

Understanding Gene Expression-to-Image Learning: Spatial Structure and Information Fidelity

Xiaoqing Lian¹, Pei Liu¹, Jiahui Tan¹, Quan Zou², Chen Lin³, Xiangxiang Zeng^{4,5}

¹ College of Computer Science and Electronic Engineering,
Hunan University, Changsha, China

² University of Electronic Science and Technology of China,
Chengdu, China

³ Xiamen University, Xiamen, China

⁴ State Key Laboratory of Chemo and Biosensing,
Institute of AI4Science, Hunan University,
Changsha 410082, P. R. China

⁵ Yuelushan Laboratory, China

Abstract

Single-cell RNA sequencing (scRNA-seq) enables cell-level analysis of biological heterogeneity, but learning effective representations from high-dimensional, unordered gene expression profiles remains challenging. Gene Expression-to-Image (GE2I) offers an effective learning strategy by arranging gene values on pixel grids for convolutional neural networks, but its underlying mechanisms remain unclear. We study GE2I on full-transcriptome data while keeping the network and training procedure fixed. Our results reveal a trade-off between spatial smoothness and gene-level fidelity. Structured mappings produce smooth, low-frequency patterns that support faster early optimization, whereas assigning multiple genes to the same pixel obscures expression differences and can reduce predictive performance. This trade-off explains why random mappings can compete with structured layouts when many genes are represented. One-to-one gene-to-pixel assignment improves prediction across the tested visual backbones even when image smoothness decreases. We further find that the benefits of pretraining and augmentation depend on the backbone and mapping. These findings provide practical guidance for pixel allocation and the transfer of vision priors to gene expression data.

1 Introduction

Single-cell RNA sequencing (scRNA-seq) measures gene expression in individual cells and helps researchers study how cells differ from one another (Bunne et al., 2024; Eraslan et al., 2022). The resulting data form a high-dimensional matrix: each row represents a cell, and each column records the expression level of a gene (Tang et al., 2009). Unlike pixels in an image or words in a sentence, the genes in each row have no natural spatial order. Computational models therefore need a useful way to represent these gene expression vectors (Gunawan et al., 2023).

Common representation learning methods include probabilistic latent-variable models, such as Variational Autoencoders (VAEs) (Kingma & Welling, 2013; Grønbech et al., 2020), and self-attention-based sequence models, such as Transformers (Vaswani et al., 2017; Cui et al., 2024; Zeng et al., 2025). Another approach places the unordered gene values on a 2D pixel grid (Figure 1). We call this approach Gene Expression-to-Image (GE2I). Once the values are arranged as an image, a Convolutional Neural Network (CNN) can learn high-order statistics and nonlinear correlations without manually designed features (Islam & Xing, 2023). Placing correlated genes together may also combine weak signals into clearer

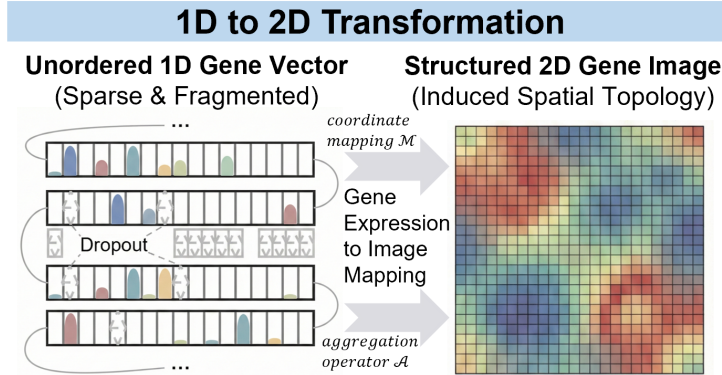


Figure 1: Gene expression-to-image (GE2I). A coordinate mapping $\mathcal{M}$ and aggregation operator $\mathcal{A}$ convert gene expression vectors into 2D images for a CNN. Colored gene labels illustrate local neighborhoods in the layout.

biological patterns. In addition, researchers can inspect the images directly to compare phenotypic differences across samples (Sharma et al., 2019).

Most GE2I methods place functionally related genes near one another because they assume that biological neighborhoods improve the representation. Many mapping algorithms are designed around this idea. However, a structured layout changes two things at once: it preserves some gene relationships and gives the data a smoother spatial pattern. It is therefore unclear which property accounts for better learning. We ask:

Which properties of gene expression images explain differences in optimization and predictive performance across mappings?

We investigate this question on full-transcriptome data while holding the network and training conditions fixed. We measure image properties, track training behavior, and use beta regression to relate each property to prediction. The results show a trade-off between spatial smoothness and gene-level fidelity. A structured layout can produce patterns that a CNN learns quickly, but placing several genes in the same pixel merges their values and can remove useful differences. Faster early optimization may therefore lead to little or no improvement in final prediction. We test this explanation by changing how genes are assigned to pixels, then use the results to guide image construction and the use of pretrained vision models. Our contributions are:

- Mechanistic analysis. Image measurements, training dynamics, and beta regression show how spatial smoothness and gene-level fidelity affect early optimization and final prediction differently.
- Assignment interventions. We test one-to-one gene-to-pixel assignments and introduce a greedy allocation procedure. Prediction improves even when these changes make the images less smooth.
- Backbone evaluation and transfer guidelines. We compare assignment effects across visual backbones and test pretraining and augmentation. The results clarify when vision priors help with transcriptomic learning.

2 Related Work

2.1 Representation Learning in Single-Cell Genomics

Representation learning in single-cell genomics aims to capture biological variation in sparse, high-dimensional expression profiles. Variational autoencoder models, including scVI (Lopez

et al., 2018) and scANVI (Xu et al., 2021), learn probabilistic cell representations, with scANVI additionally incorporating available cell-type annotations. More recently, foundation models such as scGPT (Cui et al., 2024), scBERT (Yang et al., 2022), and CellFM (Zeng et al., 2025) have used large-scale pretraining to learn representations for downstream single-cell analysis. These approaches model expression profiles without first assigning genes to a two-dimensional pixel grid. GE2I introduces this spatial representation before applying a visual model, making the arrangement of gene measurements an additional part of representation design.

