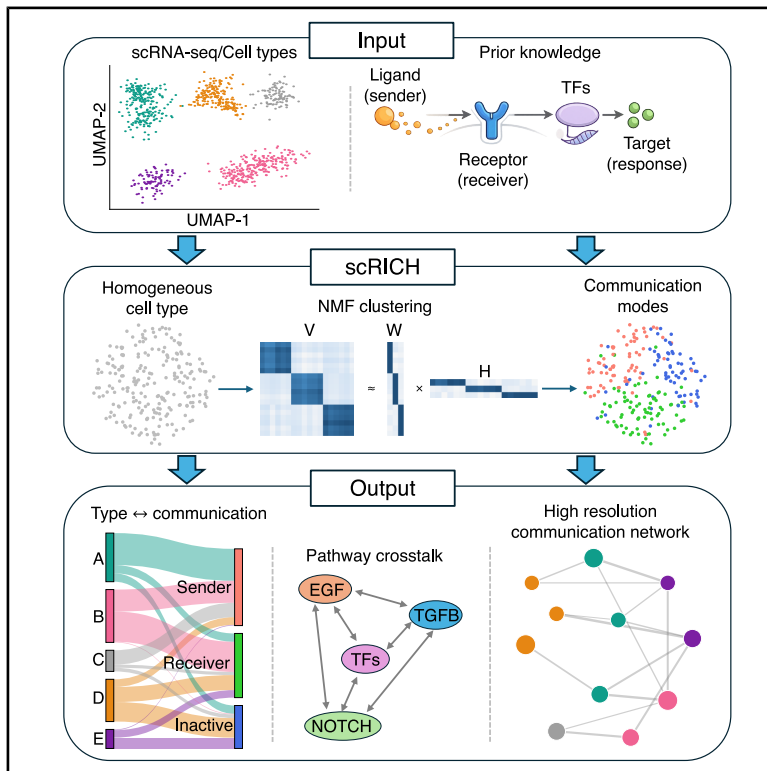


Robust identification of cell-cell communication heterogeneity in single cells

Graphical abstract



Authors

Federico Bocci, Yunlong Jia,
Scott Atwood, Qing Nie

Correspondence

federico.bocci@ru.nl (F.B.),
qnie@uci.edu (Q.N.)

In brief

Bocci et al. developed a computational method, scRICH, which leverages single-cell transcriptomics data to quantify cell-cell communication heterogeneity within cell populations. By integrating cell transition and gene regulatory network inference, scRICH identifies emerging patterns of communication along differentiation lineages and in spatial systems.

Highlights

- scRICH identifies cell-cell communication modes within scRNA-seq data
- scRICH quantifies communication heterogeneity within cell types
- scRICH identifies relationships between communication pathways
- scRICH constructs regulatory networks connecting cell-cell communication pathways

Article

Robust identification of cell-cell communication heterogeneity in single cells

Federico Bocci,^{1,*} Yunlong Jia,² Scott Atwood,³ and Qing Nie^{3,4,5,6,*}

¹Radboud Institute for Molecular Life Science, Radboud University, Nijmegen, the Netherlands

²Toulouse Cancer Research Center, Toulouse, France

³Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA, USA

⁴Department of Mathematics, University of California, Irvine, Irvine, CA, USA

⁵NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, Irvine, CA, USA

⁶Lead contact

*Correspondence: federico.bocci@ru.nl (F.B.), qnie@uci.edu (Q.N.)

<https://doi.org/10.1016/j.cels.2026.101703>

SUMMARY

Cell-cell communication modulates cell fate decisions by relaying information across tissues and inducing intracellular responses mediated by gene regulatory networks. Although the inference of cell-cell communication from high-throughput data is gaining popularity, studying how communication pathways operate across biological scales and influence cell fate decisions remains challenging. Here, we present scRICH (robust identification of cell-cell communication heterogeneity in single cells), a computational framework that leverages single-cell and spatial transcriptomics data to unravel the heterogeneity of communication behavior within cell types, link cell-cell communication to cell fate decisions by incorporating dynamical information on RNA splicing, and connect cell-cell interactions with intracellular responses by constructing multilayer regulatory networks. We validate scRICH with new experiments on epidermal growth factor (EGF) ligand/receptor co-expression in keratinocytes, comparing these predictions against those of existing communication inference methods. Applying scRICH to multiple biological scenarios demonstrates its ability to capture relationships between distinct communication pathways and emerging trends along cell differentiation lineages and in space. A record of this paper's transparent peer review process is included in the supplemental information.

INTRODUCTION

Cell-cell communication (CCC) inference from single-cell sequencing data has now gained popularity as a highly automated, hypothesis-free approach to discover new molecular interactions.^{1,2} Despite the rapid progress in the field, several existing limitations hamper the interpretability and scope of these methodologies.

First, CCC inference strikes a delicate balance between high resolution, achieved with models at the scale of individual cells, and robustness, achieved by averaging gene expression over large groups of cells—typically the annotated cell types in a dataset. Several methods exploit the single-cell resolution offered by state-of-the-art transcriptomics to predict CCC between individual cells, thus resulting in dense interaction networks.^{3–7} A major drawback of this approach is the sparsity in single-cell transcriptomics data,⁸ which can result in large cell-to-cell fluctuations in gene counts and missing information for important signaling pathways. To ensure robust predictions, many popular methodologies leverage cell annotations to average gene expression and/or the CCC connections over cell types.^{9–13} The major pitfall of this approach is the potential loss of informa-

tion about heterogeneity within individual cell types. In other words, sharing the same cell type does not necessarily imply a similar role in CCC, and cells of a given type can either send, receive, or not participate in a given signaling pathway.

Second, it is currently difficult to inject dynamical information about cell fate specification and its mutual dependence on CCC. Despite the abundance of methodologies to infer lineages and cell transitions from single-cell data, cells are typically modeled in an autonomous manner.^{14,15} A static view of CCC, where cell states are treated as stable (i.e., not transitioning toward new cell states) and homogeneous (i.e., all cells behave similarly given a ligand/receptor pair) prevents a complete understanding of how CCC is modulated, and, in turn, contributes to cell fate decision events. For such an approach to be possible, the cell states and “CCC propensity” need to be treated as distinct observations, whereby, for example, the former is defined by well-defined marker genes and morphological changes, whereas the latter is defined by ligand/receptor expression.

Third, CCC is multiscale as signal transduction networks process external information, modulating secondary CCC pathways in a constant back-and-forth interaction between the

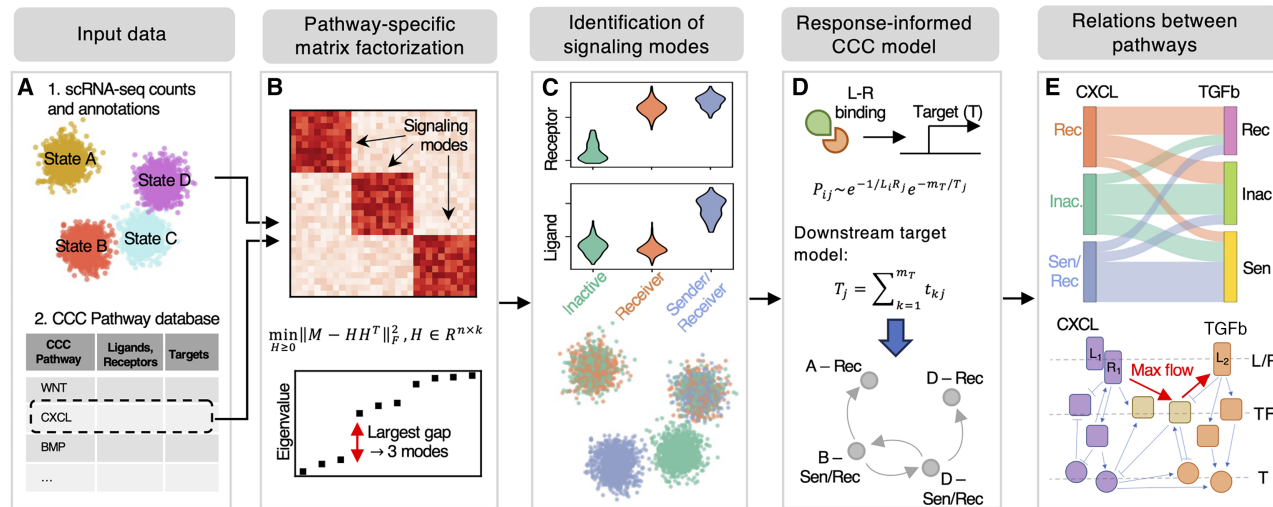


Figure 1. Overview of scRICH

(A) The inputs to scRICH are the well-annotated scRNA-seq counts and database information on pathways, ligand/receptor interactions, transcription factors (TFs), and downstream targets (Ts).

(B) For each CCC pathway, scRICH constructs a cell-cell similarity matrix. The optimal number of CCC modes is computed via nonnegative matrix factorization and spectral analysis.

(C) The CCC signaling modes are annotated based on the expression of receptors and ligands, and their correspondence to cell types is quantified.

(D) A CCC model is constructed in which the propensity of communication depends on ligand and receptor abundance in the senders and receivers, as well as downstream target activation, which is quantified either by gene expression or RNA velocity.

(E) The relationship between CCC signaling modes of distinct pathways is identified, and the molecular connection between pathways is discovered by constructing the joint GRN. Finally, a maximum-flow strategy identifies the regulatory connections in the GRN.

intercellular and intracellular scales. Recently, a few methods included information about receptor downstream targets, which can be used as a benchmark to confirm pathway activation.^{3,16–23} For example, SoptSC refines the single-cell-level CCC networks by weighting communication edges based on the expression of a downstream target gene set that is provided by the user.³ Similarly, NicheNet and scMLnet feature manually curated databases to link ligand/receptor pairs to candidate downstream targets,^{16,17} whereas CytoTalk identifies putative response genes *de novo* using a mutual information approach,¹⁸ and LIANA+ integrates CCC inference with database knowledge.¹⁹ Although these methods took decisive steps forward toward connecting the ligand/receptor and downstream signaling levels, we propose that the recent advances in gene regulatory network (GRN) reconstruction from single-cell transcriptomics²⁴ could now allow a more holistic and detailed description of the cellular flow that starts from ligand/receptor interactions, leads to a downstream response, and, in turn, regulates other CCC pathways.

Here, we introduce scRICH (robust identification of cell-cell communication heterogeneity in single cells), a methodology that addresses these existing bottlenecks by integrating CCC, cell-fate transition directionality via RNA velocity, and gene regulatory network (GRN) inference to achieve a multiscale representation of CCC. scRICH takes single-cell RNA sequencing (scRNA-seq) counts and well-annotated cell labels as input to (1) identify robust CCC modes across cell types; (2) integrate downstream target RNA velocity into a multilayered model of cell-cell signaling; (3) enable the simultaneous visualization of transition trajectories and CCC networks; and (4) connect

ligand/receptor interactions to their downstream gene regulatory network (GRN) to predict the connections between distinct CCC pathways. We apply scRICH to multiple biological scenarios, including keratinocyte differentiation, erythroid cell differentiation, and cancer epithelial-mesenchymal transition (EMT). Specifically, we confirm experimentally scRICH's prediction about epidermal growth factor (EGF) signaling heterogeneity and its relationship with other CCC pathways, including TGFb, NOTCH, and CDH4. Finally, we demonstrate scRICH's ability to identify the spatial localization of cells with coherent CCC behavior in heterogeneous tissues via analysis of spatial transcriptomics (ST) data.

RESULTS

Overview of scRICH

scRICH takes preprocessed, log-normalized scRNA-seq counts and well-annotated cell labels as input (Figure 1A, top). Although “regular” scRNA-seq counts can be used, some of the downstream analysis tools explicitly require unspliced and spliced count matrices for CCC validation and projection onto cell trajectories.

Quantification of intra-type CCC heterogeneity

First, given a specific CCC pathway, scRICH dissects CCC heterogeneity with unsupervised clustering based exclusively on the transcriptomic counts of ligands, receptors, and downstream targets (if known) of the pathway (Figure 1A, bottom). To compile a comprehensive gene set for each CCC pathway, scRICH assembles ligand/receptor sets from CellChat⁹ and downstream targets from exFinder,²⁵ which are, in turn,

assembled based on several existing databases^{16,26,27} (Figures S1A–S1H). Alternatively, pathway gene signatures compiled from the gene set enrichment analysis (GSEA) molecular signature database²⁸ or custom-made gene sets can be provided as input. scRICH determines the optimal number of CCC modes in the dataset with nonnegative matrix factorization of the pathway-specific cell-cell similarity matrix (Figure 1B), an approach that has been previously employed to identify cell states in single-cell transcriptomics data.²⁹ For clarity, the terms *cell state* and *cell type* indicate the cell annotation based on marker analysis, which is provided as input to scRICH, whereas the term “mode” indicates the pathway-specific CCC clusters identified by scRICH. To achieve robust, sub-type resolution, the cell types and CCC modes are integrated into meta-cells, whereby a dataset with N cell types and M signaling modes will be divided into $N \times M$ meta-cells, with the distinctive advantage that cells within each meta-cell exhibit a consistent transcriptomic profile and similar CCC behavior (Figure 1C). A heterogeneity score is computed to quantify whether cells of the same type have consistent CCC behavior (low score) or exhibit heterogeneous CCC behaviors (high score). This pathway-specific clustering is repeated in parallel for all pathways in the scRICH database, whereby user-defined parameters can filter the number of analyzed pathways. Finally, a statistical test based on permutation is implemented at the pathway-specific level to quantify the merit of splitting cell types into meta-cells for exploration of CCC behavior heterogeneity and downstream analysis.

Construction of communication networks and transitions

Starting from the CCC meta-cells defined in the previous step, scRICH builds a model of CCC between meta-cells through the given pathway (Figure 1D). In the model, the signaling between two meta-cells is a function of ligand and receptor expression in the sender and receiver meta-cells, respectively, as well as the activation of downstream targets in the receiver meta-cell. Specifically, scRICH offers two mathematical models of ligand/receptor interaction. First, a *diffusion-based* model, where the communication propensity depends exponentially on the expression of ligands and receptors, similar to the CCC model at individual cell scale employed by SoptSC³; alternatively, a *mass action-based* model similar to CellChat, in which the communication propensity depends on ligand and receptor expression via a Michaelis-Menten-like model.⁹ Although both models provide the same qualitative output, the diffusion-based model amplifies the strongest interactions and could thus be a better choice to simplify complex systems with several cell states and CCC signaling modes. Conversely, the mass action-based model is more permissive and highlights weaker interactions between meta-cells, perhaps being better suited to study smaller systems (Figures S1I–S1L). The activation of downstream targets is quantified by target RNA velocity, which indicates whether the gene is being upregulated or downregulated, respectively.^{30,31} Alternatively, regular RNA counts are used if unspliced/spliced counts are not provided in the input dataset. The inclusion of downstream targets is optional, and the method can be run using exclusively the information on receptors and ligands. The heterogeneity of CCC within a cell type can be visualized in a low-dimensional embedding by color-cod-

ing cells based on CCC signaling mode membership (Figure 1C, bottom).

Cellular flow between co-dependent pathways

Finally, scRICH integrates information on cell clustering based on individual CCC pathways to extrapolate emerging biological trends and predict the “cellular flow” between CCC pathways (Figure 1E). First, scRICH uses mutual information to identify CCC pathways that show similar trends across cell states (for example, being a WNT sender implies being a BMP sender, etc.) and employs Sankey-style diagrams to represent the overlap between signaling modes of distinct CCC pathways. Moreover, given two distinct CCC pathways, scRICH implements GRN inference using the spliceJAC package³² to reconstruct the joint regulatory network, including the ligands, receptors, and downstream transcription factors (TFs) connecting the two pathways, and uses a maximal-flow algorithm to identify the regulatory connections between two distinct pathways.

