# Inferring metabolite states from spatial transcriptomes using multiple graph neural network

Jiaxu Luo[1], Daosheng Ai[2,3]*, Wenzhi Sun[1,3]*

[1] School of Basic Medical Sciences, Capital Medical University, Beijing 100069, China

[2]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[3]Chinese Institute for Brain Research, Beijing 102206, China

* To whom the correspondence should be addressed:

**Wenzhi Sun, Ph.D (Lead contact)**

Email: sunwenzhi@cibr.ac.cn

**Daosheng Ai, Ph.D**

Email: aidaosheng@cibr.ac.cn

## Abstract

Metabolism serves as the pivotal interface connecting genotype and phenotype. While various methods are available for predicting metabolite levels from transcriptomic data, their efficacy remains poor. We developed an efficient and adaptable algorithm known as Multiple Graph-based Flux Estimation Analysis (MGFEA). By leveraging single-cell and spatial transcriptomic data, MGFEA enables the inference of metabolite levels. Furthermore, MGFEA can further improve the accuracy of these inferences through exploiting additional metabolome data.

## Main

In the intricate realm of cellular biology, common cells diligently uphold metabolic homeostasis within their internal milieu, a crucial state that ensures the proper functioning of distinct functional proteins. Notably, the diverse nature of cells entails varied modes of sustaining metabolic equilibrium. For instance, neoplastic cells[1–5] exhibit metabolic reprogramming, altering their metabolic profiles to adapt to their environment, and immune cells[6–8] dynamically adjust to effectively adapt and respond to their microenvironment.

Over the last decade, there has been a significant advancement in single-cell transcriptomics technology, leading to the accumulation of a substantial number of precise single-cell databases. The advancement of single-cell metabolomics has lagged behind single-cell transcriptomics largely due to its inherent technical bottlenecks[9,10]. Though progress has been slow, contemporary computational tools are now capable of characterizing metabolism via transcription. Constraint-based models in particular have shown significant promise by their ability to deduce the rate of metabolic reactions without the prerequisite detection of numerous kinetic parameters[11,12]. As a consequence, novel computational models have emerged for predicting the flux state in the resolution of single cell. These include, but are not limited to, scFBA[13], scFEA[14], Compass[15], and METAFlux[16]. Despite their innovative nature, the efficiency of these models still falls short of optimal levels.

Building upon the foundations established by scFEA and a series of constraint-based models, we introduce a novel modeling framework for inferring metabolic flux based on both metabolic network guided gene interaction graph and spatial information graph (Fig. 1a). This framework aims to estimate the divergences in metabolic reactions among cells, utilizing the known

gene-reaction relationships in the Genome Scale Metabolism model (GSMM) and single-cell transcriptomics data.

Our framework exhibits remarkable flexibility, featuring multiple modules capable of accepting inputs in diverse data formats, including single-cell transcriptome datasets, spatial transcriptome datasets, and spatial multimodal analysis (SMA) datasets[17]. Leveraging Graph Variational Autoencoder (GVAE) or Variational Autoencoder (VAE), our framework extracts cellular embeddings from the data. Additionally, another GVAE is employed to extract inter-gene metabolic information derived from the knowledge-guided graph of Genome-Scale Metabolic Models (GSMM), thereby generating gene embeddings. The original transcriptome matrix is then enhanced using a combined weight matrix that integrates both types of embeddings. Subsequently, the improved matrix serves as input, allowing for the neural network to autonomously learn without fixed gene-metabolic reaction relationships. Through unsupervised metabolic flux inference, our framework derives metabolic flux, enabling the characterization of metabolite imbalance levels across different cells akin to metabolomics analysis. Furthermore, an additional reference module harnesses the metabolic information present in the SMA dataset to further optimize the algorithm's performance and enhance its analytical capabilities.

Our framework demonstrated superior performance in terms of faster computation speed, reduced memory usage, and improved result accuracy compared to scFEA, especially evident in handling larger datasets. The code implementation and relevant data are available on GitHub (https://github.com/Sunwenzhilab/MGFEA).

