CytoFlow: A Novel Computational Method to Construct Signal Transduction Networks at Single-Cell Resolution based on Flow Networks

Yi Dai
Department of Computer Science
University of California, Irvine
Irvine, USA
ydai12@uci.edu

Ziheng Duan
Department of Computer Science
University of California, Irvine
Irvine, USA
zihend1@uci.edu

Siwei Xu
Department of Computer Science
University of California, Irvine
Irvine, USA
s.xu@uci.edu

Jing Zhang*
Department of Computer Science
University of California, Irvine
Irvine, USA
jingz31@uci.edu

2024 IEEE International Conference on Medical Artificial Intelligence (MedAI) | 979-8-3503-7761-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/MEDAI62885.2024.00034

***Abstract*—The signal transduction network is essential for eukaryotic cellular communication and response to environmental signals, with disruptions leading to various diseases. As our understanding of intracellular signaling expands, the demand for computational tools that can efficiently synthesize this information into comprehensive networks grows. To meet this need, we developed CytoFlow, a novel computational model designed to construct cell-type-specific signal transduction networks by leveraging protein-protein interaction and single-cell transcriptomics data. Specifically, it models the network from receptors to transcription factors as a penalized flow network, optimized using linear programming techniques. We validated CytoFlow's superior precision against existing methods in reconstructing the yeast mitogen-activated protein kinase pathway, and demonstrated its ability to identify cell-type-specific signaling patterns in the human prefrontal cortex and peripheral blood mononuclear cells. In summary, CytoFlow offers a precise and cost-effective solution for constructing detailed signal transduction networks, advancing our capacity to understand and analyze cellular signaling processes.**

***Index Terms*—signal transduction network, single-cell RNA sequencing, protein-protein interaction, flow network**

I. Introduction

Cellular signal transduction is a fundamental biological process in multicellular systems, wherein environmental signals initially detected by receptors on individual cells are propagated via a cascade of intracellular molecular interactions [1], [2]. It allows cells to communicate and respond to environmental signals to orchestrate essential cellular activities such as cell growth, differentiation, and apoptosis [3]–[5]. These activities are pivotal in maintaining the delicate balance of biological systems, while aberrations in signal transduction can lead to a plethora of diseases, including cancer, diabetes, and autoimmune disorders [6]–[8]. The exploration of these pathways not only sheds light on the basic principles of cellular function but also provides crucial insights into the pathogenesis of a broad spectrum of diseases.

Over the past decades, several methods have been proposed to dissect the signal transduction network, which can be summarized into two main classes: manual curation from published experimental results and computational reconstruction from genomic data. Many curated databases of signal transduction pathways have summarized extensive experimental literature, leading to numerous resources for research and education. Key products among these are online databases like the Kyoto Encyclopedia of Genes and Genomes (KEGG) [9], which provides detailed maps of molecular interaction networks, including signaling pathways. The Protein Data Bank (PDB) offers structural data of proteins involved in signaling processes, enhancing our understanding of their mechanisms [10]. While such well-known databases offer invaluable information and are warmly embraced by the scientific community, a major problem severely limit our understanding of signal transduction. It is impossible to experimentally validate all possible molecular interactions. Hence, the complexity of signaling pathways, involving numerous interactions and crosstalks, can be oversimplified, potentially overlooking critical nuances.

In addition to public data curated from previous experiments, the rapid development of high-throughput sequencing technologies has facilitated the in silico construction of signal transduction networks [11]–[13]. Several methods have been developed to utilize transcriptomics and protein-protein interaction (PPI) data for signal transduction network construction. For instance, Steffen et al. built a likelihood model to assess every possible network based on PPI data [14]. Zhao et al. and Ren et al. optimized the network using bulk RNA sequencing and PPI data [15], [16]. Moreover, reaction-contingency-

*Corresponding author

979-8-3503-7761-3/24/$31.00 ©2024 IEEE
DOI 10.1109/MedAI62885.2024.00034

based models simulate mechanistic protein interactions from biochemical knowledge bases [17], [18]. However, signal transduction is highly cell-type-specific because of cell types' distinct active proteins, functions, and environments. This limits methods designed for traditional bulk sequencing data from discovering the intricate variations and specific signaling states of individual cell types.