2.2 The Gene Expression-to-Image (GE2I) Paradigm

GE2I maps gene expression values onto two-dimensional grids for CNNs (Sharma et al., 2019; Islam & Xing, 2023). Existing methods differ in how they construct these layouts. Projection-based approaches such as DeepInsight (Sharma et al., 2019) derive feature coordinates through dimensionality reduction, while MR-DeepInsight (Sharma et al., 2024) combines multiple image representations. Fotomics (Zandavi et al., 2023) instead uses Fourier transforms to determine feature coordinates. Other methods optimize spatial arrangements according to feature relationships: REFINED (Bazgir et al., 2020) refines a low-dimensional embedding through local search, IGTD (Zhu et al., 2021) matches the rankings of feature distances and pixel distances, and Genomap (Islam & Xing, 2023) uses optimal transport to encode gene relationships in a spatial representation.

Pixel allocation and visual training choices have also been considered in prior work. IGTD assigns each feature to a distinct pixel while optimizing neighborhood structure (Zhu et al., 2021). MR-DeepInsight uses pretrained CNNs and examines conventional image augmentation (Sharma et al., 2024). These developments address how to construct and learn from gene images, but comparing final prediction scores alone does not explain how the construction choices affect learning.

Our work complements these methods by examining spatial smoothness and gene-level fidelity. We relate these properties to early optimization and final prediction under shared network and training conditions, with pixel-assignment interventions providing a way to test the explanations.

3 Analytical Framework

3.1 Problem Formulation

Let $\mathbf{x} \in \mathbb{R}^G$ be a gene expression vector. A coordinate-based GE2I transformation uses a mapping $\mathcal{M} : \{1, \dots, G\} \rightarrow \Omega$ onto an $H \times W$ grid and a permutation-invariant aggregation operator $\mathcal{A}$, such as the mean. At each occupied pixel, $\mathbf{I}(p) = \mathcal{A}(\{x_i \mid \mathcal{M}(i) = p\})$. The mapping is shared across cells. Here, “topology” refers to gene placement and neighborhood structure in the grid. Even when $HW \geq G$, a layout may assign multiple genes to the same pixel, combining their expression values. We therefore examine both spatial organization and gene-level fidelity.

3.2 Spectral Descriptors (Quantifying Smoothness)

We characterize spatial structure in the frequency domain (Cao & Wu, 2022). Averaging the centered two-dimensional power spectrum within integer-radius rings gives a radial profile $\hat{r} \in \mathbb{R}^L$, where $L = \min(\lfloor W/2 \rfloor, \lfloor H/2 \rfloor)$. We normalize this profile so that $\sum_k \hat{r}_k = 1$.

Spectral Entropy (SE) measures the spread of power across radial frequencies:

$$\text{SE} = - \sum_k \hat{r}_k \log(\hat{r}_k + \epsilon), \quad (1)$$

where ϵ ensures numerical stability. Higher SE indicates a broader frequency distribution.

Low-Frequency Energy Ratio (LFER) measures low-frequency concentration:

$$\text{LFER} = \sum_{k=0}^{\lfloor L/4 \rfloor} \hat{r}_k. \quad (2)$$

Higher LFER indicates greater relative low-frequency power and smoother spatial patterns.

3.3 Information Fidelity (Quantifying Resolution)

We describe gene-level allocation using the occupied pixels $P_{used} \subseteq \Omega$ and the number of genes $g(p)$ assigned to each pixel.

Average Pixels Per Gene (APPG) measures occupied pixels relative to gene count:

$$\text{APPG} = \frac{|P_{used}|}{G}. \quad (3)$$

For a single-view coordinate assignment, $\text{APPG} = 1$ means that each gene occupies a unique pixel.

Collision Rate (CR) measures the proportion of occupied pixels shared by multiple genes:

$$\text{CR} = \frac{|\{p \in P_{used} : g(p) > 1\}|}{|P_{used}|}. \quad (4)$$

A value of zero indicates that no genes share a pixel. For Genomap’s soft transport coupling, uniform target marginals occupy all G transport targets ($\text{APPG} = 1$), while entropic regularization gives each target contributions from multiple genes ($\text{CR} = 1$). Here, CR therefore denotes soft mixing rather than hard coordinate collisions. APPG and CR describe potential information loss through allocation, rather than directly measuring task-relevant information.

3.4 The Convolutional Probe

We use ResNet-18 as a common learner for cell-type annotation. Early loss reduction characterizes optimization, while accuracy and F1-Macro measure final prediction. We relate the four image descriptors jointly to these outcomes using beta regression (Ferrari & Cribari-Neto, 2004). Each observation is a dataset–mapping–seed run within one input scope; Full and HVG inputs and the two prediction metrics are modeled separately:

$$Y_{dms} \sim \text{Beta}(\mu_{dms}\phi, (1 - \mu_{dms})\phi), \quad (5)$$

$$\text{logit}(\mu_{dms}) = \alpha_{ds} + \beta_1 \text{APPG}_{dms} + \beta_2 \text{CR}_{dms} + \beta_3 \text{LFER}_{dms} + \beta_4 \text{SE}_{dms}, \quad (6)$$

where d , m , and s index dataset, mapping, and seed. The descriptors enter on their original scales, α_{ds} denotes dataset-by-seed fixed effects, and ϕ is constant within each model. All outcomes lie strictly in $(0, 1)$. Standard errors are clustered by dataset–method, and we report two-sided p -values together with Benjamini–Hochberg-adjusted q -values. Appendix G provides the inference, coverage, imputation, and diagnostic details. Pixel-assignment interventions then test whether changing allocation alters prediction.

4 Experimental Protocols

4.1 Datasets and Benchmarks

We evaluate cell-type classification on five datasets: Diabetic Kidney Disease (DKD) (Wilson et al., 2022), GTEx v9 (Eraslan et al., 2022), HypoMap (Steuernagel et al., 2022), Immune Cell Atlas (Domínguez Conde et al., 2022), and Tabula Sapiens (Consortium* et al., 2022). These span disease-specific cohorts and cross-tissue atlases with varying class imbalance (Table 7). Our main analyses use full-transcriptome inputs, defined as all genes retained after preprocessing; complementary HVG evaluations are reported in the appendix. We also evaluate gene-to-protein prediction on NeurIPS 2021 CITE-Seq (Stoeckius et al., 2017). Dataset specifications and evaluation protocols are provided in Appendix B.