**Result:**

Compared with scFEA, our model demonstrates significant performance advantages,

particularly in its compatibility with the growing scale of single-cell datasets and its efficiency in terms of memory usage and computational speed (Fig. 1b, c). To assess the effect of dataset size to model accuracy, we utilized a bulk RNA-seq dataset paired with metabolomics data from cancer cell lines provided by the Cancer Dependency Map (DepMap) Project[18,19]. We selected overlapped metabolites among the two models and the metabolomics dataset. Then we calculated the mean square error between the predictions and metabolomics result. We presented the results of the metabolome, the prediction of scFEA, the prediction of MGFEA, and the prediction of MGFEA with our simplified metabolic network guided gene interaction graph based on the IMM1865 model[20,21]. As the dataset size increases, the performance of predicted results of MGFEA consistently outperformed that of scFEA (Fig. 1d). For the SMA dataset, we selected one of the slices to compare the prediction results of different models against MALDI results (Fig. 1e, f). The results showed that MGFEA captured spatial similarity of different spots and showed better performance in the situation which took raw matrix as input. Consistent with scFEA results, imputation improved the performance of FEA method. Despite displaying larger variance in small dataset, MGFEA demonstrated comparable upper bound with scFEA. In an application utilizing single-cell transcriptome data from gliomas sampled across different regions, as published by Yu et al.,[22], we observed a significant difference in lactate levels between cells in the tumor core and those in the peripheral regions (Fig. S1a). This finding aligns with FMRI results [23].

Considering the scalability of stereo-seq datasets[24], we selected five thresholds, resulting in the generation of five different sizes of datasets, each with varying counts and sizes of spots. We compared memory usage and time cost in the different formats of datasets (Fig. 2a). Based on the metabolomics dataset of different brain regions from Shao et al[25] and visium sagittal spatial

RNAseq dataset, we presented the results of the metabolome, the prediction of scFEA, the prediction of MGFEA, and the prediction of MGFEA -IMM (Fig. 2b). We presented several examples of prediction results (Fig. 2c). MGFEA exhibited comparable linear correlation with scFEA in prediction accuracy (Fig. 2d). Considering the paired nature of the SMA dataset, leveraging the metabolomics results as a reference for MGFEA prediction can improve performance (Fig. 2e). We observed an improvement in accuracy and a reduction in variance when using the MALDI reference, although statistical significance was not achieved. (Fig. 2f).

**Discussion:**

Based on the flux balance loss function and non-negative flux loss function initially proposed by scFEA, we have further developed the framework with three main innovations: (1) Intergration of highly variable genes. We integrated the highly variable genes into the metabolic network guided gene interaction graphs. Utilizing graph variational autoencoders, MGFEA learned the intrinsic relationships between genes within the metabolic network guided gene interaction graphs, as well as the spatial correlations between spots in spatial transcriptome dataset. (2) Unsupervised mapping of genes to metabolic flux. Instead of adhering to the fixed mapping relationships between specific gene expression and specific reaction fluxes, we employed the multilayer perceptron (MLP) to directly map from the expression data to metabolic flux. Our model adopted the loss function with the global metabolites balance punishment. (3) Incorporation of expert knowledge and deep learning. We developed the metabolic prior knowledge guided gene interaction graph and further enhanced the data based on the gene and cell information by softmax transformation. This pipeline reflects our attempt to integrate biological prior knowledge with the

existing deep learning methods. These innovations collectively improve the performance of MGFEA.

Our approach also has some limitations. (1) Training dataset availability and quality. Appropriate datasets are crucial for the improvement and development of methodological research. However, spatially paired spatial transcriptomic and spatial metabolomic databases are still rare, lacking a sufficiently reliable basis for developing methods for metabolism inference from transcriptome. In addition, metabolite detection with high spatial resolution remains technically deficient. This situation also hinders us from assessing the predictive results of our algorithms in larger datasets. (2) Model consistency. The fully-connected structure we employed demonstrated a better upper bound. However, its model estimation exhibited inconsistency and instability. To alleviate the limitation, we prefer to estimate several times and average the multiple repetitions to reduce the instability. The instability is also reflected in the facts that prediction of results is sometimes opposite to the gene expression trends. This phenomenon arised from the flux balance loss function, which is mathematically guaranteed to yield a set of solutions within the vicinity of the zero space of the matrix formed by the metabolic network, but the solutions, whether positive or negative, do not affect the convergence of total loss. Hence, we believe that the nature of the flux estimated by MGFEA differs from the classically perceived metabolic flux. Instead, it primarily reflects the variations in reaction rates between different cells. (3) The constrained assumption. Considering the close correlation between mRNAs and proteins[26], we utilize the expression abundance of mRNAs as a surrogate for enzyme concentration, and draw upon existing knowledge of enzyme kinetics[27] to model metabolic reaction rates as nonlinear functions of gene expressions. However, this approach depends on the assumption of biological steady state. Our

model, therefore, may not be suitable for cells in dynamic states. This limitation affects the accuracy and interpretability of our predictions.