The recent breakthrough in single-cell sequencing has revolutionized the genomic field by allowing parallel molecular profiling in tens of thousands of individual cells simultaneously, opening new avenues for dissecting the complexity and diversity in cell signaling transduction [19]–[21]. As a step forward, several recent computational methods have been developed to construct signal transduction networks in a cell-type-specific manner. For example, NicheNet integrates ligand-receptor interaction and PPI with single-cell RNA sequencing (scRNA-seq) data to predict receptor-target regulatory potentials via a personalized PageRank algorithm [22]. However, the use of public PPIs ignores cell-type specificity, failing to fully exploit the potential of scRNA-seq data. Additionally, CytoTalk [23] and SoptSC [24] directly infer intracellular signal transduction from receptors to target genes based on noisy gene co-expression inferences, leading to noise accumulation in large networks and often failing to provide biological insights for signaling transduction.

To address these problems, we proposed CytoFlow, a novel computational model for predicting cell-type-specific signal transduction networks based on penalized network flow. We assume that cellular signals are transmitted by a directed network of proteins that start from receptors and terminate at transcription factors (TFs). Given a baseline network constructed from PPI and scRNA-seq data, we model the signal transduction network with a penalized maximum flow framework [25] based on a straightforward intuition – cells try to process maximal information to key regulators (TFs) using minimal energy. With this framework, CytoFlow has three major advantages over previous methods: *(i)* it can fully leverage transcriptomics and PPI data, getting cell-type-specificity while maintaining biological significance of the predicted network; *(ii)* it is highly flexible with the input number of receptors and TFs, allowing for both single-pathway inference and panoramic network analysis; and *(iii)* the size and density of the predicted network is fully controllable. To prove the effectiveness of our model, we benchmarked CytoFlow against existing methods based on the accuracy of the predicted network on known pathways with yeast PPI and RNA-seq data. We then illustrated CytoFlow's ability to discover cell-type-specific signaling patterns by applying it to human brain scRNA-seq data. Additionally, we applied CytoFlow to scRNA-seq data of human human peripheral blood mononuclear cells (PBMCs) to further demonstrate the cell-type specificity of receptor-TF flows. We have implemented CytoFlow as a free software[1] available for the community to uncover signal transduction networks. Supplementary information and tables can also be found on our Github website.

[1]https://github.com/aicb-ZhangLabs/CytoFlow

II. Materials and Methods

A. Overview of CytoFlow

CytoFlow is a computational method for reconstructing cell-type-specific signal transduction networks using transcriptomics data. To increase model interpretability, we dissect intracellular information processing into two distinct networks: signal transduction network from receptors to TFs, and gene regulatory network. **Fig. 1** exemplifies the signal transduction flow network predicted by CytoFlow, where nodes represent signal-transduction-involved genes and directed edges represent the information flow between genes. The goal of CytoFlow is to reconstruct the signal transduction network based on a straightforward intuition: cells tend to transmit the maximum amount of information with the least energy consumption. This is achieved through a penalized maximum flow optimization process, as discussed in the following section.

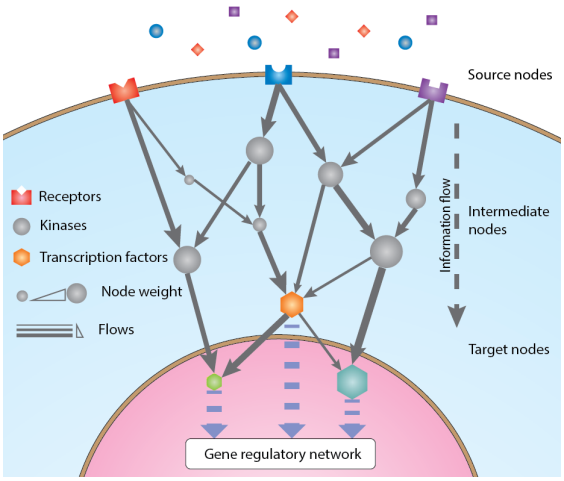

Fig. 1. *Signal transduction network predicted by CytoFlow.* Signal transduction initiates from receptors on the plasma membrane and transmits through a cascade of protein interactions. We focus on the signal transduction activities that reach transcription factors, which have a downstream impact on the expression of other genes through the gene regulatory network.

B. Optimizing flow network to construct cell-type-specific signal transduction network

We denote the input network of our model as $G = \{V, E\}$ with each undirected edge $(i, j) \in E$ associated with a weight $c_{(i,j)}$. The network and edge weights are derived from PPI, bulk transcriptomics, or scRNA-seq data as described in *Supplementary Information*. For a particular sample or cell type, every node $v_i \in V$ represents a gene associated with its expression e_i, and every undirected edge in E represents a potential molecular interaction between v_i and v_j, denoted as (i, j). The flow network defined upon G attributes the bidirectional flows f_{ij}, f_{ji} to every undirected edge $(i, j) \in E$. The signs of the flows determine their directionalities. We also