4.2 Gene Expression-to-Image Transformation Strategies

The DeepInsight family (t-SNE, UMAP, PCA) (Sharma et al., 2019) and MR-DeepInsight (Sharma et al., 2024) project genes onto fixed grids and average expression values at shared pixels. Fotomics (Zandavi et al., 2023) uses FFT-based assignment with the same aggregation rule. Adaptive methods use grids with side length $\lceil \sqrt{G} \rceil$: Genomap (Islam & Xing, 2023) constructs optimal transport-weighted pixel sums, whereas IGTD (Zhu et al., 2021) and REFINED (Bazgir et al., 2020) optimize one-to-one assignments. Random Mapping assigns genes to unique coordinates without using gene similarity and keeps these coordinates fixed across cells.

Spectral descriptors are computed on native images; CNN inputs are resized to 224×224 . Appendix C.1 details preprocessing, mapping, and normalization.

4.3 Baseline Comparisons

External methods provide performance context rather than form part of the controlled GE2I comparison. We include XGBoost on PCA features (Chen, 2016), Seurat (Hao et al., 2021), scANVI with and without scArches (Lotfollahi et al., 2020), and the foundation models scGPT (Cui et al., 2024) and CellFM (Zeng et al., 2025). These models retain their respective input and training protocols; the foundation models are evaluated through zero-shot embeddings. A tuned MLP on raw expression vectors provides an additional non-image reference in the appendix. Implementation details appear in Appendix C.3.

4.4 Training Protocol

All main GE2I comparisons use the same ResNet-18 backbone (He et al., 2016) and training hyperparameters, without mapping-specific network tuning. We train with AdamW (Loshchilov & Hutter, 2017), linear warmup, and cosine annealing, and report means and standard deviations across three random seeds. Hyperparameters are provided in Appendix C.4. Mapping construction is timed separately with a six-hour fitting budget. Additional experiments assess assignment effects across visual backbones and examine ImageNet pretraining and image augmentation.

5 Results

We first evaluate prediction and mapping cost, then examine image properties and pixel assignment, and finally study visual backbones, pretraining, augmentation, and spatial interpretation.

5.1 How Well Do Gene Expression Images Support Cell-Type Prediction?

Gene expression images support competitive cell-type prediction. Under the shared ResNet-18 setting, several GE2I representations perform comparably to established latent-variable methods, including MR-DeepInsight on the Immune Cell Atlas (Table 1). Single-view mappings also provide useful predictions, although scGPT remains stronger on several HVG datasets. External methods retain their recommended protocols, so this comparison evaluates the viability of GE2I rather than ranking fully optimized pipelines.

Full-transcriptome inputs make smoothness and gene-collision differences more pronounced than HVG inputs. For the main mechanistic analysis, we use all genes retained after preprocessing, avoiding dependence on a particular HVG selection rule. Each mapping must organize more expression values while the downstream CNN and input resolution remain fixed, making some construction characteristics more pronounced. Figure 2 shows how spectral properties change with gene scope; complete measurements appear in Appendix Table 14.

Table 2 reports full-transcriptome prediction, and Figure 3 reports construction time. DeepInsight and Fotomics fit within minutes to hours; IGTD timed out on all datasets,

Table 1: Cell-type prediction in the HVG setting. GE2I methods use a shared ResNet-18 and training configuration. External methods follow their recommended implementations and settings. Entries report means and standard deviations over three seeds.