Benchmarking and comparison with other methods

We benchmark scRICH’s ability to predict CCC networks by extensively comparing it against existing tools for CCC inference, including CellChat⁹, CellPhoneDB,¹¹ and LIANA+.¹⁹ First, we demonstrate similar accuracy when scRICH is used to infer CCC at the cell-type level, defined as the pre-existing cell annotation in the dataset. Second, we highlight novel insights uniquely provided by scRICH by reconstructing CCC at meta-cell resolution in our own scRNA-seq dataset for fibroblast-containing human skin equivalents³³ (FibHSEs). We start by identifying the top CCC pathways based on the average expression of receptors and ligands using scRICH’s built-in CCC pathway exploration function (Figure 2A). For a given CCC pathway, scRICH first identifies the number of CCC communication modes. In the example of the EGF pathway, scRICH identifies three signaling modes corresponding to EGF inactive (low expression of receptors and ligands), EGF receiver (high receptor, low ligand), and EGF sender/receiver (high receptor, high ligand). These EGF CCC modes can be visualized in a low-dimensional embedding as well as via violin plots (Figures 2B and 2C). To mitigate sparsity in signaling mode identification, scRICH utilizes imputed counts for genes associated with ligands, receptors, and targets, potentially introducing a dependence on imputation parameters, including the number of principal components (PCs) and cell neighbors (N). We evaluated this dependence by computing the Wasserstein distance between signaling mode memberships obtained for different (PC, N) parameter combinations for the top 10 identified pathways in the FibHSE dataset. Although the number of PCs does not influence signaling mode identification, cell membership in different signaling modes depends, albeit weakly, on the number of cell neighbors (Figure S2A). Specifically, some pathways, including annexin, APP, insulin growth factor (IGF), and NRG, exhibit different numbers of signaling modes for varying parameter combinations (Figure S2B), with occurrences of signaling modes splitting or merging (Figures S2C–S2F). Repeating this analysis on other datasets with varying numbers of cells and cell types confirmed the robustness with respect to the number of PCs and cell neighbors (Figures S2G–S2J). Afterward, scRICH constructs the EGF communication network detailing how the inactive, receiver, and sender/receiver subsets of the epidermal

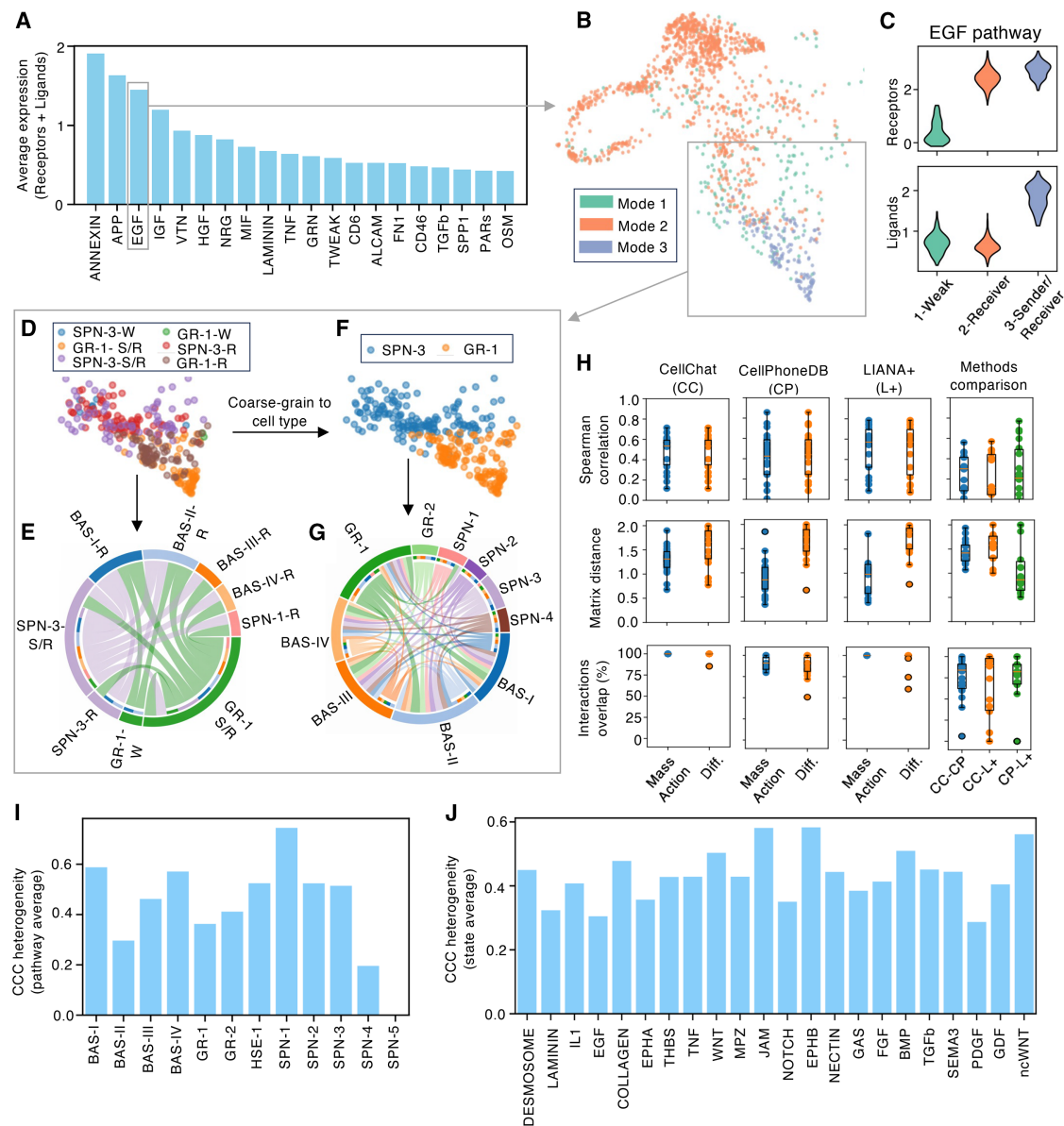


Figure 2. Comparison of scRICH against existing CCC inference methods

(A) The top CCC signaling pathways identified by scRICH in FibHSE organoids based on the average expression of ligands and receptors. (B) Low-dimensional embedding of the FibHSE dataset from Stabell et al.,³³ in which FibHSE cells are color coded with EGF CCC communication modes. (C) The expression (normalized and log transformed) of EGF receptors and ligands in the identified communication modes corresponding to inactive (green), receiver (orange), and sender/receiver (blue) states. (D–G) Schematic of the quantification pipeline. (D) The UMAP at sub-cluster resolution showing how epidermal cell states split at the meta-cell level based on the EGF communication modes. (E) The EGF CCC network predicted by scRICH at meta-cell resolution. (F) Cell-state-level UMAP representation obtained by aggregating meta-cells over epidermal cell states. (G) The EGF CCC network at cell-type resolution. (H) Detailed comparison between scRICH and the CCC networks predicted by CellChat, CellPhoneDB, and LIANA+. The rightmost column shows the comparison between the three methods as a benchmark. The line, box, and whiskers depict the average, first quartile (Q1) to third quartile (Q3), and the furthest point within $1.5 \times$ the interquartile range, respectively. (I) The cell-type CCC heterogeneity when aggregating results from all considered CCC pathways. (J) The CCC signaling pathway heterogeneity when aggregating over all cell types.

cells communicate through EGF (Figures 2D and 2E). Given that CellChat, CellPhoneDB, and LIANA+ infer communication networks at the cell-type level, we coarse-grain scRICH's prediction by averaging over CCC modes, thus predicting a CCC network that can be directly compared against these existing methods

(Figures 2F and 2G). We repeat this procedure over several CCC pathways presented in Figure 2A and compare the predicted CCC matrix based on (1) Spearman correlation, (2) element-wise matrix distance, and (3) overlap of predicted interactions. Overall, we found a strong overlap between scRICH and

the CCC networks predicted by CellChat, CellPhoneDB, and LIANA+ when applying both the mass-action and diffusion models of CCC (Figure 2H). scRICH can estimate CCC interactions based on (1) standalone ligand/receptor expression; (2) ligand/receptor and target gene expression; or (3) ligand/receptor expression plus RNA velocity of target genes. Testing these different CCC models across three datasets suggests that the relation between these models could be dataset-specific, likely due to context-specific target gene selection (Figures S2K–S2M). Furthermore, we compared different RNA velocity inference models, scVelo³⁰ and Unitvelo,³⁴ finding that scVelo-based prediction agrees closely with target-free and target-expression-based models (Figure S2M). These results highlight the importance of context-specific target gene selection and the RNA velocity model—both customizable options—when applying scRICH.

Finally, we focus on the information loss resulting from averaging CCC networks at the cell-type level by defining and computing a CCC heterogeneity score bound within [0,1]. This score quantifies whether cells with different roles in the CCC pathway (senders, receivers, etc.) are uniformly distributed across cell types (high score) or clearly divided among cell types (low score). Several keratinocyte states, such as basal keratinocytes (BAS)-I and spinous keratinocytes (SPN)-1, exhibit a high score, i.e., cells within these cell states tend to behave differently in terms of CCC, thus leading to a greater information loss when averaging CCC over cell type (Figure 2I). Similarly, several CCC pathways, such as non-canonical WNT and EPHB, have high heterogeneity scores, implying that cells with different CCC behavior are spread heterogeneously across different cell states (Figure 2J). In general, the higher the heterogeneity score, the higher the loss of information when averaging CCC over cell types, thus motivating the modeling of CCC at meta-cell resolution.

Experimental validation by analysis of downstream regulation of the EGF pathway in skin-equivalent organoids

Next, we use the FibHSE keratinocytes³³ to validate scRICH's predictions about CCC heterogeneity and CCC pathway co-dependence experimentally. These organoids reproduce the stratification of BAS that differentiate into SPN, granular (GR), and cornified (CRN) keratinocytes found in the epidermis (Figure 3A). We focus on the EGF CCC pathway, which was previously implicated in keratinocyte differentiation³³ and exhibits sender, sender/receiver, and weak signaling modes (see Figures 2B and 2C). We identify multiple cell states, including SPN-1 and SPN-3, which exhibit high EGF heterogeneity scores. Therefore, these cell states are mixtures of EGF sender, sender/receiver, and weak cells (Figures 3B and 3C).

In particular, cells in the EGF sender/receiver mode are found across several different keratinocyte cell states and are especially represented in the terminal states, including SPN-3 and GR-1. Conversely, the EGF receiver mode is mainly observed at earlier stages of the keratinocyte lineage, particularly in basal states, including BAS-III, BAS-IV, and SPN-1 (Figures 3D and 3E). To confirm the predicted role of EGF sender/receiver cells, we performed immunostaining for the EGF ligand and the EGFR receptor in FibHSE organoids. EGF and EGFR were lowly expressed without EGF induction but became much

more highly expressed following treatment with soluble EGF (Figures 3F and 3H). Furthermore, exposure to external EGF ligand increased the co-localization of EGF and EGFR in the organoids, showing that their expression correlates both in the basal layer and across the whole epidermis (Figures 3G, 3I, and 3J). Overall, these observations support a model in which late-stage keratinocytes relay information via EGF to early-stage keratinocytes, which is specifically visualized in scRICH with a superposition plot that overlays the cell-state transition field and EGF-CCC (Figure 3K).

Next, we apply scRICH to computationally model the connection between EGF and other CCC pathways involved in EMT regulation, including TGFb, NOTCH, and CDH. First, we identify three TGFb CCC signaling modes, including a sender, a strong receiver, and a weak receiver (Figure 3L), and compare cell behavior in the two pathways. Notably, strong TGFb receivers tend to be EGF receivers, whereas TGFb senders are mostly weak with respect to EGF. Finally, the TGFb weak receptor mode is almost equally spread among the three EGF signaling modes (Figure 3M). We further investigate the strong correlation between several EGF and TGFb signaling modes by inferring the intracellular gene regulatory network. Specifically, we reconstruct the network connecting the EGF ligands/receptors, including EGFR, AREG, EREG, TGFA, HBEGF, and ERBB2, with the TGFb ligands/receptors, including TGFb1, TGFb2, TGFbR1, and TGFbR2. By applying a maximum-flow algorithm to the predicted GRN, scRICH predicts the top connections and intermediate mediators connecting the EGF and TGFb CCC signaling modes, which include HRAS and INHBA (Figure 3N), both of which were previously linked to EMT.^{35–37} Consistently, the expression of HRAS and INHBA increased upon exposure to exogenous EGF in our experiment (Figure S3), suggesting a connection between the EGF sender/receiver communication mode, HRAS/INHBA activation, and TGFb signaling. To further test the relation between EGF signaling and EMT during keratinocyte differentiation, we evaluate CDH (cadherin) signaling that directly regulates cell adhesion, identifying three CCC modes corresponding to weak, medium, and strong adhesion (Figure S4A). EGF communication modes were strongly correlated with cell adhesion: whereas EGF weak cells are mostly CDH weak as well, EGF sender/receivers mostly exhibited strong adhesion (Figure S4B), and EGF signaling was predicted to directly induce cell adhesion via CDH4 activation in the inferred EGF-CDH regulatory network (Figure S4C). This prediction resonates well with the co-localization of EGF and CDH4 observed experimentally via immunostaining, whereby CDH4 expression is upregulated upon EGF induction (Figure S4D–S4F). A similar analysis of the Notch CCC pathway highlights, once again, the strong correspondence between EGF and NOTCH CCC signaling modes. Specifically, EGF sender/receivers are almost exclusively NOTCH receivers, whereas NOTCH sender/receivers are almost exclusively EGF receivers. Finally, low activity for the two pathways goes hand in hand (Figures S4G and S4H). The emerging correlation between the EGF and NOTCH receiver modes was supported by multiple activation paths identified in the predicted EGF-NOTCH gene regulatory network (Figure S4I), including the intermediate regulator STAT1 that increases along with both NOTCH and EGF (Figures S4J–S4M).

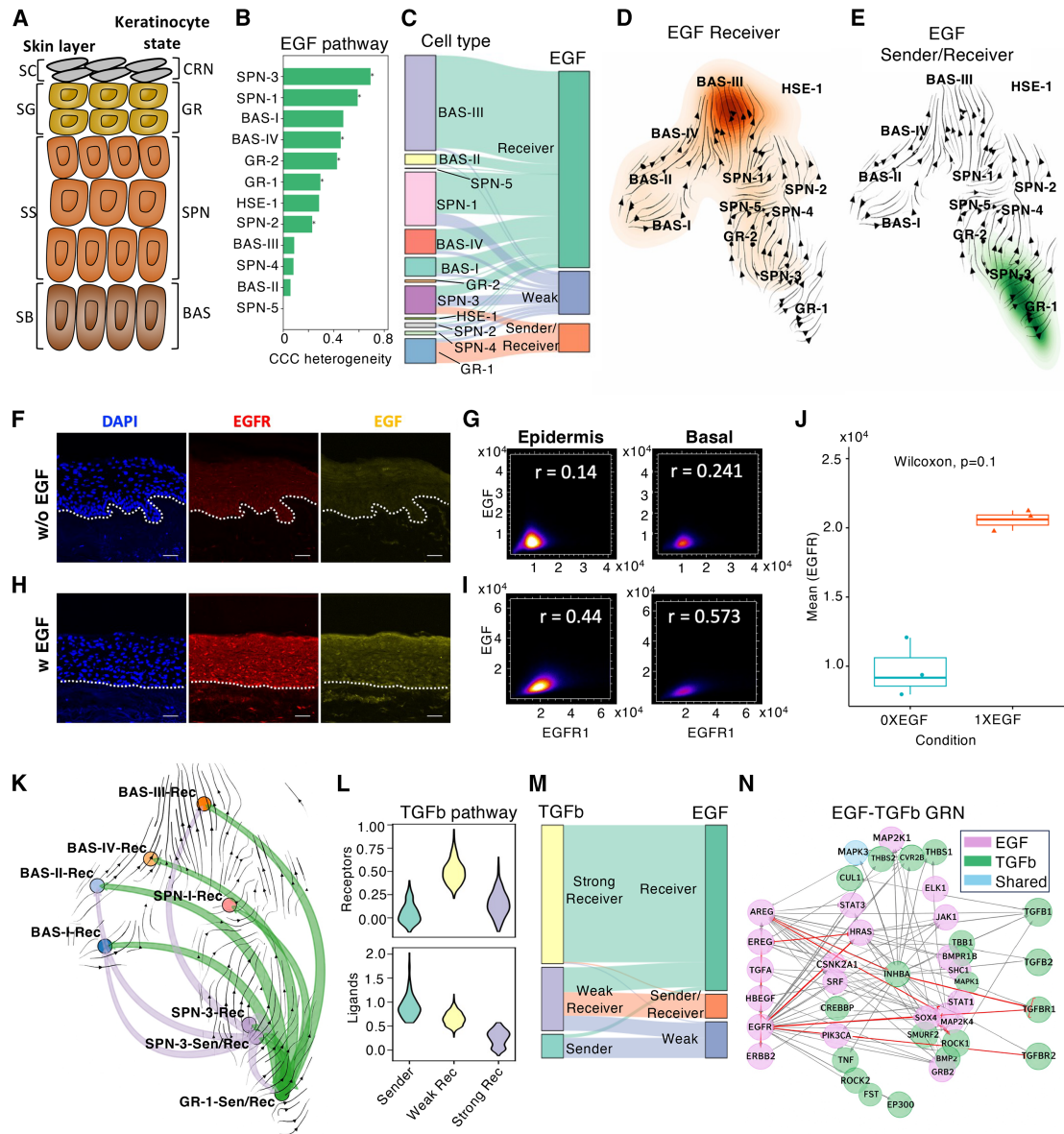


Figure 3. The EGF and TGFb CCC in FibHSEs

(A) Epidermal differentiation in human skin. Keratinocytes transition from basal (BAS; early) to spinous (SPN) and granular (GR; late) states as they migrate from the stratum basale (SB) through the stratum spinosum (SS) and stratum granulosum (SG), ultimately reaching the cornified (CRN; terminal) stage in the stratum corneum (SC).

(B) The CCC heterogeneity of the EGF pathway in the keratinocyte cell states. An asterisk indicates statistical significance based on permutation testing.

(C) Mapping between keratinocyte states and EGF CCC signaling modes in the dataset from Stabell et al.³³

(D and E) Density of EGF receiver and EGF sender/receiver cells in a low-dimensional UMAP space superimposed with RNA velocity. Color scales indicate cell density (unitless).

(F) Immunofluorescence of the epidermal layer in FibHSE showing cell nuclei, EGF, and EGFR layer without EGF treatment.

(G) Correlation between EGF and EGFR in the epidermis and the basal layer without EGF treatment. r indicates Pearson's correlation coefficient.

(H and I) Same as (F) and (G) with EGF treatment. Scale bars, 50 μ m. Dashed lines denote the epidermal-dermal junction.

(J) Quantification of EGFR expression in the epidermis based on replicates and EGF dosage ($n = 3$).

(K) The EGF CCC network between keratinocyte subsets and its relation to cell-state transitions.