define node weight w_i for every node $v_i \in V$ to represent the relative amount of protein molecules involved in the signal transduction process. Assume that the network has N_n nodes in $V = \{v_i\}_{i=1}^{N_n}$. We further divide the node set into three subsets: $R = \{v_i\}_{i=1}^{N_r}$ representing N_r receptors, $T = \{v_i\}_{i=N_n-N_t+1}^{N_t}$ representing N_t TFs, and $K = \{v_i\}_{i=N_r+1}^{N_n-N_t}$ representing $N_n - N_r - N_k$ other genes. Receptors serve as the source nodes in traditional flow networks, while TFs serve as the sink nodes. CytoFlow incorporates the input networks derived from biological data that indicate the possibility of two proteins interacting with each other (e.g. PPI and gene co-expression networks) by optimizing a linear programming problem to enable the information flow from receptors to TFs. We now introduce the five components of our model.

a) Optimization objective: As mentioned before, CytoFlow seeks to maximize the information flow in the network while minimizing energy consumption. The total flow in the network can be calculated by summing over all incoming flows of all TFs, which can be written as $\sum_{v_i \in T} \sum_{(i,j) \in E} f_{it}$. We model energy consumption with two components: node penalty and edge penalty. The node penalty simulates the energy consumption required to translate transcripts into proteins, while the edge penalty captures the energy consumed in protein-protein interaction events. By combining these components together, we obtain the objective function in (1)

$$\max \sum_{v_i \in T} \sum_{(i,j) \in E} f_{ji} - \lambda_n \sum_{v_i \in V} w_i - \lambda_e \sum_{(i,j) \in E} |f_{ij}| \qquad (1)$$

where λ_n and λ_e are parameters scaling the penalty.

b) Edge capacity constraint: We assume that the information transmitted between two proteins is influenced by both their gene expressions and their interaction strength. Additionally, there should not be a large flow if one of the nodes has very limited expression. Therefore, we write the capacity of edge $(i,j) \in E$ as $c_{(i,j)} \min\{w_i, w_j\}$. To maintain the linearity of the problem, we reformulate the capacity into the constraints in (2).

$$f_{ij} \le c_{(i,j)} w_i, \quad f_{ij} \le c_{(i,j)} w_j; \qquad \forall (i,j) \in E \qquad (2)$$

c) Basic flow constraints: The skew symmetry constraint (3) and flow conservation constraint (4) are two fundamental constraints of a flow network. The skew symmetry constraint ensures the validity of the flow on each edge. The flow conservation constraint maintains the balance between incoming and outgoing flows of intermediate nodes, representing the equilibrium between their activation and deactivation.

$$f_{ij} + f_{ji} = 0; \qquad \forall (i,j) \in E \qquad (3)$$

$$\sum_{j:(i,j) \in E} f_{ij} = 0; \qquad \forall i \in K \qquad (4)$$

d) Node weight constraint: As node weights represent the relative number of proteins involved in the network, we require them not to exceed their corresponding gene expressions, as formalized in (5).

$$0 \le w_i \le e_i; \qquad \forall i \in V \qquad (5)$$

e) Linearity trick: The absolute value function included in (1) breaks the linearity. To maintain linearity of the problem, we approximate $|f_{ij}|$ with $\widehat{f_{(i,j)}}$ under constraints (6). Its negative gradient in the objective function ensures $\widehat{f_{(i,j)}}$ to converges to $|f_{ij}|$ at the optimal solution.

$$f_{ij} \le \widehat{f_{(i,j)}}, \quad f_{ji} \le \widehat{f_{(i,j)}}; \qquad \forall (i,j) \in E \qquad (6)$$

To sum up, CytoFlow models the signal transduction network with the following linear programming problem:

$$\max \quad \sum_{v_i \in T} \sum_{(i,j) \in E} f_{ji} - \lambda_n \sum_{v_i \in V} w_i - \lambda_e \sum_{(i,j) \in E} \widehat{f_{(i,j)}}$$

$$\begin{aligned}
s.t. \quad & f_{ij} \le c_{(i,j)} w_i, f_{ij} \le c_{(i,j)} w_j && \forall (i,j) \in E \\
& f_{ij} + f_{ji} = 0 && \forall (i,j) \in E \\
& \sum_{j:(i,j) \in E} f_{ij} = 0, && \forall i \in K \\
& 0 \le w_i \le e_i && \forall i \in V \\
& f_{ij} \le \widehat{f_{(i,j)}}, f_{ji} \le \widehat{f_{(i,j)}} && \forall (i,j) \in E
\end{aligned}$$

where f_{ij} and w_i are variables to optimize, and λ_n and λ_e are tunable parameters.