Method	DKD		GTEX v9		HypoMap		Immune Cell Atlas		Tabula Sapiens	
	ACC	F1-M	ACC	F1-M	ACC	F1-M	ACC	F1-M	ACC	F1-M
XGBoost	0.8646 $\pm$ 0.01	0.6860 $\pm$ 0.01	0.4509 $\pm$ 0.04	0.1249 $\pm$ 0.04	0.9661 $\pm$ 0.00	0.3933 $\pm$ 0.19	0.6040 $\pm$ 0.03	0.4350 $\pm$ 0.02	0.2387 $\pm$ 0.08	0.0925 $\pm$ 0.03
scANVI	0.9107 $\pm$ 0.01	0.8082 $\pm$ 0.03	0.5608 $\pm$ 0.05	0.1694 $\pm$ 0.06	0.9863 $\pm$ 0.00	0.5061 $\pm$ 0.23	0.7529 $\pm$ 0.02	0.6207 $\pm$ 0.02	0.3856 $\pm$ 0.13	0.1484 $\pm$ 0.06
scANVI (scArches)	0.9242 $\pm$ 0.00	0.8424 $\pm$ 0.02	0.5691 $\pm$ 0.05	0.1539 $\pm$ 0.03	0.9866 $\pm$ 0.00	0.5105 $\pm$ 0.23	0.7606 $\pm$ 0.02	0.6468 $\pm$ 0.01	0.3236 $\pm$ 0.02	0.1594 $\pm$ 0.07
Seurat	0.8648 $\pm$ 0.00	0.5908 $\pm$ 0.03	0.5308 $\pm$ 0.07	0.1305 $\pm$ 0.05	0.9843 $\pm$ 0.00	0.4782 $\pm$ 0.23	0.5458 $\pm$ 0.02	0.4208 $\pm$ 0.02	0.3154 $\pm$ 0.10	0.1287 $\pm$ 0.05
scGPT	0.9421 $\pm$ 0.01	0.8286 $\pm$ 0.02	0.7588 $\pm$ 0.01	0.2896 $\pm$ 0.07	-	-	0.7713 $\pm$ 0.02	0.7050 $\pm$ 0.01	0.4192 $\pm$ 0.16	0.2123 $\pm$ 0.03
CellFM	0.6594 $\pm$ 0.09	0.3744 $\pm$ 0.03	0.4016 $\pm$ 0.12	0.1143 $\pm$ 0.05	0.6810 $\pm$ 0.16	0.0852 $\pm$ 0.00	0.4088 $\pm$ 0.01	0.2101 $\pm$ 0.01	0.0027 $\pm$ 0.00	0.0011 $\pm$ 0.00
Random (HVG)	0.8996 $\pm$ 0.00	0.7734 $\pm$ 0.01	0.6530 $\pm$ 0.05	0.2106 $\pm$ 0.01	0.9795 $\pm$ 0.01	0.5571 $\pm$ 0.28	0.6828 $\pm$ 0.03	0.4842 $\pm$ 0.02	0.3837 $\pm$ 0.08	0.2147 $\pm$ 0.07
DeepInsight (t-SNE) (HVG)	0.8924 $\pm$ 0.00	0.7424 $\pm$ 0.01	0.6262 $\pm$ 0.01	0.1943 $\pm$ 0.01	0.9833 $\pm$ 0.00	0.5720 $\pm$ 0.22	0.6928 $\pm$ 0.03	0.5040 $\pm$ 0.01	0.3711 $\pm$ 0.08	0.1844 $\pm$ 0.04
DeepInsight (UMAP) (HVG)	0.9000 $\pm$ 0.00	0.7587 $\pm$ 0.02	0.6765 $\pm$ 0.04	0.2209 $\pm$ 0.01	0.9811 $\pm$ 0.01	0.5783 $\pm$ 0.25	0.7045 $\pm$ 0.02	0.5100 $\pm$ 0.00	0.3919 $\pm$ 0.08	0.1967 $\pm$ 0.06
DeepInsight (PCA) (HVG)	0.9029 $\pm$ 0.00	0.7808 $\pm$ 0.01	0.6230 $\pm$ 0.02	0.1759 $\pm$ 0.00	0.9833 $\pm$ 0.00	0.5723 $\pm$ 0.18	0.6858 $\pm$ 0.03	0.4810 $\pm$ 0.01	0.3742 $\pm$ 0.06	0.2016 $\pm$ 0.07
GenoMAP (HVG)	0.8808 $\pm$ 0.03	0.7717 $\pm$ 0.02	0.6439 $\pm$ 0.04	0.2073 $\pm$ 0.02	0.9766 $\pm$ 0.01	0.5422 $\pm$ 0.14	0.6683 $\pm$ 0.02	0.4738 $\pm$ 0.00	0.3611 $\pm$ 0.09	0.2097 $\pm$ 0.07
IGTD (HVG)	0.9022 $\pm$ 0.00	0.7328 $\pm$ 0.04	0.6660 $\pm$ 0.04	0.2112 $\pm$ 0.03	0.9836 $\pm$ 0.00	0.5390 $\pm$ 0.21	0.6900 $\pm$ 0.02	0.4998 $\pm$ 0.01	0.3928 $\pm$ 0.08	0.2067 $\pm$ 0.06
Fotomics (HVG)	0.9015 $\pm$ 0.00	0.7424 $\pm$ 0.02	0.6577 $\pm$ 0.02	0.2007 $\pm$ 0.00	0.9814 $\pm$ 0.01	0.5200 $\pm$ 0.23	0.7070 $\pm$ 0.03	0.5104 $\pm$ 0.00	0.3830 $\pm$ 0.08	0.2086 $\pm$ 0.07
Refined (HVG)	0.9023 $\pm$ 0.01	0.7527 $\pm$ 0.01	0.6369 $\pm$ 0.07	0.1936 $\pm$ 0.02	0.9810 $\pm$ 0.00	0.5595 $\pm$ 0.18	0.6627 $\pm$ 0.00	0.4840 $\pm$ 0.01	0.3725 $\pm$ 0.05	0.2154 $\pm$ 0.07
MR-DeepInsight (HVG)	0.9085 $\pm$ 0.01	0.7967 $\pm$ 0.02	0.6842 $\pm$ 0.04	0.1989 $\pm$ 0.05	0.9922 $\pm$ 0.01	0.5932 $\pm$ 0.24	0.7482 $\pm$ 0.02	0.6396 $\pm$ 0.02	0.3872 $\pm$ 0.08	0.2097 $\pm$ 0.10

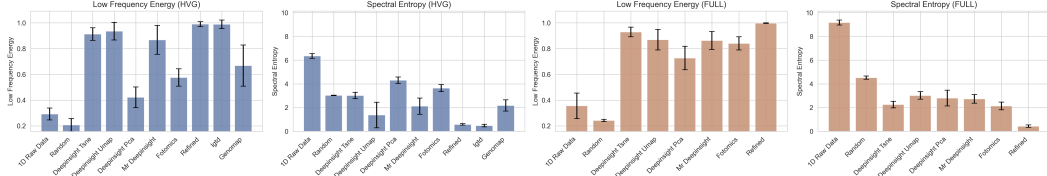


Figure 2: Spectral properties of GE2I under HVG and full-transcriptome inputs. LFER and SE characterize the native images produced by each mapping. Several structured mappings concentrate more power at low spatial frequencies than Random Mapping.

Table 2: Full-transcriptome prediction. Absolute accuracy and F1-Macro are reported under the shared GE2I learning configuration. scGPT is an external reference; the raw-expression MLP is reported in Appendix Table 9. T/O denotes a fit exceeding the six-hour construction budget; dashes denote unreported results.

Method	DKD		GTEX v9		HypoMap		Immune Cell Atlas		Tabula Sapiens	
	Acc	F1-M	Acc	F1-M	Acc	F1-M	Acc	F1-M	Acc	F1-M
scGPT	0.9462 $\pm$ 0.00	0.8553 $\pm$ 0.02	0.8091 $\pm$ 0.02	0.3415 $\pm$ 0.09	-	-	0.8061 $\pm$ 0.03	0.7462 $\pm$ 0.00	0.3930 $\pm$ 0.15	0.1977 $\pm$ 0.06
Random	0.9421 $\pm$ 0.00	0.8053 $\pm$ 0.01	0.8489 $\pm$ 0.03	0.4601 $\pm$ 0.03	0.9901 $\pm$ 0.00	0.6553 $\pm$ 0.23	0.8785 $\pm$ 0.00	0.7305 $\pm$ 0.01	0.4569 $\pm$ 0.10	0.3453 $\pm$ 0.11
DeepInsight (t-SNE)	0.9537 $\pm$ 0.00	0.8868 $\pm$ 0.02	0.8548 $\pm$ 0.04	0.5064 $\pm$ 0.04	0.9918 $\pm$ 0.00	0.7487 $\pm$ 0.17	0.8766 $\pm$ 0.01	0.7480 $\pm$ 0.01	0.4270 $\pm$ 0.13	0.2467 $\pm$ 0.02
DeepInsight (UMAP)	0.9572 $\pm$ 0.00	0.8924 $\pm$ 0.00	0.8462 $\pm$ 0.04	0.4885 $\pm$ 0.03	0.9935 $\pm$ 0.00	0.7713 $\pm$ 0.15	0.8820 $\pm$ 0.00	0.7625 $\pm$ 0.02	0.4275 $\pm$ 0.21	0.2672 $\pm$ 0.06
DeepInsight (PCA)	0.9575 $\pm$ 0.00	0.8458 $\pm$ 0.00	0.8416 $\pm$ 0.03	0.4415 $\pm$ 0.01	0.9926 $\pm$ 0.00	0.7276 $\pm$ 0.19	0.8716 $\pm$ 0.01	0.7483 $\pm$ 0.03	0.4423 $\pm$ 0.09	0.3054 $\pm$ 0.11
MR-DeepInsight	0.9700 $\pm$ 0.00	0.9271 $\pm$ 0.01	0.8715 $\pm$ 0.04	0.3656 $\pm$ 0.05	0.9994 $\pm$ 0.00	0.8021 $\pm$ 0.13	0.8990 $\pm$ 0.01	0.8224 $\pm$ 0.00	0.4097 $\pm$ 0.16	0.2525 $\pm$ 0.06
Fotomics	0.9543 $\pm$ 0.00	0.8518 $\pm$ 0.01	0.8327 $\pm$ 0.03	0.4184 $\pm$ 0.03	0.9913 $\pm$ 0.00	0.7191 $\pm$ 0.19	0.8613 $\pm$ 0.01	0.7195 $\pm$ 0.01	0.4601 $\pm$ 0.13	0.2947 $\pm$ 0.05
Refined	0.9566 $\pm$ 0.00	0.8704 $\pm$ 0.02	0.8467 $\pm$ 0.05	0.4875 $\pm$ 0.04	0.9920 $\pm$ 0.00	0.7274 $\pm$ 0.18	0.8827 $\pm$ 0.00	0.7302 $\pm$ 0.01	T/O	T/O