(L) The TGFb signaling modes in the FibHSE keratinocytes (gene expression normalized and log transformed).

(M) Mapping between TGFb and EGF CCC modes.

(N) The gene regulatory network connecting the EGF and TGFb pathways, in which red edges indicate the path of maximum flow connecting EGF and TGFb receptors.

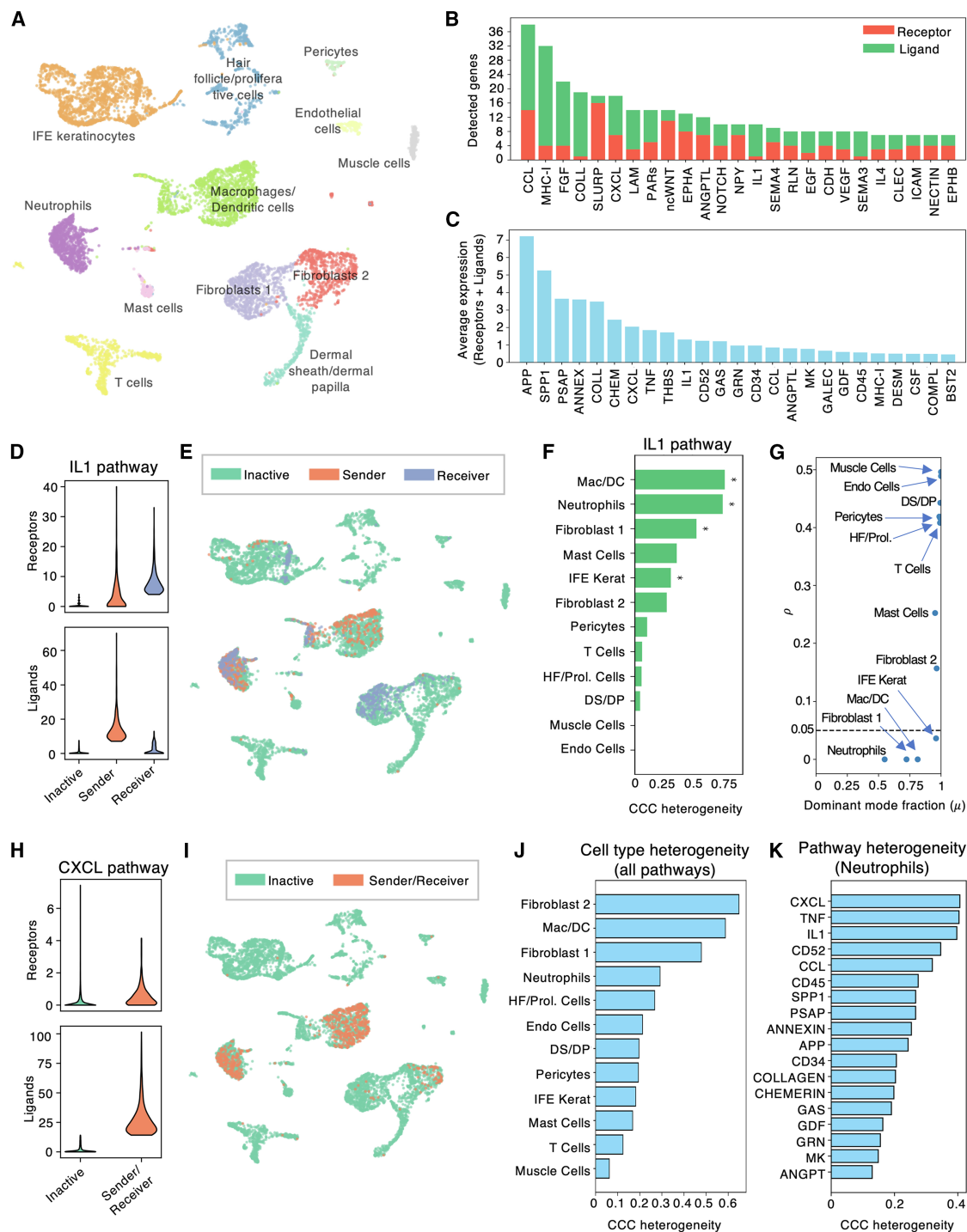


Figure 4. Functional separation of cell types into robust CCC modes in skin wound healing

(A) UMAP representation of the wound-healing cell types in the dataset by Haensel et al. ³⁸

(B and C) The top CCC signaling pathways identified in skin wound healing based on the number of detected genes (B) and the average expression of ligands and receptors (C).

(D) The IL-1 communication modes corresponding to an inactive (green), sender (orange), and receiver (blue) states.

(E) Low-dimensional embedding in which cells are color coded with IL-1 CCC communication modes.

(F) The CCC heterogeneity of the IL-1 pathway. An asterisk indicates statistical significance based on permutation testing.

(legend continued on next page)

Overall, these results highlight that CCC behavior in FibHSEs is heterogeneous, leading to a loss of information when evaluating CCC at the level of cell types. Conversely, scRICH identifies well-defined signaling modes of cells with coherent communication behavior and clear relationships between distinct CCC pathways.

scRICH distinguishes subsets of sender and receiver cells within cell types

To demonstrate scRICH's ability to identify within-type CCC heterogeneity, we analyze a scRNA-seq skin wound-healing dataset that includes multiple distinct types of fibroblasts, keratinocytes, and immune cells³⁸ (Figure 4A). An overview based on either ligand/receptor average expression or the number of detected ligands/receptors in the dataset suggests the activation of several CCC signaling pathways (Figures 4B and 4C).

To highlight a representative workflow, we focus on the interleukin (IL)-1 CCC pathway. The IL-1 pathway is associated with the inflammatory response and thus plays a role in coordinating cell behavior during wound healing.³⁹ When clustering cells based on IL-1 ligand/receptor expression, we identify 3 clusters associated with a receiver mode (high receptor, low ligand expression), a sender mode (low receptor, high ligand expression), and an inactive mode (low expression of both) (Figure 4D). Given scRNA-seq data sparsity, it could be hypothesized that the inactive mode originates from cells with low RNA counts. Visualizing cell membership in the 3 IL-1 modes in the uniform manifold approximation and projection (UMAP) embedding, however, highlights a specific pattern in which 4 cell types (fibroblast 1, neutrophils, macrophage/dendritic cells [DCs], and IFE keratinocytes) possess subgroups of sender and receiver cells, whereas all other cell types are almost exclusively inactive (Figure 4E). Therefore, these cell types exhibit high IL-1 heterogeneity scores (Figure 4F). To proceed with the functional separation of these cell types between signaling modes, we perform permutation testing (Figure 4G), indicating that these 4 cell types should be split for functional characterization of IL-1 CCC. Specifically, the fibroblast 1 and keratinocyte cell types are composed of large populations of IL-1-inactive cells and smaller but well-separated populations of IL-1-receiver cells. Therefore, averaging IL-1-based communication inference over cell types would weaken the IL-1-receiver behavior. Furthermore, the neutrophil and macrophage/DC cell types include well-separated subsets of IL1-receiver and IL-1-sender cells. Therefore, cell-type-level averaging would miss this separation, instead predicting sender/receiver behavior. Permutation testing allows us to select robust patterns in CCC heterogeneity. For example, mast cells exhibit a high IL-1 heterogeneity score (see Figure 4F). The low mast cell count, however, prevents a statistically robust functional separation within this cell type. In other words, the null hypothesis that a small subset of IL-1-sender cells exhibits an IL-1-sender, rather than an IL-1-inactive, mode by random chance cannot be rejected with sta-

tistical significance. To showcase the relevance of intra-type heterogeneity, we consider the CXCL pathway that is exclusively activated in neutrophil and macrophage/DC cell types. Compared with IL-1, CXCL features a true sender/receiver state in which individual cells co-express CXCL receptors and ligands (Figures 4H and 4I).

After identifying CCC signaling modes for all the most highly expressed CCC pathways, each cell type can be characterized by a CCC heterogeneity score that summarizes the coherence of CCC behavior of cells of the same type across many CCC pathways (Figure 4J). Fibroblasts (both type 1 and type 2), macrophages/DC cell types, and neutrophils exhibit more heterogeneous behavior (high CCC heterogeneity score), whereas muscle cells, T cells, and mast cells exhibit more coherent CCC behavior (low CCC heterogeneity score). Individual cell types can be further characterized based on their CCC heterogeneity score for each distinct pathway. For example, neutrophils behave heterogeneously along specific pathways, including CXCL, tumor necrosis factor (TNF), IL-1, and CD52 (Figure 4K).

scRICH identifies CCC gradients along cell differentiation lineages

To further test scRICH's ability to capture CCC evolution during cell fate transitions, we investigate cell differentiation during erythroid cell maturation, where blood progenitor cells differentiate into erythroid cells⁴⁰ (Figure S5A). As blood progenitor cells undergo enucleation, we first inspected the presence of unspliced RNA, confirming active transcription in all cell types (Figure S5B). We identify several CCC pathways that are highly expressed and showcase heterogeneous behavior across erythroid cell states, including IGF and VEGF (Figures 5A and S5C–S5E). Specifically, we identify three IGF CCC modes, including a sender, a receiver, and an inactive mode (Figures S5F–S5H).

Juxtaposing the IGF communication modes with the directions of cell-state transition computed with UnitVelo³⁴ suggests a connection between the IGF CCC network and cell fate. Here, UnitVelo is chosen because it was previously shown to generate more robust cell-transition trajectories compared with scVelo in the context of blood cell differentiation. scRICH, however, is compatible with both packages. The early-stage BP-1 cells split into subpopulations of BP-1 senders and BP-1 receivers, whereas the later Er-1 and Er-2 cells are mostly IGF senders, overall suggesting a transition mechanism whereby later-stage erythroid cells communicate with early-stage cells and may contribute to their differentiation (Figures 5B and 5C). The evolution of IGF along the lineage is further captured by inspecting the fraction of cells in different IGF CCC modes along latent time, which shows a progressive decrease in the fraction of cells that behave as receivers of both IGF and VEGF (Figures 5D and 5E). This trend is further visualized by cell-density plots in which early cells tend to behave as IGF receivers, whereas cells in the intermediate lineage mostly behave as senders,

(G) Permutation testing *p* value as a function of the fraction of cells in the dominant CCC signaling mode in each cell type. Cell types below the threshold of 0.05 are selected for functional separation.

(H and I) Same as (D) and (E) for the CXCL pathway. Inactive (green) and sender/receiver (orange) communication modes are identified.

(J) Cell-type CCC heterogeneity score (average over all CCC pathways).

(K) Heterogeneity score of each CCC pathway in the neutrophil cell type.

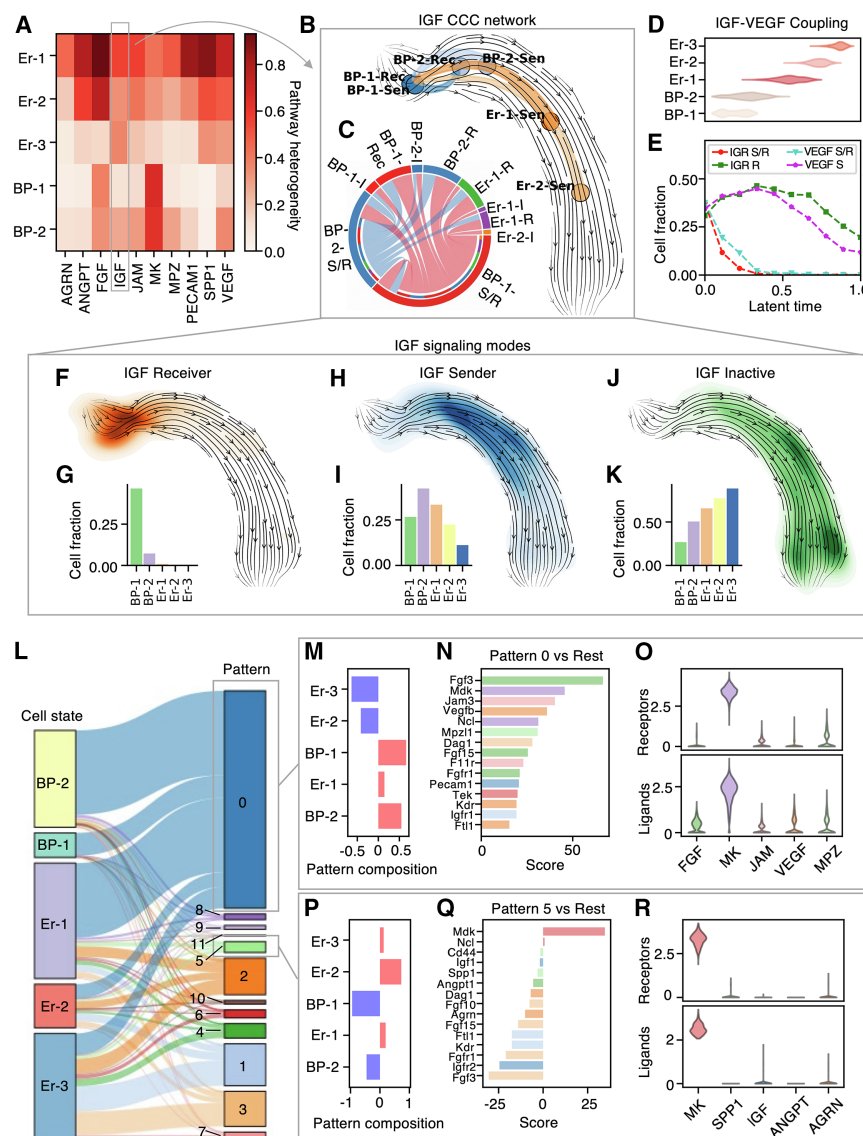


Figure 5. Application of scRICH to the erythroid lineage

(A) The CCC heterogeneity of erythroid cell states for multiple CCC pathways in the dataset by Barile et al.⁴⁰

(B and C) The IGF CCC network depicted over RNA velocity (B) and as a chord diagram (C).

(D) The distribution of cell states over latent time.

(E) The fraction of IGF and VEGF CCC modes as a function of latent time.

(F) The cell-density plot of IGF receivers.

(G) The fraction of IGF receivers in the erythroid cell states.

(H and I) Same as (F) and (G) for IGF senders.

(J and K) Same as (F) and (G) for IGF inactive cells. In (F), (H), and (J), the color scale indicates cell density (unitless).

(L) Sankey diagram of the matching between erythroid states and CCC patterns.

(M) The representation fold change of erythroid cell states in CCC pattern 0.

(N) The top 15 ligands and receptors expressed in CCC pattern 0.

(O) Violin plot of the top 5 receptors (top) and ligands (bottom) in CCC pattern 0.

(P–R) Same as (M)–(O) for CCC pattern 5.

CCC pathways are orchestrated together at different stages of the cell lineage.

TGFb and BMP modulate parallel cell-state-transition trajectories

To test scRICH across different biological scenarios, we next consider a time course of EMT in the cancer cell line OVCA420 under exogenous transforming growth factor β (TGFb) exposure.⁴² Cells in the dataset undergo EMT from 0 to 7 days of TGFb exposure as reported based on cell morphology in the original publication as well as by examination

and finally, IGF signaling is mostly inactive in the later stages (Figures 5F–5K).