CytoFlow outputs the flow network and node weights in the optimal solution after removing non-positive-weighted directed flow edges and zero-weighted nodes (a threshold of 10^{-5} was applied to allow for numerical errors).

III. Results

To comprehensively demonstrate CytoFlow's application potential in various organisms and tissues, we applied CytoFlow to three public datasets to construct signal transduction networks. Data processing and parameter selection details can be found in *Supplementary Information*.

A. CytoFlow outperforms existing methods in reconstructing known pathway

As a proof-of-concept experiment, we incorporated two types of networks to re-construct the signal transduction network in *Saccharomyces cerevisiae*: the PPI network and the gene co-expression network derived from microarray data [26] (*Supplementary Information*). Our results indicate that CytoFlow's reconstructed networks, using both types of inputs, accurately reflect the documented yeast mitogen-activated protein kinase (MAPK) pheromone response pathway in Kyoto Encyclopedia of genes and genomes (KEGG) (**Fig. 2B&D** vs **Fig. 2A)** [9], [27]–[29].

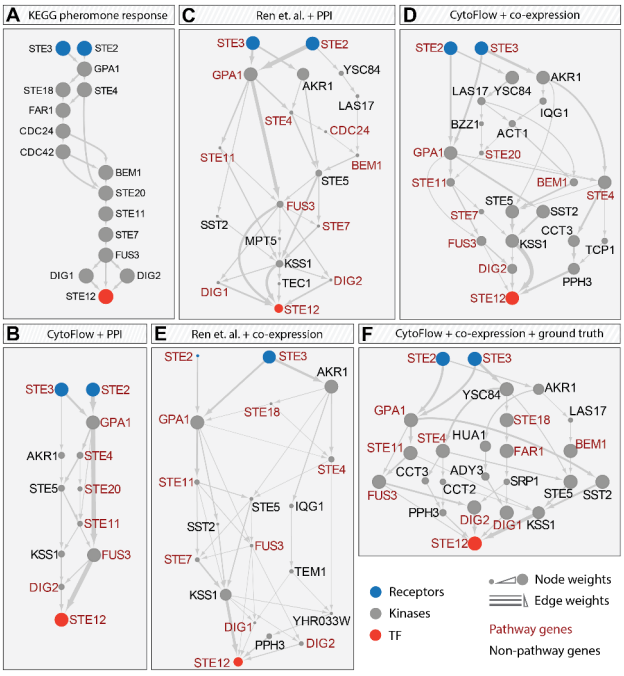

Fig. 2. *Network visualizations and performance statistics.* Node sizes and edge widths are proportional to their corresponding weights. Genes labelled in red are documented pheromone response pathway genes. (**A**) MAPK pheromone response pathway documented in KEGG. (**B**) and (**D**) Predicted pathways with CytoFlow. (**C**) and (**E**) Predicted pathway skeletons with NF. (nodes that transmit less than 0.1 units of flow are omitted). (**F**) Predicted network with CytoFlow given ground truth nodes.

TABLE I
NODE PREDICTION PERFORMANCE BENCHMARK STATISTICS

		TP	TPR	Prec	FPR	P
PPI network	CytoFlow	9	0.56	**0.75**	**0.03**	12
	ILP	**16**	**1.00**	0.21	0.63	78
	NF	15	0.93	0.38	0.24	39
Co-expression	CytoFlow	11	0.68	**0.47**	**0.10**	23
	ILP	12	0.75	0.29	0.25	42
	NF	**16**	**1.00**	0.31	0.30	52

Firstly, we tested the PPI network as the initial input. For example, 9 out of 12 genes in our predicted network were also documented in KEGG (**Fig. 2A**). We also benchmarked against two existing methods using integer linear programming (ILP) [15] and network flow (NF) [16], and found that CytoFlow outperformed both methods in precision and false positive rate (FPR). For instance, CytoFlow's FPR is only 0.032, significantly lower than ILP's 0.63 and NF's 0.24 (**Table I**). Notably, CytoFlow's penalty mechanism effectively reduces the size of the output network while maintaining a high precision of 0.75. This demonstrates CytoFlow's ability to capture important hub genes in the network, such as *guanine nucleotide-binding protein subunit alpha* (*GPA1*) and *mitogen-activated protein kinase FUS3*. Beyond the visualized networks in **Fig. 2B&C**, the full networks predicted by all three methods are available in **Table S1**.

