Stem Cell-Parenchymal Fusion: Communication and Gene Regulation

JoAnne Huang[1,#], Fateme Nazaryabrbekoh[1,#], Joohyun Kim[2] and Jangwook P. Jung[1,*]

1. Department of Biological Engineering, Louisiana State University, Baton Rouge, LA
2. Vanderbilt Genetics Institute and Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN

[#]Equal contribution

*Corresponding author:
Jangwook P. Jung, Ph.D.
Department of Biological Engineering
Louisiana State University
167 E.B. Doran Hall
Baton Rouge, LA 70803
Phone: (225) 578-2919
Fax: (225) 578-3492
E-mail: jjung1@lsu.edu

Abstract
We investigated the cellular and molecular consequences of accidental cell fusion between mHL1 and mMSC cell lines using previously published single-cell RNA sequencing (scRNA-seq) data. We characterized the resulting cell types, their communication patterns, and underlying gene regulatory networks. Initial analyses identified four distinct cell types (mHL1, mHL1 fusion, mMSC, and mMSC fusion), which were also consolidated into three clusters via unsupervised learning. Differential gene expression analysis revealed diverse cell type associations with shared and exclusive gene expression. Single-cell compatible weighted gene co-expression network analysis (WGCNA) linked specific gene modules to distinct cell types and associated biological processes, including muscle contraction and metabolism in mHL1/mHL1 fusion, and extracellular matrix (ECM) formation and lipid metabolism in mMSC fusion. Cell-cell communication analysis revealed distinct intercellular communication patterns, with mMSC primarily acting as a ligand and mHL1 as both ligand and receptor. Reducing cell type complexity progressively simplified communication networks and altered pathway enrichment, while ECM-receptor pathways remained largely unchanged, with only minor shifts in ligand-receptor pairs. Interestingly, removing parental cells from the analysis altered cell type assignments, underscoring the influence of parental cell presence on fusion outcomes. Finally, gene regulatory network analysis identified key transcription factors (TFs) driving cell identity, distinguishing between master regulators and TFs with high cell-type specificity. These findings demonstrate the power of integrating multiple inter- and intra-cellular communication analysis approaches to dissect the complex interplay between cell type, communication, and gene regulation in cell fusion.

Introduction
Cell-cell fusion is a biological process where two cells merge their membranes, forming a single hybrid cell[1]. This fundamental event plays crucial roles in development, reproduction, and tissue repair across various organisms[2]. Heterotypic fusion, the merging of cells with distinct origins, presents a complex interplay of benefits and risks. Essential for processes such as fertilization, it can also disrupt cellular homeostasis. The fusion of distinct cellular environments can have profound effects on nuclear function, including competition for gene regulation, chromosomal reorganization, and even nuclear ejection[3-5]. Mesenchymal stromal cells (MSCs) exhibit a remarkable capacity to fuse with diverse cell types within various organs, including the brain, liver, and heart[6-13]. This phenomenon of cell fusion plays a pivotal role in development and regeneration, while also contributing to cancer progression. Notably, MSC fusion has demonstrated beneficial effects in certain contexts, such as promoting liver regeneration through fusion with hepatocytes[6,14,15] and contributing to muscle regeneration through fusion with myocytes[16,17]. However, the consequences of MSC fusion in many other contexts remain uncertain. A growing body of evidence links heterotypic fusion involving MSCs to the development of cancer and metastasis, raising significant concerns about the safety of MSC-based therapies[18-23]. Understanding the underlying mechanisms of cell fusion is crucial to distinguish between beneficial and detrimental outcomes. This knowledge will pave the way for therapeutic strategies that harness controlled cell fusion for tissue repair and regeneration, while simultaneously preventing pathological fusion events.

This study revisits previously published data on the fusion of mouse HL-1 cardiomyocytes (mHL1) and mouse MSCs (mMSC)[23]. While the original study utilized bulk RNA-seq analysis methods, we now leverage more recent single-cell RNA-seq (scRNA-seq) analysis tools to gain deeper insights into cell-cell communication and the underlying gene networks orchestrating this specific fusion process. By applying contemporary scRNA-seq analysis tools, we aim to extend beyond the previous limitations of principal component analysis (PCA), hierarchical clustering (HC), and Gene Ontology (GO) analyses[23]. Specifically, we investigate two key aspects of cellular communication (**Figure 1**): (a) intercellular

communication using dedicated cell-cell communication tools and (b) intracellular communication inferred through gene regulatory network inference tools. We further differentiate between communication occurring within fused cells and between parental cell types. Finally, we discuss these new findings in the context of the original results, providing a more comprehensive understanding of the molecular mechanisms driving mHL1-mMSC fusion.

Methods

Data acquisition and samples

The original FPKM gene expression matrix obtained from the Gene Expression Omnibus (GEO, accession number GSE69926) was converted to TPM for subsequent analysis. TPM is generally considered more robust and preferred for most analyses, especially when comparing gene expression across different samples[24]. For pre-fusion cells, parental controls consisted of 15 cells of each parental type (mMSC_1–15 and mHL1_1–15) isolated prior to co-culture. After fusion, each cell type was harvested and subjected to single-cell RNA sequencing[23]. Fusion cells comprised 28 fusion products identified using BiFC (BiFC_D1_F1–5, 24 h) and dual color (DC) expression of GFP and mCherry (DC_D1_F1–16, 24 h; DC_D3_F1–7, 72 h). Additionally, 10 parental cells were exposed to co-culture conditions, but did not fuse were included: 5 cells of each parental cell type isolated 24 h after co-culture (mMSC_D1_1–5 and HL1cm_D1_1–5).

Marker gene analysis of distinct cell types

To generate AnnData format files, we converted TPM values from the available FPKM gene matrix and metadata using Scanpy (ver. 1.10.1) in Python (ver. 3.12.2). Then, we filtered the data, retaining only genes expressed in at least one cell and cells expressing at least 200 genes. Next, we computed the K-nearest neighbors of cells, considering the 10 nearest neighbors of each cell in a 40-dimensional principal component space. Finally, we visualized the high-dimensional embedding using Uniform Manifold Approximation and Projection (UMAP) in a 12-dimensional space, overlaying our known annotations from the metadata.

Clustree: To compare our known annotations with unsupervised clustering, we applied Clustree[25]. The raw AnnData was converted into a Seurat object and processed through normalization, feature selection and scaling. PCA was employed to reduce dimensionality, followed by clustering at multiple resolutions (0.2, 0.4, 0.6, 0.8, 1.0) to identify cell clusters. Furthermore, we visualized Clustree graphs using the clustree function, where resolutions were specified with the prefix RNA_snn_res, nodes are clusters, and edges indicate overlap between clusters.

FindAllMarkers: To find differentially expressed genes, the FindAllMarkers command was ran by adjusting the parameters to only return positive genes which have a $\log_2$ fold change of 2 or greater, while keeping all other default parameters. From there, a csv file of the list of markers was saved and used for analysis.

Venn diagram: We compared the number of differentially expressed gene markers identified by our known annotation and unsupervised clustering. We visualized these results using a Venn diagram, where numbers represent gene counts in each specific cell type and cluster. The Venn diagrams were generated with the matplotlib_venn library by importing venn2 and venn3 in Python.

Cell Markers: To identify cell types, relevant gene names from our annotation and Clustree were collected and used to query the EnrichR analysis tool[26]. This identified cell types associated with the annotated four cell types, the three cell types from Clustree, and all hub genes in each module eigengenes (ME). Subsequently, specific cell types were further identified from the Tabula Muris atlas[27].

High dimensional Weighted Gene Co-Expression Network Analysis (hdWGCNA)

To identify modules of interconnected genes with high co-expression in all cell types, hdWGCNA was carried out using the hdWGCNA package[28-30] using the AnVIL platform (R 4.4.0). From there, the hub genes in each module could be identified and were used to visualize how each module was connected to each other. A Seurat object was first created using all the cell types. The metacells were constructed using a k-nearest neighbors value of 2, a minimum of 5 shared cells between two metacells and a maximum of 25 shared cells between two metacells. From there, the rest of the pipeline was run using the default parameters.

Intercellular communications: CellCall and CellChat

Intercellular communication between cell types was inferred using the R packages CellCall[31] and CellChat[32] with R (ver. 4.3.3). Both packages utilize known ligand-receptor (LR) interactions to build the intercellular communication network. CellCall combines its inferences of intercellular and intracellular communication by integrating the expression of LR pairs and transcription factors (TF) activities to build an L-R-TF axis. Like CellCall, CellChat can combine previous knowledge of LR genes, however, lacks information on the associated TF.

CellCall was run using default parameters for all analyses. A circle plot showing the signaling between each cell type as well as which cell was the ligand and which was the receptor, was first generated. From there, a bubble plot containing the associated pathways for each permutation of cell type combinations, which were colored based off their normalized enrichment score (NES). The L-R-TF axes were then established for each permutation of the cell type names, which was used to define the sending and receiving cell. A TF enrichment plot was also drawn to show the enrichment score (ES) for significantly activated TFs in receiver cells.

CellChat was employed to further dissect cell-cell communication using three distinct modes. With a minimum of two cells required for communication and all other parameters set to their default values, the major signaling pathways involved in cell-cell interactions were identified. Three modes of interaction were investigated: Secreted Signaling, ECM-Receptor, and Cell-Cell Contact. For each signaling pathway, a heatmap, circle plot, chord plot, and bubble plot were generated. CellChat analysis[33] typically requires a minimum of 10 cells per cell type[34]. However, due to the limited number of total cells in our analysis, we adjusted this parameter to a minimum of 2 cells per cell type.

Intracellular communications: Gene regulatory network

We employed the pySCENIC (Single-Cell rEgulatory Network Inference and Clustering) workflow (ver. 0.12.1) on the Anvil Platform to analyze gene regulatory networks (GRNs) and assess regulon activity. Our analysis leveraged cisTarget databases for ranking and scoring, a TF list and a motif enrichment table. First, we identified co-expressed genes based on enriched DNA motifs within 10 kb or 500 bp of the transcription start site (TSS) using RcisTarget databases. The pySCENIC's GRNBoost2 algorithm then inferred GRNs to identify co-expression modules. Next, cisTarget motif enrichment analysis was performed on these modules to detect regulons. We used motif rankings and scores to identify TF binding motifs, which were subsequently matched to TFs using the motif enrichment table. Regulon activity was quantified using AUCell. This method ranks cells based on their regulon gene expression and calculates the area under the recovery curve (AUC) to assess activity. Activity scores were transformed into binary matrices, and regulon specificity scores (RSS) were computed using the default settings. These scores provide insights into regulon specificity and activity across cellular contexts. To find master regulators we used binarize.matrix function on its pySCENIC default threshold. We extracted master regulators and their gene targets with the importance score greater than two list from adjacencies.tsv file generated by the GRNboost2 algorithm of pySCENIC. Finally, we imported our extracted gene list in Cytoscape (ver. 3.10.1) and visualized the predicted intracellular GRN by defining nodes and edges.

Results

We confirmed cell types from the annotation and unsupervised learning of 68 cells.

We reproduced the old PCA results (**Figure 2A**) and performed UMAP (**Figure 2B**) to distinguish four cell types associated in the fusion experiments.

We performed Clustree, an unsupervised clustering method, which resulted in three major cell types (**Figure 2C**). To build a clustering tree, we tracked how cells group as clustering resolution increases. Each cluster forms a node, and edges connect clusters from lower to higher resolutions, showing how clusters are related. This helps identify distinct and unstable clusters. Our unsupervised learning identified three distinct cell types.