With the development of MALDI technology, we are able to measure the intricate relationships between different metabolites and the gene expression, however, according to SMA data[17], the correlation between metabolites and the actual gene expression value is weak, which is related to the accuracy of the measurement technology, but also means that we need to be able to adapt the computational method to the more complex scenarios to help further explore the complex non-linear relationship between different metabolites and genes. Our reference based framework attempted to further improve the predictive power of the model for specific metabolite by providing additional metabolomics data, which also means that our flexible framework deserves further development.

## Figure legends

Fig. 1: An algorithmic framework to estimate metabolic states based on spatial and single cell transcriptomic data, and computational performance measurement.

a, MGFEA algorithmic framework. b, Comparison of peak memory usage between scFEA and MGFEA for different sized datasets. c, Comparison of time cost between scFEA and MGFEA for different sized datasets. d, Metabolic prediction of the same bulk RNA-seq dataset of the cancer cell lines by scFEA, MGFEA, and MGFEA-IMM. e, Results of spatial metabolomics and the prediction by scFEA, MGFEA and MGFEA-IMM for the relative metabolite level of SMA dataset with raw matrix and imputated matrix. f. Mean square error of the models with the different imputation methods.

Fig. 2: Comparison of MGFEA and scFEA in terms of computational performance, accuracy of computation, and the performance of reference module.

a, Comparison of peak memory usage and computation time between scFEA and MGFEA for different formats of stereo-seq dataset. b, Results of mean square error between Shao datasets and the prediction of scFEA, MGFEA and MGFEA-IMM for the relative metabolite level of mouse brain. c, Detection of log fold change of residual metabolites from hippocampus and cortex using metabolome, scFEA and MGFEA. d, The correlation of log-fold changes in predicted results with log-fold changes in metabolomes in hippocampus and cortex. e, Comparison of dopamine prediction basic MGFEA and referenced MGFEA with raw matrix and imputated matrix as input. f, Statistics of 50 repeats of two models' mean square error.

Fig. S1: Comparison of MGFEA and scFEA in terms of cancer metabolites imbalance level.

a, Predicted relative levels of lactate in different glioma regions by MGFEA (a) and scFEA (b).

# Methods

MGFEA framework

$X, x_k$ : expression matrix, expression vector of kth cell

$X_r$ : matrix of maldi

$T^g, T^c$ : the extracted Eigen vector matrix in the gene and cell dimension of the expression matrix using PCA

$u, v$ : cell embedding, gene embedding

$G^m, G^s, A^m, A^s$ : metabolic network guided gene interaction graph, spatial information graph, metabolic network guided gene interaction graph adjacent matrix, spatial information graph adjacent matrix

$\varphi_k, \theta_k$ : cell GVAE encoder parameter, decoder parameter

$\varphi_G, \theta_G$ : gene GVAE encoder parameter, decoder parameter

$F$ : Flux matrix of all cells in dataset

$f$ : flux vector of single cell

$S$ : stoichiometry matrix of GSMM model

MGFEA data augment

MGFEA takes a preprocessed expression matrix $X$ and a preprocessed GSMM model as input. Inspired by the GLUE[28] framework developed by Gao et al, we employed two separate VGAEs[29] to learn cell embeddings in the spatial transcriptomic data and gene embeddings in metabolic networks separately. For the cell embedding $u$, the input consists of the expression matrix $X$ and spatial coordinate information of spots $A^s$. The obtained latent layer embedding $u$ serves as

the representation of different cells.

For scRNAseq input, the training of VGAE satisfies the following loss function:

$$L_{cell}(\varphi_k, \theta_k) = E_{x_k \sim p_{data}(x_k)} \left[ E_{u \sim q(u|x_k;\varphi_k)} logp(x_k \mid u; \theta_k) - KL\left(q(u \mid x_k; \varphi_k) \parallel p(u)\right)\right]$$

For spatial RNAseq input, the training of VGAE satisfies the following loss function:

$$L_{cell}(\varphi_k, \theta_k) = E_{i,j \sim A^s} \left[ E_{u \sim q(u|x_k, A^s;\varphi_k)} logp(A^s \mid u; \theta_k) - KL\left(q(u \mid x_k, A^s; \varphi_k) \parallel p(u)\right)\right]$$

For gene embedding $v$, we transform the metabolic network graph into a gene interaction graph. The graph incorporated metabolic relationships from the GSMM model, along with the intrinsic gene co-expression information in the expression matrix, as the edges between genes. For metabolic edges, we construct gene interaction between genes connected by reaction or metabolite. For gene co-expression edges, we calculated the expression correlation of metabolic genes with highly variable genes. Edge connections are established between the top k hvg genes and the metabolic genes with the highest correlation in their expression. To ensure the information density of HVG is comparable to that of metabolic genes, we compute the normalized dispersion of all metabolic genes. Then, the sum of normalized dispersion is used as the threshold value to select the top k HVG genes whose corresponding statistics equals to that of the metabolic genes.