Next, we benchmarked CytoFlow using the gene co-expression network as input. **Fig. 2D&E** show the networks predicted by CytoFlow and NF, respectively (complete networks of all three methods are in **Table S1**). Compared to other methods [15], [16], CytoFlow maintained the highest precision (0.47) and the lowest false positive rate (0.1), as shown in **Table I**. Ideally, CytoFlow interprets edge weights of the input network as the interaction strength between proteins, but the gene co-expression network provides this information at a low confidence level. Despite this gap between model assumptions and real data, CytoFlow can still recover 11 out of 16 ground truth genes (marked in red in **Fig. 2D**), demonstrating its ability to recover information from noisy data.

Beyond de novo network reconstruction, CytoFlow exhibited predictive capability with prior knowledge. **Fig. 2F** illustrates the pathway predicted when all ground truth nodes are known, simulating scenarios where additional pathway information is available or certain nodes are forced to be included to elucidate their roles in the signaling pathway.

In summary, our findings confirmed that CytoFlow outperforms existing methods in both PPI and transcriptomics data scenarios, highlighting its enhanced stability and efficacy in signal transduction network reconstruction.

B. CytoFlow reveals cell type heterogeneity in human prefrontal cortex

To demonstrate CytoFlow's ability to identify cell-type-specific signal transduction patterns, we applied it to scRNA-seq data from human prefrontal cortex (PFC) [30]. The cell type distribution is visualized in **Fig. 3A**. All major cell types are well-separated in this dataset, indicating that distinct signaling patterns are expected between different cell types. As described in *Supplementary Information*, we selected cell-type-specific receptors and TFs to analyze their signaling patterns (**Table S2**). Their expression profiles are shown in **Fig. 3B**. The low co-expression values of receptors and TFs across different cell types highlight the cell type heterogeneity in the human PFC.

Upon constructing cell-type-specific gene co-expression networks (*Supplementary information*), we optimized the flow networks and calculated the total flows between each pair of receptors and TFs. The blockwise heatmaps in **Fig. 3C** illustrate CytoFlow's ability to detect cell-type-specific pairwise flow patterns in the human brain. Diagonal blocks, representing cell-type-specific receptor-TF pairs in their corresponding cell types, show significantly larger flows than off-diagonal blocks. Notably, some off-diagonal receptor-TF pairs also transmit considerable flow, such as *odium/potassium-transporting ATPase subunit alpha-3* (*ATP1A3*, first inhibitory-neuron-specific receptor) and *cut like homeobox 2* (*CUX2*, fifth excitatory-neuron-specific TF), explained by the high activity of *CUX2* is both neuron types. These results sup-

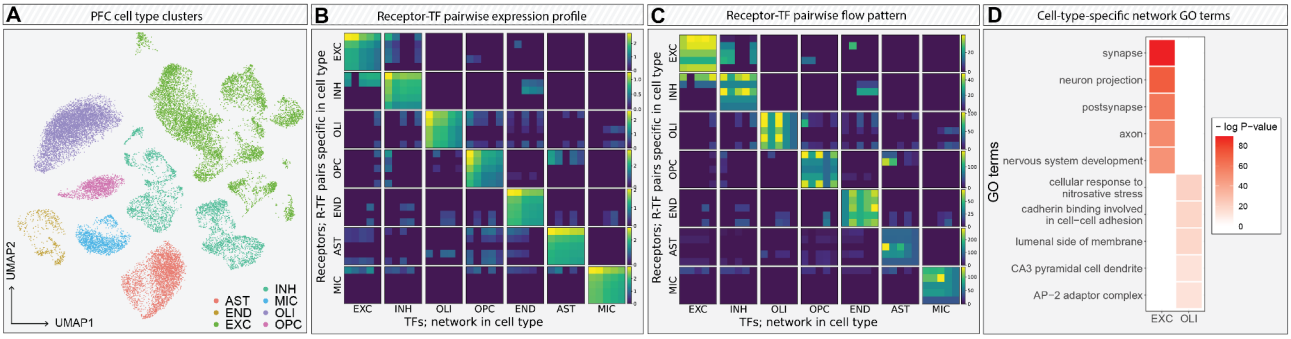

Fig. 3. *Pairwise receptor-TF flow analysis with human brain scRNA-seq data.* (**A**) Cell type clusters of all cells in the brain sample. (**B**) Expression profiles of cell-type-specific receptor-TF pairs. Entry (i, j) in block (m, n) indicates the geometric mean of cell type m's specific receptor i's and TF j's expression in cell type n. Blocks in the same row (the same receptors and TFs) share the same color scale. (**C**) Pairwise flows between cell-type-specific receptor-TF pairs. Entry (i, j) in block (m, n) indicates the flow from cell type m's specific receptor i to TF j inferred with cell type n's co-expression network. (**D**) Gene ontology (GO) enrichment analysis p-value heatmap. Columns are GO terms. Rows are cell types, where we took as GO enrichment query the union of all nodes predicted of every receptor-TF pair of the cell type with the expression of that cell type.

port CytoFlow's capability to recover cell-type-specific signal transduction patterns.

