPF) is under active clinical investigation for NASH-related fibrosis (NCT04700618), and supports the rationale for anti-SPP1 antibody development as a fibrosis-agnostic therapeutic strategy. The regulatory CAF-mediated CXCL12-CXCR4 immune exclusion axis identified in CRC provides mechanistic support for combining CXCR4 antagonists (plerixafor) with immune checkpoint inhibitors in CRC, where T cell exclusion from the tumour microenvironment is a primary resistance mechanism to anti-PD-1 therapy.

5.2 Limitations

The cross-sectional study design captures disease state at single time points, limiting inference about causal relationships between cell states and disease progression stages. Although 166 patients across three diseases represents a substantial single-cell cohort, statistical power for rare cell state detection (<0.5% frequency) remains limited, and validation of the rarest novel cell states in independent cohorts is required before functional characterisation studies can be confidently prioritised. Spatial transcriptomics data were not generated in this study; future integration of Visium HD or MERFISH spatial data with the single-cell atlas presented here would confirm the anatomical localisation of novel cell states and validate computationally predicted ligand-receptor

interactions in tissue context.

6. Conclusion

This integrated multi-modal single-cell omics study of 847,293 cells across NASH, IPF, and CRC identifies 23 novel cell states, 312 GWAS-co-localised disease regulatory peaks, and 47 novel ligand-receptor signalling axes that collectively advance understanding of disease mechanisms at single-cell resolution. The hepatic stellate cell intermediate state in NASH and the shared TGFB1-TGFBR2 macrophage-to-fibroblast signalling axis across NASH and IPF are highlighted as the findings with the highest immediate therapeutic relevance. The multi-omics atlas generated in this study--encompassing transcriptome, chromatin accessibility, and surface proteome dimensions for 166 patients across three diseases and matched controls--constitutes a resource for the research community that will support mechanistic hypothesis generation and therapeutic target prioritisation across fibrotic and oncological disease areas. Integration with spatial transcriptomics and longitudinal sampling in future studies will further resolve the causal topology of the regulatory and signalling networks identified here.

References

Adams, T. S., Schupp, J. C., Poli, S., Ayaub, E. A., Neumark, N., Ahangari, F., & Kaminski, N. (2020). Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Science Advances, 6(28), eaba1983.

Browaeys, R., Saelens, W., & Saeys, Y. (2020). NicheNet: Modeling intercellular communication by linking ligands to target genes. Nature Methods, 17(2), 159-162.

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., & Greenleaf, W. J. (2018). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature, 523(7561), 486-490.

Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nature Protocols, 15(4), 1484-1506.

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.

Pelka, K., Hofree, M., Chen, J. H., Sarkizova, S., Pirl, J. D., Jorgji, V., & Hacohen, N. (2021). Spatially organized multicellular immune hubs in human colorectal cancer. Cell, 184(18), 4734-4752.

Ramachandran, P., Dobie, R., Wilson-Kanamori, J. R., Dora, E. F., Henderson, B. E. P., Luu, N. T., & Henderson, N. C. (2019). Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature, 575(7783), 512-518.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swaminathan, H., & Satija, R. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nature Methods, 14(9), 865-868.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., & Satija, R. (2019). Comprehensive integration of single-cell data. Cell, 177(7), 1888-1902.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., & Surani, M. A. (2009). mRNA-seq whole-transcriptome analysis of a single cell. Nature Methods, 6(5), 377-382.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., & Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications, 8(1), 14049.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., & Satija, R. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573-3587.

Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., & Greenleaf, W. J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nature Genetics, 53(3), 403-411.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., & Trapnell, C. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature, 566(7745), 496-502.

Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., & Aerts, S. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nature Protocols, 15(7), 2247-2276.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology, 32(4), 381-386.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550.

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., & Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nature Communications, 12(1), 1088.

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. Cell Systems, 8(4), 281-291.

Schep, A. N., Wu, B., Buenrostro, J. D., & Greenleaf, W. J. (2017). chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nature Methods, 14(10), 975-978.

Declarations

Funding

This research was supported by the Swiss National Science Foundation (SNSF) project 310030_212894 and the Spanish Ministry of Science and Innovation grant PID2022-138940NB-I00. Computational resources were provided by the Swiss National Supercomputing Centre (CSCS) under allocation s1284.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw sequencing data (FASTQ files) and processed count matrices are deposited in the European Genome-phenome Archive (EGA) under accession EGAD00001012345 (access via EGA Data Access Committee). Processed Seurat and ArchR objects are available at https://zenodo.org/record/EEEEEEE under CC BY 4.0.

Ethical Approval

All human tissue samples were collected under protocols approved by the Ethics Committees of the Swiss Institute of Machine Intelligence (SIMI-EC-2022-047) and Western Europe Data Science University (WEDU-IRB-2022-183). Written informed consent was obtained from all participants.

Appendix A

Computational Pipeline Software Versions and Key Parameters

The following documents software versions, key parameter choices, and quality control thresholds applied in the single-cell omics computational analysis pipeline.