Supplementary materials

A Dual-Graph-Driven Non-negative Matrix Factorization Model for Single-Cell Omics Analysis

Junjie Lan[1], Nizhuan Wang[2], Jin Deng[1]*

[1] *College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China.*

[2]*The Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, HongKong 100872 China*

** Corresponding Author*

1. Bioinformatics analyses

To assess the ability of ADGNMF to identify marker genes, this study conducted experiments using the Pollen dataset. Specifically, the most significantly differentially expressed genes within each rank were calculated from the gene weight matrix W learned by ADGNMF, which were then designated as marker genes. Supplementary Fig. 2A illustrates the expression patterns of these identified marker genes across different cell clusters in the Pollen dataset. This figure demonstrates that genes identified by ADGNMF are significantly expressed exclusively in specific cell clusters with minimal cross-cluster leakage, indicating its ability to accurately identify marker genes for each cluster and underscoring its strong potential for deep biological insights.

Next, this study further validates whether the marker genes identified by ADGNMF can reveal the underlying biological information of single-cell datasets. To this end, 72 marker genes derived from the Pollen dataset that profiles cerebral cortex sequencing and identified by ADGNMF were subjected to GO-KEGG enrichment analysis. The analysis was performed using clusterProfiler with strict multiple-testing correction (FDR < 0.05) to avoid spurious associations, ensuring robust results. The results of the enrichment analysis are presented in Supplementary Fig. 2B, which displays significantly enriched pathways (P-value < 0.05). Pathway IDs, their official descriptions, Ontology (BP/CC/MF), and counts are provided in Supplementary Table 2.

Building on the enrichment analysis results, this study further investigates the underlying biological information of the Pollen dataset. By systematically cross-referencing with landmark neural development studies, this study uses Supplementary Table 3 to demonstrate that enrichment-identified pathways are significantly associated with cerebral cortex development. This alignment with established biological frameworks not only validates the identified pathways but also highlights ability of ADGNMF to effectively uncover the underlying biological processes within single-cell datasets.

2. Supplementary figure

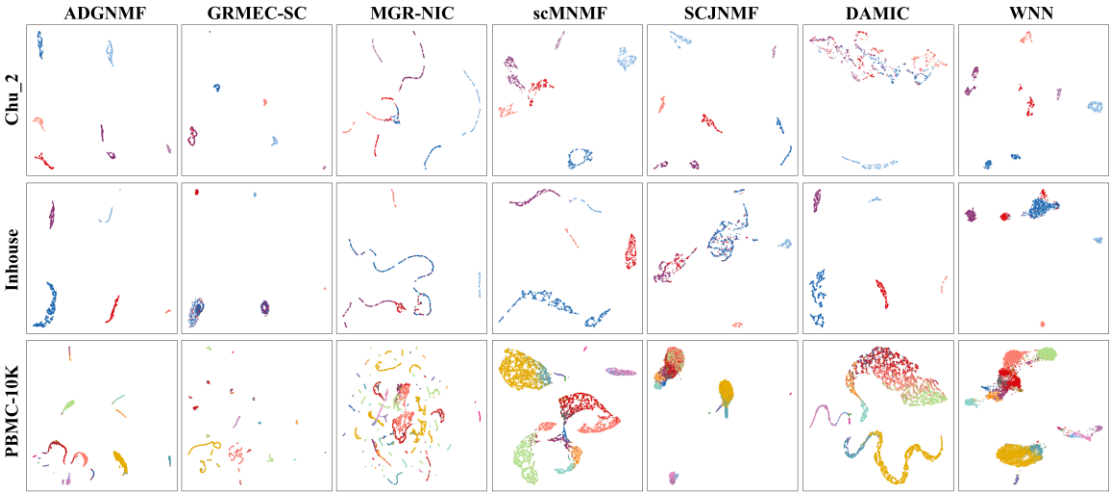

Fig. S1. Visualization of various methods applied to the Chu_2, Inhouse, and PBMC-10K datasets.

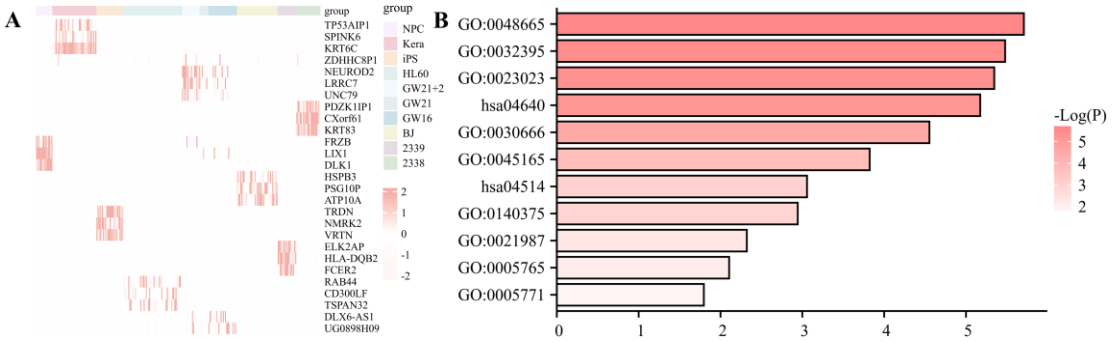

Fig. S2. (A) Heatmap depicting expression patterns of multiple genes across different cell types in the Pollen dataset: each row represents a gene and each column corresponds to a cell, visually illustrating the expression patterns of specific genes within each cell cluster. (B) Enrichment analysis based on marker genes identified by ADGNMF from the Pollen dataset.