Finally, we investigate the emerging trend in CCC using an unsupervised learning strategy, leading to the identification of 12 CCC patterns with different levels of representation in the erythroid cell states (Figure 5L). These patterns exemplify how different CCC pathways are orchestrated at different stages of the erythroid lineage. For example, pattern 0 represents a CCC behavior overrepresented among early-stage cells, including BP-1, BP-2, and Er-1, featuring strong expression of the midkine (MK) receptor—a growth factor related to several cellular programs⁴¹—and the FGF, MK, and VEGF ligands, among others (Figures 5M–5O). In other words, albeit belonging to different cell states, cells in pattern 0 share a common CCC behavior when evaluating their communication profile across several different CCC pathways. Conversely, pattern #5 is enriched for cells in the later Er-2 stage and features co-expression of the MK ligand and receptor (Figures 5P–5R). Overall, these CCC patterns provide a comprehensive representation of how different

of signatures of epithelial and mesenchymal gene sets (Figures S6A and S6B). Clustering cells based on gene expression rather than the time point suggested 3 cell states corresponding to epithelial (E), intermediate (I), and mesenchymal (M) cell states, with the mesenchymal state becoming increasingly populated at later time points, in good agreement with the reported EMT induction (Figures S6C–S6E). The TGFbb pathway, a well-known driver of EMT, splits into 3 CCC signaling modes, including TGFb-receiver, TGFb-sender, and TGFb-inactive (Figure 6A). Although the TGFb CCC modes were spread heterogeneously across the E, I, and M cell states, the fraction of TGFb sender cells increased along the pseudotime axis, suggesting that cells undergoing EMT could, in turn, induce EMT in neighboring cells via TGFb signaling (Figures S6F–S6H). By overlaying the TGFb signaling states and cell-state-transition trajectories marked by RNA velocity, scRICH suggested that the cells in the early stage of the EMT trajectory tend to be mostly (epithelial, TGFb-inactive) or (epithelial, TGFb-receiver); conversely, at later stages, cells switch to the (intermediate,

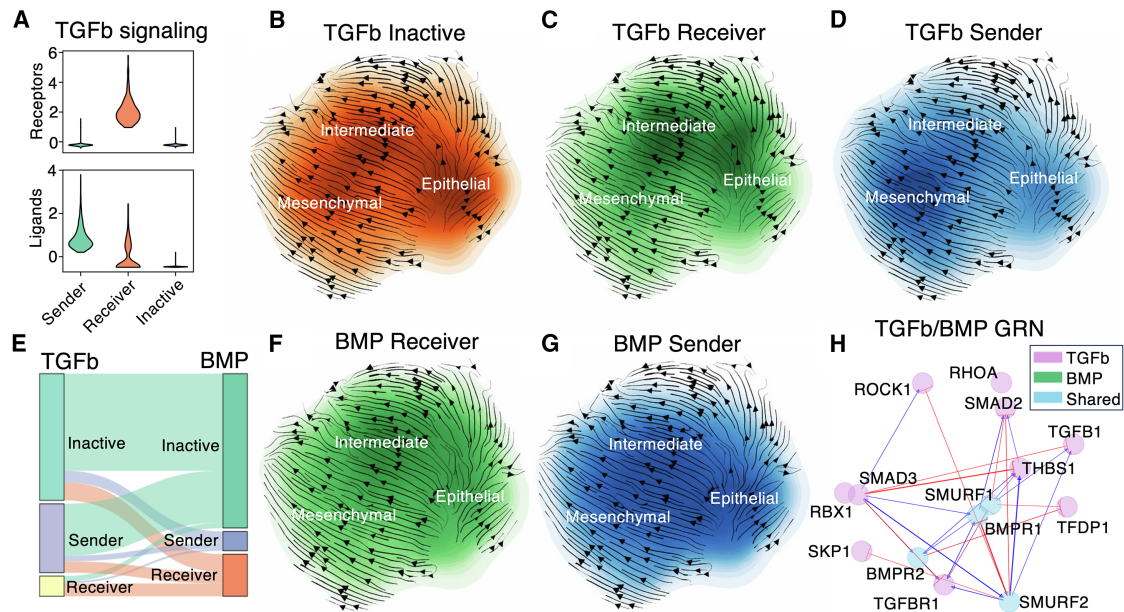


Figure 6. Interplay of TGFb and BMP during EMT

(A) The TGFb signaling modes in OVCA420 cells (expression is normalized and log transformed) in the dataset by Cook and Vanderhyden.⁴² (B–D) The cell density of TGFb CCC modes overlaid on RNA velocity for TGFb inactive (B), receivers (C), and senders (D). Color scales indicate cell density (unitless). (E) Sankey diagram of the correspondence between TGFb and BMP CCC modes. (F and G) Cell density of BMP receivers and BMP senders. Color scales indicate cell density (unitless). (H) The joint TGFb-BMP gene regulatory network. Purple and blue shading represent nodes associated with the TGFb pathway or shared between the two pathways, respectively.

TGFb-receiver) and finally to the (mesenchymal, TGFb-sender) states (Figures 6B–6D).

RNA velocity suggests that epithelial cells can either transition directly to the mesenchymal state or traverse a “longer” path through the intermediate state. To characterize the signaling mechanism along these trajectories, we evaluated the pathway similarity, finding a high mutual information between TGFb and BMP, another pathway usually associated with EMT (Figure S6I). Similar to TGFb, BMP exhibited both a sender and receiver signaling mode in the dataset (Figures S6J and S6K), which were highly correlated with the TGFb signaling modes. Specifically, TGFb-inactive and TGFb-receiver cells had a high probability of being BMP-inactive and BMP-receiver, respectively (Figure 6E). Interestingly, TGFb and BMP showed different activation patterns along the EMT trajectory. The (mesenchymal, TGFb-sender) cells tend to be largely BMP-inactive, whereas the (intermediate, TGFb-receiver) cells tend to be BMP-receivers as well. Combined with the RNA velocity transition trajectories, these results suggest alternative EMT trajectories. First, a direct E-M trajectory that relies on TGFb but not BMP activation; second, an E-I-M trajectory where both TGFb and BMP are activated (Figures 6F and 6G). To characterize the relationship between TGFb and BMP further, we reconstructed the intracellular GRN connecting the TGFb and BMP pathways, which identified several TFs commonly associated with EMT, including SMAD3, SMAD6, and SMURF2 (Figures 6H, S6M, and S6N).

scRICH identifies spatial patterns of sender and receiver cells from ST data

Finally, we consider an ST dataset of chicken heart development⁴³ originally obtained via the 10× Genomics Spatial RNA-seq Visium platform to leverage the spatial organization of CCC. Although the spatial distribution of cell types in this dataset is highly heterogeneous (Figure S7A), we hypothesize that CCC should still exhibit spatial organization due to the limited diffusion of ligands across the tissue. Among several CCC pathways (Figure S7B), we focus specifically on the highly expressed MK pathway, which is heavily implicated in the regulation of stem cell renewal, angiogenesis, and inflammation.⁴¹ Strikingly, we observe a robust spatial organization when classifying cells based on their MK communication behavior instead of cell type. Specifically, the three identified MK communication modes, including a receiver, a sender/receiver, and a weak sender/receiver mode, are distributed heterogeneously across cell types (Figure 7A) but are mainly localized to distinct spatial locations (Figures 7B–7D), resulting in a CCC network in which multiple cell types contribute to the signaling (Figure 7E). To further summarize the spatial organization of the MK CCC pathway, we define a spatial heterogeneity index, which quantifies whether cells assume a CCC communication mode similar to that of their spatial neighbors. Interestingly, cells in different spatial locations exhibit varying degrees of spatial heterogeneity, with the region corresponding to the strong sender/receiver MK mode being the more robust/less heterogeneous, whereas

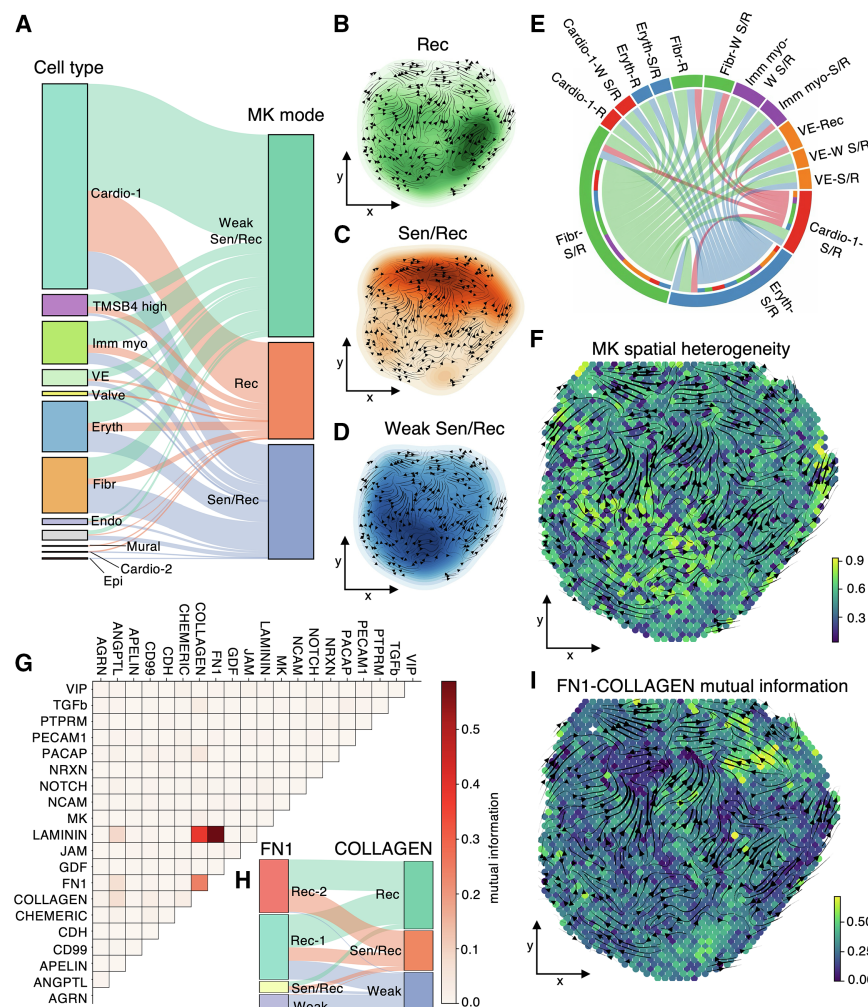


Figure 7. The spatial organization of CCC modes in the developing chicken heart

(A) Composition of cell types and MK CCC modes in the dataset by Mantri et al.⁴³ (B–D) Spatial distribution of the identified MK signaling modes, including MK receiver (B), MK sender/receiver (C), and MK weak sender/receiver (D). Color scales indicate cell density (unitless). (E) The MK communication network. (F) The spatial heterogeneity of MK CCC signaling indicated with a unitless color scale. (G) Mutual information between CCC pathways. (H) Correspondence of CCC signaling modes between the FN1 and COLLAGEN pathways. (I) Spatial distribution of the FN1-collagen mutual information in the developing chicken heart indicated with a unitless color scale.

regions, potentially suggesting distinct roles for these pathways locally in tissue development.

DISCUSSION

CCC enables the flow of information in multicellular systems and connects the tissue and cellular scales through intricate gene regulatory networks (GRNs) that process external information and inform cellular responses, for example, by activating secondary CCC pathways. High-throughput single-cell and ST technologies offer unprecedented opportunity to observe these biological processes at single-cell resolution.^{15,46,47}

Yet, existing models mostly focus on the scales of CCC mediated by ligand/receptor binding and intracellular GRNs separately,^{1,48,49} thus failing to provide a holistic picture of how CCC and cellular identity interact and evolve together in complex cell lineages.

To address this existing gap, we have developed *scRICH*, a mathematical framework and bioinformatics package that leverages single-cell transcriptomics, prior biological information on CCC pathways, and mathematical modeling of inter- and intracellular signaling dynamics. *scRICH* takes advantage of the unspliced and spliced RNA counts from scRNA-seq data to (1) quantify the diverse CCC behaviors and heterogeneity within cell types; (2) build a model of CCC that couples ligand/receptor binding with intracellular GRN response, thus enabling us to investigate the association between distinct CCC pathways; and (3) identify emerging global communication patterns and their evolution during cell-fate-transition events.

Recently, FlowSig proposed a learning framework to discover relationships between signaling pathway activation and expression of new ligands/receptors using conditional independence.⁵⁰ Compared with this method, *scRICH* focuses on (1) the heterogeneity of CCC behaviors within cell types and (2) the relationship between CCC and cell fate by incorporating RNA splicing dynamics, thus providing unique and complementary insights.

regions corresponding to weakly interacting cells are more heterogeneous (Figure 7F).

The spatial segregation of cells with distinct MK communication behaviors suggests that cell-cell interactions play a role in the development of the tissue. Thus, we investigate whether this spatial organization is also observed in the correlations between CCC pathways. Based on mutual information analysis, we identify pairs of pathways that are highly correlated, such as FN1 and COLLAGEN (Figures 7G and 7H). To test the relationship between these CCC pathways in different anatomical regions of the tissue, we define a local, space-dependent mutual information measure that only considers the relationship between the pathways in the neighborhood of a specific spatial location. This analysis showcases different anatomical locations of the tissue where the relationship between the CCC pathways differs (Figure 7I). For example, certain regions are characterized by high mutual information, thus implying that it is possible to predict a cell's communication behavior in the COLLAGEN pathway if FN1 behavior is known or vice versa. Generally, the connection between these pathways is not surprising, as both FN1 and COLLAGEN play a key role in the processing of interactions between cells and the extracellular matrix.^{44,45} These pathways, however, have low mutual information in specific spatial

First, scRICH treats cell types and CCC behavior as independent observations, thus enabling us to quantify how cells of the same type operate via different (sets of) CCC pathways and build CCC networks at the intermediate scale of meta-cells, i.e., groups of cells that share the same type and communication behavior. This approach enabled us to identify cell subpopulations that operate via distinct sets of cell-cell interactions despite sharing the same set of core marker genes, such as groups of EGF sender/receiver spinous cells that were validated experimentally based on EGF ligand/receptor co-expression in FibHSE organoids. Second, scRICH employs multilayer networks and mutual information to predict the connection between CCC pathways, for example, when predicting intermediate TFs HRAS and INHBA that orchestrate a joint (EGF receiver, TGFb receiver) cellular state in basal keratinocytes. When coupled with cell fate transition information, this approach can clarify how CCC pathways work together to modulate early decisions in cell fate. For example, scRICH predicts parallel EMT trajectories driven by TGFb in ovarian cancer-derived cells, whereby the first trajectory shows strong coupling between the TGFb and BMP pathways and passes through an intermediate epithelial/mesenchymal cell state, whereas the second, BMP-independent trajectory directly connects the E and M cell states.

Despite these encouraging results, this method could be generalized in the future to interface with other omics data modalities besides scRNA-seq to improve predictive power. First, we follow a common assumption in the field and utilize RNA counts to estimate ligand/receptor expression, thus disregarding several biological steps, including translation and ligand/receptor transport to the cell surface. This approximation poses specific challenges in biological contexts or datasets with poor protein-RNA correlation. Conversely, proteomics⁵¹ or cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)—which measures protein expression at the cell surface^{52,53}—could provide better estimations. Similarly, GRN inference based exclusively on single-cell transcriptomics, albeit by integrating unspliced and spliced RNA layers, would benefit in the future from integration with DNA accessibility data. For example, single cell assay for transposase-accessible chromatin (scATAC-seq) has been used in combination with regular (i.e., spliced) scRNA-seq for robust GRN inference by providing a prior on whether DNA sequences associated with specific genes are accessible.^{54,55} Moreover, scRICH leverages gene expression and/or RNA velocity (when available) of downstream targets to control for false positives, thus detecting CCC if and only if a downstream response is observed in the receiver cell. More strategies, however, should be investigated in the future to further alleviate false positives. For example, scRICH currently employs spatial information (ST) exclusively to investigate the spatial organization of CCC modes within a tissue, whereas other methods employ ST to cut off interactions between faraway cells,^{56–60} and the inferred results could be applied to non-spatial scRNA-seq datasets with more cells and deeper sequencing. For example, stMLnet combines ST for false-positive correction and intracellular gene signatures for CCC prediction.⁶¹ Finally, an additional limitation of ST is the increased data sparsity, as sequencing-based, whole-genome data are needed as scRICH considers ligands, receptors, and downstream targets over several CCC pathways. Future applications of scRICH will

benefit from the integration of ST with non-spatial data for improved data quality.⁶²

In addition to multi-modal and/or spatial high-throughput data, scRICH could be integrated with dynamical modeling of cell fate^{63–66} and/or temporal information, including time-course and lineage-tracing data, to further explore the relationship between cell fate and CCC. Existing methods leverage these types of data to estimate cell-state transitions under the assumption of cell-autonomous processes. For example, Cospar integrates lineage tracing and single-cell transcriptomics to build a Markov chain Monte Carlo model of cell-cell and state-state transitions,⁶⁷ whereas TIGON implements optimal transport to estimate transitions and cell growth between consecutive time points in time-course data.⁶⁸ Developing non-autonomous models that integrate multiscale CCC in the style of scRICH with cell-state transitions would provide a unifying framework to understand the influence of cell-cell interactions on cell-fate decisions.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact, Qing Nie (qnie@uci.edu).

Materials availability

This study did not generate new materials.

Data and code availability

This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the [key resources table](#).

All original code has been deposited at <https://github.com/federicobocci/Robust-Identification-of-Cell-cell-Communication-Heterogeneity-in-single-cells> and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

The work was partly supported by National Science Foundation grants DMS1763272 and CBET2134916.

AUTHOR CONTRIBUTIONS

F.B. and Q.N. conceived the project. F.B. designed the computational model, wrote the code, and applied the model to all analyzed single-cell datasets. Y.J. and S.A. designed the benchmarking experiments. Y.J. performed the benchmarking experiments and analyzed the newly generated data. F.B. wrote the manuscript with inputs from all authors. All authors revised and approved the paper. Q.N. supervised the project and acquired funding.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - FibHSE organotypic culture
 - Immunofluorescence staining

● **METHOD DETAILS**

- Input data and preprocessing
- Cell-cell signaling database
- Average expression within single cells
- Identification of cell-cell signaling modes
- Quantification of cell-cell signaling heterogeneity
- Statistical analysis of signaling mode robustness
- Model of cell-cell communication
- Emerging CCC patterns
- Pathway redundancy
- Inference of gene regulatory network
- Comparison with other CCC inference methods
- Data preprocessing
- Data availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cels.2026.101703>.