We used the FindAllMarkers function in the Seurat package to assign the most suitable cell types. First, we listed all marker genes for each cell type. The top 10 representative genes were visualized in bubble plots. HC was used to visualize differentially expressed genes of four annotated cell types (**Figure 2D**) and three Clustree-defined cell types (**Figure 2E**).

Using Venn diagrams, we counted the overlapping and exclusive genes, and the numbers are presented in **Figure 2F**. Cluster 0 contained genes similar to those expressed in both mHL1 and mHL1 fusion. However, 246 genes were unique to neither cell type. In contrast, mMSC fusion and mMSC shared a large fraction of genes with Cluster 1 and Cluster 2, respectively. As demonstrated in **Figure 2F**, we found that mHL1 is less adaptable to cell fusion than mMSC, which readily acquired hybrid cells. Clustree aggregated the gene expression from both mHL1 fusion and mHL1, treating these two cell types as a single cluster. Certain genes like *Myl7*, *Tnni3*, *Tnnt2* and *Myl1* showed predominantly higher expression in Cluster 0. Overall, these findings suggest that mHL1 cells exhibit a less integrated and adaptable response to fusion compared to mMSC cells, which demonstrate a smoother transition and more successful integration of the fusion process.

Using EnrichR and the specific Tabula Muris atlas, enrichment analysis revealed that the gene list for mHL1 was enriched for cardiac muscle cell heart (CL:0000746). In contrast, mHL1 fusion was enriched for bladder cell bladder (CL:1001319), epithelial cell fat (CL:0000066), and epithelial cell trachea (CL:0000066). mMSC was enriched for mesenchymal stem cell of adipose fat (CL:0002570) and pancreatic stellate cell pancreas (CL:0002410). Finally, mMSC fusion was enriched for mesenchymal cell bladder (CL:0008019), mesenchymal stem cell muscle (CL:0000134), and stromal cell trachea (CL:0000499). As expected, Cluster 0 was identified as cardiac muscle cell heart (CL:0000746). Cluster 1 was identified as mesenchymal stem cell muscle (CL:0000134), mesenchymal cell bladder (CL:0008019), and fibroblast heart (CL:0000057). Cluster 2 was identified as mesenchymal stem cell of adipose fat (CL:0002570), pancreatic stellate cell pancreas (CL:0002410), mesenchymal stem cell muscle (CL:0000134), and fibroblast heart (CL:0000057). **Table S1** and **S2** provide a comprehensive list of the genes utilized to identify cell types using EnrichR for the four cell types and the three cell types, respectively.

We also applied our scRNA-seq data to WGCNA, a systems biology method that identifies highly interconnected groups of genes (called "modules") within a dataset. These modules often represent genes involved in the same biological processes. Specifically, we applied high-dimensional WGCNA (hdWGCNA)[29,30], as it is better suited for handling the complexity and sparsity of single-cell data, where many genes have low expression levels in most cells. From the list of Module Eigengenes (MEs) (**Figure 2G**), we observed distinct and overlapping MEs across all four cell types. Hub genes within M13, associated with muscle contraction and metabolism (e.g., *Art3*, *Myl1*, *Mlf1*, *Myl7*, *Tnni3*, and *Tnnt2*), were linked to both mHL1 and mHL1 fusion. In M20, hub genes were associated with extracellular matrix (ECM) formation (*Bgn*, *Col1a2*, *Ecm1*), antioxidant defense (*Gpx8*, *Tm4sf1*), and lipid metabolism (*Serpine2*, *Anxa1*, *S100a6*). M13 (**Figure 2H**) was primarily represented by cardiac muscle cell heart (CL:0000746), aligning with mHL1 and Cluster 0. In contrast, M20 (**Figure 2H**) was primarily represented by mesenchymal cell bladder (CL:0008019), aligning with Cluster 1, which

exhibited a greater distance from Cluster 2. In addition to M13 and M20, we found that both mHL1 fusion and mMSC fusion are associated with M10 and M15, both of which represent skeletal muscle satellite cells (CL:0000594). **Figure 2I** displays the gene co-expression networks for each hdWGCNA-detected module, constructed using the top three hub genes. Modules M13 (hub genes: *Fabp3*, *Art3*, and *Sh3bgr*) and M20 (hub genes: *Bgn*, *Gpx8*, and *Col1a2*) are specifically highlighted. **Table S3** provides a comprehensive list of the genes utilized in EnrichR.

Since 30 parental cells (15 of each mHL1 and mMSC) were never co-cultured, we hypothesized that excluding these 30 cells from the 68-cell dataset would reveal distinct cell type identities with less pronounced gene expression differences between mHL1 and mMSCs. As shown in **Figure S1A**, unsupervised Clustree analysis resulted in two clusters: Cluster 0, identified as fibroblast heart (CL:0000057), and Cluster 1, identified as cardiac muscle cell heart (CL:0000746). Both mHL1 and mHL1 fusion cells were identified as cardiac muscle cell heart (CL:0000746). Upon removal of the 30 non-co-cultured parental cells, mMSCs were newly identified as stem cells of epidermis skin (CL:1000428), while mMSC fusion cells remained classified as fibroblast heart (CL:0000057). **Table S4** and **S5** provide a comprehensive list of the genes utilized to identify cell types using EnrichR for the four cell types and the two cell types, respectively. These changes were also reflected in the top 10 differentially expressed genes (DGEs) (**Figure S1B** and **S1C**). Venn diagrams revealed altered numbers of overlapping and exclusive genes compared to **Figure 2F**. As demonstrated in **Figure S1D**, the distinction of cell types became less clear with this grouping, likely due to the removal of parental cells. Therefore, in the following sections, we investigated the intricate interactions of these cell types, analyzing 38 cells using both our annotation and Clustree clustering, and 28 cells using only our annotation (as Clustree clustering was not possible with this smaller dataset).

<u>Using CellCall, we identified unique intercellular communication patterns and associated TFs specific to each cell type.</u>

All cells in the co-cultures (38 cells) were considered for intercellular communications. While neither type of parental cell was induced for cell fusion[23], communication between these groups of cells is still feasible, as depicted in **Figure 1**. First, we investigated all types of cell-cell communication, including their associated TFs. CellCall incorporates a scoring system that reflects the combined influence of ligand, receptor, and TF activity[35]. As demonstrated in **Figure 3A**, not all cell types played the role of receptors or ligands. Surprisingly, in terms of receptor-ligand interactions, mMSC only contributed as a ligand for other cell receptors. **Figures 3B** and **3C** demonstrate ligand- and receptor-dependent enrichment of the Wnt signaling pathway. When mMSC fusion acts as the receiver, there is significant enrichment of "signaling pathways regulating pluripotency of stem cells," with the exception of self-interaction (mMSC fusion to mMSC fusion). Notably, all interactions resulted in enrichment of "proteoglycans in cancer" pathways. While "pathways in cancer" were enriched when mHL1 fusion served as the receiver, this enrichment was not observed when mMSC fusion was the receiver.

"FoxO signaling pathways" show a similar pattern to "human T-cell leukemia virus 1 (HTLV-1) infection," "cellular senescence" and "human cytomegalovirus infection (HCMV)" in mMSC fusion self-interaction, mHL1 fusion – mMSC fusion and mMSC – mMSC fusion. Both HTLV-1 and HCMV infections can impact the FoxO signaling pathway. HTLV-1 primarily disrupts FoxO activity, which may contribute to viral pathogenesis[36]. HCMV infection may modulate FoxO activity, potentially influencing antiviral responses[37]. These disruptions, whether through inhibition or activation of FoxO, can contribute to cellular senescence. "Endocrine resistance" pathways were enriched in all cell-cell interactions where mMSC fusion acts as the receiver.

Similarly, "EGFR tyrosine kinase inhibitor resistance" was enriched when mHL1 is a receiver. EGFR tyrosine kinase inhibitor resistance occurs when cancer cells, particularly in lung

cancer, develop ways to evade drugs designed to block the EGFR protein[38]. The resistance typically develops when cancer cells either mutate the EGFR receptor so drugs cannot bind properly or when they activate alternative growth pathways to bypass the blocked EGFR signal, which poses a significant challenge in cancer treatment. Interestingly, mMSC fusion self-interaction also moderately enriched the "ErbB signaling pathway," indicative of increased cell growth/division, resistance to death, altered differentiation, and potentially leading to the development and progression of cancer[39]. Additionally, mHL1 fusion – mMSC fusion moderately enriched "breast cancer" pathways as well.

Analysis revealed distinct transcriptional profiles between parental cells and fusion products. Parental cells exhibited limited (mHL1, **Figure 3D**) or no TF association (mMSC), while fusion products displayed an enrichment of TFs involved in cell-cell communication. Enrichment analysis identified PPARδ, a key regulator of cardiac metabolism in HL1 cardiomyocytes[40], as a prominent TF in mHL-1 fusions. Furthermore, Creb, a key mediator of cardiac hypertrophy[41], and Stat3, a critical regulator of antioxidant defense, anti-inflammation, and cell survival[42,43], showed a biphasic pattern in enrichment analysis. Creb and Stat3 may exert dual regulatory roles on mHL1, activating genes that promote the mHL1 phenotype (as indicated by the initial positive ES) and subsequently repressing genes that inhibit it (reflected in the later negative ES). This unexpected pattern may be attributed to the limited sample size of parental cells (5 of each mHL1 and mMSC) used in the CellCall analysis.

In mHL1 fusion cells, a group of TFs, including E2f4, Egr1, Elk4, Ep300, Fos, Fosl1, Jun, and Trp53, exhibited similar patterns of upregulation with relatively small variations in enrichment scores (ES) (**Figure 3E**). These TFs, as assessed through TFCheckpoint 2.0[44], primarily regulate cell cycle progression (E2f4), mediate cellular responses to stimuli (Egr1), influence cell growth and differentiation (Elk4), and regulate gene expression through chromatin remodeling (Ep300). Additionally, they include key components of the AP-1 complex (Fos, Fosl1, Jun) and the tumor suppressor p53. In contrast, TFs associated with mMSC fusion cells (Ctnnb1, Fos, Lrf1, Jund, Myc, Pbx1, Smad3/4, Sp1, and Stat6) demonstrated similar patterns of random walks (**Figure 3F**) in comparison to mHL1 fusion ones. This group exhibits a broader range of functions, encompassing roles in Wnt signaling (Ctnnb1), TGF-β signaling (Smad3/4), immune responses (Stat6), development and organogenesis (Pbx1), and cell growth and proliferation (Myc, Fos, Jund). These findings suggest that mHL1 fusions may prioritize cell cycle regulation and growth, while mMSC fusions exhibit a more diverse transcriptional landscape, potentially reflecting the stem cell-like characteristics of mMSCs.

Reducing intercellular communication between cell types (**Figure S1**) simplified receptor-ligand interactions, limiting the roles of both ligands and receptors (**Figure S2A**). **Figure S2B** shows a low NES for the "Wnt signaling pathway" in Cluster 0 self-interaction. "Tight junction" pathways, enriched when Cluster 1 acted as the receiver in the two-cell-type condition (Clustree), were absent not in the four cell types conditions (annotation, **Figure 3**). "Signaling pathways regulating pluripotency of stem cells" remained enriched with Cluster 0 as the receiver in both conditions. A key difference in the two-cell-type scenario (Clustree) was the emergence of a significant, albeit low-NES, enrichment of the JAK-STAT signaling pathway, not prominent in the four-cell-type analysis (annotation, **Figure 3**). All "endocrine resistance" pathways were significantly enriched in the two-cell-type condition. In the four-cell-type context (annotation, **Figure 3**), "endocrine resistance" was enriched when either fusion cell acted as the receiver. Reducing cell-cell communication interactions from sixteen to four yielded more distinct cell-cell communication scores (**Figure S2C**) compared to **Figure 3C**, enabling clearer discrimination of cell-cell communication patterns across cell types.