According to the above method, we construct a gene adjacency matrix $A^m$. Using this gene interaction graph as input, the VGAE learns the intrinsic relationships of genes and acquires gene embeddings to represent different genes.

The loss function of VGAE which exported gene embeddings satisfies:

$$L_{gene}(\varphi_G, \theta_G) = E_{i,j \sim A^m} \left[ logp(A^m \mid v, \theta_{G_m}) - KL\left[q(v \mid A^m, T^g, \varphi_{G_m}) \parallel p(v)\right]\right]$$

The loss of two VGAE is the latent loss of MGFEA.

$$Ll(\varphi_{G_s}, \theta_{G_s}, \varphi_{G_m}, \theta_{G_m}) = \frac{1}{K}\sum_{k=1}^{K} L_{cell} + L_{gene}$$

The two VGAEs learned the representations of the transcriptional cell states $u$ and gene representations $v$ defined jointly by the metabolic network and co-expression information. The former represents a conventional cellular state and incorporates both the expression matrix and spatial coordinate information of spots of spatial transcriptomics. The latter means the gene representations defined by gene interaction graph. The gene representations mean the genes' location in metabolic space. The dot product of two representations shares the same matrix form as the original expression matrix. The form is used for data enhancement.

Taking into account the inherent projection nature of the dot product, we interpret the dot product of the two as a projection of the cellular state representation vector onto the metabolic space. For scRNAseq, by utilizing the difference between this projection and the original transcriptional expression matrix as a loss function, we enable VAE to rationalize the cell representations it learns. For spRNAseq, VGAE learns to reconstruct the spatial coordinate graph and gene interaction graph to rationalize the obtained cell embeddings and gene embeddings.

With the above framework, we are able to generate matrix containing both metabolic gene relationships and cell transcriptional states. Then we conducted softmax transformation between different genes within a cell based on the generated matrix. This transformation yields a matrix which contains genes' weights in cells. Subsequently, we performed element-wise multiplication (Hadamard product) between weights matrix and the original expression matrix $X$. Through this process, we enhance the cell specific metabolic features in the original matrix $X$ to preserve the transcriptional states and the augmented matrix $X_a$ improved the process of MGFEA flux-estimation.

$$X_a = Softmax_i(v^T u) \otimes X$$
$$F = MLP(X_a)$$

MGFEA-Flux estimation

The balance of metabolites in reaction network is influenced by both in-flux and out-flux. Considering that the efficiency of enzymes in metabolic networks is regulated by the regulatory genes, the transcriptional state of the cell plays a crucial role in influencing metabolic balance. Building upon this premise, we utilized the transcriptional expression matrices of metabolic genes, along with a restricted set of highly variable genes, as input. We finally employed a Multilayer perceptron to estimate the fluxes of all metabolic modules (reactions).

$$L_{flux} = \sum_{i}^{reactions} \sum_{j}^{metabolites} [|S \cdot f| + |f| - f + |sgn(f)| - sgn(f)]$$

$$L_{total} = \lambda_f L_{flux} + (1 - \lambda_f) L_l \left(\varphi_{G_s}, \theta_{G_s}, \varphi_{G_m}, \theta_{G_m}\right)$$

Reference based framework

We use paired dataset[17] to offer a reference for spatialGraphFEA learning. We modify flux loss and add a reference loss for our model. We add reciprocal of flux to prohibit maintained decrease of flux loss. We use metabolism quantitative information as a reference and use a mean square error formula to forced predicted metabolism quantification into reference result.

$$L_{reference} = \frac{1}{K} mse(S \cdot f - X_r)$$

$$L_{total} = \lambda_f L_{flux} + \lambda_l L_l \left(\varphi_{G_s}, \theta_{G_s}, \varphi_{G_m}, \theta_{G_m}\right) + (1 - \lambda_f - \lambda_l) L_{reference}$$

GSMM model preprocess

We employed two published GSMM model Recon3D[21] and IMM1865 for MGFEA prediction of relative metabolites level. The original models have large numbers of metabolites consists of the

same metabolites located in different organelles. In our modified models, we removed duplicated metabolites and used function find_blocked_reaction from package cobrapy to remove most blocked reactions.