Received: July 22, 2025

Revised: April 29, 2026

Accepted: July 17, 2026

REFERENCES

1. Armingol, E., Baghdassarian, H.M., and Lewis, N.E. (2024). The diversification of methods for studying cell-cell interactions and communication. *Nat. Rev. Genet.* 25, 381–400. <https://doi.org/10.1038/s41576-023-00685-8>.
2. Almet, A.A., Cang, Z., Jin, S., and Nie, Q. (2021). The landscape of cell-cell communication through single-cell transcriptomics. *Curr. Opin. Syst. Biol.* 26, 12–23. <https://doi.org/10.1016/j.coisb.2021.03.007>.
3. Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019). Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* 47, e66. <https://doi.org/10.1093/nar/gkz204>.
4. Wilk, A.J., Shalek, A.K., Holmes, S., and Blish, C.A. (2024). Comparative analysis of cell-cell communication at single-cell resolution. *Nat. Biotechnol.* 42, 470–483. <https://doi.org/10.1038/s41587-023-01782-z>.
5. Subedi, S., and Park, Y.P. (2023). Single-cell pair-wise relationships untangled by composite embedding model. *iScience* 26, 106025. <https://doi.org/10.1016/j.isci.2023.106025>.
6. Raredon, M.S.B., Yang, J., Kothapalli, N., Lewis, W., Kaminski, N., Niklasen, L.E., and Kluger, Y. (2023). Comprehensive visualization of cell-cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* 39, btac775. <https://doi.org/10.1093/bioinformatics/btac775>.
7. Kojima, Y., Mii, S., Hayashi, S., Hirose, H., Ishikawa, M., Akiyama, M., Enomoto, A., and Shimamura, T. (2024). Single-cell colocalization analysis using a deep generative model. *Cell Syst.* 15, 180–192.e7. <https://doi.org/10.1016/j.cels.2024.01.007>.
8. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21, 31. <https://doi.org/10.1186/s13059-020-1926-6>.
9. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* 12, 1088. <https://doi.org/10.1038/s41467-021-21246-9>.
10. Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaude, M., Trichot, C., Kieffer, Y., Mechta-Grigoriou, F., and Soumelis, V. (2021). Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* 12, 1089. <https://doi.org/10.1038/s41467-021-21244-x>.
11. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* 15, 1484–1506. <https://doi.org/10.1038/s41596-020-0292-x>.
12. Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* 48, e55. <https://doi.org/10.1093/nar/gkaa183>.
13. Jin, Z., Zhang, X., Dai, X., Huang, J., Hu, X., Zhang, J., and Shi, L. (2022). InterCellDB: A User-Defined Database for Inferring Intercellular Networks. *Adv. Sci. (Weinh)* 9, e2200045. <https://doi.org/10.1002/adv.202200045>.
14. Deconinck, L., Cannoodt, R., Saelens, W., Deplancke, B., and Saeys, Y. (2021). Recent advances in trajectory inference from single-cell omics data. *Curr. Opin. Syst. Biol.* 27, 100344. <https://doi.org/10.1016/j.coisb.2021.05.005>.
15. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* 24, 695–713. <https://doi.org/10.1038/s41580-023-00615-w>.
16. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 17, 159–162. <https://doi.org/10.1038/s41592-019-0667-5>.
17. Cheng, J., Zhang, J., Wu, Z., and Sun, X. (2021). Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Brief. Bioinform.* 22, 988–1005. <https://doi.org/10.1093/bib/bbaa327>.
18. Hu, Y., Peng, T., Gao, L., and Tan, K. (2021). CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data. *Sci. Adv.* 7, eabf1356. <https://doi.org/10.1126/sciadv.abf1356>.
19. Dimitrov, D., Schäfer, P.S.L., Farr, E., Rodriguez Mier, P., Lobentanzer, S., Dugourd, A., Tanevski, J., Ramirez Flores, R.O., and Saez-Rodriguez, J. (2023). LIANA+: an all-in-one cell-cell communication framework. *Nat. Cell Biol.* 26, 1613–1622. <https://doi.org/10.1101/2023.08.19.553863>.
20. Zhang, Y., Liu, T., Hu, X., Wang, M., Wang, J., Zou, B., Tan, P., Cui, T., Dou, Y., Ning, L., et al. (2021). CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res.* 49, 8520–8534. <https://doi.org/10.1093/nar/gkab638>.
21. Lummertz da Rocha, E., Kubaczka, C., Sugden, W.W., Najia, M.A., Jing, R., Markel, A., LeBlanc, Z.C., Dos Santos Peixoto, R., Falchetti, M., Collins, J.J., et al. (2022). CellComm infers cellular crosstalk that drives haematopoietic stem and progenitor cell development. *Nat. Cell Biol.* 24, 579–589. <https://doi.org/10.1038/s41556-022-00884-1>.
22. Xin, Y., Lyu, P., Jiang, J., Zhou, F., Wang, J., Blackshaw, S., and Qian, J. (2022). LRLoop: a method to predict feedback loops in cell-cell communication. *Bioinformatics* 38, 4117–4126. <https://doi.org/10.1093/bioinformatics/btac447>.
23. Baruzzo, G., Cesaro, G., and Di Camillo, B. (2022). Identify, quantify and characterize cellular communication from single-cell RNA sequencing data with scSeqComm. *Bioinformatics* 38, 1920–1929. <https://doi.org/10.1093/bioinformatics/btac036>.
24. Badia-i-Mompel, P., Wessels, L., Müller-Dott, S., Trimbou, R., Ramirez Flores, R.O., Argelaguet, R., and Saez-Rodriguez, J. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet.* 24, 739–754. <https://doi.org/10.1038/s41576-023-00618-5>.
25. He, C., Zhou, P., and Nie, Q. (2023). exFINDER: identify external communication signals using single-cell transcriptomics data. *Nucleic Acids Res.* 51, e58. <https://doi.org/10.1093/nar/gkad262>.
26. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 29, 1363–1375. <https://doi.org/10.1101/gr.240663.118>.
27. Harmor, A.J., Hills, R.A., Rosser, E.M., Jones, M., Buneman, O.P., Dunbar, D.R., Greenhill, S.D., Hale, V.A., Sharman, J.L., Bonner, T.I., et al. (2009). IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and

- ion channels. *Nucleic Acids Res.* 37, D680–D685. <https://doi.org/10.1093/nar/gkn728>.
28. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
29. Sha, Y., Wang, S., Zhou, P., and Nie, Q. (2020). Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. *Nucleic Acids Res.* 48, 9505–9520. <https://doi.org/10.1093/nar/gkaa725>.
30. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* 38, 1408–1414. <https://doi.org/10.1038/s41587-020-0591-3>.
31. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* 560, 494–498. <https://doi.org/10.1038/s41586-018-0414-6>.
32. Bocci, F., Zhou, P., and Nie, Q. (2022). spliceJAC: transition genes and state-specific gene regulation from single-cell transcriptome data. *Mol. Syst. Biol.* 18, e11176. <https://doi.org/10.15252/msb.202211176>.
33. Stabell, A.R., Lee, G.E., Jia, Y., Wong, K.N., Wang, S., Ling, J., Nguyen, S.D., Sen, G.L., Nie, Q., and Atwood, S.X. (2023). Single-cell transcriptomics of human-skin-equivalent organoids. *Cell Rep.* 42, 112511. <https://doi.org/10.1016/j.celrep.2023.112511>.
34. Gao, M., Qiao, C., and Huang, Y. (2022). UniTVelo: temporally unified RNA velocity reinforces single-cell trajectory inference. *Nat. Commun.* 13, 6586. <https://doi.org/10.1038/s41467-022-34188-7>.
35. Vuoriluoto, K., Haugen, H., Kiviluoto, S., Mpindi, J.P., Nevo, J., Gjerdrum, C., Tiron, C., Lorens, J.B., and Ivaska, J. (2011). Vimentin regulates EMT induction by Slug and oncogenic H-Ras and migration by governing Axl expression in breast cancer. *Oncogene* 30, 1436–1448. <https://doi.org/10.1038/onc.2010.509>.
36. Yoh, K.E., Regunath, K., Guzman, A., Lee, S.-M., Pfister, N.T., Akanni, O., Kaufman, L.J., Prives, C., and Prywes, R. (2016). Repression of p63 and induction of EMT by mutant Ras in mammary epithelial cells. *Proc. Natl. Acad. Sci. USA* 113, E6107–E6116. <https://doi.org/10.1073/pnas.1613417113>.
37. Yu, Y., Wang, W., Lu, W., Chen, W., and Shang, A. (2021). Inhibin β -A (INHBA) induces epithelial-mesenchymal transition and accelerates the motility of breast cancer cells by activating the TGF- β signaling pathway. *Bioengineered* 12, 4681–4696. <https://doi.org/10.1080/21655979.2021.1957754>.
38. Haensel, D., Jin, S., Sun, P., Cinco, R., Dragan, M., Nguyen, Q., Cang, Z., Gong, Y., Vu, R., MacLean, A.L., et al. (2020). Defining Epidermal Basal Cell States during Skin Homeostasis and Wound Healing Using Single-Cell Transcriptomics. *Cell Rep.* 30, 3932–3947.e6. <https://doi.org/10.1016/j.celrep.2020.02.091>.
39. Peña, O.A., and Martin, P. (2024). Cellular and molecular mechanisms of skin wound healing. *Nat. Rev. Mol. Cell Biol.* 25, 599–616. <https://doi.org/10.1038/s41580-024-00715-1>.
40. Barile, M., Imaz-Rosshandler, I., Inzani, I., Ghazanfar, S., Nichols, J., Marioni, J.C., Guibentif, C., and Göttgens, B. (2021). Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. *Genome Biol.* 22, 197. <https://doi.org/10.1186/s13059-021-02414-y>.
41. Xu, C., Zhu, S., Wu, M., Han, W., and Yu, Y. (2014). Functional Receptors and Intracellular Signal Pathways of Midkine (MK) and Pleiotrophin (PTN). *Biol. Pharm. Bull.* 37, 511–520. <https://doi.org/10.1248/bpb.b13-00845>.
42. Cook, D.P., and Vanderhyden, B.C. (2020). Context specificity of the EMT transcriptional response. *Nat. Commun.* 11, 2142. <https://doi.org/10.1038/s41467-020-16066-2>.
43. Mantri, M., Scuderi, G.J., Abedini-Nassab, R., Wang, M.F.Z., McKellar, D., Shi, H., Grodner, B., Butcher, J.T., and De Vlaminc, I. (2021). Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. *Nat. Commun.* 12, 1771. <https://doi.org/10.1038/s41467-021-21892-z>.
44. Singh, P., Carraher, C., and Schwarzbauer, J.E. (2010). Assembly of Fibronectin Extracellular Matrix. *Annu. Rev. Cell Dev. Biol.* 26, 397–419. <https://doi.org/10.1146/annurev-cellbio-100109-104020>.
45. Hoffmann, G.A., and Smith, M.L. (2019). New Insights into Collagen and Fibronectin Reciprocity during Extracellular Matrix Formation. *Chem* 5, 1930–1932. <https://doi.org/10.1016/j.chempr.2019.07.012>.
46. Aldridge, S., and Teichmann, S.A. (2020). Single cell transcriptomics comes of age. *Nat. Commun.* 11, 4307. <https://doi.org/10.1038/s41467-020-18158-5>.
47. Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. *Nat. Methods* 19, 534–546. <https://doi.org/10.1038/s41592-022-01409-2>.
48. Pratapa, A., Jaliha, A.P., Law, J.N., Bharadwaj, A., and Murali, T.M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* 17, 147–154. <https://doi.org/10.1038/s41592-019-0690-6>.
49. Wang, L., Zhang, Q., Qin, Q., Trasanidis, N., Vinyard, M., Chen, H., and Pinello, L. (2021). Current progress and potential opportunities to infer single-cell developmental trajectory and cell fate. *Curr. Opin. Syst. Biol.* 26, 1–11. <https://doi.org/10.1016/j.coisb.2021.03.006>.
50. Almet, A.A., Tsai, Y.-C., Watanabe, M., and Nie, Q. (2024). Inferring pattern-driving intercellular flows from single-cell and spatial transcriptomics. *Nat. Methods* 21, 1806–1817. <https://doi.org/10.1038/s41592-024-02380-w>.
51. Bennett, H.M., Stephenson, W., Rose, C.M., and Darmanis, S. (2023). Single-cell proteomics enabled by next-generation sequencing or mass spectrometry. *Nat. Methods* 20, 363–374. <https://doi.org/10.1038/s41592-023-01791-5>.
52. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14, 865–868. <https://doi.org/10.1038/nmeth.4380>.
53. Liu, Y., DiStasio, M., Su, G., Asashima, H., Ennifou, A., Qin, X., Deng, Y., Nam, J., Gao, F., Bordignon, P., et al. (2023). High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat. Biotechnol.* 41, 1405–1409. <https://doi.org/10.1038/s41587-023-01676-0>.
54. Wang, L., Trasanidis, N., Wu, T., Dong, G., Hu, M., Bauer, D.E., and Pinello, L. (2023). Dictys: dynamic gene regulatory network dissects developmental continuum with single-cell multiomics. *Nat. Methods* 20, 1368–1378. <https://doi.org/10.1038/s41592-023-01971-3>.
55. Kamimoto, K., Stringa, B., Hoffmann, C.M., Jindal, K., Solnica-Krezel, L., and Morris, S.A. (2023). Dissecting cell identity via network inference and in silico gene perturbation. *Nature* 614, 742–751. <https://doi.org/10.1038/s41586-022-05688-9>.
56. Cang, Z., Zhao, Y., Almet, A.A., Stabell, A., Ramos, R., Plikus, M.V., Atwood, S.X., and Nie, Q. (2023). Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* 20, 218–228. <https://doi.org/10.1038/s41592-022-01728-4>.
57. Ru, B., Huang, J., Zhang, Y., Aldape, K., and Jiang, P. (2023). Estimation of cell lineages in tumors from spatial transcriptomics data. *Nat. Commun.* 14, 568. <https://doi.org/10.1038/s41467-023-36062-6>.
58. Garcia-Alonso, L., Handfield, L.-F., Roberts, K., Nikolakopoulou, K., Fernando, R.C., Gardner, L., Woodhams, B., Arutyunyan, A., Polanski, K., Hoo, R., et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* 53, 1698–1711. <https://doi.org/10.1038/s41588-021-00972-2>.
59. Rao, N., Kumar, T., Pai, R., Mishra, A., Ginhoux, F., Chan, J., Sharma, A., and Zafar, H. (2023). Charting spatial ligand-target activity using Renoir. *Nat. Commun.* 17, 3983. <https://doi.org/10.1038/s41467-026-72388-7>.

60. Shao, X., Li, C., Yang, H., Lu, X., Liao, J., Qian, J., Wang, K., Cheng, J., Yang, P., Chen, H., et al. (2022). Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat. Commun.* 13, 4429. <https://doi.org/10.1038/s41467-022-32111-8>.
61. Yan, L., Cheng, J., Nie, Q., and Sun, X. (2025). Dissecting multilayer cell-cell communications with signaling feedback loops from spatial transcriptomics data. *Genome Res.* 35, 1400–1414. <https://doi.org/10.1101/gr.279857.124>.
62. Khan, M., Arslanturk, S., and Draghici, S. (2025). A comprehensive review of spatial transcriptomics data alignment and integration. *Nucleic Acids Res.* 53, gkaf536. <https://doi.org/10.1093/nar/gkaf536>.
63. Qiu, X., Zhang, Y., Martin-Rufino, J.D., Weng, C., Hosseinzadeh, S., Yang, D., Pogson, A.N., Hein, M.Y., Hoi, Hoi Joseph Min, K., Wang, L., et al. (2022). Mapping transcriptomic vector fields of single cells. *Cell* 185, 690–711.e45. <https://doi.org/10.1016/j.cell.2021.12.045>.
64. Barcenas, M., Bocci, F., and Nie, Q. (2024). Tipping points in epithelial-mesenchymal lineages from single-cell transcriptomics data. *Biophys. J.* 123, 2849–2859. <https://doi.org/10.1016/j.bpj.2024.03.021>.
65. Zhou, P., Bocci, F., Li, T., and Nie, Q. (2024). Spatial transition tensor of single cells. *Nat. Methods* 21, 1053–1062. <https://doi.org/10.1038/s41592-024-02266-x>.
66. Wang, W., Ni, K., Poe, D., and Xing, J. (2024). Transiently Increased Coordination in Gene Regulation During Cell Phenotypic Transitions. *PRX Life* 2, 043009. <https://doi.org/10.1103/PRXLife.2.043009>.
67. Wang, S.-W., Herriges, M.J., Hurley, K., Kotton, D.N., and Klein, A.M. (2022). CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nat. Biotechnol.* 40, 1066–1074. <https://doi.org/10.1038/s41587-022-01209-1>.
68. Sha, Y., Qiu, Y., Zhou, P., and Nie, Q. (2023). Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nat. Mach. Intell.* 6, 25–39. <https://doi.org/10.1038/s42256-023-00763-w>.
69. Roth-Carter, Q.R., Koetsier, J.L., Broussard, J.A., and Green, K.J. (2022). Organotypic Human Skin Cultures Incorporating Primary Melanocytes. *Curr. Protoc.* 2, e536. <https://doi.org/10.1002/cpz1.536>.
70. Lutz-Bonengel, S., Niederstätter, H., Naue, J., Koziel, R., Yang, F., Sängner, T., Huber, G., Berger, C., Pflugradt, R., Strobl, C., et al. (2021). Evidence for multi-copy Mega-NUMTs in the human genome. *Nucleic Acids Res.* 49, 1517–1531. <https://doi.org/10.1093/nar/gkaa1271>.
71. Li, J., and Sen, G.L. (2015). Generation of Genetically Modified Organotypic Skin Cultures Using Devitalized Human Dermis. *J. Vis. Exp.* 53280, e53280. <https://doi.org/10.3791/53280>.
72. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
73. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. <https://doi.org/10.1038/nbt.4096>.
74. Melsted, P., Boeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H.J., Da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nat. Biotechnol.* 39, 813–818. <https://doi.org/10.1038/s41587-021-00870-2>.
75. He, Z., Xie, S., Zdunek, R., Zhou, G., and Cichocki, A. (2011). Symmetric Nonnegative Matrix Factorization: Algorithms and Applications to Probabilistic Clustering. *IEEE Trans. Neural Netw.* 22, 2117–2131. <https://doi.org/10.1109/TNN.2011.2172457>.
76. Sha, Y., Wang, S., Bocci, F., Zhou, P., and Nie, Q. (2021). Inference of Intercellular Communications and Multilayer Gene-Regulations of Epithelial-Mesenchymal Transition From Single-Cell Transcriptomic Data. *Front. Genet.* 11, 604585. <https://doi.org/10.3389/fgene.2020.604585>.
77. Abdelaal, T., Grossouw, L.M., Pasterkamp, R.J., Lelieveldt, B.P.F., Reinders, M.J.T., and Mahfouz, A. (2024). SIRV: spatial inference of RNA velocity at the single-cell resolution. *NAR Genom. Bioinform.* 6, lqae100. <https://doi.org/10.1093/nargab/lqae100>.
78. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 490–495. <https://doi.org/10.1038/s41586-019-0933-9>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-EGF	Proteintech	27141-1-AP; RRID: AB_2880771
Mouse anti-EGFR	Proteintech	66455-1-Ig; RRID: AB_2881824
Rabbit anti-HRAS	Proteintech	18295-1-AP; RRID: AB_2121046
Mouse anti-INHBA	Proteintech	60352-1-Ig; RRID: AB_2881461
Rabbit anti-CDH4	Proteintech	19795-1-AP; RRID: AB_10642434
Mouse anti-STAT1	Proteintech	66545-1-Ig; RRID: AB_144695
Rabbit anti-NOTCH1	Proteintech	20687-1-AP; RRID: AB_2535812
Goat anti-Mouse, Alexa Fluor 546	Invitrogen	A-11030
Goat anti-Rabbit, Alexa Fluor 647	Invitrogen	A-21244
Biological samples		
Human skin	New York Firefighters Skin Bank	http://www.cornellsurgery.org/pro/services/burn-surgery/skin-bank.html ; RRID:SCR_004604
Chemicals, peptides, and recombinant proteins		
Penicillin-Streptomycin	GIBCO	15140-122
Keratinocyte-SFM medium	GIBCO	10724-011
Bovine Pituitary Extract	GIBCO	13028-014
Human recombinant Epidermal Growth Factor	GIBCO	10450-013
DMEM	GIBCO	11995
Ham's F12	Cambrex	12-615F
FBS	GIBCO	10437-028
Adenine	Sigma	A-9795
Cholera Toxin	Sigma	C-8052
Hydrocortisone	Calbiochem	3896
Insulin	Sigma	I-1882
Transferrin	Sigma	T-0665
Ciprofloxacin Hydrochloride	Serologicals	89-001-1
Experimental models: Cell lines		
Primary human keratinocytes	Hospital Maternity Ward	N/A
Primary human fibroblasts	Hospital Maternity Ward	N/A
Software and algorithms		
R	R core	https://www.r-project.org/
scVelo	Bergen et al. ³⁰	https://scvelo.readthedocs.io/en/stable/
UnitVelo	Gao et al. ³⁴	https://unitvelo.readthedocs.io/en/latest/
Cellchat	Jin et al. ⁹	https://github.com/jinworks/CellChat
CellphoneDB	Efremova et al. ¹¹	https://www.cellphonedb.org
Liana+	Dimitrov et al. ¹⁹	https://liana-py.readthedocs.io/en/latest/
exFinder	He et al. ²⁵	https://github.com/ChanghanGitHub/exFINDER
spliceJAC	Bocci et al. ³²	https://splicejac.readthedocs.io/en/latest/
scRICH	This paper	https://github.com/federicobocci/Robust-Identification-of-Cell-cell-Communication-Heterogeneity-in-single-cells ; DOI: 10.5281/zenodo.20394265
External (published) datasets		
human-skin-equivalent organoids (FibHSE)	Stabell et al. ³³	GSE190695