The association of Cluster 0 (**Figure S2D**) with a diverse array of TFs suggests a complex and multifaceted regulatory landscape. Specifically, the presence of Ctnnb (involved in Wnt signaling), Fos and Myc (immediate early response and cell proliferation), Foxo1 and Hif1a (stress response and metabolism, including hypoxia), Mitf (melanocyte development, though

potentially broader roles exist), Pbx1 and Sox9 (development and differentiation), Rela, Stat1, and Stat6 (immune response and inflammation), Smad2 and Smad4 (TGF-beta signaling, impacting cell growth and differentiation), and Sp1 (general transcription regulation) implies that Cluster 0 cells may exhibit active Wnt signaling, responsiveness to environmental stimuli, metabolic adaptation (potentially including stress responses like hypoxia), a capacity for differentiation, and potential roles in immune modulation. The TFs associated with Cluster 1 also suggest a distinct regulatory profile. Specifically, E2F4 (a transcriptional repressor involved in cell cycle control), Mybl2 (involved in DNA replication and cell cycle progression), and Rb1 (a major cell cycle checkpoint protein) point to active cell cycle regulation. Gata4 (important for differentiation) suggests potential for lineage specification. Fos (immediate early response and cell proliferation) indicates responsiveness to stimuli. NCoa3 (a transcriptional coactivator) suggests enhanced activity of other TFs. Srebf1 (a key regulator of lipid metabolism) implies a distinct metabolic profile. Finally, Stat1, Stat3, and Stat5a (members of the STAT family involved in signal transduction, particularly cytokine and growth factor signaling) point to active signaling pathways. The combined presence of these TFs suggests Cluster 1 cells are characterized by active cell cycle regulation, differentiation potential, responsiveness to stimuli, a distinct metabolic signature (particularly lipid-related), and active signal transduction.

Analysis of 28 cell-cell communication networks, comprised of two fusion product types (**Figure S3**), revealed significant shifts in communication patterns compared to both the four-cell-type (annotation, **Figure 3**) and two-cell-type (Clustree, **Figure S2**) conditions. These alterations may be attributed to the limited number of interaction permutations. As shown in **Figure S3A**, mHL1 fusion cells lack a receptor function, while mMSC fusion cells lack a ligand function. Notably, five pathways (**Figure S3B**) were removed in: "human cytomegalovirus infection," "growth hormone synthesis, secretion and action," "EGFR tyrosine kinase inhibitor resistance," "chemokine signaling pathway," and "cGMP-PKG signaling pathway." In mMSC fusion self-interactions, the "Jak-STAT signaling pathway" became more significant (smaller p-values), while "cellular senescence" exhibited a lower NES. In mMSC fusion-mHL1 fusion interactions, the "MAPK signaling pathway" became less significant. Conversely, in mHL1 fusion to mMSC fusion interactions, the "breast cancer" pathway demonstrated an increase in NES. Cell-cell communication analysis (**Figure S3C**) revealed distinct patterns for each of the four cell-cell communication types, clearly distinguishable from both each other and the patterns observed in the 38-cell, two-cell-type condition (Clustree, **Figure 3**).

TF associations varied between mHL1 fusion and mMSC fusion cells. In mHL1 fusion cells (**Figure S3D**), E2f4, Egr1, Ep300, Fos, Fosl1, Jun, Nr1h2, Pbx1, and Trp53 were identified. Elk4 appeared in **Figure 3E** was removed from this group, while Nr1h2 and Pbx1 were newly associated. In mMSC fusion cells (**Figure S3E**), Ctnnb1, Fos, Hif1a, Irf9, Jund, Myc, Pbx1, Smad3/4, Sp1, Srebf1, and Stat6 were identified. Hif1a, Irf9, and Srebf1 were newly associated, while Lrf1 (**Figure 3F**) was removed. Notably, Pbx1 was associated with both mHL1 and mMSC fusions. In mHL1 fusions, the loss of Elk4, involved in cell growth and differentiation, and the gain of Nr1h2, a regulator of lipid metabolism and inflammation, suggest potential shifts in cellular behavior. The retention of Pbx1 in both fusion cell types suggests its importance in the fusion process. In contrast, mMSC fusions exhibited the gain of Hif1a (hypoxia response), Irf9 (immune signaling), and Srebf1 (lipid metabolism), and the loss of Lrf1, involved in cell cycle regulation and differentiation. These findings suggest that mHL1 fusion cells may exhibit altered lipid metabolism and inflammatory responses, while mMSC fusion cells may exhibit changes in hypoxia response, immune function, lipid metabolism, and cell cycle control.

<u>CellChat Analysis Reveals Distinct Modes of Intercellular Communication Among All Cell Types</u>
All ligand-receptor pairs were probed, and the most prominent cell-cell communication interactions were selected based on their relative contributions (**Figure S4-6**). **Figure 4A** highlights secreted molecules such as Spp1, Galectin, and Angptl, which bind to specific

receptors like ITGB and CD44 on target cells. Notably, the Spp1 (osteopontin) pathway lacks cell-cell communication between mHL1 and mHL1 fusion cells. In contrast, Angptl (Angiopoietin-like proteins) demonstrates exclusive activity as a ligand from mMSC fusion cells to all other cell types. The finding that Spp1 communication is absent between mHL1 and mHL1 fusion cells suggests that these cells may have altered bone remodeling or immune responses compared to other cell types. The observation that Angptl acts exclusively as a ligand from mMSC fusion cells to other cell types indicates that mMSC fusions may play a significant role in regulating lipid metabolism, angiogenesis, or other processes influenced by Angptl proteins. Interestingly, Galectin was more prominently involved in communication between mHL1 and mHL1 fusion cells.

Figure 4B highlights key ECM components and their associated cell surface receptor pathways, focusing on Collagen, Thrombospondin (Thbs), Fibronectin (Fn1), Tenascin, and Laminin. An expanded network of collagen-receptor interactions is visualized in **Figure 4C**. Molecules such as Collagen, Fn1, and Laminin interact with integrin receptors, enabling cell adhesion and signal transduction. While mMSC fusion cells primarily drove all of these five ECM-Receptor interactions, the collagen pathway exhibited complex interactions involving all four cell types, suggesting that all cell types contribute to collagen secretion and mediate fusion events. Notably, Thbs and Laminin pathways also demonstrated involvement of mHL1 and mHL1 fusion cells.

We identified six distinct types of Cell-Cell Contact (**Figure 4D**), representing direct physical interactions between cells mediated by surface proteins. These pathways play crucial roles in intercellular communication. Jam and Desmosomes are essential for maintaining cell-cell adhesion and tissue integrity, particularly critical for cardiomyocytes. EphA/EphB receptors regulate cell migration and signaling, influencing interactions within and between cell populations. Cadm proteins contribute to diverse functions, including cell adhesion and signal transduction. While Icos is primarily involved in T cell activation[45] and not typically expressed in cardiomyocytes or MSCs under normal conditions, its expression and function may be altered in fusion cells.

Unsupervised clustering of the 38 cells into two clusters using Clustree revealed significant alterations in cell-cell communication patterns (**Figure S7**), despite largely unchanged ECM-Receptor pathway associations (**Figure S7B**). While the relative contributions of ligand-receptor pairs remained somewhat similar, Secreted Signaling pathways (**Figure S7A**) exhibited a shift in pathway ordering (Galectin, Spp1, Angptl) and the addition of Ptn (Pleiotrophin) and Fgf (Fibroblast Growth Factor). Ptn, a pleiotropic growth factor, plays crucial roles in various biological processes, including hematopoietic stem cell maintenance, adipocyte differentiation, chondrocyte development, endothelial cell migration, and leukocyte activation[46]. FGFs exert their effects through binding to Fibroblast Growth Factor Receptors (FGFRs) on the cell surface, activating downstream signaling pathways such as MAPK and PI3K/Akt[47]. ECM-Receptor pathways and their order, however, remained consistent (**Figure S7B**). Cell-Cell Contact pathways (**Figure S7C**) were reduced from six to two (Jam and Cdh), likely due to the change in cell type annotation from four to two clusters. The reduction in cell types consequently decreased the number of identified ligands and receptors. In contrast to the relatively diverse ligand-receptor interactions observed previously (**Figure 4**), cell-cell communication in this two-cluster configuration (**Figure S7**) was primarily driven by Cluster 0 in Secreted Signaling and ECM-Receptor interactions, and by Cluster 1 in Cell-Cell Contact. While exhibiting some alterations, the complex Collagen pathway remained the most intricate ligand-receptor interaction in both cell type annotations.

To evaluate the influence of parental cells, we re-analyzed the data with CellChat after their removal. Further reducing the fusion products to two cell types (28 cells) resulted in additional shifts in the cell-cell communication landscape (**Figure S8**), exhibiting some similarities to the two-cell-type configuration of the 38-cell analysis (**Figure S7**). In Secreted

Signaling (**Figure S8A**), Galectin was removed, while Ptn and Fgf pathways, driven by mMSC fusion, were identified. Notably, the significance of Spp1 and Angptl pathways was significantly diminished. In ECM-Receptor interactions (**Figure S8B**), the overall pathway landscape remained largely unchanged, although the complexity of cell-cell communication was simplified or removed. Notably, mMSC fusion cells continued to drive the majority of ECM-Receptor interactions. In Cell-Cell Contact (**Figure S8C**), the analysis was significantly simplified, with only Jam and Cadm pathways retained, primarily driven by mHL1 fusion cells. The appearance of Cdh (cadherin) is noteworthy, as classical cadherins mediate strong cell-cell adhesion through homophilic interactions, while Cadm proteins exhibit more diverse functions[48]. Importantly, EphA/B, Desmosome, and Icos pathways were not observed in the context of two fusion cell type analysis.

SCENIC Identified Gene Regulatory Networks and Cell Type-Specific TFs

Before the accidental fusion event, the 30 parental cells (mHL1 and mMSC) exhibited distinct transcriptional profiles. UMAP analysis revealed clear separation between these cell types (**Figure S9A**). While some target genes were shared between TFs such as Nfia, Hmga2, Elf1, Gata4, Sp8, Sox11 and Fos (**Figure S9B**), five TFs associated with each cell type demonstrated non-overlapping expression patterns (**Figure S9C-D**). The RSS assesses the specificity of a regulon (genes regulated by a specific TF) to a particular cell type. High RSS values indicate that the regulon is predominantly active within that cell type, suggesting a crucial role in defining its unique characteristics. Master regulators, identified based on their influence and connectivity within the GRN, often exhibit high connectivity and drive the expression of many other genes. However, RSS emphasizes the cell-type specificity of a TF's regulon. A TF can be a master regulator due to its network influence but may not exhibit high cell-type specificity (lower RSS values). Conversely, a TF with a highly cell-type-specific regulon (high RSS values) might not possess the same level of overall network influence as a master regulator. In **Figure S9E**, all five mHL1 master regulators were identified, while Hmga2 and Rlf1, despite their master regulator status, were absent from the top RSS list. This observation highlights the distinction between TFs with high network influence (master regulators) and those with high cell-type specificity, as defined by their RSS values.