Data preprocess

Expression matrix $X$ is loaded in h5ad format and is normalized and log-scale transformed. We used magic[30] and stagate[31] for data imputation. We removed reactions without expression in expression matrix, and used the modified reaction network for MGFEA reaction prediction.

Mean square error

In all situations where the SMA dataset was used, we first scaled the metabolomics data and metabolites stock level output from the model, and then calculated the mean squared error. For all cases involving the neural cell metabolomics dataset, as the vectors are not paired, so we first computed the mean metabolites level in different cell types. Then we computed the mean square error for different metabolites in overlapped metabolites.

Metabolites' correlation

We scaled the relative levels of metabolites overlapped in all samples in metabolomics and predicted results from models. Then we computed the mean value across different brain region. Then we computed the log fold change of different metabolites between the hippocampus and cortex in different predicted results. The pearson correlation is computed based on the fold change predicted in different models and the fold change detected in the metabolomics.

Memory usage and time cost

We used python package memory-profiler to measure peak memory and total time cost of different models.

## Data availability

All datasets used in this study have been published and can be obtained in h5ad format from https://cellxgene.cziscience.com/datasets.

We used Recon3D and IMM1865 as human and mouse GSMM model. Our raw file and preprocess code can be obtained from our github site (https://github.com/Sunwenzhilab/MGFEA). Detailed message and URLs of datasets is recorded in Table 1

Table 1

| datasets | Source | website |
|---|---|---|
| 4M brain dataset | Cao et al. [32] | https://cellxgene.cziscience.com/collections/c114c20f-1ef4-49a5-9c2e-d965787fb90c |
| 1M GBM dataset | Ruiz Moreno et al. [33] | https://cellxgene.cziscience.com/collections/999f2a15-3d7e-440b-96ae-2c806799c08c |
| 200k brain dataset | Siletti et al.[34] | https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443 |
| 50k brain dataset | Smith et al. [35] | https://cellxgene.cziscience.com/collections/e02201d7-f49f-401f-baf0-1eb1406546c0 |
| 10k brain dataset | Siletti et al. [34] | https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443 |
| Yu dataset | Yu et al. [22] | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117891 |
| Stereo-seq dataset | Stereo-seq | http://116.6.21.110:8090/share/dd965cba-7c1f-40b2-a275-0150890e005f |

| | | |
|---|---|---|
| SMA dataset | Vicari et al. [17] | https://data.mendeley.com/datasets/w7nw4km7xd/1 |
| Visium sagittal brain section | 10X Genomics | https://www.10xgenomics.com/datasets/preservation-method-comparison-on-visium-cytassist-fixed-frozen-mouse-brain-sagittal-11-mm-capture-area-2-standard |
| Shao dataset | Shao et al.[25] | https://www.nature.com/articles/s41392-023-01552-y |
| DepMap transcriptome | Ghandi et al. [18] | https://depmap.org/portal/download/custom/ |
| DepMap metabolome | Li et al. [19] | https://depmap.org/portal/download/custom/ |

## Code availability

The code and related data can be accessible from the following GitHub respiratory (https://github.com/Sunwenzhilab/MGFEA).

**Author Contributions**

D.A and J.L. conceived the project. W.S. supervised the whole project. J.L. implemented all model codes. J.L. and D.A. performed all data analysis. All authors reviewed and edited the manuscript.

**Acknowledgement**

We thank all the member of Sun Laboratory and other colleagues from CIBR for their feedback on this work. This project is financially supported by STI2030-Major Projects (2022ZD0204700) and CIBR Internal Fund to W.S..