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Murine skin wound healing	Haensel et al. ³⁸	GSE142471
Erythroid differentiation	Pijuan-Sala et al. ⁷⁹	GSE87038
OVCA420 time course	Cook and Vanderhyden ⁴²	GSE147405
Chicken heart	Mantri et al. ⁴³	GSE149457

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

FibHSE organotypic culture

Human primary keratinocytes and dermal fibroblasts were isolated from discarded neonatal foreskin, as described previously.^{69,70} Primary human keratinocytes were maintained below 70% confluency in K-SFM medium (Keratinocyte-SFM medium supplemented with 50 µg/ml Bovine Pituitary Extract, 5 ng/ml human recombinant Epidermal Growth Factor 1–53 and 100 U/mL penicillin-streptomycin). Primary fibroblasts were maintained in complete DMEM (DMEM supplemented with 10% fetal bovine serum and 100 U/mL penicillin-streptomycin). Split-thickness human skin was purchased from the New York Firefighters Skin Bank (New York, New York, USA). FibHSEs were generated using a modified protocol as outlined Li and Sen.⁷¹ Briefly, primary fibroblasts (1×10^6 cells) were seeded onto the reticular side of devitalized de-epidermal human dermis (DED) in 1.5 mL of complete DMEM, followed by centrifugation at 600 g for 1 hour. A suspension of primary keratinocytes (1×10^6 cells in 100 µL) was then seeded onto the papillary side of the DED and cultured at an air-liquid interface (ALI) to promote differentiation and stratification over 14 days, with media changes every other day. To assess the impact of EGF supplementation, an EGF-free culture medium was introduced on day 7 of ALI.

Immunofluorescence staining

Immunofluorescence (IF) staining were performed as previously described³³ with slight modifications. In brief, frozen slides containing 20 µm organoid sections were fixed with 4% paraformaldehyde (PFA) for 15 min, blocked in PBS containing 2% bovine serum albumin (BSA) and 0.1% Triton X-100 for 1h, and co-stained with either rabbit anti-EGF (1:500; Proteintech; 27141-1-AP) and mouse anti-EGFR (1:500; Proteintech; 66455-1-Ig), or rabbit anti-HRAS (1:500; Proteintech; 18295-1-AP) and mouse anti-INHBA (1:500; Proteintech; 60352-1-Ig). Slides were washed and then incubated with Alexa Fluor 546 Goat Anti-Mouse IgG (H+L) (Invitrogen) and Alexa Fluor 647 Goat Anti-Rabbit IgG (H+L) (Invitrogen). Mounting medium (abcam) containing 4,6-diamidino-2-phenylindole (DAPI) was used to visualize nuclei. All slides were immunostained in a single batch to ensure consistency across samples. Images were captured during the same experimental session using consistent settings on a ZEISS LSM780 confocal microscope at 20x magnification. For each condition, three different fields of tissue sections were acquired from two independent FibHSE replicates. Colocalization analysis was performed using the *Colocalization Finder* plugin in ImageJ.

METHOD DETAILS

Input data and preprocessing

The input to scRICH is the scRNA-seq data object, which is assumed to be preprocessed following standard preprocessing pipelines,^{72,73} and includes log-normalized RNA counts, filtered genes and cell annotations. In addition, the unspliced and spliced RNA layers are required for the downstream modeling, GRN inference, and visualization. The unspliced/spliced counts are obtainable via compatible python packages including *velocyto*³¹ and *kallisto bustool*⁷⁴ to allow efficient streamline of the analysis in a single notebook.

Cell-cell signaling database

scRICH includes a pre-built database comprising a list of ligands and receptors for cell-cell communication pathways updated based on CellChat database,⁹ spanning both human and mouse. Moreover, a database of downstream transcription factors (TF) was constructed by merging information about ligand-receptor-TF connections from NicheNet¹⁶ and exFinder.²⁵ All scRICH's analysis steps can be executed from the pre-built database. Additionally, several package applications can be integrated with custom-built lists of upregulated/downregulated downstream genes for individual CCC pathways.

Average expression within single cells

Several modeling steps of scRICH use average expression of ligands, receptors, and/or downstream targets within individual cells. scRICH offers multiple possibilities to compute these averages, which can be chosen by the user, including: (1) Log-normalized RNA counts (standard output of scanpy's preprocessing⁷²); (2) moments of RNA counts; and (3) Average over the cell neighborhood graph.

Identification of cell-cell signaling modes

For each pathway included in the database, scRICH identifies the number of communication modes by applying symmetric non-negative factorization of a cell-cell similarity matrix,⁷⁵ which has been previously applied for unbiased clustering of scRNA-seq

data.^{29,76} First, different K-means cell clustering solutions are computed based solely on the average expression of ligands and receptors in each cell. By default, 5 clustering solutions with randomized seed are computed for $k=3,4,5,6$, whereby all parameters can be modified by the user. Second, a cell-cell similarity matrix is computed where a matrix element $S_{ij} \in [0, 1]$ is the fraction of clustering solutions where cells i and j are grouped in the same cluster. The similarity matrix is then decomposed using symmetric non-negative factorization:

$$\min_{H \geq 0} \|S - HH^T\|_F^2, H \in R^{n \times k}, \quad (\text{Equation 1})$$

where S is the cell-cell similarity matrix, H is a non-negative low-rank matrix, and n, k are the number of cells and clusters. The optimal number of signaling modes is identified based on the largest gap between the sorted eigenvalues of the matrix H .

Quantification of cell-cell signaling heterogeneity

After identification of CCC modes of a given pathway P , the dataset can be categorized into N cell types and K CCC modes, resulting in $N \times K$ sub-clusters. The heterogeneity (h) of cell type (i) with respect to the CCC pathway P is defined as:

$$h_i = 1 - \frac{1}{K} \sum_{j=1}^K \frac{|n_{ij} - \rho_j|}{s_{ij}}, \quad (\text{Equation 2})$$

where n_{ij} is the fraction of cells in cell type i that were clustered in CCC mode j , and ρ_j is the overall fraction of cells in signaling mode j in the entire dataset. Finally, s_{ij} is a normalization factor to ensure that h_i is bound between 0 and 1:

$$s_{ij} = \begin{cases} 1 - \rho_j & \text{if } n_{ij} - \rho_j > 0 \\ \rho_j & \text{otherwise} \end{cases}. \quad (\text{Equation 3})$$

Based on this definition, the cell type heterogeneity score h_i is minimized ($h_i = 0$) when all cells within cell type i assume the same CCC mode; conversely, h_i is maximized ($h_i = 1$) when cells within cell type i are divided between all CCC modes with same proportions observed in the whole dataset.

An overall heterogeneity score for the CCC pathway P is obtained by averaging the heterogeneity scores h_i over all cell types in the dataset. This aggregated score quantifies whether cells with different roles in the CCC pathway (senders, receivers, etc.) are uniformly distributed across cell types (low score) or clearly divided along cell types (high score). Conversely, the overall heterogeneity score of a cell type is obtained by averaging the heterogeneity scores of cell type i over all signaling pathways and informs whether cells of that type assume different communication behaviors across many CCC pathways (low score) or mostly behave consistently (high score).

To define a spatial heterogeneity score, the definition of Equation 3 is modified to sum over all cells within a cutoff distance from a given spatial coordinate (x_i, y_i) .

Statistical analysis of signaling mode robustness

Given a CCC pathway P with K associated CCC signaling modes, cells within any cell type will split between these K modes (for example: the SPN-3 cell type splits between $K = 3$ signaling modes in the EGF pathway including with sender, sender/receiver, and weak modes – see Figure 3C). We seek to discriminate between situations where (1) the cell type is largely composed of cells from a single mode, and it is thus possible to average CCC behavior over the entire cell type, or (2) the cell type is a mixture of multiple modes and cell type-level averaging would lead to significant loss of information. Let N_i be the number of cells in cell type i , and N_{ij} the fraction of cells in cell type i that were clustered in CCC mode j . Let n_{ij} be a Poisson-distributed random variable with mean and standard deviation equal to N_{ij} : $P(n_{ij} = x) = \frac{N_{ij}^x e^{-N_{ij}}}{x!}$. Then, we define a p-value as the probability to randomly sample a number of cells equal or larger than N_i . By default, the threshold value is set to a default value $\tau = 0.05$ following standard convention for p-value thresholding and based on a sample of 10000 random numbers, but the parameters can be modified by users.

Model of cell-cell communication

First, a score of the communication via a specific CCC pathway S_{ij} is defined between each pair of cells in the dataset as a product of signaling intensity (I_{ij}) and intracellular response (K_j):

$$S_{ij} = I_{ij}(L_i, R_j) K_j(T_j). \quad (\text{Equation 4})$$

Based on this definition, communication from cell i to cell j requires both (1) expression of ligands/receptors as well as (2) activation of the pathway's downstream targets in the receiving cell. First, $I_{ij}(L_i, R_j)$ is the intensity whereby L_i and R_j represent the average expression of ligand and receiver in cells i and j , respectively. scRICH implements two possible models for signaling intensity, including a diffusion model:

$$I_{ij}(L_i, R_j) = e^{-1/L_i R_j} \quad (\text{Equation 5})$$

and a Michaelis-Menten model:

$$I_{ij}(L_i, R_j) = \frac{L_i R_j}{C + L_i R_j} \quad (\text{Equation 6})$$

To normalize the variable copy numbers associated to different pathways, the product $L_i R_j$ is rescaled within $[0, 1]$ by normalizing to the largest value of $L_i R_j$ among all cell pairs in the dataset. Based on this normalization, we set $C = 0.5$ in Equation 6. The user can choose between the diffusion and Michaelis-Menten models, and further comparison is presented in Figure S2. Second, $K_j(T_j)$ quantifies the intracellular response in the receiving cell as:

$$K(T_j) = e^{-m_T/T_j}, \quad (\text{Equation 7})$$

where T_j is the activation of downstream pathways in cell j and m_T is the number of downstream targets. The activation is defined either as the average expression of downstream pathways in the receiving cell:

$$T_j = \sum_{k=1}^{m_T} t_{kj}, \quad (\text{Equation 8})$$

or, in the “RNA velocity communication model”, as the average RNA velocity of the downstream targets in the receiving cell:

$$T_j = \sum_{k=1}^{m_T} \beta U_{kj} - \gamma S_{kj}, \quad (\text{Equation 9})$$

where U_{kj} and S_{kj} are the unspliced and spliced copy numbers of the k -th downstream target in cell j , and β and γ are the splicing and degradation rate constants computed using the scVelo package.³⁰

Finally, the dense cell-to-cell communication network is coarse-grained to a meta-cell to meta-cell network by aggregating the cell-pair communication scores based on the meta-cell assignment that summarizes both cell type and CCC signaling modes that was described in a previous section.

Emerging CCC patterns

After CCC modes are identified for multiple (K) CCC signaling pathways, each cell in the dataset is described by a K -long vector whereby the i^{th} element represents the cell's membership to the i^{th} CCC signaling pathway's mode. The K observables are used for unsupervised clustering using the Leiden algorithm, whereby each cluster represents a CCC pattern.

Pathway redundancy

Given a system where cells are clustered in N_1 and N_2 CCC modes along two CCC signaling pathways (S_1 and S_2), let $P_1(j)$ and $P_2(k)$ be the probabilities to observe a cell in the j -th CCC mode of pathway S_1 and the k -th mode of pathway S_2 , respectively, and $P(j, k)$ be the joint probability to observe a cell in the j -th CCC mode for pathway S_1 and the k -th CCC mode for pathway S_2 . The redundancy between the two pathways is defined as:

$$R_{P_1, P_2} = \frac{I(P_1, P_2)}{H(P_1) + H(P_2)}, \quad (\text{Equation 10})$$

where $I(P_1, P_2)$ is the mutual information between P_1 and P_2 , and $H(P_1)$ and $H(P_2)$ are the entropies of P_1 and P_2 . With the definition of Equation 10, R_{P_1, P_2} is semi-positive and assumes a minimum value of 0 if and only if the membership to CCC modes for the two pathways are completely uncorrelated.

Inference of gene regulatory network

To construct the coupled gene regulatory network linking two CCC pathways (i.e., EGF and TGFb), we first compiled a gene set including: (1) the set of ligands and receptor of the two signaling pathways and (2) the set of downstream genes associated with the two pathways. scRICH includes a built-in database for ligand, receptor, and intracellular mediators, but offers the option to provide user inputs. Then, the GRN is inferred from the scRNA-seq data object including the unspliced and spliced layers and the specified gene set using the spliceJAC package.³²

Comparison with other CCC inference methods

The benchmarking against existing CCC inference methods includes 3 steps: (1) Calculation of scRICH's CCC signaling matrix; (2) calculation of other method(s)'s CCC signaling matrix; and (3) comparison via multiple scoring metrics. First, scRICH is run on a given CCC signaling pathways. The resulting CCC network, however, is averaged over the CCC signaling modes, resulting in a CCC matrix between cell types. Second, we compute the same CCC matrix using existing methods using CellChat, CellPhoneDB and Liana+ using their standard pipeline and default settings. Liana+ integrates multiple CCC inference methods to provide a consensus prediction for each distinct ligand-receptor pair.¹⁹ These pairs were aggregated into CCC pathways using the Cellchat dataset (for example, interactions between Jag1-Notch1 and Jag2-Notch2 are aggregated into a unique interaction for Notch pathway). The output CCC

matrices from these methods describe interactions between cell types and are thus directly comparable with scRICH's coarse-grained output. Finally, the CCC matrices are compared based on 3 scoring metrics. (1) Spearman correlation is computed on the one-dimensional vectors obtained by flattening the two-dimensional CCC matrices. (2) Matrix distance is a squared sum of pair-wise distances between matrix elements:

$$D_{M_1, M_2} = \sum_{ij} |M_1^{(ij)} - M_2^{(ij)}|^2, \quad (\text{Equation 11})$$

where the sum runs over the matrix's rows (i) and columns (j). (3) Interaction overlap is the fraction of connection between cell types that are identified in both models, regardless of interaction amplitude.