3. Supplementary table

TABLE 1 Details of single cell Datasets.

Dataset	Species	Cell	Type	Omic1	Omic2
Pollen	Homo	259	10	23730	59033
Ning	Homo	460	4	15388	45731
Deng	Mouse	286	10	13709	50622
Chu_1	Homo	758	6	14441	71537
Chu_2	Homo	806	6	16020	81855
Biase	Mouse	56	5	13532	23536
Zanini	Mouse	4052	15	16030	75232
Darmanis	Homo	466	9	17617	109804
Inhouse	Homo	1182	6	33538	10
10X10K	Homo	7865	8	33525	14
Ma-2020	Mouse	5692	6	21478	340341
PMBC-10K	Homo	9631	19	29095	107194

Note: Omic1 and Omic2 represent feature counts for different omics.

TABLE 2 Functional Annotation Terms.

Term ID	Description	Ontology	Count
GO:0048665	neuron fate specification	BP	4
GO:0045165	cell fate commitment	BP	6
GO:0021987	cerebral cortex development	BP	3
GO:0030666	endocytic vesicle membrane	CC	6
GO:0005765	lysosomal membrane	CC	5
GO:0005771	multivesicular body	CC	2
GO:0032395	MHC class II receptor activity	MF	3
GO:0023023	MHC protein complex binding	MF	4
GO:0140375	immune receptor activity	MF	4
hsa04640	Hematopoietic cell lineage	MF	5
hsa04514	Cell adhesion molecules	MF	4

TABLE 3 Functional mechanisms of marker gene enrichment analysis in the Pollen dataset

Pathway ID	Mechanism of Action	Ref
GO:0005771	Multivesicular bodies are involved in key biological processes during cerebral cortex development, such as substance sorting, mitochondrial maintenance, and synaptic regulation.	[1]
GO:0005765	Cortical neurogenesis promotes cell fate remodeling by enhancing autophagy to generate new auto phagolysosomes.	[2]
GO:0021987	The progression of the cerebral cortex over time from its initial formation until its mature state.	[3]
GO:0140375	Inflammatory responses associated with immunodeficiency may lead to pathological conditions during cortical development.	[4]
hsa04514	Cells at the neural plate border form homodimers via calcium-dependent cadherins, driving the elevation and closure of the neural folds.	[5]
GO:0045165	Neural stem cell asymmetric division significantly regulates distinct daughter cell fates.	[6]
GO:0030666	Endocytosis plays a crucial role in the internalization of key signaling molecules and receptors in neural stem cells during neural development.	[7]
hsa04640	Hematopoietic stem and progenitor cells play a critical role in the expansion, distribution, and myeloid differentiation of microglia-like phenotypes.	[8]
GO:0023023	Pathological endothelial cells exhibit upregulation of MHC class II molecules.	[9]
GO:0032395	Expression of MHC-II in human neural progenitor cells promotes central nervous system development.	[10]
GO:0048665	The temporal characteristics of neurogenesis exhibit a significant association with human brain evolution.	[11]

References

[1] Turegano-Lopez M, Santuy A, DeFelipe J, et al. Size, shape, and distribution of multivesicular bodies in the juvenile rat somatosensory cortex: a 3D electron microscopy study[J]. Cerebral Cortex, 2020, 30(3): 1887-1901.

[2] Zou W, Lv Y, Zhang S, et al. Lysosomal dynamics regulate mammalian cortical neurogenesis[J]. Developmental Cell, 2024, 59(1): 64-78. e5.

[3] Di Bella D J, Domínguez-Iturza N, Brown J R, et al. Making Ramón y Cajal proud: Development of cell identity and diversity in the cerebral cortex[J]. Neuron, 2024, 112(13): 2091-2111.

[4] Nisticò R, Salter E, Nicolas C, et al. Synaptoimmunology-roles in health and disease[J]. Molecular Brain, 2017, 10(1): 26.

[5] Taneyhill L A. To adhere or not to adhere: the role of Cadherins in neural crest cell development[J]. Cell adhesion & migration, 2008, 2(4): 223-230.

[6] Liu K, Shen D, Shen J, et al. The super elongation complex drives neural stem cell fate commitment[J]. Developmental cell, 2017, 40(6): 537-551. e6.

[7] Lv Y, Zou W, Li L, et al. IFITM2 Modulates Endocytosis Maintaining Neural Stem Cells in Developing Neocortex[J]. Advanced Science, 2025, 12(17): 2501593.

[8] Milazzo R, Montepeloso A, Kumar R, et al. Therapeutic efficacy of intracerebral hematopoietic stem cell gene therapy in an Alzheimer's disease mouse model[J]. Nature Communications, 2024, 15(1): 8024.

[9] Wälchli T, Ghobrial M, Schwab M, et al. Single-cell atlas of the human brain vasculature across development, adulthood and disease[J]. Nature, 2024, 632(8025): 603-613.

[10] Vagaska B, New S E P, Alvarez-Gonzalez C, et al. MHC-class-II are expressed in a subpopulation of human neural stem cells in vitro in an IFNγ-independent fashion and during development, Sci. Rep., 6, 24251[EB/OL].(2016)

[11] Bonnefont J, Vanderhaeghen P. Neuronal fate acquisition and specification: time for a change[J]. Current opinion in neurobiology, 2021, 66: 195-204.