## References


1. Finley, L. W. S. What is cancer metabolism? *Cell* **0**, (2023).
2. Martínez-Reyes, I. & Chandel, N. S. Cancer metabolism: looking forward. *Nat. Rev. Cancer* **21**, 669–680 (2021).
3. Melone, M. A. B. *et al.* The carnitine system and cancer metabolic plasticity. *Cell Death Dis.* **9**, 1–12 (2018).
4. Bunse, L. *et al.* Suppression of antitumor T cell immunity by the oncometabolite (R)-2-hydroxyglutarate. *Nat. Med.* **24**, 1192–1203 (2018).
5. Notarangelo, G. *et al.* Oncometabolite d-2HG alters T cell metabolism to impair CD8+ T cell function. *Science* **377**, 1519–1529 (2022).
6. Panieri, E. & Santoro, M. M. ROS signaling and redox biology in endothelial cells. *Cell. Mol. Life Sci.* **72**, 3281–3303 (2015).
7. Panieri, E. & Santoro, M. M. ROS homeostasis and metabolism: a dangerous liason in cancer cells. *Cell Death Dis.* **7**, e2253–e2253 (2016).
8. O'Neill, L. A. J., Kishton, R. J. & Rathmell, J. A guide to immunometabolism for immunologists. *Nat. Rev. Immunol.* **16**, 553–565 (2016).
9. Lanekoff, I., Sharma, V. V. & Marques, C. Single-cell metabolomics: where are we and where are we going? *Curr. Opin. Biotechnol.* **75**, 102693 (2022).
10. Guo, S., Zhang, C. & Le, A. The limitless applications of single-cell metabolomics. *Curr. Opin. Biotechnol.* **71**, 115–122 (2021).
11. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
12. Hrovatin, K., Fischer, D. S. & Theis, F. J. Toward modeling metabolic state from single-cell transcriptomics. *Mol. Metab.* **57**, 101396 (2022).
13. Damiani, C. *et al.* Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. *PLOS Comput. Biol.* **15**, e1006733 (2019).
14. Alghamdi, N. *et al.* A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. *Genome Res.* gr.271205.120 (2021) doi:10.1101/gr.271205.120.
15. Wagner, A. *et al.* Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. *Cell* **184**, 4168-4185.e21 (2021).
16. Huang, Y. *et al.* Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlux. *Nat. Commun.* **14**, 4883 (2023).
17. Vicari, M. *et al.* Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nat. Biotechnol.* 1–5 (2023) doi:10.1038/s41587-023-01937-y.
18. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
19. Li, H. *et al.* The landscape of cancer cell line metabolism. *Nat. Med.* **25**, 850–860 (2019).
20. Khodaee, S., Asgari, Y., Totonchi, M. & Karimi-Jafari, M. H. iMM1865: A New Reconstruction of Mouse Genome-Scale Metabolic Model. *Sci. Rep.* **10**, 6177 (2020).
21. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
22. Yu, K. *et al.* Surveying brain tumor heterogeneity by single-cell RNA-sequencing of

multi-sector biopsies. *Natl. Sci. Rev.* **7**, 1306–1318 (2020).

23. Kubelt, C. *et al.* Intratumoral Distribution of Lactate and the Monocarboxylate Transporters 1 and 4 in Human Glioblastoma Multiforme and Their Relationships to Tumor Progression-Associated Markers. *Int. J. Mol. Sci.* **21**, 6254 (2020).

24. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792.e21 (2022).

25. Shao, Y. *et al.* A metabolome atlas of mouse brain on the global metabolic signature dynamics following short-term fasting. *Signal Transduct. Target. Ther.* **8**, 1–12 (2023).

26. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).

27. Cornish-Bowden, A. Current IUBMB recommendations on enzyme nomenclature and kinetics. *Perspect. Sci.* **1**, 74–87 (2014).

28. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).

29. Kipf, T. N. & Welling, M. Variational Graph Auto-Encoders. Preprint at https://doi.org/10.48550/arXiv.1611.07308 (2016).