Data preprocessing

OVCA420 time course data was preprocessed, normalized, regressed and scaled using scanpy's standard workflow⁷² (min_genes=200, min_cells=3, genes_by_counts<3500, total_counts<1500, target_sum=1e4, max_value=10). Highly variable genes were identified (min_mean=0.0125, max_mean=3, min_disp=0.5), followed by clustering with n_pcs=40, n_neighbors=10. The erythroid differentiation dataset was downloaded and preprocessed using scvelo's built-in functions after setting min_shared_counts=20. For skin organoid and OVCA420 datasets, RNA velocity was computed following scvelo's standard workflow with n_pcs=30, n_neighbors=30. For the erythroid differentiation dataset, RNA velocity was computed following Unitvelo's tutorial on the same dataset maintaining the same parameter values.³⁴ For the developing chicken heart spatial transcriptomics dataset, RNA velocity were computed following SIRV's tutorial on the same dataset maintaining the same parameter values.⁷⁷

Data availability

All sequencing data analyzed in this project is publicly available through NCBI's Gene Expression Omnibus (GEO) database with accession numbers GSE190695 (human-skin-equivalent organoids), GSE147405 (time course of OVCA420 cells), GSE149457 (Spatial transcriptomics of chicken heart development). The erythroid differentiation dataset was originally published by Pijuan-Sala et al.,⁷⁸ but was accessed through scvelo's built-in *datasets.gastrulation_erythroid()* function.³⁰

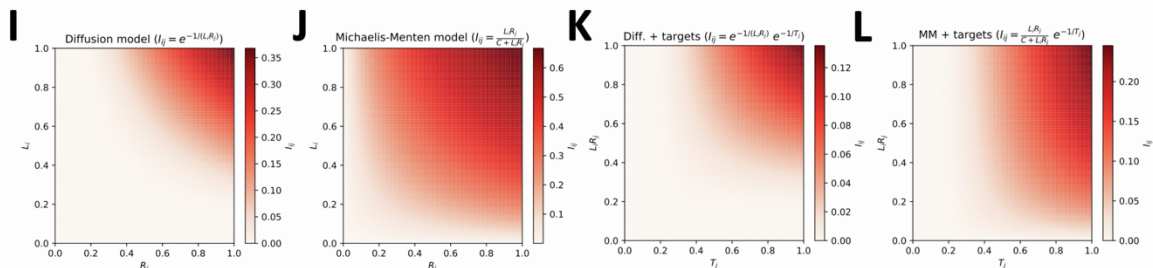
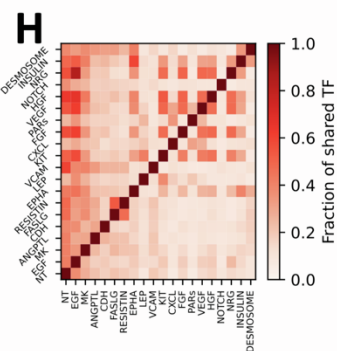
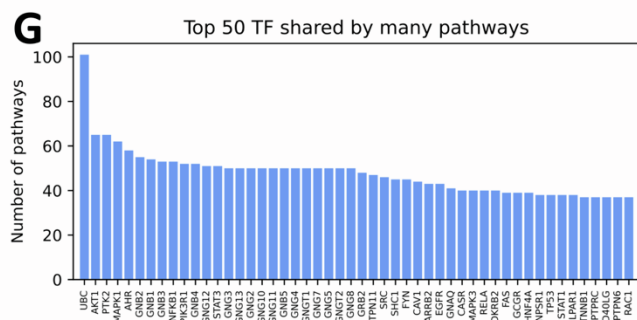
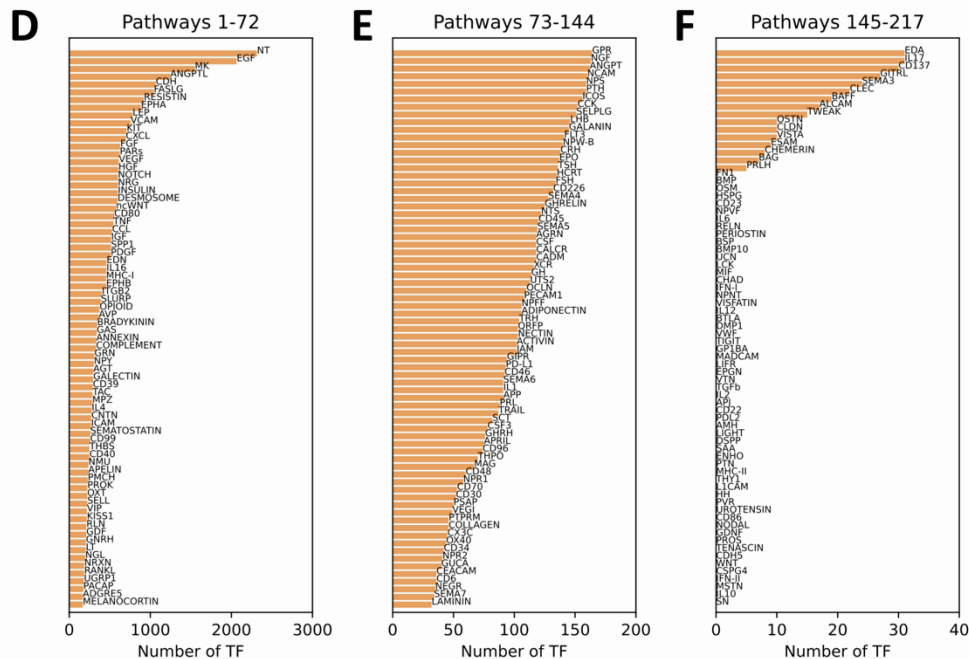
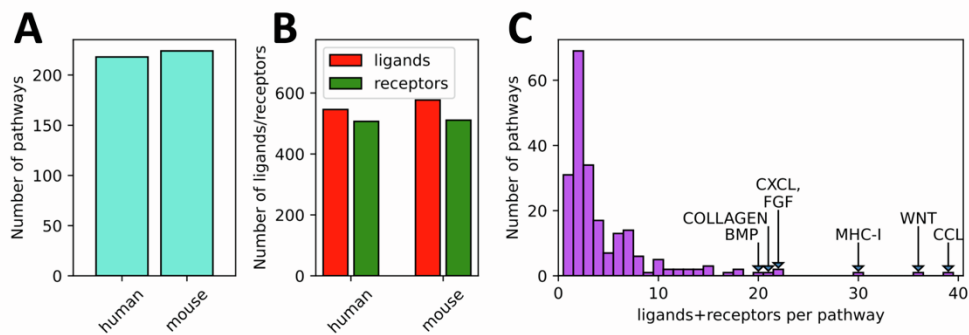
Cell Systems, Volume 17

Supplemental information

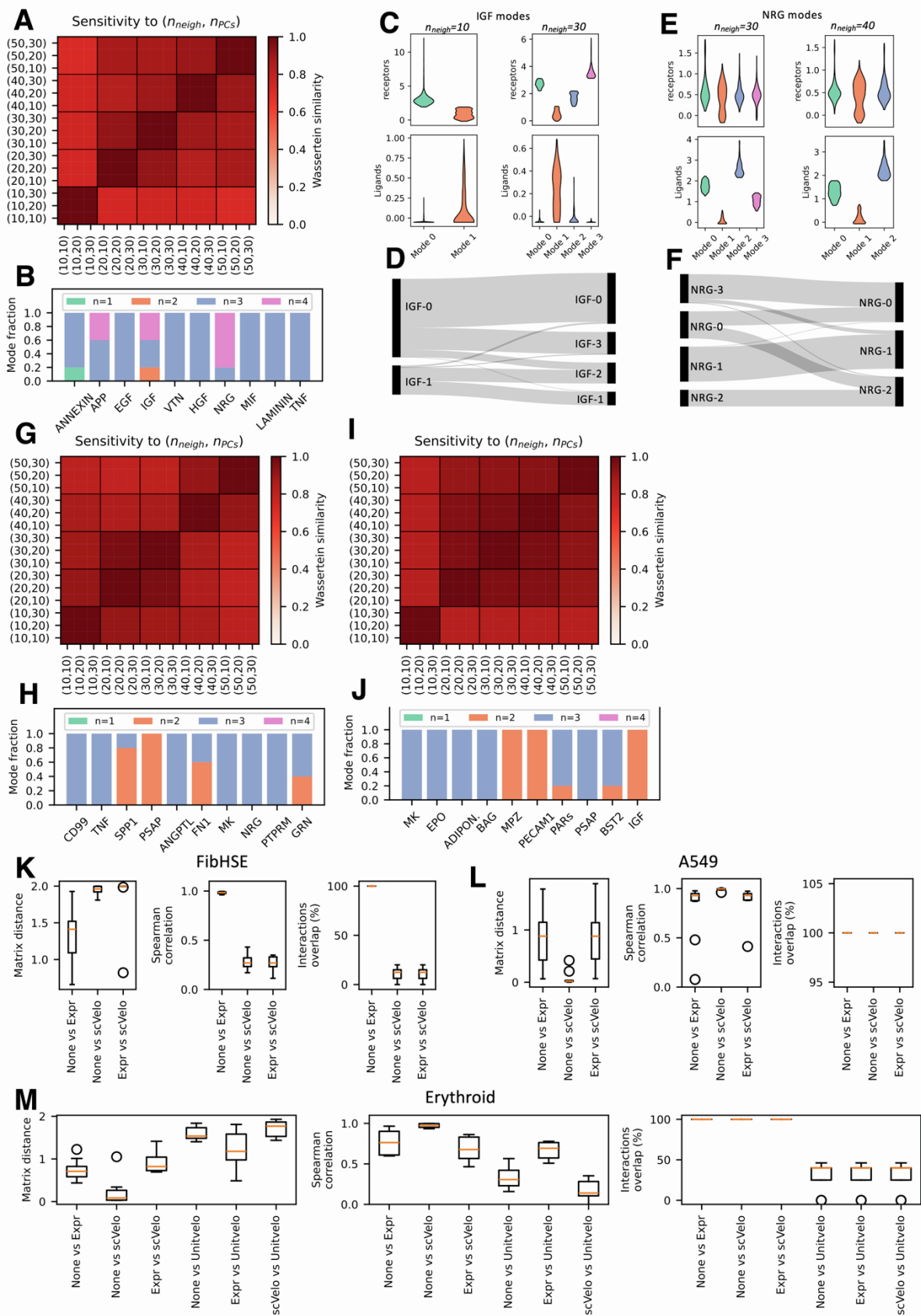
**Robust identification of cell-cell communication
heterogeneity in single cells**

Federico Bocci, Yunlong Jia, Scott Atwood, and Qing Nie

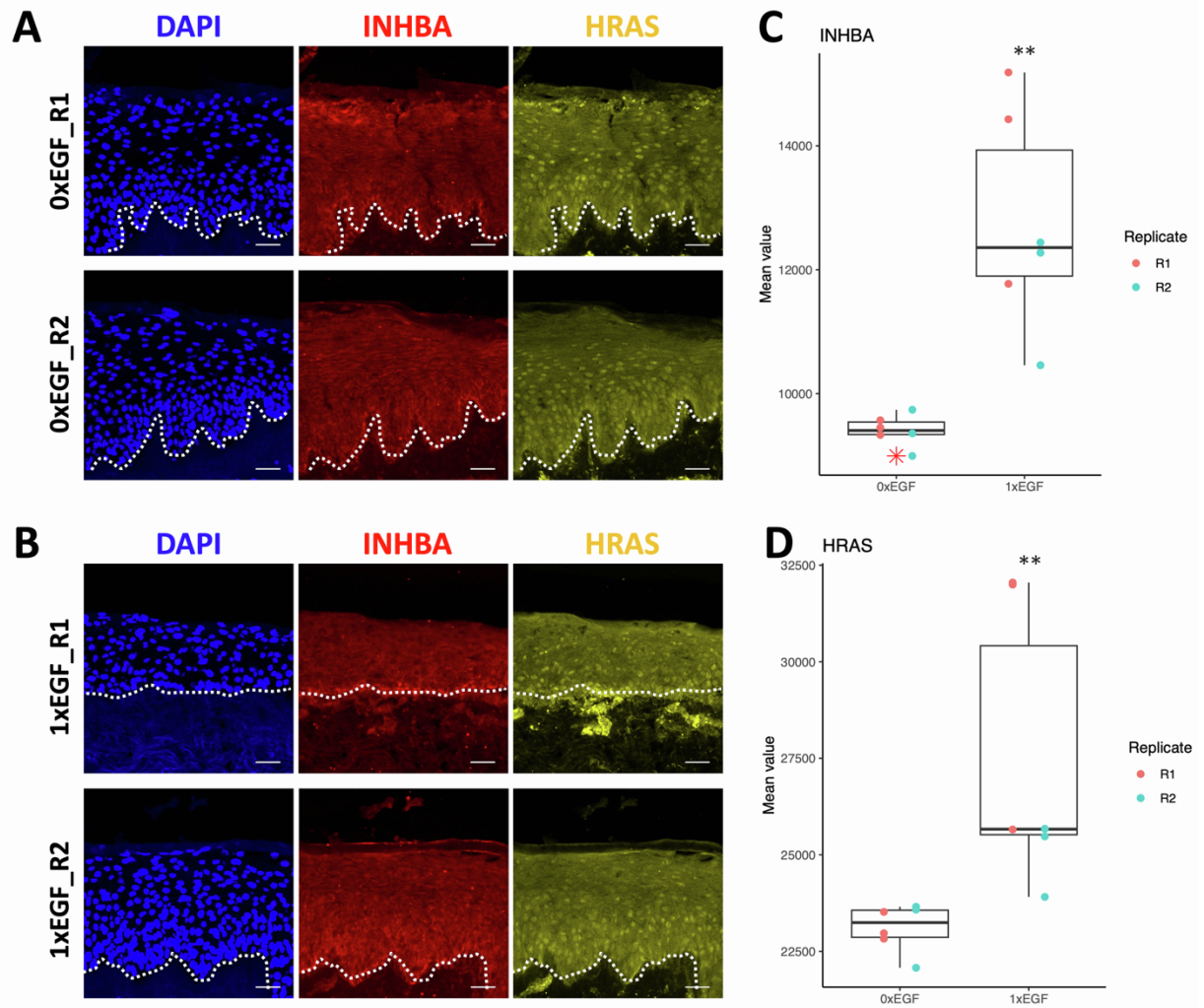
Supplementary Figures



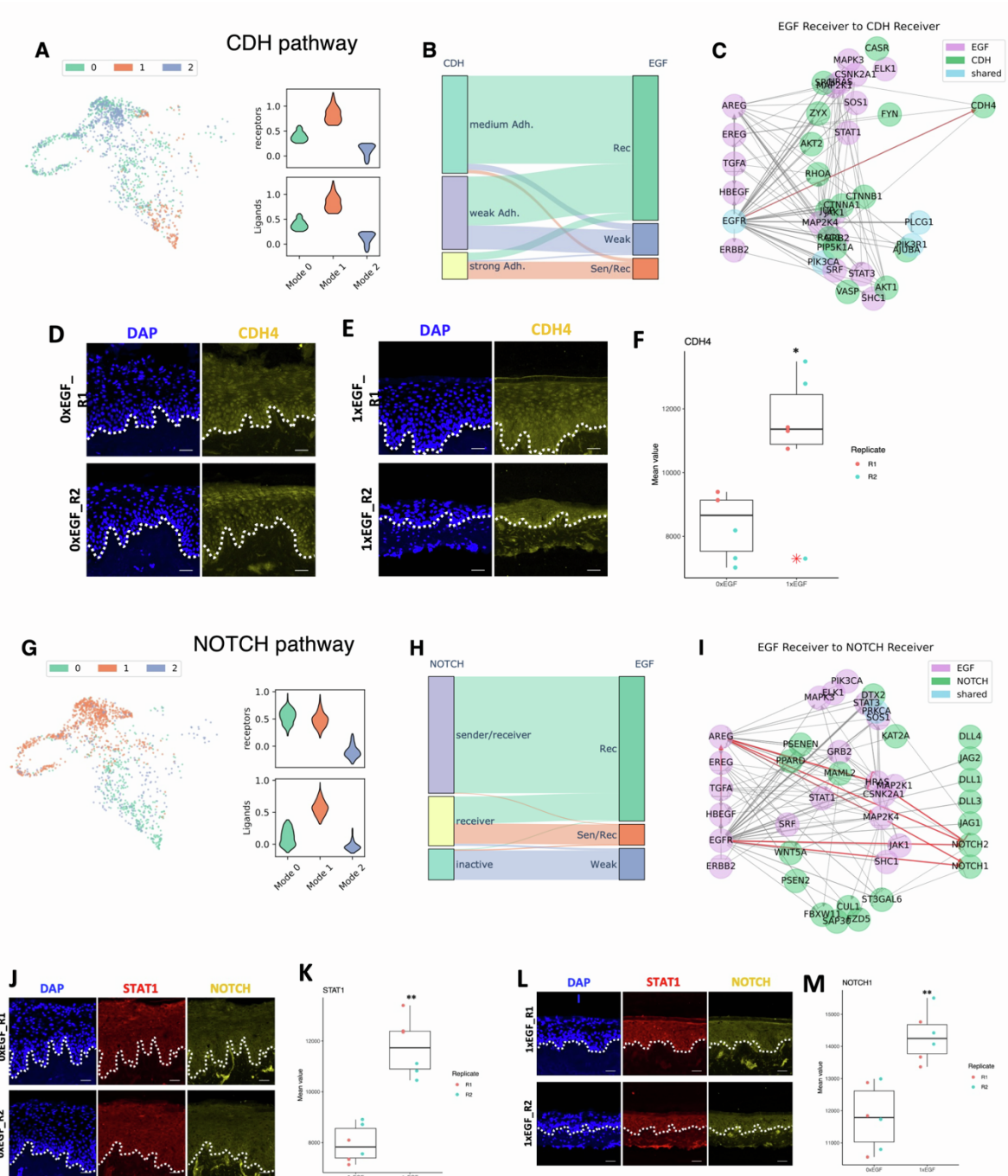
Supplementary figure 1. Overview of the scRICH database (related to Figure 1). (A) Number of cell-cell communication signaling pathways for human and mouse. (B) Total number of ligands and receptors for human and mouse. (C) Histogram of number of ligands + receptors per pathway. Annotations highlight the pathways with higher number of ligands + receptor. (D-E-F) Number of transcription factors (TF) directly connected to each pathway's set of receptors. The pathways are split between the 3 panels for ease of visualization. (G) The top 50 TF that are directly connected to the most pathways. (H) The TF overlap between the top 20 pathways (based on panel D). The overlap is quantified as the fraction of TF that are shared between the pair of pathways. (I-L) Comparison of signaling models between meta-cells. (I) The signaling between meta-cells i (the sender) and meta-cell j (the receiver) as a function of receiver and ligand levels in the simplified diffusion model without downstream target. (J) The signaling between meta-cells in the simplified Michaelis-Menten model without downstream target. While both models are symmetrical with respect to L_i and R_j , the signaling in the Michaelis-Menten model increases gradually, whereas the signaling exhibits a drastic cut-off when L_i and R_j are below threshold values. (K) The signaling between meta-cells as a function of receiver-ligand product $L_i R_j$ and downstream target levels in the receiver meta-cell T_j in the diffusion model. (L) Same as panel K in the Michaelis-Menten model with downstream target. For simplicity, in I-L it is assumed here that only one species of ligand, receptor, and target exist. ScRICH averages expression levels in presence of multiple species. $C = 1$ in both panels J and L, since it is assumed that ligand and receptor expression are re-normalized in $[0, 1]$.