Our GRN analysis of 38 cells, including 10 parental cells, identified four distinct cell types, despite the presence of two main clusters (**Figure 5A**). Notably, fusion products retained associations with their parental cell types. We identified 12 unique master regulators (**Figure 5B-D**). SCENIC identifies master regulators through a three-step process: 1) identifying co-expressed gene modules between TFs and potential target genes; 2) refining these modules by analyzing motif enrichment within the target genes to pinpoint direct TF-target relationships; and finally, 3) defining "regulons," which represent a TF and its set of direct target genes. A master regulator is considered a TF with a highly active regulon across a significant portion of the cell population, indicating its substantial influence on gene expression within that specific cell type[49]. Cell-type-specific master regulators included Gata4 and Zfp764 in mHL1, Zfp35 and Sox7 in mMSC, and Fli1 in mMSC fusion products. The remaining master regulators were associated with more than two cell types.

Among these, Pax3[50] and Nkx2-1[51], known for their roles in neural crest and respiratory system development, respectively, likely have minimal roles in mHL1. Zbtb33[52] might potentially influence gene expression within these cells, although its specific function remains unclear. In contrast, Gata4 and Zfp764 are crucial for cardiac development and function. Gata4 is a master regulator essential for cardiomyocyte differentiation and regulates genes involved in contractility, hypertrophy, and development[53]. Zfp764, encoding the transcription factor ZNF764, functions as both a DNA-binding regulator of gene expression and a cofactor that modulates glucocorticoid receptor (GR) activity, impacting GR binding, gene expression, and distinct biological pathways[54].

mHL1 fusion cells did not exhibit a distinct master regulator, suggesting potential deviation from the developmental and differentiation pathways observed in mMSC or mMSC fusion. Given their primary roles in neural crest development (Pax3), respiratory system development (Nkx2-1), and other developmental processes (Dlx1)[55], Pax3, Nkx2-1, and potentially Dlx1 may have limited roles in mHL1 fusion cells. Zbtb33 might influence gene expression related to cell growth and survival, while Gabpa could play a role in regulating cellular energy metabolism[56].

TFs such as Tcf4, Rarg, Zfp35, Dlx1, and Sox7 play crucial roles in regulating various aspects of mMSC biology. Tcf4, a key component of the Wnt/β-catenin signaling pathway, influences self-renewal, differentiation, and proliferation[57]. Rarg mediates the effects of retinoic acid, impacting adipogenic and osteogenic differentiation[58]. Zfp35[59] is involved in cellular processes such as cell growth, differentiation, and apoptosis, potentially influencing MSC proliferation and differentiation. Dlx1[55], known for its role in craniofacial development and neurogenesis, may be involved in regulating MSC differentiation, particularly towards osteogenic lineages. Sox7[60], implicated in neurogenesis and chondrogenesis, may play a role in chondrogenic differentiation of MSCs.

In mMSC fusion, Fli1 was the only non-overlapping TF[61]. Fli1, crucial for hematopoiesis and angiogenesis, may regulate vascularization and influence cell proliferation within the fused cell. As demonstrated in **Figure 5B**, the extent of target gene overlap was significantly less pronounced compared to that observed with 30 parental cells (**Figure S9B**). Notably, shared target genes included Aspn between Tcf4 and Fli1, Pkfp between Rarg and Gata4, and Hectd2 between Zfp35 and Dlx1.

As shown in **Figure 5E**, mMSC fusion cells generally exhibited high RSS values compared to other cell types. Notably, mHL1 and mHL1 fusion cells shared a higher number of regulons with mid-to-high RSS values, albeit with some distinctions. This discrepancy is evident in our analysis, where many overlapping master regulators (e.g., Zbtb33, Nkx2-1, Dlx1, Gabpa, Tcf4) were not found among the top RSS TFs, with the exception of Fli1, highlighting the complex interplay between TF influence and cell-type specificity. While some master regulators may exhibit high network influence but low cell-type specificity, and vice versa, this analysis demonstrates a discrepancy between these two sets of TFs.

Reducing the number of cell types from four to two resulted in changes across all analyzed TF regulators. While the master regulators of Cluster 0 aligned with those in mMSC fusion, Cluster 1's regulators aligned with mHL1 fusion (**Figure S10A**). Although the GRNs were altered (**Figure S10B**), the UMAP projection, based on individual cells, remained consistent (**Figure S10C**). Specifically, Zfp35, Dlx1, Zfp764, and Sox7 were removed. Loss of Zfp35, potentially involved in neuronal development, could disrupt neuronal function or differentiation. Absence of Dlx1, crucial for forebrain development and GABAergic interneuron formation, would likely have significant developmental consequences. Zfp764, functioning as both a DNA-binding regulator of gene expression and a cofactor modulating GR activity[54], was also lost. Similarly, removal of Sox7, involved in vascular development, endoderm differentiation, and other processes, could disrupt these developmental pathways. Consequently, the top 10 RSS also changed (**Figure S10D**). Therefore, this reduction in cell types and the subsequent TF landscape changes suggest a substantial influence of parental cells on *in vitro* development and differentiation.

Upon excluding the 10 parental cells for SCENIC analysis, distinct clusters were still evident in UMAP (**Figure S11A**). While similar patterns of GRNs emerged, only one shared target gene (Ctsz) was identified between Fli1 and Hmga1 (**Figure S11B**). Notably, multiple new master regulators emerged: Trp53, Usf1 (overlapped), Rfxap, Zfp754, and Hoxa1 in mHL1 fusion cells, and Hmga2, Elk3, and Hoxb4 in mMSC fusion cells. Fli1 remained a master regulator, as observed in **Figure 5C**. Interestingly, no mHL1 fusion master regulators were identified among the top RSS TFs (**Figure S11C-D**), and the overlapping Usf1 TF was also

absent from this list (**Figure S11E**). TFs associated with mHL1 fusion play diverse roles in cellular processes. Trp53 (Tumor Protein 53), often referred to as the "guardian of the genome," regulates cell cycle arrest and apoptosis in response to cellular stress[62]. Usf1 (Upstream Stimulatory Factor 1) is involved in a wide range of cellular processes, including cell growth and differentiation[63]. Rfxap (Regulatory Factor X Associated Protein) regulates gene expression in various tissues, including those involved in immune responses[64]. Zfp574 (Zinc Finger Protein 574) is a transcription factor that plays a crucial role in the survival and development of lymphocytes, a type of white blood cell[65]. It interacts with another protein called THAP12, and together they promote the transition of cells from the G1 to the S phase of the cell cycle, which is essential for cell division and proliferation[65]. Hoxa1 (Homeobox A1), a key player in embryonic development, regulates cell differentiation and organogenesis[66]. As demonstrated in **Figures S11D** and **S11E**, TFs associated with either of the fusion products exhibited closer relationships or higher relevance compared to those shown in **Figure S9** (all parental cells) or **Figure 5** (four cell types).

Discussion

A primary objective of this study was to characterize the cell types resulting from accidental fusion between mHL1 and mMSC. We hypothesized that fusion would rapidly induce novel phenotypes, exhibiting biased redistribution of gene expression and cell mass rather than a symmetrical exchange. This bias, we further posited, would be detectable in intercellular and intracellular communication patterns. Initial analysis using PCA and UMAP distinguished four cell types: mHL1, mHL1 fusion, mMSC, and mMSC fusion. Clustree, an unsupervised clustering method, grouped these into three clusters, combining mHL1 and mHL1 fusion while separating mMSC and mMSC fusion. EnrichR analysis, based on the four PCA/UMAP-defined groups, revealed distinct tissue and cell type annotations: cardiac muscle for mHL1, bladder and epithelial cells for mHL1 fusion, mesenchymal stem cells and pancreatic stellate cells for mMSC, and mesenchymal cells, mesenchymal stem cells, and stromal cells for mMSC fusion. These annotations were largely confirmed by EnrichR analysis of the three Clustree clusters. Recognizing the limitations of differential expression analysis in noisy, variable single-cell RNA-seq data, particularly with small sample sizes[67-69], we employed a complementary approach using WGCNA. Specifically, we utilized hdWGCNA[29,30] to identify gene modules associated with distinct biological processes. This analysis linked mHL1 and mHL1 fusion to muscle contraction and metabolism, and mMSC fusion to ECM formation, antioxidant defense, and lipid metabolism. Strikingly, removing parental cells prior to re-clustering with Clustree drastically altered cell type assignments, collapsing the four original cell types into two and highlighting the profound influence of parental cell presence on the final cellular composition and characteristics observed after fusion. These results underscore the value of integrating multiple cell type identification methods and demonstrate the significant impact of parental cell influence on fusion outcomes.

We investigated intercellular communication across three cell group configurations using CellCall: 38 cells representing four distinct cell types, 38 cells representing two cell types defined by Clustree, and 28 cells representing two fusion cell types. Parental cells, not cultured prior to fusion induction, were excluded from all communication analyses. The four-cell system, with its complex interaction network, revealed mMSC primarily acting as a ligand and mHL1 as both ligand and receptor, alongside a broad range of enriched pathways and distinct TF profiles in each fusion type. Reducing to two Clustree-defined cell types simplified interactions and altered pathway enrichment, notably impacting Wnt and JAK-STAT signaling. Further restricting to two fusion cell types dramatically limited communication, with mHL1 losing receptor function and mMSC losing ligand function. This restriction resulted in the loss of several pathways, changes in the significance of others, and distinct shifts in TF associations, suggesting altered cellular behavior. For example, mHL1 fusion cells lost Elk4 (cell growth and differentiation) and

gained Nr1h2 (lipid metabolism and inflammation), while mMSC fusion cells gained Hif1a (hypoxia), Irf9 (immune signaling), and Srebf1 (lipid metabolism) and lost Lrf1 (cell cycle and differentiation). These findings underscore the profound influence of cell type composition and interaction potential on cellular behavior and gene regulation.

CellChat analysis revealed distinct intercellular communication patterns across the three cell group configurations. In the four-cell system, a diverse repertoire of Secreted Signaling molecules, including Spp1, Galectin, and Angptl, mediated complex interactions, with mMSC primarily acting as a ligand and mHL1 as both ligand and receptor. ECM-Receptor interactions, largely driven by mMSC fusions, involved key components like Collagen. Six distinct Cell-Cell Contact pathways further contributed to this intricate communication network. Reducing to two Clustree-defined cell types altered the ordering of Secreted Signaling pathways and introduced Ptn and Fgf, which are implicated in diverse processes like cell growth[70], differentiation[71], and angiogenesis[72], while ECM-Receptor interactions remained relatively stable. Cell-cell Contact pathways, however, were substantially reduced. Further restricting to two fusion cell types dramatically simplified the communication landscape. Galectin was lost, and the significance of Spp1 and Angptl diminished in secreted signaling. While ECM-Receptor interactions remained largely unchanged in terms of pathway presence, their complexity was reduced. Most strikingly, Cell-Cell Contact pathways were drastically simplified, with the loss of several key interactions. These findings demonstrate how cell type composition and the presence of parental cells profoundly influence intercellular communication, impacting both the types of signals exchanged and the cells driving these interactions. The progressive simplification of the communication network with decreasing cell type complexity highlights the importance of considering the cellular context when studying intercellular signaling.

