**Supplementary Files**

***for***

***Network Activity Evaluation Reveals Significant Gene Regulatory Architectures during SARS-CoV-2 Viral Infection from Dynamic scRNA-seq Data***

Chuanyuan Wang, Zhi-Ping Liu

## **S1. FPCA fits time-course gene expression data**

Functional Principal Component Analysis (FPCA) is a flexible non-parametric method for analyzing continuous trajectory data [1]. It models the data as a time process, and its characteristic basis functions reflect the main change patterns underlying the data. Recently, FPCA has been utilized to identify time-course gene expression profile trajectories for subsequent analysis. Moreover, the smooth continuous curve derived by FPCA can easily obtain the derivative value at any time, which facilitates downstream ODE system modeling [2, 3]. Specifically, FPCA approximates the gene expression curve with the following equation:

, (S1)

where  is the mean expression level over time,  is an orthonormal eigenfunction,  is corresponding coefficient, and  represents unexplained temporal variation.

Taking the simulation dataset D1 in the main text as an example, FPCA analysis is conducted on Cell type I. We plotted the fitted curve of expression values for one gene, as shown in Fig. S1, where we uniformly sampled eight time points along the curve (marked by red dots). Thus, the discrete time points of gene expression values of individual genes have been fitted into a smoothed continuous function curve. For simplicity, FPCA values for all genes are computed by the same time intervals. The derivation values of each gene expression at each time point (left hand side of Equation (1) in the main text) can be obtained easily after continuous fitting. Then the differential equations in ODE system can be transformed as a linear system with algebraic equations.

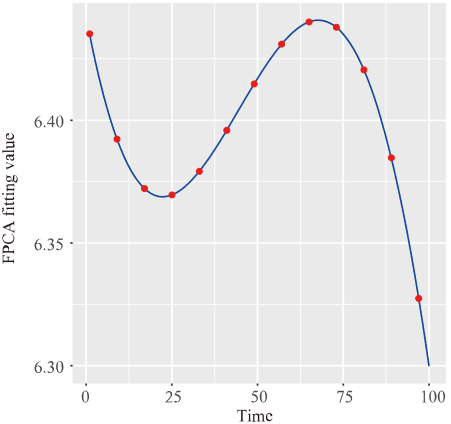


Fig. S1 An example of the fitted expression curves by FPCA on the D1 dataset. Fitted gene expression values are plotted as a continuous curve and the sampled time points are shown in red.

## **S2. Evaluation results of gene regulatory network activity in simulation datasets**

We used the SERGIO tool to generate a total of 12 relatively small-size simulation datasets D1...D12. For comparison, we ran the network activity evaluation tasks using scNAE, topologyGSA [4], SPIA [5], netGO[6] and NetGSA[7], respectively. Tables SI, SII, and SIII list the evaluation results on the simulation datasets.

TABLE SI

Experimental results in the D1-D6 datasets.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets  (Nodes, Cell Type) | scNAE with different parameters *Lp* | | | | | | topolo  gyGSA | SPIA | netGO | NetGSA |
| 0 | 1/3 | 1/2 | 2/3 | 1 | 2 |
| D1(30, I) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.599 | 0.559 | 0.061 | 0.704 |
| D1(30, II) | 0.007 | 0.008 | 0.008 | 0.005 | 0.015 | 0.008 | 0.005 | 0.272 | 0.490 | **0.005** |
| D2(50, I) | **0.001** | **0.001** | **0.001** | **0.001** | 0.006 | **0.001** | 0.008 | 0.065 | 0.019 | 0.037 |
| D2(50, II) | 0.011 | 0.017 | 0.013 | 0.018 | 0.033 | 0.021 | 0.009 | **0.001** | 0.127 | 0.008 |
| D3(100, I) | **0.001** | 0.001 | **0.001** | **0.001** | 0.015 | **0.001** | 1.000 | 0.272 | 0.218 | 0.003 |
| D3(100, II) | 0.003 | 0.002 | **0.001** | **0.001** | 0.001 | 0.002 | 0.504 | 0.047 | 0.349 | 0.645 |
| D4(30, I) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.136 | 0.786 | 0.089 | 0.349 |
| D4(30, II) | **0.001** | **0.001** | **0.001** | **0.001** | 0.005 | **0.001** | 0.011 | 0.845 | 0.107 | 0.077 |
| D4(30, III) | **0.002** | **0.002** | 0.002 | 0.005 | 0.003 | 0.003 | 0.009 | 0.562 | 0.030 | 0.631 |
| D5(50, I) | 0.002 | 0.002 | **0.001** | 0.002 | **0.001** | 0.002 | 0.006 | 0.438 | 0.053 | 0.121 |
| D5(50, II) | 0.044 | 0.054 | 0.055 | 0.052 | 0.064 | **0.042** | 0.068 | 0.330 | 0.164 | 0.926 |
| D5(50, III) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.043 | 0.0738 | 0.069 | 0.026 |
| D6(100, I) | **0.001** | **0.001** | **0.001** | 0.002 | **0.001** | **0.001** | 0.019 | 0.890 | 0.293 | 0.347 |
| D6(100, II) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.012 | 0.565 | 0.283 | 0.003 |
| D6(100, III) | 0.029 | 0.042 | 0.048 | 0.046 | 0.041 | 0.029 | 0.039 | 0.610 | 0.524 | **0.002** |