30. D, van D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, (2018).

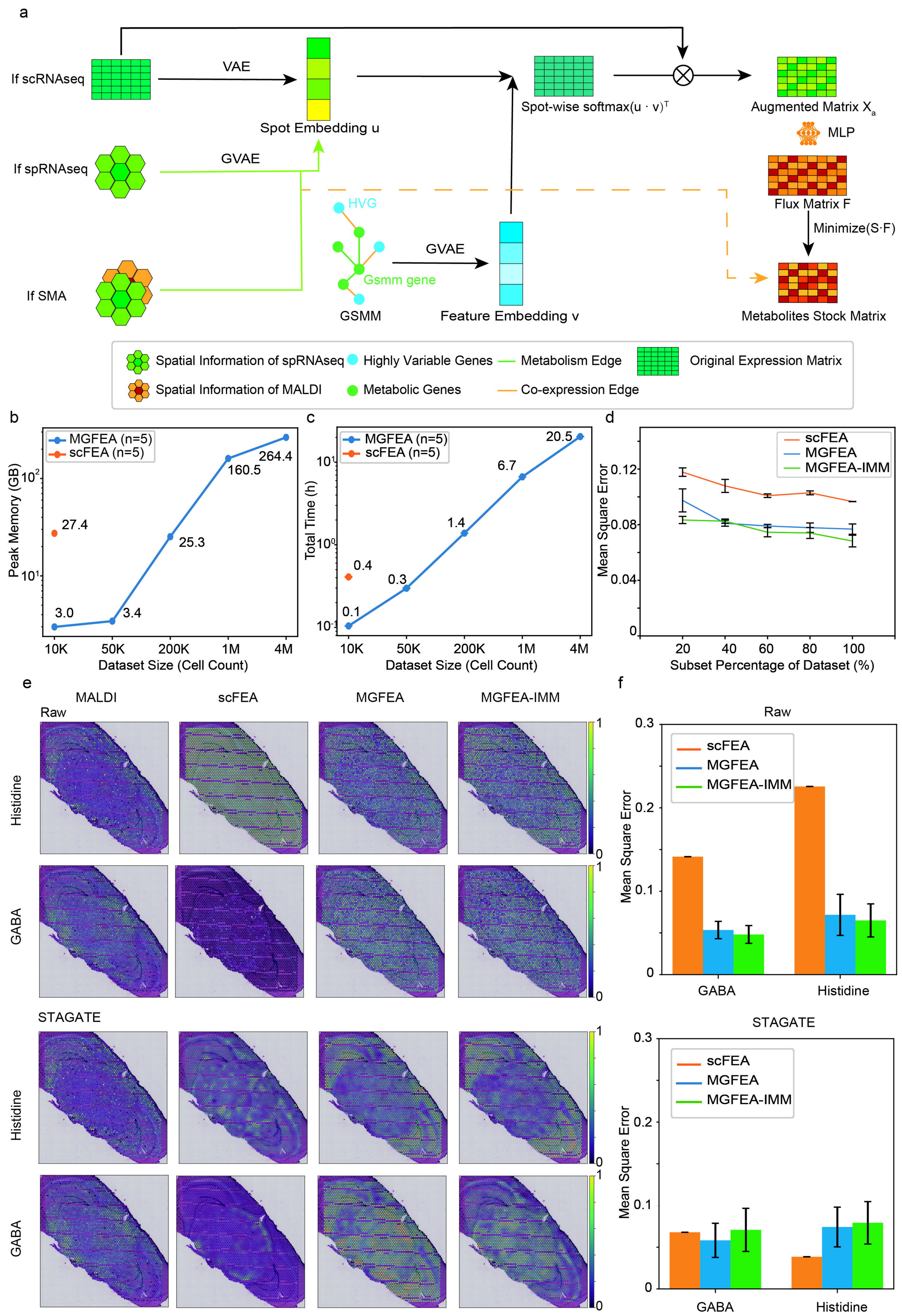
31. K, D. & S, Z. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, (2022).