This study investigated the dynamic changes in TF regulation across different cell populations, progressing from distinct parental cells to fused cell types. Analysis of the parental cells revealed distinct TF profiles, highlighting the inherent differences between mHL1 and mMSC. In the four-cell system, which included both parental and fusion cells, fusion products retained TF associations with their parental origins, and twelve master regulators were identified, including cell-type-specific ones like Gata4 and Zfp764 in mHL1, Zfp35 and Sox7 in mMSC, and Fli1 in mMSC fusion. However, this analysis also highlighted the distinction between master regulators and TFs with high regulon specificity (RSS), demonstrating that network influence and cell-type specificity are not always coupled. Reducing the cell types to two using Clustree resulted in the loss of several key TFs, including Zfp35, Dlx1, Zfp764, and Sox7, and consequent changes in the top RSS TFs, suggesting a substantial influence of parental cells on the TF landscape. Most strikingly, removing parental cells entirely and analyzing only the fusion products revealed a completely reshaped TF landscape, with the emergence of new master regulators, such as Trp53, Usf1, Rfxap, Zfp754, and Hoxa1 in mHL1 fusion cells, and Hmga2, Elk3, and Hoxb4 in mMSC fusion cells. Fli1 remained a master regulator in mMSC fusion. This analysis also showed that no mHL1 fusion master regulators were among the top RSS TFs. The observed TF shifts across the different cell populations, including the specific loss and gain of certain TFs in each condition, underscore the dynamic nature of gene regulatory networks and the profound influence of cell type composition, particularly the presence or absence of parental cells, on shaping cellular identity and behavior. This highlights the importance of carefully considering the cellular context when studying TF regulation and its impact on cell fate.

Our previous analysis of cell fusion[23] revealed a significant upregulation of ribosomal protein genes in hybrid cells with unique transcriptomes, including 14 Rpl and 9 Rps family genes. This observation aligns with the involvement of key TFs known to regulate ribosomal protein expression such as Myc, the Fos/Jun (AP-1) complex, E2F[73-75]. Additionally, we observed increased expression of Ubb, a component of the ubiquitin-proteasome system, consistent with

the role of p53, a tumor suppressor protein, in regulating ubiquitination pathways[76]. Notably, our previous publication[23] highlighted the enrichment of cancer-related gene groups, including Trp53, Fos, and Jun, in certain hybrid cells. These findings are now more prominent in cell types associated with mHL1 fusion, suggesting that accidental cell fusion can lead to the acquisition of novel cellular properties, including potentially cancerous phenotypes, through random reprogramming events.

Cellular fusion initiates with the merger of two distinct parental cells. The resultant fusion products exhibited gene expression signatures associated with both stemness (Ly6a, CD44, Itgb1, Itgav, and Gnl3) and mesodermal differentiation (Spp1, Sox9, Col1a1, Scd1, and Fn1)[23]. While CD44, Itgb1, and Itgav expression was observed in Spp1-mediated Secreted Signaling (**Figure 4A**), Ly6a (encoding Sca-1, crucial for stem cell maintenance[77]) and Gnl3 (nucleostemin, involved in stemness and ribosome assembly)[78] were not detected in our primary analysis. Mesodermal gene expression was evident in both secreted (Spp1, with mMSC as sender and mMSC fusion as receiver, **Figure 4A**) and -receptor modes (Col1a1: most of the ligand-receptor interactions are between mMSC and mMSC fusion, and Fn1: mMSC fusion as sender and the other three cell types as receivers, **Figure 4B**), suggesting the significance of non-cell-cell contact signaling in fusion and differentiation. Although Sox9 (SRY-related orphan-9, related to skeletal development during embryonic development) and Scd1 (Stearoyl-CoA Desaturase 1) gene expression was not detected, the expression of Nanog was observed in fusion cells. Notably, Myc was expressed in both mHL1 fusion and mMSC fusion.

Of the 28 fusion cells, cardiac development gene expression (Hand2, Cby1, Tbx5, Tbx20, and Gata4) was modestly increased, while contractile machinery genes (Actc1, Tnni3, Cox6a2, fabp3, Cyc1, Atp1a1, Atp1b1, Des, Tnnt2, and Myh6) exhibited greater upregulation[23]. While most of these genes were not included in the primary analysis, Gata4, a master regulator of mHL1 (**Figure 5**), and Tnni3 and Tnnt2, expressed in Cluster 0 (**Figure 2E**) and M13 (**Figure 2G**), are relevant to mHL1 function. We observed Nkx2-1 TF in both mHL1 and mHL1 fusion cells (**Figure 5B-D**), contrasting with a previous report of unexpectedly high Nkx2-5 expression in a related system[23]. Nkx2-1 and Nkx2-5 regulate distinct organ development: Nkx2-1 is crucial for thyroid[79]and lung[80] development, while Nkx2-5 plays a key role in cardiac development[81]. Cell marker analysis using Tabula Muris revealed that mHL-1 expresses Mesothelial Cell Lung (CL:0000077), Stromal Cell Lung (CL:0000499) associated with mMSCs, and Endothelial Cell Lung (CL:0000115) within Cluster 0 and M13, albeit with moderate scores. Fusion cells exhibited gene expression profiles associated with developmental processes (Ncoa6, Dvl1) and contractility (Cox7a1, Des, Tnnt2, and Myh6)[23]. However, with the exception of Tnnt2, expression of these genes was not detected in our primary analysis.

Conclusions

This study provides a multi-faceted characterization of the cellular and molecular changes induced by accidental cell fusion between mHL1 and mMSC cell lines. Our initial hypothesis, that fusion would generate novel phenotypes with biased gene expression and cell mass redistribution, was supported by the observed heterogeneity in cell types and communication patterns. The initial identification of four distinct cell types (mHL1, mHL1 fusion, mMSC, and mMSC fusion) by PCA/UMAP, subsequently refined into three clusters by Clustree, highlighted the complex cellular landscape resulting from fusion. EnrichR analysis further emphasized the functional diversity of these populations, revealing distinct cell type associations. The frequent co-clustering of mHL1 and mHL1 fusion, in contrast to the separate identities maintained by mMSC and mMSC fusion, suggested differential adaptability to the fusion process between the parental cell lines. hdWGCNA analysis provided valuable insights into the biological processes associated with each cell type, linking specific gene modules to key cellular functions, such as muscle contraction and metabolism in mHL1/mHL1 fusion, and ECM formation and lipid metabolism in mMSC fusion.

A key finding was the dramatic impact of parental cell presence on observed cell type identities. Removing parental cells prior to re-clustering drastically altered cell type assignments, collapsing the original four groups into two and underscoring the significant influence of the initial cell populations on the final fused cell characteristics. This observation highlights the critical importance of considering the cellular context, particularly the presence and influence of parental cells, when studying cell fusion. It suggests that parental cells may exert a dominant influence during early fusion stages, masking or modulating the emergence of truly novel fusion-specific cell identities, which only become more apparent upon their removal.

CellCall analysis mapped intercellular communication patterns, revealing distinct interaction modes. The observation that mMSC primarily acted as a ligand, while mHL1 functioned as both ligand and receptor, suggested a directional flow of information within the fused cell population. The progressive simplification of communication networks with decreasing cell type complexity further emphasized the importance of cell-cell interactions in shaping cellular behavior. Changes in pathway enrichment and TF associations across different cell groups underscored the dynamic nature of gene regulation in response to fusion. Finally, SCENIC analysis provided insights into the transcriptional regulation underlying cell identity, highlighting the complex interplay between network influence and cell-type specificity. The emergence of new master regulators upon parental cell removal further reinforces the notion that parental cell presence can mask or modulate the expression of fusion-specific TFs. This study underscores the value of integrating multiple bioinformatics approaches and emphasizes the profound influence of parental cell presence on fusion outcomes, providing a foundation for future investigations into the mechanisms and functional roles of key regulators in fused cell behavior.

Limitation of the current study

The allure of fusion lies in its potential to facilitate reprogramming to an early mesodermal state. However, it's crucial to acknowledge that single-cell sequencing technologies typically capture only a fraction (approximately 5-10%) of the total cellular transcripts, potentially missing low-abundance genes[82], including transcription factors crucial for cell differentiation, especially in hybrid cells where obtaining and confirming bona fide replicates is challenging. Moreover, our sample sizes fall below the recommended 500-1,000 cells per sample for typical scRNA-seq analysis.

For better understanding of cell-cell fusion, future studies can incorporate RNA splicing information (RNA velocity) and pseudotime analysis. Pseudotime analysis[83], by ordering cells based on gene expression similarity, can reveal the temporal order of gene expression changes as fused cells transition to a new state. RNA velocity analysis[84], which estimates future gene expression by analyzing unspliced and spliced mRNA ratios, can predict the direction of these changes, providing insights into immediate transcriptional responses and potential cell fates. Both methods can identify key genes and regulatory pathways driving the fusion process and subsequent reprogramming. Furthermore, these analyses can uncover heterogeneity among fused cell populations, revealing distinct subpopulations with unique transcriptional trajectories and potential fates. RNA velocity can also predict future transcriptional states, providing valuable insights into potential differentiation pathways and functional outcomes, crucial for understanding the long-term consequences of cell fusion in development, disease, and regenerative medicine.

Declaration of competing interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement:

JoAnne Huang: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Review & Editing.
Fateme Nazaryabrbekoh: Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Review & Editing.
Joohyun Kim: Conceptualization, Methodology, Writing – Review & Editing.
Jangwook P. Jung: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing.

Funding: This work was supported by the National Science Foundation CAREER DMR 2047018 (JPJ).

Figures

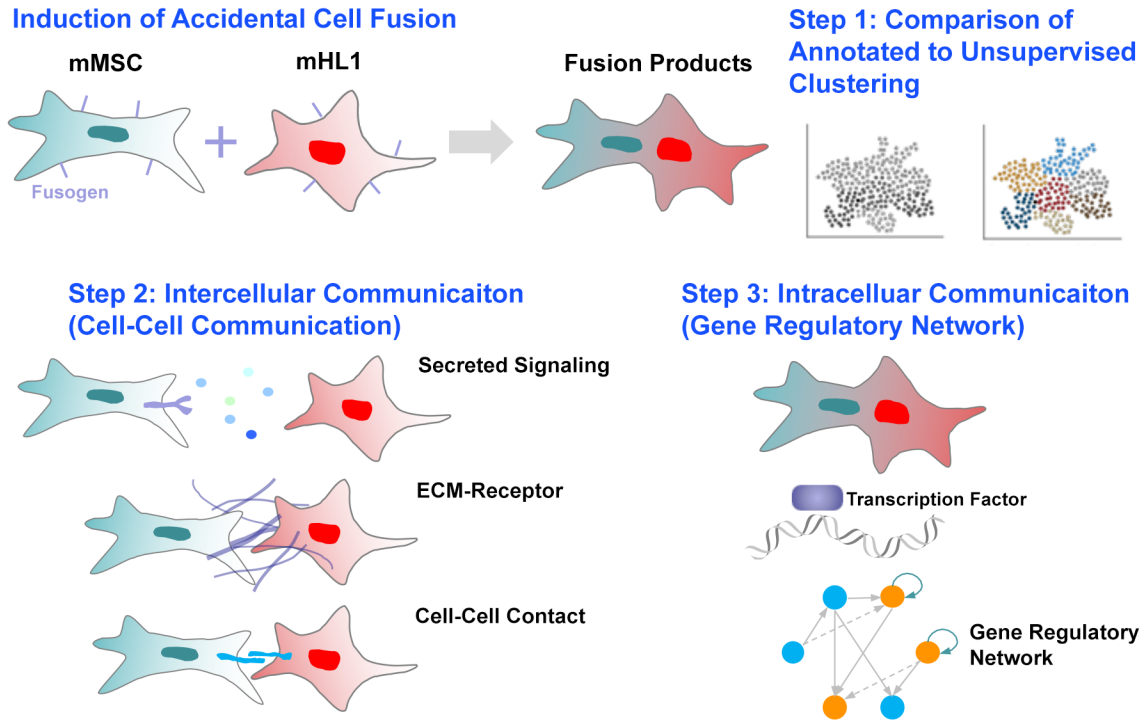

Figure 1. A schematic overview of the current study on intercellular and intracellular communications between mMSC and mHL1 upon fusion. First, we compared annotated cell types to unsupervised cell types using single-cell RNA-seq data from all samples. Subsequently, intercellular communications were investigated in co-cultured samples. Finally, intracellular communication upon fusion will be elucidated by inferring gene regulatory networks.