TABLE SII

Experimental results in the D7-D9 datasets.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets  (Nodes, Cell Type) | scNAE with different parameters *Lp* | | | | | | topolo  gyGSA | SPIA | netGO | NetGSA |
| 0 | 1/3 | 1/2 | 2/3 | 1 | 2 |
| D7(30, I) | 0.011 | 0.013 | 0.011 | 0.013 | 0.004 | 0.008 | **0.001** | 0.991 | 0.072 | 0.412 |
| D7(30, II) | 0.014 | 0.004 | 0.007 | 0.008 | **0.007** | **0.007** | 0.059 | 0.987 | 0.142 | 0.145 |
| D7(30, III) | 0.003 | 0.001 | 0.001 | 0.001 | **0.001** | **0.001** | 0.82 | 0.489 | 0.140 | 0.820 |
| D7(30, IV) | 0.005 | **0.004** | 0.049 | 0.052 | 0.059 | 0.045 | 0.015 | 0.538 | 0.207 | 0.748 |
| D8(50, I) | 0.032 | 0.03 | 0.029 | 0.022 | **0.020** | 0.033 | 0.302 | 0.698 | 0.049 | 0.041 |
| D8(50, II) | 0.012 | 0.023 | 0.027 | 0.022 | **0.012** | 0.025 | 0.048 | 0.565 | 0.234 | 0.063 |
| D8(50, III) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.017 | 0.272 | 0.289 | 0.015 |
| D8(50, IV) | 0.006 | 0.011 | 0.007 | 0.008 | 0.01 | **0.004** | 0.01 | 0.272 | 0.332 | 0.009 |
| D9(100, I) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.039 | 0.992 | 0.093 | 0.410 |
| D9(100, II) | 0.369 | 0.384 | 0.344 | 0.312 | 0.335 | 0.399 | **0.241** | 0.998 | 0.108 | 0.131 |
| D9(100, III) | **0.001** | 0.002 | 0.002 | **0.001** | **0.001** | **0.001** | 0.081 | 0.321 | 0.167 | 0.652 |
| D9(100, IV) | 0.011 | 0.017 | 0.013 | **0.009** | 0.013 | 0.013 | 1.000 | 0.485 | 0.288 | 0.020 |

## **S3. Evaluation results for gene network activity in large-size datasets**

For validating the performance of scNAE on large networks, we also generated single-cell gene expression profiling data for gene regulatory networks with 300, 500 and 1000 genes in three cell types. These datasets are denoted as D13, D14, and D15 for the networks with 300, 500, and 1000 genes, respectively. The results of network activity evaluation *P*-values are listed in TABLE SIV.

As shown, it achieves relatively satisfactory results when the model using L1-norm regularization term. In most of the results, scNAE demonstrates significant evaluation performance as expected. For comparison purpose, we simultaneously conduced the four alternative network-based algorithms. The results indicate that topologyGSA and SPIA did not work well in large-scale datasets, with no significant *P*-values. NetGO also demonstrates no performance advantages on large-scale networks. It is found that NetGSA provides relatively stable evaluations for larger networks with different sizes. However, its effectiveness was limited to certain few conditions. In contrast, our proposed scNAE method accurately estimastes the activity of gene regulatory networks in large-scale single-cell datasets, especially when setting the suitable sparse penalty for yielding optimal results.

TABLE SIV

Experimental results in the D13-D15 datasets.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets  (Nodes, Cell Type) | scNAE with different parameters *Lp* | | | | | topolo  gyGSA | SPIA | netGO | NetGSA |
| 0 | 1/3 | 1/2 | 2/3 | 1 |
| D13(300, I) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 1.000 | 1.000 | 0.291 | 0.077 |
| D13(300, II) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 1.000 | 1.000 | 0.193 | 0.031 |
| D13(300, III) | 0.002 | 0.002 | 0.002 | **0.001** | **0.001** | 1.000 | 1.000 | 0.818 | 0.086 |
| D14(500, I) | 0.032 | 0.035 | 0.035 | 0.253 | **0.001** | 1.000 | 1.000 | 0.793 | 0.117 |
| D14(500, II) | 0.273 | 0.273 | 0.273 | 0.787 | **0.016** | 1.000 | 1.000 | 0.698 | 0.081 |
| D14(500, III) | **0.001** | 0.003 | 0.003 | **0.001** | **0.001** | 1.000 | 1.000 | 0.943 | 0.830 |
| D15(1000, I) | 0.672 | 0.546 | 0.885 | 0.885 | 0.340 | 1.000 | 1.000 | 0.469 | **0.162** |
| D15(1000, II) | 0.087 | 0.051 | 0.218 | 0.218 | 0.030 | 1.000 | 1.000 | 0.709 | **0.002** |
| D15(1000, III) | **0.001** | **0.001** | **0.001** | 0.002 | **0.001** | 1.000 | 1.000 | 0.344 | 0.006 |

Note: The model with L2-norm penalty term has been omitted due to the significant time and memory costs needed for the solver of corresponding quadratic programming optimization.