We then conducted gene ontology (GO) enrichment analysis to functionally analyze the nodes predicted in the pairwise flow networks [31]. The union of nodes from all pairwise networks within a diagonal block was used as the query gene list for GO enrichment analysis. **Fig. 3D** presents the significant GO terms and their p-values of excitatory neurons (*EXC*) and oligodendrocytes (*OLI*). As an example, *synapse* is a key component of neurons, while *cellular response to nitrosative stress* is a crucial function of *OLI*s [32]. This GO enrichment analysis further demonstrates CytoFlow's ability to capture cell-type-specific signal transduction patterns.

C. Cell-type-specific signal transduction patterns in human peripheral blood mononuclear cells

In the realm of immune system research, peripheral blood mononuclear cells (PBMCs) are crucial elements extensively utilized in the development of pharmaceuticals and therapeutic strategies. We applied CytoFlow to scRNA-seq data of human PBMCs [33]. **Fig. 4A** illustrates the distribution of all cells in the dataset, and **Fig. 4B** shows the expression profiles of cell-type-specific receptor-TF pairs. Unlike human PFC, PBMC cell types are not as clearly separated. Lymphocytes are highly developmental, with subtypes often displaying continuous trajectories. Due to this characteristic, we expect similar signaling patterns between closely-related lymphocyte subtypes, such as CD14-positive monocytes (*CD14 Mono*) and CD16-positive monocytes (*CD16 Mono*) [34]. To demonstrate the cell-type specificity of signal transduction, we pruned the PBMC cell types as described in *Supplementary Information*.

Similar with the previous experiment, we optimized the flows between each pair of receptors and TFs based on cell-type-specific gene co-expression networks (**Table S2**). **Fig. 4C** displays the flow patterns of selected cell types. Consistent with the previous experiment, flows within diagonal blocks dominate over those in off-diagonal blocks, confirming the

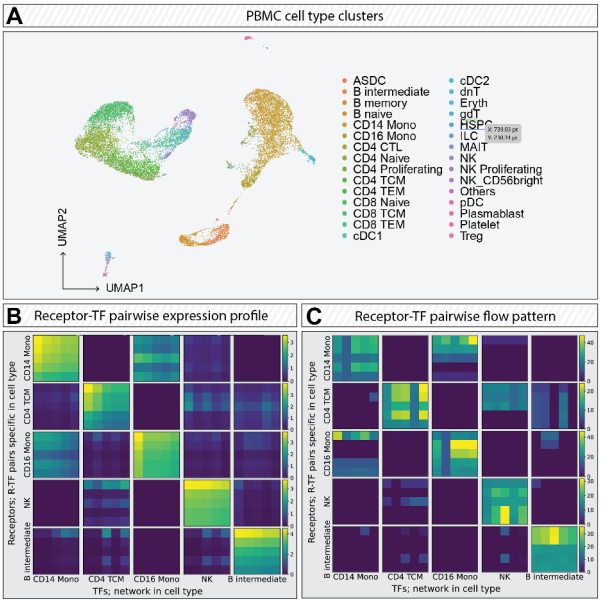

Fig. 4. *Pairwise receptor-TF flow analysis with human PBMC scRNA-seq data.* (**A**) Cell type clusters of all cells in the PBMC sample. (**B**) Expression profiles of cell-type-specific receptor-TF pairs. Entry (i, j) in block (m, n) indicates the geometric mean of cell type m's specific receptor i's and TF j's expression in cell type n. Blocks in the same row (the same receptors and TFs) share the same color scale. (**C**) Pairwise flows between cell-type-specific receptor-TF pairs. Entry (i, j) in block (m, n) indicates the flow from cell type m's specific receptor i to TF j inferred with cell type n's co-expression network.

presence of cell-type-specific receptor-TF signal transduction patterns. Interestingly, the off-diagonal block of *CD14 Mono* and *CD16 Mono* shows similar flow amounts compared to diagonal blocks. Since the monocyte subtypes are closely related to each other [34], their receptors and TFs exhibit similar activities, leading to analogous signaling patterns. These results provide further evidence that CytoFlow can identify cell-type-specific signal transduction patterns.