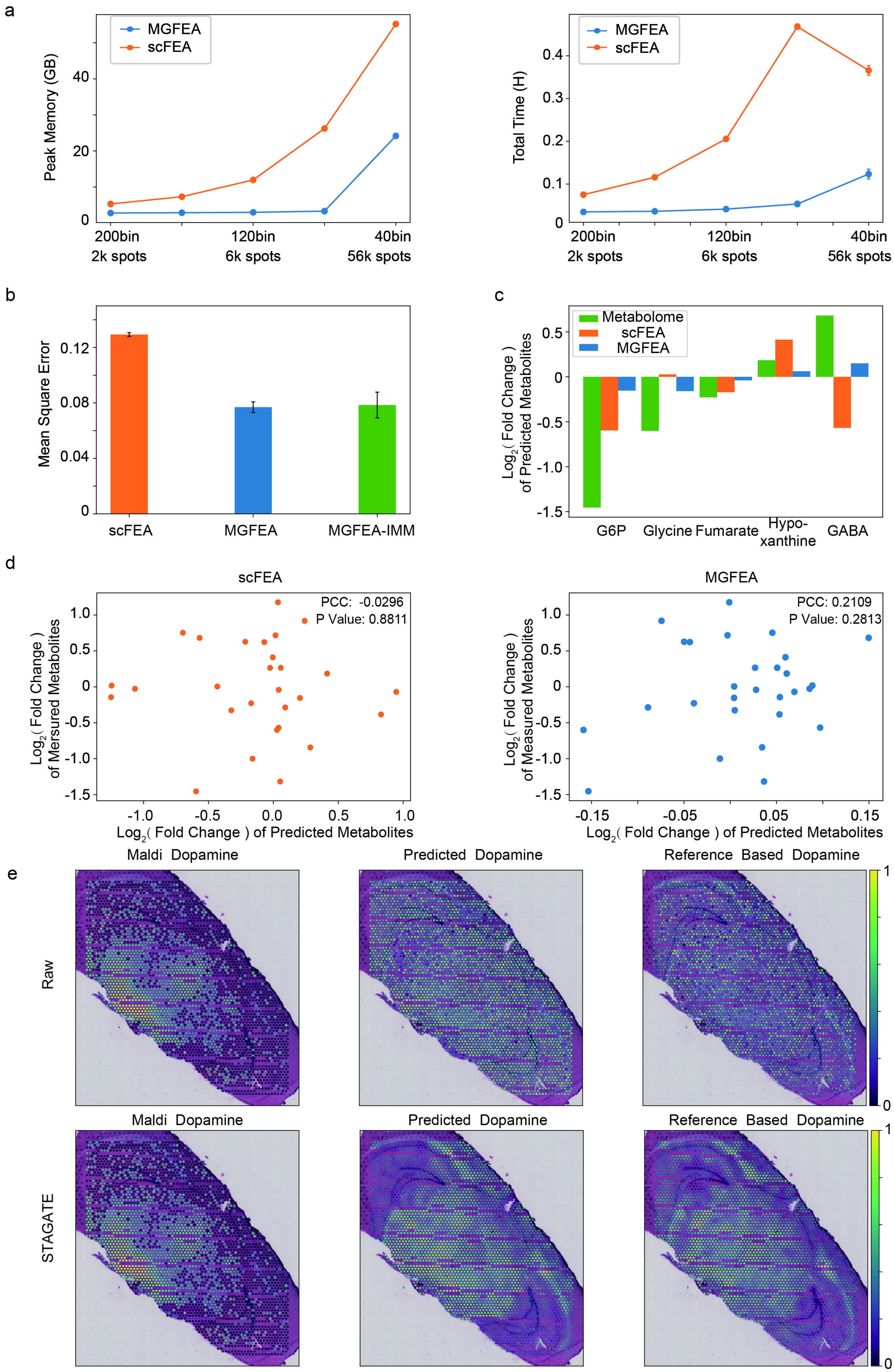
32. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).

33. Ruiz-Moreno, C. *et al.* Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma. 2022.08.27.505439 Preprint at https://doi.org/10.1101/2022.08.27.505439 (2022).

34. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).

35. Smith, R. S. *et al.* Early role for a Na+,K+-ATPase (ATP1A3) in brain development. *Proc. Natl. Acad. Sci.* **118**, e2023333118 (2021).


a
If scRNAseq
VAE
Spot Embedding u
Spot-wise softmax(u · v)^T
Augmented Matrix X_a
MLP
If spRNAseq
GVAE
Flux Matrix F
Minimize(S·F)
HVG
GVAE
Gsmm gene
If SMA
GSMM
Feature Embedding v
Metabolites Stock Matrix
Spatial Information of spRNAseq
Highly Variable Genes
Metabolism Edge
Original Expression Matrix
Spatial Information of MALDI
Metabolic Genes
Co-expression Edge
b
MGFEA (n=5)
scFEA (n=5)
Peak Memory (GB)
27.4
3.0
3.4
25.3
160.5
264.4
10K
50K
200K
1M
4M
Dataset Size (Cell Count)
c
MGFEA (n=5)
scFEA (n=5)
Total Time (h)
0.4
0.1
0.3
1.4
6.7
20.5
10K
50K
200K
1M
4M
Dataset Size (Cell Count)
d
scFEA
MGFEA
MGFEA-IMM
Mean Square Error
0.12
0.08
0.04
0
20
40
60
80
100
Subset Percentage of Dataset (%)
e
MALDI
scFEA
MGFEA
MGFEA-IMM
Raw
Histidine
GABA
STAGATE
Histidine
GABA
f
Raw
scFEA
MGFEA
MGFEA-IMM
Mean Square Error
0.3
0.2
0.1
0
GABA
Histidine
STAGATE
scFEA
MGFEA
MGFEA-IMM
Mean Square Error
0.3
0.2
0.1
0
GABA
Histidine




f

| Imputation | Raw | | STAGATE | |
|---|---|---|---|---|
| Model | No Reference | Reference | No Reference | Reference |
| Mean | 0.1629 | 0.1656 | 0.1789 | 0.1707 |
| Variance | 0.0010 | 0.0007 | 0.0019 | 0.0014 |
| T test P Value | 0.6548 | | 0.3182 | |