TABLE SIII

Experimental results in the D10-D12 datasets.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets  (Nodes, Cell Type) | scNAE with different parameters *Lp* | | | | | | topolo  gyGSA | SPIA | netGO | NetGSA |
| 0 | 1/3 | 1/2 | 2/3 | 1 | 2 |
| D10(30, I) | 0.106 | 0.145 | 0.125 | 0.123 | 0.120 | 0.136 | **0.045** | 0.963 | 0.085 | 0.512 |
| D10(30, II) | 0.013 | 0.066 | 0.028 | 0.017 | **0.006** | 0.081 | 0.714 | 0.811 | 0.127 | 0.505 |
| D10(30, III) | 0.012 | **0.001** | 0.006 | 0.007 | 0.015 | **0.001** | 0.007 | 0.927 | 0.216 | 0.095 |
| D10(30, IV) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.005 | 0.594 | 0.207 | 0.945 |
| D10(30, V) | 0.006 | **0.001** | 0.006 | 0.005 | 0.008 | **0.001** | 0.016 | 0.908 | 0.359 | 0.001 |
| D11(50, I) | 0.468 | 0.408 | 0.402 | 0.413 | 0.156 | 0.123 | **0.030** | 0.995 | 0.043 | 0.129 |
| D11(50, II) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.089 | 0.993 | 0.243 | 0.020 |
| D11(50, III) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.007 | 0.919 | 0.299 | 0.160 |
| D11(50, IV) | 0.004 | 0.002 | **0.001** | 0.004 | 0.005 | 0.004 | 0.044 | 0.562 | 0.310 | 0.018 |
| D11(50, V) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 0.032 | 0.999 | 0.300 | 0.006 |
| D12(100, I) | 0.010 | **0.007** | 0.009 | 0.013 | 0.011 | 0.010 | 1.000 | 0.928 | 0.063 | 0.241 |
| D12(100, II) | 0.372 | 0.412 | 0.353 | 0.322 | 0.414 | 0.388 | **0.018** | 0.984 | 0.164 | 0.097 |
| D12(100, III) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 1.000 | 0.687 | 0.130 | **0.001** |
| D12(100, IV) | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | **0.001** | 1.000 | 0.738 | 0.221 | **0.001** |
| D12(100, V) | 0.012 | 0.007 | 0.008 | 0.010 | **0.006** | 0.010 | 0.014 | 0.924 | 0.135 | 0.404 |

## **Supplementary References**

[1] S. Wu, and H. Wu, “More powerful significant testing for time course gene expression data using functional principal component analysis approaches,” *BMC Bioinformatics,* vol. 14, no. 1, pp. 6, 2013.

[2] S. Wu, Z.-P. Liu, X. Qiu, and H. Wu, “Modeling genome-wide dynamic regulatory network in mouse lungs with influenza infection using high-dimensional ordinary differential equations,” *PLoS One,* vol. 9, no. 5, pp. e95276, 2014.

[3] Y. Wang, C. Liu, X. Qiao, X. Han, and Z.-P. Liu, “PKI: A bioinformatics method of quantifying the importance of nodes in gene regulatory network via a pseudo knockout index,” *Biochimica et Biophysica Acta -Gene Regulatory Mechanisms,* vol. 1866, no. 2, pp. 194911, 2023.

[4] M. S. Massa, M. Chiogna, and C. Romualdi, “Gene set analysis exploiting the topology of a pathway,” *BMC Systems Biology,* vol. 4, no. 1, pp. 1-15, 2010.

[5] A. L. Tarca, S. Draghici, P. Khatri, S. S. Hassan, P. Mittal, J.-s. Kim, C. J. Kim, J. P. Kusanovic, and R. Romero, “A novel signaling pathway impact analysis,” *Bioinformatics,* vol. 25, no. 1, pp. 75-82, 2009.

[6] J. Kim, S. Yoon, and D. Nam, “netGO: R-Shiny package for network-integrated pathway enrichment analysis,” *Bioinformatics,* vol. 36, no. 10, pp. 3283-3285, 2020.

[7] M. Hellstern, J. Ma, K. Yue, and A. Shojaie, “NetGSA: Fast computation and interactive visualization for topology-based pathway enrichment analysis,” *PLoS Computational Biology,* vol. 17, no. 6, pp. e1008979, 2021.