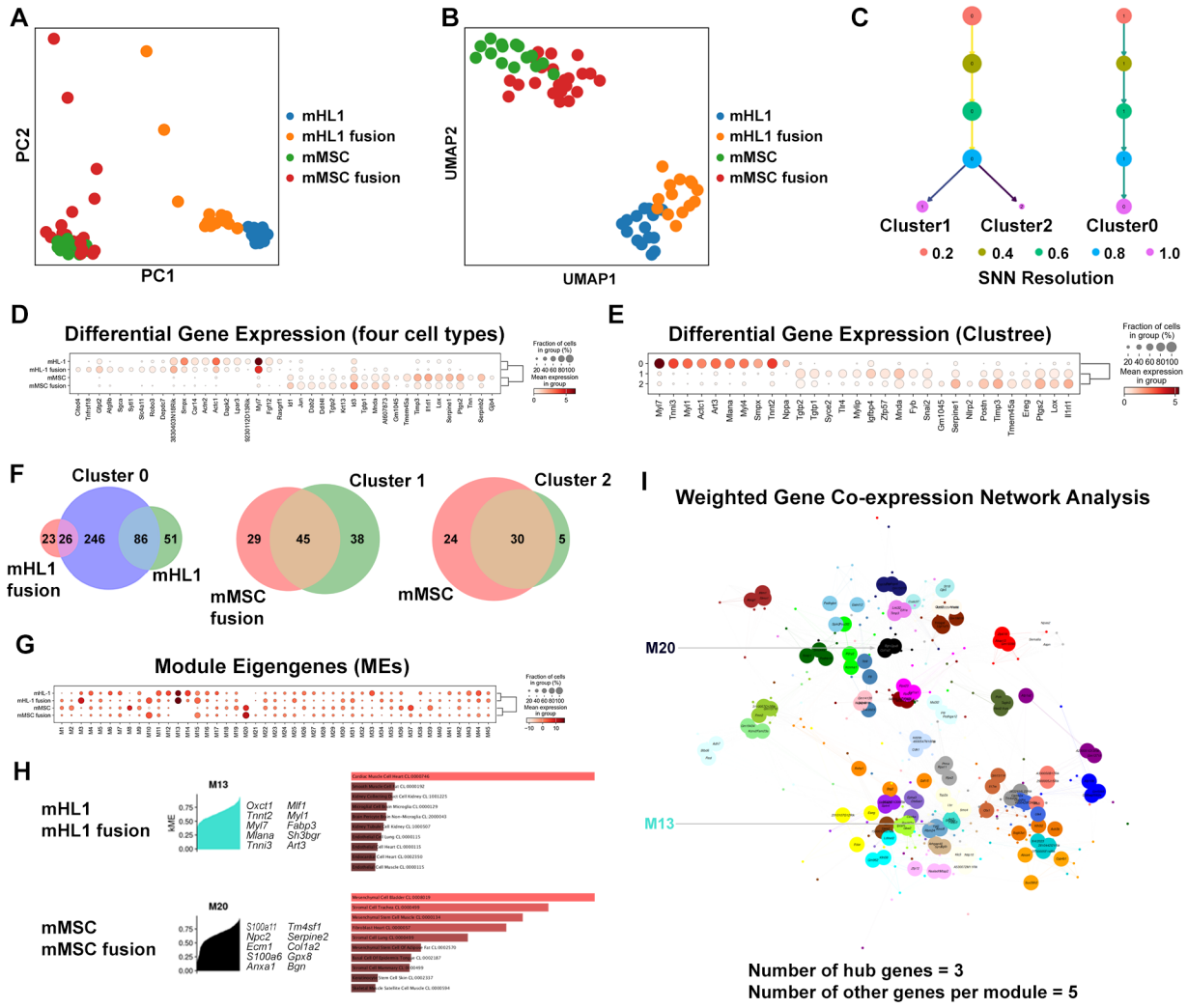

Figure 2. Analysis was performed on all 68 samples, including 30 prefusion and 38 fusion samples. Labeled cell types were identified using PCA (A) and UMAP (B). Clustree analysis revealed three distinct cell types (C). DGE analysis was performed for these four (D) and three (E) cell types. Based on these analyses, mHL1 and mHL1 fusion are distinguished but largely grouped together in Cluster 0, but mMSC fusion and mMSC are relatively closer to Clustree Cluster 1 and Cluster 2, respectively, as evidenced in the Venn diagrams (F). WGCNA identified MEs (G) and associated genes in top MEs (H). The overall WGCNA network, highlighting modules M13 and M20, is depicted in (I).

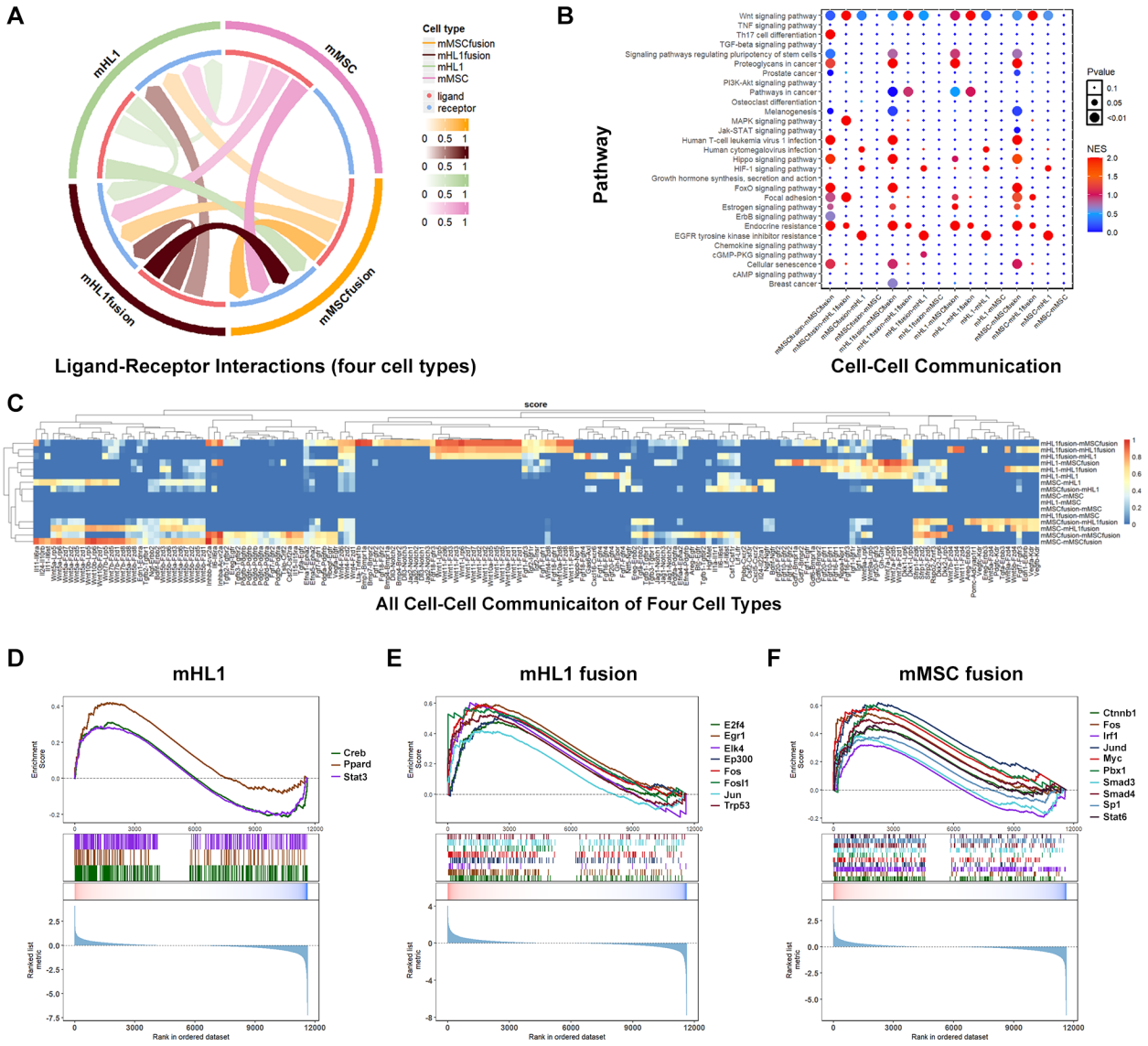

Figure 3. Intercellular communication after fusion, including associated pathways and TFs. Ligand (sender)-receptor (receiver) interactions are plotted between all four cell types (A) and relevant pathways (B). All 16 interactions among the four cell types and their ligand-receptor interaction scores are plotted in (C). TFs associated with specific cell types and their interactions are demonstrated through enrichment analysis (D-F). NES (normalized enrichment score).