Genomap on all but DKD, and REFINED on Tabula Sapiens. Although the shared CNN processes images at the same resolution, mapping construction remains a separate computational constraint. We also evaluate GE2I on CITE-Seq gene-to-protein prediction (Appendix D).

No layout performs best across datasets. DeepInsight (t-SNE) outperforms Random Mapping in F1-Macro on GTEX v9, whereas Random Mapping achieves the highest mean F1-Macro among the reported GE2I methods on Tabula Sapiens. A tuned expression-space MLP is competitive on DKD and the Immune Cell Atlas, while structured GE2I methods achieve higher F1-Macro on GTEX v9 and HypoMap (Appendix Table 9). These differences motivate examining image properties and their relation to learning, rather than inferring the value of spatial organization from final scores alone.

5.2 How Do Image Properties Relate to Optimization and Prediction?

Smoother images optimize faster, but preserving separate gene values is also important for final prediction. We characterize these two effects using LFER and SE for spatial structure and APPG and CR for gene-level allocation. Appendix Figures 6 and 7 show the resulting image patterns, and Appendix Table 14 provides the complete measurements.

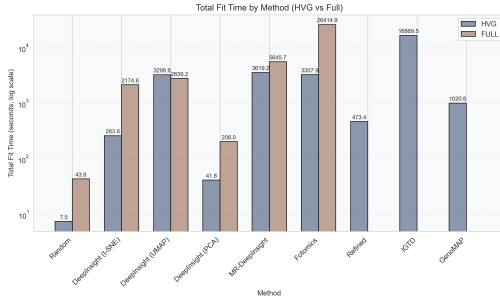


Figure 3: Mapping construction cost. Runtime to fit $\mathcal{M}$, aggregated across datasets for the plotted methods. Iterative strategies such as REFINED and Genomap are excluded from this summary because they exceeded the fitting budget on large benchmarks.

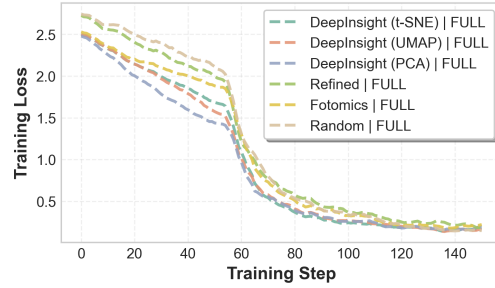


Figure 4: Early optimization on full-transcriptome DKD. Training-loss curves over the first 60 steps under the shared CNN. Structured mappings show faster early loss reduction than Random Mapping.

Structured mappings produce smoother, lower-frequency images than random placement. DeepInsight (t-SNE/UMAP) and REFINED have full-transcriptome LFER values of approximately 0.8–1.0, whereas Random Mapping remains near 0.24 (Figure 2). Placing similarly expressed genes nearby produces coherent regions with gradually varying intensity. Random placement yields more irregular neighboring values and a broader frequency distribution. The CNN’s local filters receive smoother patterns under structured layouts and more abrupt spatial changes under random placement.

Low-frequency structure accelerates early optimization but does not ensure better final prediction. On DKD, high-LFER mappings such as DeepInsight reduce loss faster over the first 60 training steps than Random Mapping (Figure 4). This is consistent with the preference for low-frequency patterns described in work on neural-network spectral bias (Rahaman et al., 2019). Spatial organization can therefore aid optimization by presenting expression values in a form the CNN learns more readily. However, early loss reduction does not establish final predictive performance, motivating analysis of how individual gene values are represented.

Joint regression links gene mixing to worse prediction and organized spatial structure to better prediction. We fit beta regressions relating APPG, CR, LFER, and SE jointly to accuracy and F1-Macro in the Full setting (Table 3). Dataset-by-seed fixed effects account for baseline performance differences, while clustered standard errors account for repeated seeds within each dataset–method pair. Both outcomes show the same pattern: APPG and CR have negative coefficients, whereas LFER and SE have positive coefficients, with all eight associations remaining significant after multiple-testing adjustment. These results connect the two sides of the trade-off: gene mixing is associated with lower prediction quality, while organized spatial structure is associated with higher prediction quality, consistent with the representation-level mechanism derived in Appendix A.

More gene mixing is associated with worse prediction. For F1-Macro, CR has a coefficient of -3.349 ($q = 0.001$). When several genes share one pixel, their distinct expression values are replaced by a shared average, reducing gene-level separation.

Organized low-frequency structure is associated with better prediction. LFER has a coefficient of $+5.334$ ($q = 0.003$), consistent with the smoother layouts and faster early optimization observed above. SE is also positive ($+0.995$, $q = 0.003$). Together, these results favor images that are spatially smooth but still contain variation across scales, rather than images that are irregular or nearly uniform.