IV. Discussion

In conclusion, this paper presents CytoFlow, a novel methodology for constructing cell-type-specific signal transduction networks using protein-protein interaction and transcriptomics data. CytoFlow utilizes penalized flow networks to model signal transduction from receptors to transcription factors. It has demonstrated superior predictive precision compared to other methods [15], [16] in reconstructing known pathways in both PPI and transcriptomics data scenarios. Furthermore, the application of pairwise flow analysis using scRNA-seq data from the human brain and PBMCs highlighted CytoFlow's ability to construct cell-type-specific signal transduction networks. CytoFlow has been implemented as open-source software, freely available to the public. In summary, we anticipate that CytoFlow will be a valuable tool for constructing cell-type-specific signal transduction networks, providing a promising avenue for understanding signaling processes.

Acknowledgment

We thank the UCI OIT department for managing computational resources.

References

[1] K. B. Lengeler, R. C. Davidson, C. D'Souza, T. Harashima, W. C. Shen, P. Wang, X. Pan, M. Waugh, and J. Heitman, "Signal transduction cascades regulating fungal development and virulence," *Microbiol Mol Biol Rev*, vol. 64, no. 4, pp. 746–85, 2000.

[2] J. Lee, J. T. and J. A. McCubrey, "The raf/mek/erk signal transduction cascade as a target for chemotherapeutic intervention in leukemia," *Leukemia*, vol. 16, no. 4, pp. 486–507, 2002.

[3] S. Liu, J. Ren, and P. Ten Dijke, "Targeting tgfbeta signal transduction for cancer therapy," *Signal Transduct Target Ther*, vol. 6, no. 1, p. 8, 2021.

[4] P. S. Ward and C. B. Thompson, "Signaling in control of cell growth and metabolism," *Cold Spring Harb Perspect Biol*, vol. 4, no. 7, p. a006783, 2012.

[5] A. Glaviano, A. S. C. Foo, H. Y. Lam, K. C. H. Yap, W. Jacot, R. H. Jones, H. Eng, M. G. Nair, P. Makvandi, B. Geoerger, M. H. Kulke, R. D. Baird, J. S. Prabhu, D. Carbone, C. Pecoraro, D. B. L. Teh, G. Sethi, V. Cavalieri, K. H. Lin, N. R. Javidi-Sharifi, E. Toska, M. S. Davids, J. R. Brown, P. Diana, J. Stebbing, D. A. Fruman, and A. P. Kumar, "Pi3k/akt/mtor signaling transduction pathway and targeted therapies in cancer," *Mol Cancer*, vol. 22, no. 1, p. 138, 2023.

[6] W. Kolch, M. Halasz, M. Granovskaya, and B. N. Kholodenko, "The dynamic control of signal transduction networks in cancer cells," *Nat Rev Cancer*, vol. 15, no. 9, pp. 515–27, 2015.

[7] T. Takamura, "Hepatokine selenoprotein p-mediated reductive stress causes resistance to intracellular signal transduction," *Antioxid Redox Signal*, vol. 33, no. 7, pp. 517–524, 2020.

[8] Z. Duan, C. Lee, and J. Zhang, "Exad-gnn: Explainable graph neural network for alzheimer's disease state prediction from single-cell data," *Apsipa Transactions on Signal and Information Processing*, vol. 12, no. 5, 2023.

[9] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, and M. Ishiguro-Watanabe, "Kegg for taxonomy-based analysis of pathways and genomes," *Nucleic Acids Res*, vol. 51, no. D1, pp. D587–D592, 2023.

[10] H. Berman, K. Henrick, H. Nakamura, and J. L. Markley, "The worldwide protein data bank (wwpdb): ensuring a single, uniform archive of pdb data," *Nucleic Acids Res*, vol. 35, no. Database issue, pp. D301–3, 2007.

[11] Z. Wang, M. Gerstein, and M. Snyder, "Rna-seq: a revolutionary tool for transcriptomics," *Nat Rev Genet*, vol. 10, no. 1, pp. 57–63, 2009.

[12] L. David, W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, and L. M. Steinmetz, "A high-resolution map of transcription in the yeast genome," *Proc Natl Acad Sci U S A*, vol. 103, no. 14, pp. 5320–5, 2006.

[13] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by rna-seq," *Nat Methods*, vol. 5, no. 7, pp. 621–8, 2008.

[14] M. Steffen, A. Petti, J. Aach, P. D'Haeseleer, and G. Church, "Automated modelling of signal transduction networks," *BMC Bioinformatics*, vol. 3, p. 34, 2002.

[15] X. M. Zhao, R. S. Wang, L. Chen, and K. Aihara, "Uncovering signal transduction networks from high-throughput data by integer linear programming," *Nucleic Acids Res*, vol. 36, no. 9, p. e48, 2008.