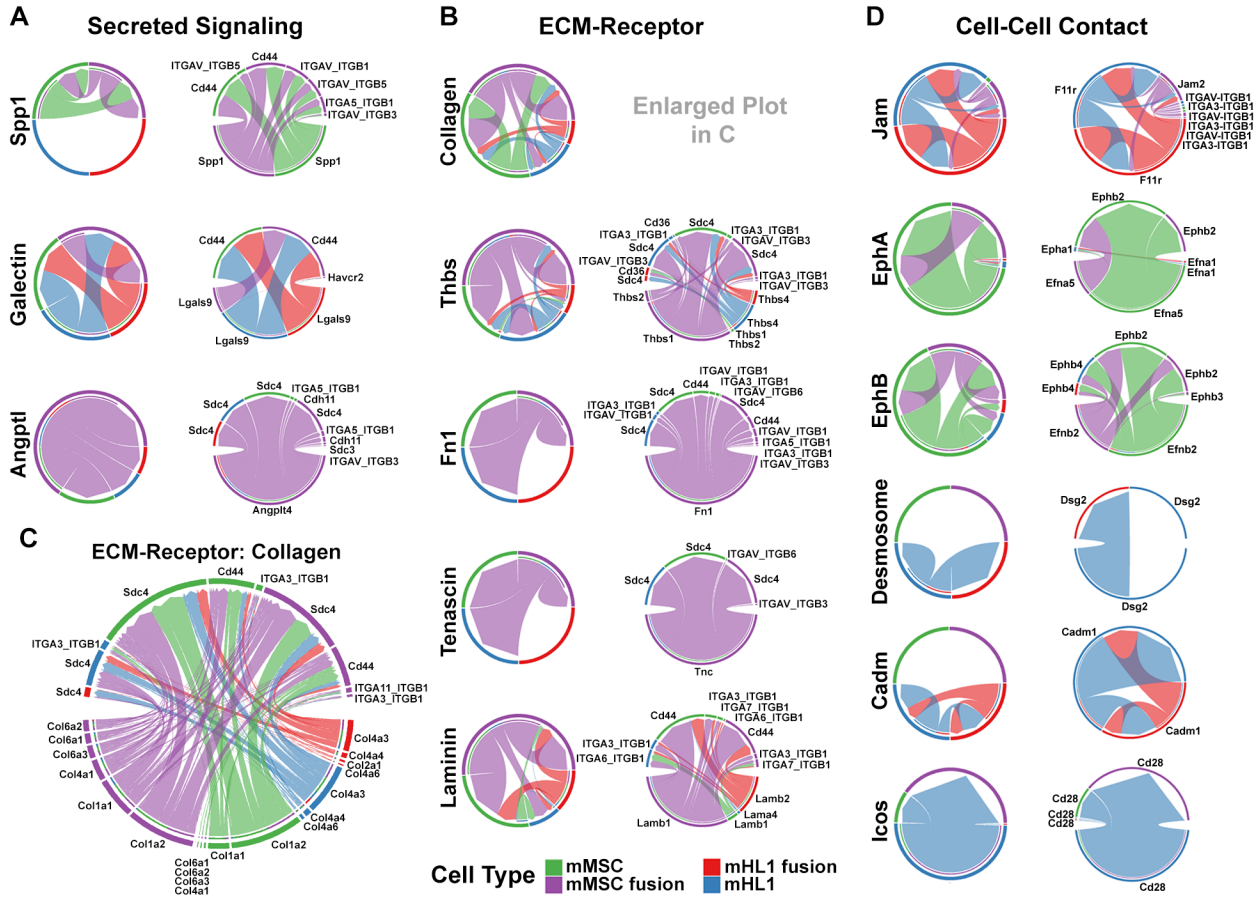

Figure 4. Intercellular communication after fusion with three different modes of cell-cell communication. Chord plots on the left depict associated pathways, while those on the right illustrate ligand-receptor interactions. To better understand the specific modes of the cell-cell communication, Secreted Signaling (A), ECM-Receptor interactions (B) and Cell-Cell contact (D) modes are analyzed. In (C), interactions with collagen ligands are enlarged due to the complex nature of the chord plot.

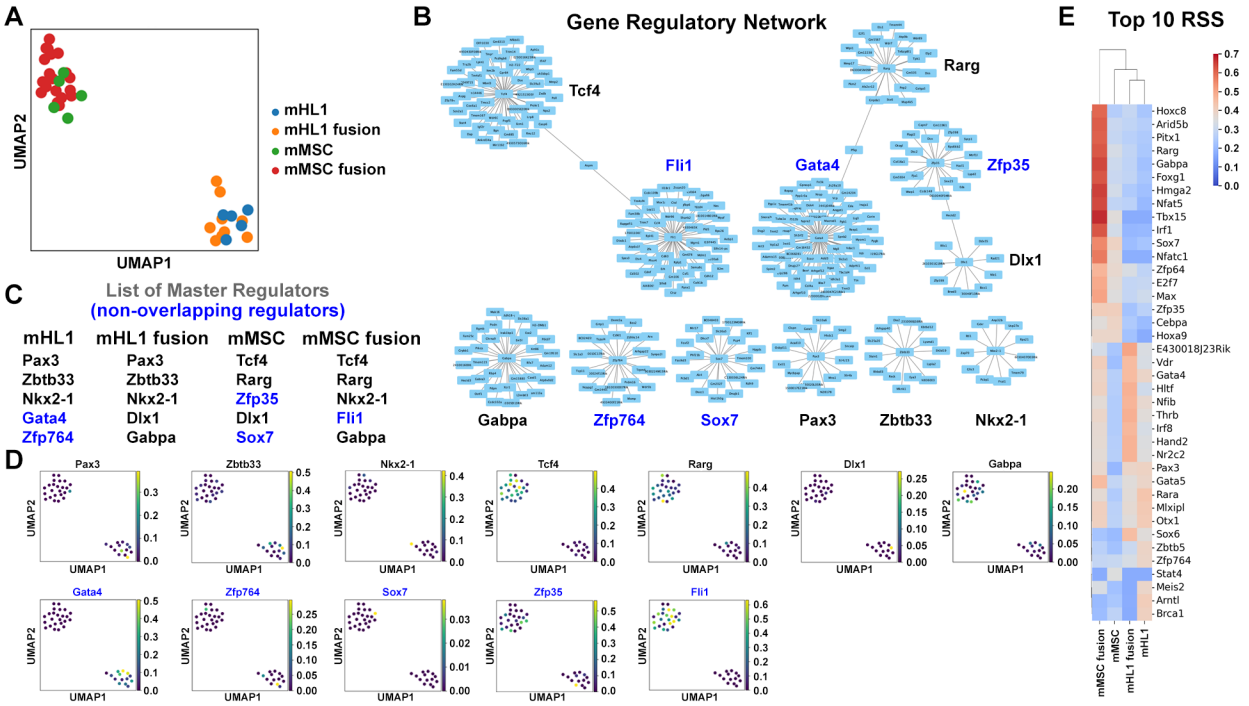

Figure 5. Intracellular Communication probed by GRNs, Master Regulators, and TF-Target Genes. The 38 fusion samples are visualized using UMAP based on SCENIC results (A). GRNs (B) identified 12 master regulators (C), whose distributions across the 38 samples are shown in (D). The top 10 regulatory TFs, ranked by RSS, are plotted using a heatmap with HC in (E).

REFERENCES

1 Searles, S. C., Santosa, E. K. & Bui, J. D. Cell-cell fusion as a mechanism of DNA exchange in cancer. *Oncotarget* **9**, 6156-6173 (2018). https://doi.org:10.18632/oncotarget.23715

2 Brukman, N. G., Uygur, B., Podbilewicz, B. & Chernomordik, L. V. How cells fuse. *The Journal of cell biology* **218**, 1436-1451 (2019). https://doi.org:10.1083/jcb.201901017

3 Ogle, B. M. & Platt, J. L. The Biology of Cell Fusion: Cells of different types and from different species can fuse, potentially transferring disease, repairing tissues and taking part in development. *American scientist* **92**, 420-427 (2004).

4 Ogle, B. M., Cascalho, M. & Platt, J. L. Biological implications of cell fusion. *Nat Rev Mol Cell Biol* **6**, 567-575 (2005). https://doi.org:10.1038/nrm1678

5 Palermo, A. *et al.* Nuclear reprogramming in heterokaryons is rapid, extensive, and bidirectional. *Faseb j* **23**, 1431-1440 (2009). https://doi.org:10.1096/fj.08-122903

6 Vassilopoulos, G., Wang, P. R. & Russell, D. W. Transplanted bone marrow regenerates liver by cell fusion. *Nature* **422**, 901-904 (2003). https://doi.org:10.1038/nature01539

7 Noiseux, N. *et al.* Mesenchymal stem cells overexpressing Akt dramatically repair infarcted myocardium and improve cardiac function despite infrequent cellular fusion or differentiation. *Mol Ther* **14**, 840-850 (2006). https://doi.org:10.1016/j.ymthe.2006.05.016

8 Kemp, K. *et al.* Fusion between human mesenchymal stem cells and rodent cerebellar Purkinje cells. *Neuropathol Appl Neurobiol* **37**, 166-178 (2011). https://doi.org:10.1111/j.1365-2990.2010.01122.x

9 Huda, F. *et al.* Fusion of Human Fetal Mesenchymal Stem Cells with "Degenerating" Cerebellar Neurons in Spinocerebellar Ataxia Type 1 Model Mice. *PloS one* **11**, e0164202 (2016). https://doi.org:10.1371/journal.pone.0164202

10 Gonçalves, M. A. *et al.* Human mesenchymal stem cells ectopically expressing full-length dystrophin can complement Duchenne muscular dystrophy myotubes by cell fusion. *Hum Mol Genet* **15**, 213-221 (2006). https://doi.org:10.1093/hmg/ddi438

11 Freeman, B. T., Kouris, N. A. & Ogle, B. M. Tracking fusion of human mesenchymal stem cells after transplantation to the heart. *Stem Cells Transl Med* **4**, 685-694 (2015). https://doi.org:10.5966/sctm.2014-0198

12 Ferrand, J. *et al.* Human bone marrow-derived stem cells acquire epithelial characteristics through fusion with gastrointestinal epithelial cells. *PloS one* **6**, e19569 (2011). https://doi.org:10.1371/journal.pone.0019569

13 Freeman, B. T. & Ogle, B. M. Viral-mediated fusion of mesenchymal stem cells with cells of the infarcted heart hinders healing via decreased vascularization and immune modulation. *Scientific reports* **6**, 20283 (2016). https://doi.org:10.1038/srep20283

14 Wang, X. *et al.* Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* **422**, 897-901 (2003). https://doi.org:10.1038/nature01531

15 Willenbring, H. & Grompe, M. Delineating the hepatocyte's hematopoietic fusion partner. *Cell Cycle* **3**, 1489-1491 (2004). https://doi.org:10.4161/cc.3.12.1304

16 de la Garza-Rodea, A. S. *et al.* Myogenic properties of human mesenchymal stem cells derived from three different sources. *Cell transplantation* **21**, 153-173 (2012). https://doi.org:10.3727/096368911x580554

17 Ding, Z. *et al.* Galectin-1-induced skeletal muscle cell differentiation of mesenchymal stem cells seeded on an acellular dermal matrix improves injured anal sphincter. *Discov Med* **21**, 331-340 (2016).

18 Pawelek, J. M. Fusion of bone marrow-derived cells with cancer cells: metastasis as a secondary disease in cancer. *Chin J Cancer* **33**, 133-139 (2014). https://doi.org:10.5732/cjc.013.10243

19 Li, H. *et al.* Fusion of HepG2 cells with mesenchymal stem cells increases cancer-associated and malignant properties: an in vivo metastasis model. *Oncol Rep* **32**, 539-547 (2014). https://doi.org:10.3892/or.2014.3264
20 Lazova, R. *et al.* A Melanoma Brain Metastasis with a Donor-Patient Hybrid Genome following Bone Marrow Transplantation: First Evidence for Fusion in Human Cancer. *PloS one* **8**, e66731 (2013). https://doi.org:10.1371/journal.pone.0066731
21 Sun, C. *et al.* Fusion of cancer stem cells and mesenchymal stem cells contributes to glioma neovascularization. *Oncol Rep* **34**, 2022-2030 (2015). https://doi.org:10.3892/or.2015.4135
22 Luo, F. *et al.* Bone marrow mesenchymal stem cells participate in prostate carcinogenesis and promote growth of prostate cancer by cell fusion in vivo. *Oncotarget* **7**, 30924-30934 (2016). https://doi.org:10.18632/oncotarget.9045
23 Freeman, B. T., Jung, J. P. & Ogle, B. M. Single-cell RNA-seq reveals activation of unique gene groups as a consequence of stem cell-parenchymal cell fusion. *Scientific reports* **6**, 23270 (2016). https://doi.org:10.1038/srep23270
24 Johnson, K. A. & Krishnan, A. Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. *Genome Biol* **23**, 1 (2022). https://doi.org:10.1186/s13059-021-02568-9
25 Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7** (2018). https://doi.org:10.1093/gigascience/giy083
26 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013). https://doi.org:10.1186/1471-2105-14-128
27 Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018). https://doi.org:10.1038/s41586-018-0590-4
28 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). https://doi.org:10.1186/1471-2105-9-559
29 Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet* **53**, 1143-1155 (2021). https://doi.org:10.1038/s41588-021-00894-z
30 Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods* **3**, 100498 (2023). https://doi.org:10.1016/j.crmeth.2023.100498
31 Zhang, Y. *et al.* CellCall: integrating paired ligand–receptor and transcription factor activities for cell–cell communication. *Nucleic Acids Research* **49**, 8520-8534 (2021). https://doi.org:10.1093/nar/gkab638
32 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12** (2021). https://doi.org:10.1038/s41467-021-21246-9
33 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021). https://doi.org:10.1038/s41467-021-21246-9
34 Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat Protoc* (2024). https://doi.org:10.1038/s41596-024-01045-4
35 Zhang, Y. *et al.* CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res* **49**, 8520-8534 (2021). https://doi.org:10.1093/nar/gkab638
36 Oteiza, A. & Mechti, N. The human T-cell leukemia virus type 1 oncoprotein tax controls forkhead box O4 activity through degradation by the proteasome. *J Virol* **85**, 6480-6491 (2011). https://doi.org:10.1128/jvi.00036-11
37 Sleman, S. Virus-induced FoxO factor facilitates replication of human cytomegalovirus. *Arch Virol* **167**, 109-121 (2022). https://doi.org:10.1007/s00705-021-05279-5