Using more pixels alone does not guarantee better prediction. APPG has a coefficient of -0.966 ($q = 0.041$) after holding CR and spectral descriptors fixed. Performance depends on which genes share pixels and how the values are arranged. Accuracy shows the same

Table 3: Beta regression on full-transcriptome performance. Descriptors enter jointly with dataset-by-seed fixed effects, and SE is clustered by dataset-method. The p -values use a t distribution with 34 degrees of freedom; q denotes the Benjamini-Hochberg-adjusted p -value across the 16 descriptor tests. Coefficients are on the logit-mean scale.

Target	Descriptor	Coef.	Cluster SE	z^*	p	q
Accuracy	APPG	-1.044	0.358	-2.92	0.006	0.011
	CR	-3.251	0.617	-5.27	< 0.001	< 0.001
	LFER	4.954	1.367	3.62	< 0.001	0.003
	SE	0.933	0.265	3.52	0.001	0.003
F1-Macro	APPG	-0.966	0.414	-2.33	0.026	0.041
	CR	-3.349	0.791	-4.23	< 0.001	0.001
	LFER	5.334	1.485	3.59	0.001	0.003
	SE	0.995	0.282	3.53	0.001	0.003

Table 4: Image-size and pixel-assignment experiments on DKD. (a) Prediction and allocation descriptors across the reported image sizes. (b) LSA results, with superscripts indicating relative changes from the original assignment.

(a) Image-size ablation					(b) Efficacy of linear sum assignment (LSA)				
Size	Acc	F1-M	APPG	CR	Method	LFER	CR	Acc	F1-M
64 ²	0.9608	0.8990	0.480	0.493	DeepInsight (t-SNE)	0.90 ^{-4%}	0.00 ^{-100%}	0.97 ^{+1%}	0.91 ^{+3%}
128 ²	0.9570	0.8913	0.489	0.481	DeepInsight (UMAP)	0.84 ^{-7%}	0.00 ^{-100%}	0.97 ^{+1%}	0.91 ^{+2%}
224 ²	0.9556	0.8915	0.486	0.485					
384 ²	0.9597	0.9013	0.478	0.495					

directions. Coefficient magnitudes are on the logit-mean scale; Appendix G reports the HVG results and diagnostics.

Random layouts can compete by preserving gene identity in separate pixels. Random Mapping forgoes similarity-based neighborhoods but preserves gene identity through fixed, separate pixels across cells. This helps explain its competitive performance against smoother mappings that merge expression values. Broader gene coverage accentuates this distinction: for DeepInsight (t-SNE) on HypoMap, CR rises from 0.198 with HVGs to 0.641 with full-transcriptome inputs. IGTD and REFINED show that spatial organization and collision-free allocation can coexist. If aggregation limits a structured mapping’s benefit, separating genes should improve prediction even when the image becomes less smooth.

5.3 Can Changing Pixel Assignment Improve Prediction?

Yes. Assigning each gene to a unique pixel improves prediction even when image smoothness decreases. Across the tested DeepInsight mappings and datasets, removing collisions increases F1-Macro while LFER decreases. In contrast, simply enlarging the image leaves substantial gene overlap and provides little improvement.

Larger canvases neither remove gene collisions nor meaningfully improve prediction. In the DKD ablation (Table 4a), increasing image dimensions from 64² to 384² changes F1-Macro from 0.8990 to 0.9013, with lower scores at intermediate sizes. CR remains near 0.49 throughout. Larger images thus provide little improvement while substantial gene overlap persists, motivating a change to pixel allocation with the input genes and downstream learner fixed.

One-to-one assignment improves prediction despite reducing image smoothness. Linear sum assignment (LSA) (Burkard & Derigs, 1980) gives each gene a distinct pixel. On DKD, it eliminates native-image collisions and improves F1-Macro for DeepInsight (t-SNE) and DeepInsight (UMAP) by approximately 3% and 2%, respectively, while LFER decreases by 4-7% (Table 4b). Consistent with the prediction from Section 5.2, separating gene values

improves prediction despite lower image smoothness: the benefit of revised allocation can outweigh reduced low-frequency concentration.

Greedy unique assignment removes collisions and improves all six mapping-dataset comparisons. We introduce a greedy assignment procedure that avoids a dense gene-pixel cost matrix. Genes are processed in a fixed order, each placed at the nearest available pixel to its rounded target coordinate, with an occupancy mask preventing reuse. This preserves projected coordinates as placement targets while enforcing distinct locations (Appendix C.2).

Table 5 reports reassignment results for DeepInsight (t-SNE) and DeepInsight (UMAP) on DKD, GTEx v9, and HypoMap. All six comparisons improve mean F1-Macro, with CR = 0 and APPG = 1 on reassigned native images. For example, DeepInsight (t-SNE) on GTEx v9 improves from approximately 0.506 to 0.556, while LFER falls from 0.867 to 0.756. LFER decreases in every comparison, so low-frequency concentration does not explain the gains.

These interventions support reducing unnecessary gene aggregation while using the projected layout to guide placement. Reassignment changes both gene separation and local spatial patterns, and this combination improves prediction despite lower smoothness across the tested mappings. We next examine whether the benefit extends beyond ResNet-18 and how other visual training choices interact with GE2I.

Table 5: Greedy unique assignment across datasets. Results use ResNet-18. Superscripts report relative changes from conventional binning for the corresponding dataset and mapping. The algorithm is described in Appendix C.2.