[16] X. W. Ren and X. S. Zhang, "A linear programming model based on network flow for pathway inference," *Journal of Systems Science & Complexity*, vol. 23, no. 5, pp. 971–977, 2010.

[17] M. Flottmann, F. Krause, E. Klipp, and M. Krantz, "Reaction-contingency based bipartite boolean modelling," *BMC Syst Biol*, vol. 7, p. 58, 2013.

[18] J. Romers, S. Thieme, U. Munzner, and M. Krantz, "A scalable method for parameter-free simulation and validation of mechanistic cellular signal transduction network models," *NPJ Syst Biol Appl*, vol. 6, p. 2, 2020.

[19] D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, and I. Amit, "Massively parallel single-cell rna-seq for marker-free decomposition of tissues into cell types," *Science*, vol. 343, no. 6172, pp. 776–9, 2014.

[20] D. Grun, A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, N. Sasaki, H. Clevers, and A. van Oudenaarden, "Single-cell messenger rna sequencing reveals rare intestinal cell types," *Nature*, vol. 525, no. 7568, pp. 251–5, 2015.

[21] W. Jin, Q. Tang, M. Wan, K. Cui, Y. Zhang, G. Ren, B. Ni, J. Sklar, T. M. Przytycka, R. Childs, D. Levens, and K. Zhao, "Genome-wide detection of dnase i hypersensitive sites in single cells and ffpe tissue samples," *Nature*, vol. 528, no. 7580, pp. 142–6, 2015.

[22] R. Browaeys, W. Saelens, and Y. Saeys, "Nichenet: modeling intercellular communication by linking ligands to target genes," *Nat Methods*, vol. 17, no. 2, pp. 159–162, 2020.

[23] Y. Hu, T. Peng, L. Gao, and K. Tan, "Cytotalk: De novo construction of signal transduction networks using single-cell transcriptomic data," *Sci Adv*, vol. 7, no. 16, 2021.

[24] S. Wang, M. Karikomi, A. L. MacLean, and Q. Nie, "Cell lineage and communication network inference via optimization for single-cell transcriptomics," *Nucleic Acids Res*, vol. 47, no. 11, p. e66, 2019.

[25] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *Journal of the Acm*, vol. 35, no. 4, pp. 921–940, 1988.

[26] P. Daran-Lapujade, M. L. Jansen, J. M. Daran, W. van Gulik, J. H. de Winde, and J. T. Pronk, "Role of transcriptional regulation in controlling fluxes in central carbon metabolism of saccharomyces cerevisiae. a chemostat culture study," *J Biol Chem*, vol. 279, no. 10, pp. 9125–38, 2004.

[27] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes and genomes," *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, 2000.

[28] L. Bardwell, "A walk-through of the yeast mating pheromone response pathway," *Peptides*, vol. 26, no. 2, pp. 339–50, 2005.

[29] M. Kanehisa, "Toward understanding the origin and evolution of cellular organisms," *Protein Sci*, vol. 28, no. 11, pp. 1947–1951, 2019.

[30] Z. Duan, Y. Dai, A. Hwang, C. Lee, K. Xie, C. Xiao, M. Xu, M. J. Girgenti, and J. Zhang, "iherd: an integrative hierarchical graph representation learning framework to quantify network changes and prioritize risk genes in disease," *PLoS Comput Biol*, vol. 19, no. 9, p. e1011444, 2023.

[31] L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, and H. Peterson, "gprofiler2 – an r package for gene list functional enrichment analysis and namespace conversion toolset g:profiler," *F1000Res*, vol. 9, 2020.

[32] R. L. Haynes, R. D. Folkerth, R. J. Keefe, I. Sung, L. I. Swzeda, P. A. Rosenberg, J. J. Volpe, and H. C. Kinney, "Nitrosative and oxidative injury to premyelinating oligodendrocytes in periventricular leukomalacia," *J Neuropathol Exp Neurol*, vol. 62, no. 5, pp. 441–50, 2003.

[33] J. Zhong, M. Qiu, Y. Meng, P. Wang, S. Chen, and L. Wang, "Single-cell multi-omics sequencing reveals the immunological disturbance underlying stat3-v637m hyper-ige syndrome," *Int Immunopharmacol*, vol. 122, p. 110624, 2023.

[34] R. Mukherjee, P. Kanti Barman, P. Kumar Thatoi, R. Tripathy, B. Kumar Das, and B. Ravindran, "Non-classical monocytes display inflammatory features: Validation in sepsis and systemic lupus erythematous," *Sci Rep*, vol. 5, p. 13886, 2015.