38 Koulouris, A., Tsagkaris, C., Corriero, A. C., Metro, G. & Mountzios, G. Resistance to TKIs in EGFR-Mutated Non-Small Cell Lung Cancer: From Mechanisms to New Therapeutic Strategies. *Cancers (Basel)* **14** (2022). https://doi.org:10.3390/cancers14143337
39 Wang, Z. ErbB Receptors and Cancer. *Methods Mol Biol* **1652**, 3-35 (2017). https://doi.org:10.1007/978-1-4939-7219-7_1
40 Cheng, L. *et al.* Cardiomyocyte-restricted peroxisome proliferator-activated receptor-delta deletion perturbs myocardial fatty acid oxidation and leads to cardiomyopathy. *Nature medicine* **10**, 1245-1250 (2004). https://doi.org:10.1038/nm1116
41 Gusterson, R. J., Jazrawi, E., Adcock, I. M. & Latchman, D. S. The transcriptional co-activators CREB-binding protein (CBP) and p300 play a critical role in cardiac hypertrophy that is dependent on their histone acetyltransferase activity. *The Journal of biological chemistry* **278**, 6838-6847 (2003). https://doi.org:10.1074/jbc.M211762200
42 Harhous, Z., Booz, G. W., Ovize, M., Bidaux, G. & Kurdi, M. An Update on the Multifaceted Roles of STAT3 in the Heart. *Front Cardiovasc Med* **6**, 150 (2019). https://doi.org:10.3389/fcvm.2019.00150
43 Jacoby, J. J. *et al.* Cardiomyocyte-restricted knockout of STAT3 results in higher sensitivity to inflammation, cardiac fibrosis, and heart failure with advanced age. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12929-12934 (2003). https://doi.org:10.1073/pnas.2134694100
44 Acencio, M. L., Vazquez, M., Chawla, K., Lægreid, A. & Kuiper, M. TFCheckpoint database update, a cross-referencing system for transcription factors from human, mouse and rat. *Nucleic Acids Res* **52**, D334-d344 (2024). https://doi.org:10.1093/nar/gkad1030
45 Hutloff, A. *et al.* ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* **397**, 263-266 (1999). https://doi.org:10.1038/16717
46 Wang, X. Pleiotrophin: Activity and mechanism. *Adv Clin Chem* **98**, 51-89 (2020). https://doi.org:10.1016/bs.acc.2020.02.003
47 Coutu, D. L., François, M. & Galipeau, J. Inhibition of cellular senescence by developmentally regulated FGF receptors in mesenchymal stem cells. *Blood* **117**, 6801-6812 (2011). https://doi.org:10.1182/blood-2010-12-321539
48 Ku, S. C. *et al.* Comprehensive analysis of prognostic significance of cadherin (CDH) gene family in breast cancer. *Aging (Albany NY)* **14**, 8498-8567 (2022). https://doi.org:10.18632/aging.204357
49 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017). https://doi.org:10.1038/nmeth.4463
50 Lagha, M. *et al.* Regulation of skeletal muscle stem cell behavior by Pax3 and Pax7. *Cold Spring Harb Symp Quant Biol* **73**, 307-315 (2008). https://doi.org:10.1101/sqb.2008.73.006
51 Minoo, P. *et al.* Physical and functional interactions between homeodomain NKX2.1 and winged helix/forkhead FOXA1 in lung epithelial cells. *Mol Cell Biol* **27**, 2155-2165 (2007). https://doi.org:10.1128/mcb.01133-06
52 Park, J. I. *et al.* Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Developmental cell* **8**, 843-854 (2005). https://doi.org:10.1016/j.devcel.2005.04.010
53 Pikkarainen, S., Tokola, H., Kerkelä, R. & Ruskoaho, H. GATA transcription factors in the developing and adult heart. *Cardiovascular research* **63**, 196-207 (2004). https://doi.org:10.1016/j.cardiores.2004.03.025
54 Fadda, A. *et al.* Genome-wide Regulatory Roles of the C2H2-type Zinc Finger Protein ZNF764 on the Glucocorticoid Receptor. *Scientific reports* **7**, 41598 (2017). https://doi.org:10.1038/srep41598

55 Kraus, P. & Lufkin, T. Dlx homeobox gene control of mammalian limb and craniofacial development. *Am J Med Genet A* **140**, 1366-1374 (2006). https://doi.org:10.1002/ajmg.a.31252
56 Yang, Z. F., Drumea, K., Mott, S., Wang, J. & Rosmarin, A. G. GABP transcription factor (nuclear respiratory factor 2) is required for mitochondrial biogenesis. *Mol Cell Biol* **34**, 3194-3201 (2014). https://doi.org:10.1128/mcb.00492-12
57 Feng, X. *et al.* Molecular mechanism underlying the difference in proliferation between placenta-derived and umbilical cord-derived mesenchymal stem cells. *J Cell Physiol* **235**, 6779-6793 (2020). https://doi.org:10.1002/jcp.29572
58 Kashyap, V. *et al.* RARγ is essential for retinoic acid induced chromatin remodeling and transcriptional activation in embryonic stem cells. *J Cell Sci* **126**, 999-1008 (2013). https://doi.org:10.1242/jcs.119701
59 Cunliffe, V., Williams, S. & Trowsdale, J. Genomic analysis of a mouse zinc finger gene, Zfp-35, that is up-regulated during spermatogenesis. *Genomics* **8**, 331-339 (1990). https://doi.org:10.1016/0888-7543(90)90290-b
60 Chen, Y. S. *et al.* Small molecule mesengenic induction of human induced pluripotent stem cells to generate mesenchymal stem/stromal cells. *Stem Cells Transl Med* **1**, 83-95 (2012). https://doi.org:10.5966/sctm.2011-0022
61 Badwe, C. R., Lis, R., Barcia Duran, J. G., Kunar, B. & Rafii, S. Fli1 Is Essential for the Maintenance of Hematopoietic Stem Cell Homeostasis and Function. *Blood* **130**, 3769-3769 (2017). https://doi.org:10.1182/blood.V130.Suppl_1.3769.3769
62 Castaño, B. A. *et al.* The levels of p53 govern the hierarchy of DNA damage tolerance pathway usage. *Nucleic Acids Res* **52**, 3740-3760 (2024). https://doi.org:10.1093/nar/gkae061
63 Sanchez, A. P. *et al.* Role of the USF1 transcription factor in diabetic kidney disease. *Am J Physiol Renal Physiol* **301**, F271-279 (2011). https://doi.org:10.1152/ajprenal.00221.2011
64 Durand, B. *et al.* RFXAP, a novel subunit of the RFX DNA binding complex is mutated in MHC class II deficiency. *EMBO J* **16**, 1045-1055 (1997). https://doi.org:10.1093/emboj/16.5.1045
65 Zhong, X. *et al.* Disruption of the ZFP574-THAP12 complex suppresses B cell malignancies in mice. *Proceedings of the National Academy of Sciences of the United States of America* **121**, e2409232121 (2024). https://doi.org:10.1073/pnas.2409232121
66 De Kumar, B. *et al.* Hoxa1 targets signaling pathways during neural differentiation of ES cells and mouse embryogenesis. *Developmental biology* **432**, 151-164 (2017). https://doi.org:10.1016/j.ydbio.2017.09.033
67 Baik, B., Yoon, S. & Nam, D. Benchmarking RNA-seq differential expression analysis methods using spike-in and simulation data. *PloS one* **15**, e0232271 (2020). https://doi.org:10.1371/journal.pone.0232271
68 Sánchez-Baizán, N., Ribas, L. & Piferrer, F. Improved biomarker discovery through a plot twist in transcriptomic data analysis. *BMC Biol* **20**, 208 (2022). https://doi.org:10.1186/s12915-022-01398-w
69 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8
70 Shen, D. *et al.* Pleiotrophin, a multifunctional cytokine and growth factor, induces leukocyte responses through the integrin Mac-1. *The Journal of biological chemistry* **292**, 18848-18861 (2017). https://doi.org:10.1074/jbc.M116.773713
71 Papadimitriou, E. *et al.* On the role of pleiotrophin and its receptors in development and angiogenesis. *Int J Dev Biol* **66**, 115-124 (2022). https://doi.org:10.1387/ijdb.210122ep

72 Yeh, H. J., He, Y. Y., Xu, J., Hsu, C. Y. & Deuel, T. F. Upregulation of pleiotrophin gene expression in developing microvasculature, macrophages, and astrocytes after acute ischemic brain injury. *J Neurosci* **18**, 3699-3707 (1998). https://doi.org:10.1523/jneurosci.18-10-03699.1998
73 van Riggelen, J., Yetil, A. & Felsher, D. W. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer* **10**, 301-309 (2010). https://doi.org:10.1038/nrc2819
74 Garces de Los Fayos Alonso, I. *et al.* The Role of Activator Protein-1 (AP-1) Family Members in CD30-Positive Lymphomas. *Cancers (Basel)* **10** (2018). https://doi.org:10.3390/cancers10040093
75 Ruzhynsky, V. A. *et al.* Cell cycle regulator E2F4 is essential for the development of the ventral telencephalon. *J Neurosci* **27**, 5926-5935 (2007). https://doi.org:10.1523/jneurosci.1538-07.2007
76 Brooks, C. L. & Gu, W. p53 regulation by ubiquitin. *FEBS Lett* **585**, 2803-2809 (2011). https://doi.org:10.1016/j.febslet.2011.05.022
77 Morcos, M. N. F. *et al.* SCA-1 Expression Level Identifies Quiescent Hematopoietic Stem and Progenitor Cells. *Stem Cell Reports* **8**, 1472-1478 (2017). https://doi.org:10.1016/j.stemcr.2017.04.012
78 Tsai, R. Y. Turning a new page on nucleostemin and self-renewal. *J Cell Sci* **127**, 3885-3891 (2014). https://doi.org:10.1242/jcs.154054
79 Guazzi, S. *et al.* Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J* **9**, 3631-3639 (1990). https://doi.org:10.1002/j.1460-2075.1990.tb07574.x
80 Tagne, J. B. *et al.* Genome-wide analyses of Nkx2-1 binding to transcriptional target genes uncover novel regulatory patterns conserved in lung development and tumors. *PloS one* **7**, e29907 (2012). https://doi.org:10.1371/journal.pone.0029907
81 Prall, O. W. *et al.* An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* **128**, 947-959 (2007). https://doi.org:10.1016/j.cell.2007.01.042
82 Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat Methods* **11**, 637-640 (2014). https://doi.org:10.1038/nmeth.2930
83 Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59 (2019). https://doi.org:10.1186/s13059-019-1663-x
84 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology* **38**, 1408-1414 (2020). https://doi.org:10.1038/s41587-020-0591-3