Dataset	Mapping	APPG	CR	LFER	SE	Acc	F1-M
DKD	DeepInsight (t-SNE)	1.0000 ^{+104.8%} ±0.00	0.0000 ^{-100.0%} ±0.00	0.8021 ^{-15.0%} ±0.00	2.3435 ^{+19.7%} ±0.00	0.9679 ^{+1.5%} ±0.00	0.9047 ^{+2.0%} ±0.00
DKD	DeepInsight (UMAP)	1.0000 ^{+84.8%} ±0.00	0.0000 ^{-100.0%} ±0.00	0.7811 ^{-13.5%} ±0.00	2.4673 ^{-10.0%} ±0.00	0.9667 ^{+1.0%} ±0.00	0.9006 ^{+0.9%} ±0.02
GTEx v9	DeepInsight (t-SNE)	1.0000 ^{+142.0%} ±0.00	0.0000 ^{-100.0%} ±0.00	0.7560 ^{-12.8%} ±0.04	2.7173 ^{+2.8%} ±0.30	0.8630 ^{+0.0%} ±0.03	0.5557 ^{+9.7%} ±0.02
GTEx v9	DeepInsight (UMAP)	1.0000 ^{+271.3%} ±0.00	0.0000 ^{-100.0%} ±0.00	0.7374 ^{-8.0%} ±0.05	2.8200 ^{-12.0%} ±0.32	0.8495 ^{+0.4%} ±0.04	0.5357 ^{+0.0%} ±0.05
HypoMap	DeepInsight (t-SNE)	1.0000 ^{+229.2%} ±0.00	0.0000 ^{-100.0%} ±0.00	0.8602 ^{-9.4%} ±0.00	2.0356 ^{-7.7%} ±0.07	0.9940 ^{+0.0%} ±0.00	0.7945 ^{+0.1%} ±0.13
HypoMap	DeepInsight (UMAP)	1.0000 ^{+359.5%} ±0.00	0.0000 ^{-100.0%} ±0.00	0.8085 ^{-11.7%} ±0.05	2.2516 ^{-21.8%} ±0.34	0.9938 ^{+0.0%} ±0.00	0.8137 ^{+5.5%} ±0.13

5.4 How Do Visual Backbones and Training Choices Affect GE2I?

Unique pixel assignment helps across backbones, whereas pretraining and augmentation depend on the architecture and mapping. On DKD, reassignment improves mean F1-Macro for ResNet-18, ResNet-50, ConvNeXt-Tiny, EfficientNet-B0, and ViT-B/16 by 3.4%–7.0% (Appendix Table 10). The consistent gains across convolutional and transformer models show that the allocation benefit is not specific to ResNet-18. Appendix Table 11 provides the corresponding Random Mapping results.

ImageNet pretraining accelerates early ResNet-18 learning for structured images, but not for random layouts. ImageNet initialization (Deng et al., 2009) has different effects across mappings and architectures. In the ResNet-18 training curves on DKD, pretrained weights accelerate early learning for structured DeepInsight images, whereas Random Mapping shows no clear acceleration (Figure 5). This contrast is consistent with pretrained visual features being better suited to the smoother spatial patterns of structured layouts.

Final prediction depends on the backbone (Table 6). ImageNet pretraining improves mean F1-Macro for ViT-B/16 and ResNet-50 under both DeepInsight (t-SNE) and REFINED, whereas ResNet-18 performs better with random initialization. Thus, some learners benefit from pretraining even when the smaller CNN learns effectively from expression images alone. Together with the training curves, these results indicate architecture- and mapping-dependent benefits; early acceleration alone does not establish a final predictive advantage.

Augmentation: geometry versus intensity. On DKD, geometric transformations lower mean F1-Macro for DeepInsight (t-SNE) and REFINED but improve Random Mapping (Appendix Table 12). Because these transformations change the locations associated with gene identities, their suitability depends on the mapping and cannot be inferred solely from their usefulness for natural images.

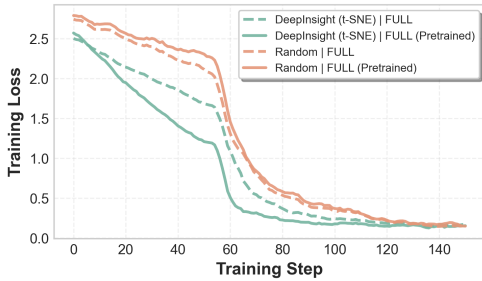


Figure 5: ImageNet initialization and early learning on DKD. ResNet-18 training curves with and without pretrained weights. Structured DeepInsight images benefit in early optimization, whereas Random Mapping shows no clear acceleration.

Table 6: Backbones and pretraining on DKD. Accuracy and F1-Macro with ImageNet initialization and random initialization for two structured mappings.

Method	Backbone	Pretrained	ACC	F1-Macro
DeepInsight (t-SNE)	ResNet-18	Yes	0.9479 $\pm$ 0.01	0.8767 $\pm$ 0.01
		No	0.9537 $\pm$ 0.00	0.8868 $\pm$ 0.02
	ResNet-50	Yes	0.9512 $\pm$ 0.00	0.8915 $\pm$ 0.02
		No	0.9457 $\pm$ 0.01	0.8679 $\pm$ 0.01
	ViT-B/16	Yes	0.9510 $\pm$ 0.00	0.8659 $\pm$ 0.03
		No	0.9401 $\pm$ 0.01	0.8252 $\pm$ 0.01
Refined	ResNet-18	Yes	0.9504 $\pm$ 0.01	0.8392 $\pm$ 0.02
		No	0.9566 $\pm$ 0.00	0.8704 $\pm$ 0.03
	ResNet-50	Yes	0.9452 $\pm$ 0.01	0.8266 $\pm$ 0.02
		No	0.9468 $\pm$ 0.00	0.8159 $\pm$ 0.02
	ViT-B/16	Yes	0.9529 $\pm$ 0.00	0.8460 $\pm$ 0.01
		No	0.9505 $\pm$ 0.01	0.8294 $\pm$ 0.04

Intensity perturbations vary brightness and contrast without changing gene coordinates and improve mean F1-Macro for all three mappings. Stronger perturbation is not uniformly better: combining all augmentations lowers F1-Macro for DeepInsight (t-SNE) and REFINED to 0.3227 and 0.1941, respectively, whereas Random Mapping improves under every tested setting. Intensity perturbation is thus the most consistent choice in this comparison; geometric and combined augmentations require mapping-specific evaluation.

Spatial interpretation beyond prediction. Competitive prediction from random layouts does not make spatial organization irrelevant to interpretation. On DKD, DeepInsight (t-SNE) has a Grad-CAM marker hit ratio of 0.50, versus 0.25 for Random Mapping (Appendix Table 13). Its salient regions contain five of five listed markers for distal convoluted tubule cells and podocytes, versus two of five for Random Mapping (Appendix Figure 8). These more marker-aligned regions support gene-level inspection in this analysis.