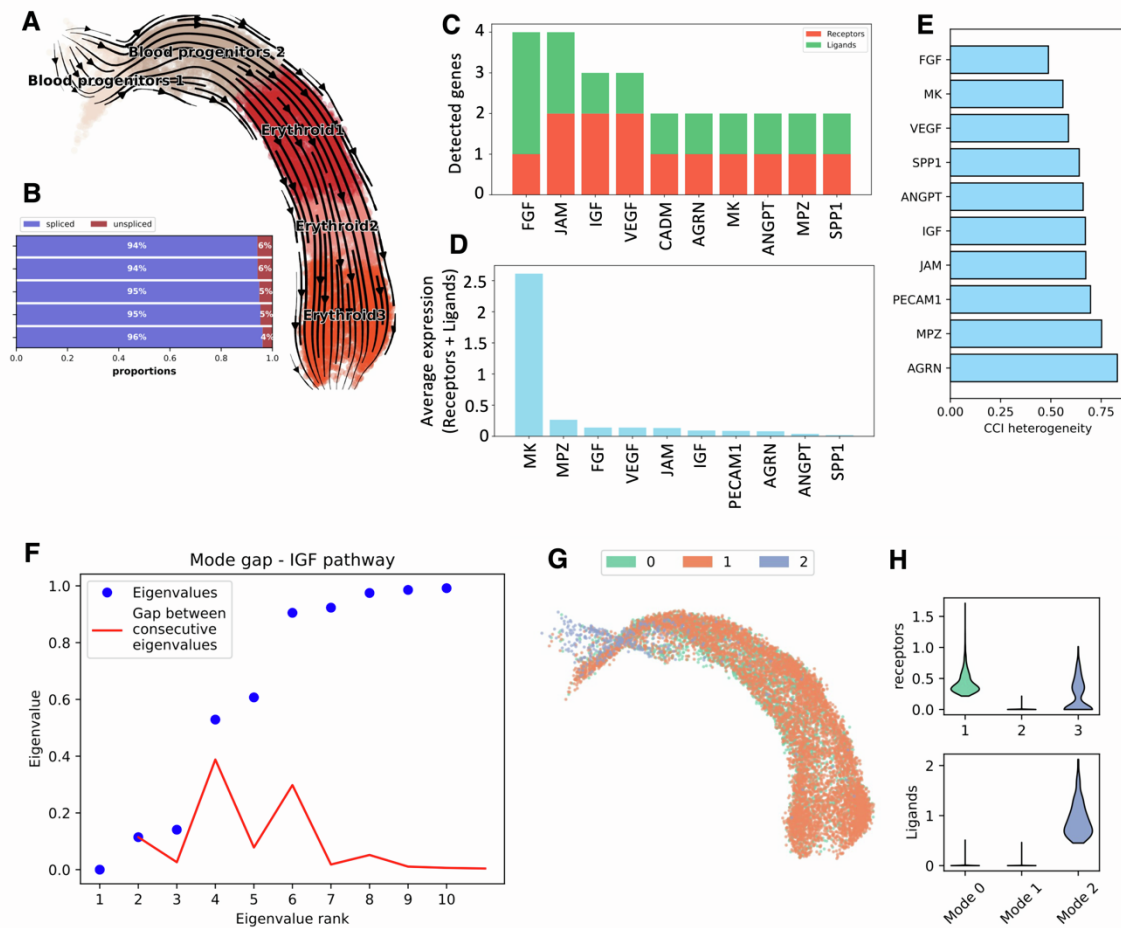
Supplementary figure 2. Benchmarking and parameter testing (related to Figure 2). (A) Similarity based on Wassertein distance of cell-cell communication modes for the top 10 pathways identified in the FibHSE dataset for different combinations of number of nearest neighbors and principal components. (B) Fraction of mode clustering solutions with different number of modes across all parameter combinations from (A) for each CCC pathway. (C) CCC modes for the IGF pathway for two parameter combinations yielding different number of modes. (D) Mapping of IGF signaling modes for the two parameter combinations. (E-F) Same as (C-D) for the NRG pathway. (G-J) Same as (A-B) for the A549 dataset analyzed in Fig. 6 (G-H) and for the erythroid differentiation dataset analyzed in Fig. 5 (I-J). (K) Comparison of ScRICH CCC communication prediction in FibHSE for the top 10 CCC pathways based on standalone ligand/receptor expression (None), ligand/receptor and target gene expression (Expr), and (3) ligand/receptor expression plus RNA velocity of target genes (scVelo). Comparison of CCC interaction matrix is based on matrix distance, Spearman correlation, and overlap similar to Fig. 2H. (L) Same as (K) for the A549 dataset. (M) Same as (K) for the erythroid differentiation dataset including a comparison between scVelo-based and Unitvelo-based RNA velocity inference.



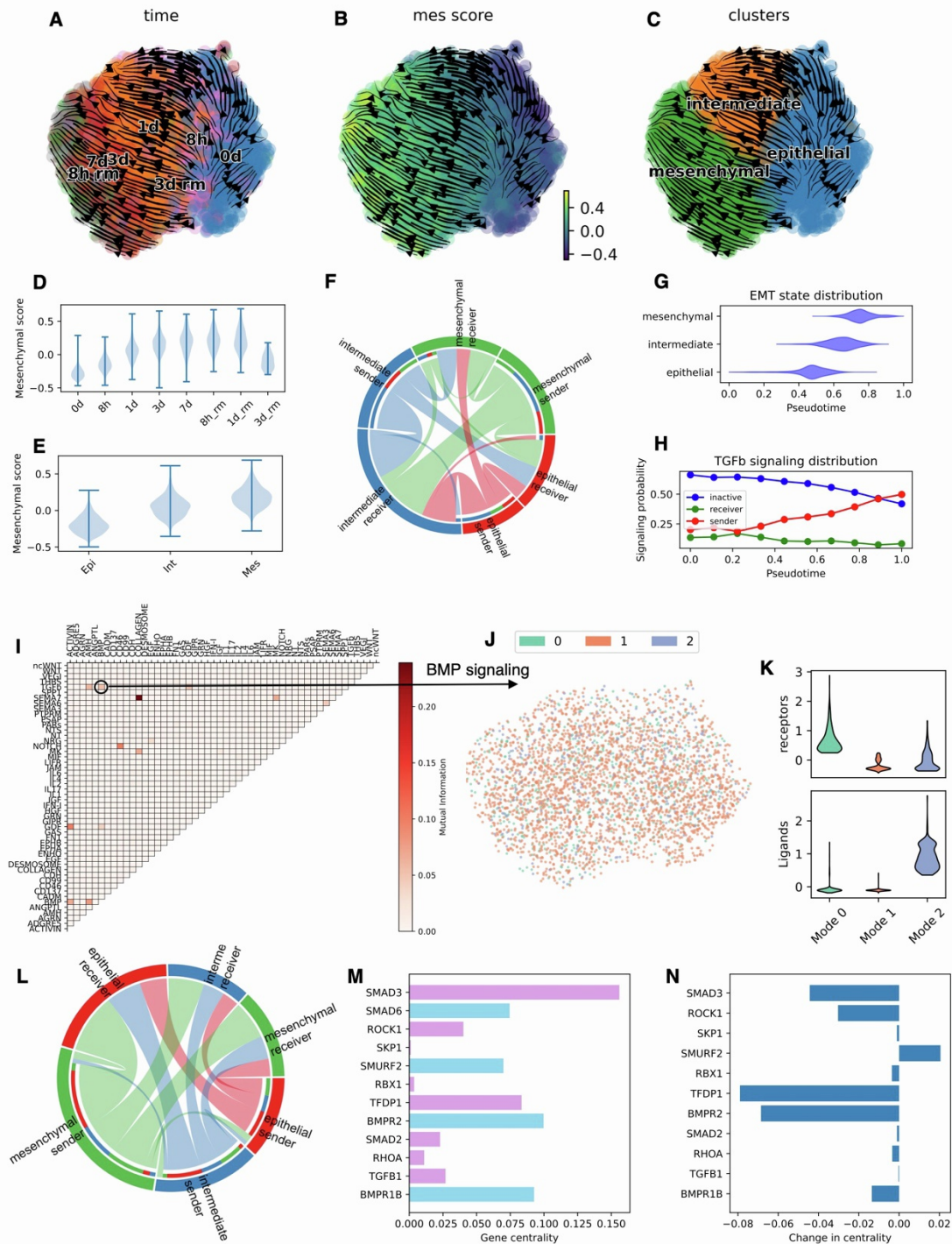
Supplementary Figure S3. Expression of HRAS and INHBA in FibHSE (related to Figure 3). (A-B) Co-localization of HRAS and INHBA in FibHSE for the cases of no exogenous EGF (A, 0xEGF) and soluble EGF in solution (B, 1xEGF). R1 and R2 indicate replicates for each condition. (C-D) Quantification of INHBA expression based on replicates in C and EGF dosage. (D) Same as (C) for HRAS.



Supplementary figure S4. Connection between EGF and other EMT-related CCC pathways (related to Figure 3). (A) The CDH CCC modes including low, medium and high adhesion. (B) Sankey diagram of correspondence between CDH and EGF CCC modes. (C) The gene regulatory network connecting the EGF and CDH pathways in the EGF receiver cells. (D-E) Co-localization of CDH4 in FibHSE for the cases of no exogenous EGF (0XEGF) and soluble EGF in solution (1XEGF). R1 and R2 indicate replicates for each condition. (F) Quantification of CDH4 expression based in replicate and EGF dosage. (G-H-I) Same as A-B-C for the Notch CCC pathway. (J, L) Co-localization of STAT1 and NOTCH in FibHSE for the cases of no exogenous EGF (0XEGF) and soluble EGF in solution (1XEGF). R1 and R2 indicate replicates for each condition. (K, M) Quantification of STAT1 and NOTCH expression based in replicate and EGF dosage.

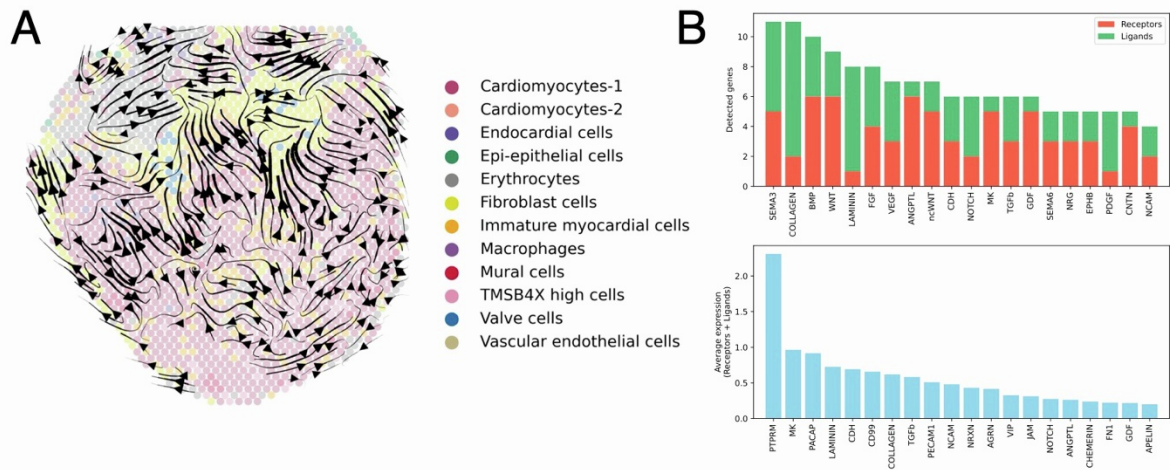


Supplementary figure S5 (related to Figure 5). Supplementary analysis of erythroid cell differentiation. (A) UMAP, cell types, and RNA velocity of erythroid cell lineage. **(B)** Fraction of unspliced RNA counts in each cell type. **(C-D)** The top CCC pathways identified based on total ligands and receptors expressed (top) and ligand/receptor expression (bottom). **(E)** CCI heterogeneity of the top CCC pathway. **(F)** Eigengap for the IGF CCC pathway highlighting three CCC modes. **(G)** UMAP embedding of the IGF CCC modes. **(H)** Expression of IGF ligands and receptors in the three IGF modes via violinplot. The three modes correspond to receiver, inactive, and sender.



Supplementary Figure 6. Supplementary analysis of CCC during EMT (related to Figure 6). (A-B-C) Low dimensional UMAP embedding and RNA velocity where cells are color-coded by time point (A), mesenchymal score (B), and EMT cluster annotation (C). (D-E) The mesenchymal score by time point (D) and cluster (E). (F) The TGFb CCC network. (G) EMT cell state distribution as a function of pseudotime. (H) Fraction of cells in the different TGFb communication modes as a function of pseudotime. (I) Pairwise mutual information between CCC pathways. (J) The CCC modes of the BMP pathway in UMAP embedding. (K) Expression of BMP receptors and ligands highlighting a BMP receiver (mode 0), inactive (mode 1) and

sender (mode 2). **(L)** The BMP CCC network. **(M)** Betweenness centrality of genes in the joint TGFb-BMP gene regulatory network. Purple and blue bars indicate genes that belong to the TGFb module and shared between the two pathways, respectively, following the same notation of Fig. 5H. **(N)** Change in gene betweenness centrality when comparing the TGFb-BMP GRN inferred at $t=0$ days and $t=7$ days.



Supplementary figure S7. Supplementary analysis of chicken earth spatial transcriptomics data (related to Figure 7). (A) The cell types and RNA velocity in two-dimensional physical space. (B) The top CCC pathways identified based on total ligands and receptors expressed (top) and ligand/receptor expression (bottom).

Supplementary Tables

EGF	CSNK2A1, EGF, EGFR, ELK1, FOS, GRB2, HRAS, JAK1, JUN, MAP2K1, MAP2K4, MAP3K1, MAPK3, MAPK8, PIK3CA, PIK3R1, PLCG1, PRKCA, PRKCB, RAF1, RASA1, SHC1, SOS1, SRF, STAT1, STAT3, STAT5A
TGFB	ACVR1, ACVR1C, ACVR2A, ACVR2B, ACVRL1, AMH, AMHR2, BMP2, BMP4, BMP5, BMP6, BMP7, BMP8A, BMP8B, BMPR1A, BMPR1B, BMPR2, CDKN2B, CHRD, COMP, CREBBP, CUL1, DCN, E2F4, E2F5, EP300, FST, GDF5, GDF6, GDF7, ID1, ID2, ID3, ID4, IFNG, INHBA, INHBB, INHBC, INHBE, LEFTY1, LEFTY2, LTBP1, MAPK1, MAPK3, MYC, NODAL, NOG, PITX2, PPP2CA, PPP2CB, PPP2R1A, PPP2R1B, RBL1, RBL2, RBX1, RHOA, ROCK1, ROCK2, RPS6KB1, RPS6KB2, SKP1, SKP1P2, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5, SMAD6, SMAD7, SMAD9, SMURF1, SMURF2, SP1, TFDP1, TGFB1, TGFB2, TGFB3, TGFB1, TGFB2, THBS1, THBS2, THBS3, THBS4, TNF, ZFYVE16, ZFYVE9

Supplementary Table T1. The gene sets of CCC pathways obtained from the GSEA database used to infer gene regulatory networks.