6 Conclusion and Limitations

Under controlled training conditions, GE2I exhibits a trade-off between spatial smoothness and gene-level fidelity. Structured mappings produce smooth patterns that can ease early CNN optimization, but assigning multiple genes to the same pixel may obscure expression differences and reduce final predictive performance. Our pixel-assignment experiments show that preserving gene-level information improves prediction even when the images become less smooth. Transcriptomic image design should therefore balance spatial organization with information preservation rather than treat smoothness or early optimization as a proxy for prediction quality.

Limitations and Future Directions. Our experiments focus mainly on cell-type prediction with a shared CNN and training protocol. This setting isolates representation-level effects but does not measure the best performance attainable through method-specific tuning. Our pixel-level descriptors also do not capture higher-order relationships among genes. Further work should learn adaptive allocations that balance spatial organization, gene-level fidelity, and interpretability, and evaluate them on larger multimodal datasets and with foundation models.

AI Use Statement

Generative AI tools were used to support research ideation, literature discovery, manuscript drafting, and language polishing. AI-assisted suggestions were reviewed against the experimental results and relevant literature. All experiments, numerical results, and statistical analyses were conducted and verified by the authors. The authors take full responsibility for the final content of this work.

Ethics Statement

This study uses previously released single-cell transcriptomic datasets and does not involve the collection of new human-subject data. The analyses are intended for computational genomics research rather than clinical diagnosis or treatment.

Reproducibility Statement

Experimental settings, preprocessing procedures, mapping parameters, training configurations, statistical analyses, and pixel-assignment methods are described in the main text and appendix. Results are reported across three random seeds where applicable. Source code will be released upon publication.

References

- Omid Bazgir, Ruibo Zhang, Saugato Rahman Dhruba, Raziur Rahman, Souparno Ghosh, and Ranadip Pal. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nature communications*, 11(1):4391, 2020.
- Charlotte Bunne, Yusuf Roohani, Yanay Rosen, Ankit Gupta, Xikun Zhang, Marcel Roed, Theo Alexandrov, Mohammed AlQuraishi, Patricia Brennan, Daniel B Burkhardt, et al. How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell*, 187(25):7045–7063, 2024.
- Rainer E Burkard and Ulrich Derigs. The linear sum assignment problem. In *Assignment and Matching Problems: Solution Methods with FORTRAN-Programs*, pp. 1–15. Springer, 1980.
- Yun-Hao Cao and Jianxin Wu. A random cnn sees objects: One inductive bias of cnn and its applications. In *Proceedings Of The AAI conference on artificial intelligence*, volume 36, pp. 194–202, 2022.
- Tianqi Chen. Xgboost: A scalable tree boosting system. Cornell University, 2016.
- The Tabula Sapiens Consortium*, Robert C Jones, Jim Karkanias, Mark A Krasnow, Angela Oliveira Pisco, Stephen R Quake, Julia Salzman, Nir Yosef, Bryan Bulthaupt, Phillip Brown, et al. The tabula sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, 376(6594):eabl4896, 2022.
- Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.
- C Domínguez Conde, Chao Xu, Louie B Jarvis, Daniel B Rainbow, Sara B Wells, Tamir Gomes, SK Howlett, O Suchanek, K Polanski, HW King, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594):eabl5197, 2022.
- Gökçen Eraslan, Eugene Drokhlyansky, Shankara Anand, Evgenij Fiskin, Ayshwarya Subramanian, Michal Slyper, Jiali Wang, Nicholas Van Wittenberghe, John M Rouhana, Julia Waldman, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*, 376(6594):eabl4290, 2022.
- Silvia Ferrari and Francisco Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815, 2004.

- Christopher Heje Grønbech, Maximillian Fornitz Vording, Pascal N Timshel, Casper Kaae Sønderby, Tune H Pers, and Ole Winther. scvae: variational auto-encoders for single-cell gene expression data. *Bioinformatics*, 36(16):4415–4422, 2020.
- Ihuan Gunawan, Fatemeh Vafaei, Erik Meijering, and John George Lock. An introduction to representation learning for single-cell data analysis. *Cell Reports Methods*, 3(8), 2023.
- Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- Md Tauhidul Islam and Lei Xing. Cartography of genomic interactions enables deep analysis of single-cell expression data. *Nature Communications*, 14(1):679, 2023.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- Mohammad Lotfollahi, Mohsen Naghipourfar, Malte D Luecken, Matin Khajavi, Maren Büttner, Ziga Avsec, Alexander V Misharin, and Fabian J Theis. Query to reference single-cell integration with transfer learning. *bioRxiv*, pp. 2020–07, 2020.
- Malte D Luecken, Scott Gigante, Daniel B Burkhardt, Robrecht Cannoodt, Daniel C Strobl, Nikolay S Markov, Luke Zappia, Giovanni Palla, Wesley Lewis, Daniel Dimitrov, et al. Defining and benchmarking open problems in single-cell analysis. *Nature Biotechnology*, pp. 1–6, 2025.
- Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International conference on machine learning*, pp. 5301–5310. PMLR, 2019.
- Alok Sharma, Edwin Vans, Daichi Shigemizu, Keith A Boroevich, and Tatsuhiko Tsunoda. Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific reports*, 9(1):11399, 2019.
- Alok Sharma, Yosvany López, Shangru Jia, Artem Lysenko, Keith A Boroevich, and Tatsuhiko Tsunoda. Enhanced analysis of tabular data through multi-representation deepinsight. *Scientific Reports*, 14(1):12851, 2024.
- Lukas Steuernagel, Brian YH Lam, Paul Klemm, Georgina KC Dowsett, Corinna A Bauder, John A Tadross, Tamara Sotelo Hitschfeld, Almudena del Rio Martin, Weiye Chen, Alain J De Solis, et al. Hypomap—a unified single-cell gene expression atlas of the murine hypothalamus. *Nature metabolism*, 4(10):1402–1419, 2022.
- Marlon Stoekius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 14(9): 865–868, 2017.
- Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, et al. mrna-seq whole-transcriptome analysis of a single cell. *Nature methods*, 6(5):377–382, 2009.

- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- Parker C Wilson, Yoshiharu Muto, Haojia Wu, Anil Karihaloo, Sushrut S Waikar, and Benjamin D Humphreys. Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. *Nature communications*, 13(1):5253, 2022.
- Chenling Xu, Romain Lopez, Edouard Mehlman, Jeffrey Regier, Michael I Jordan, and Nir Yosef. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular systems biology*, 17(1):e9620, 2021.
- Fan Yang, Wenchuan Wang, Fang Wang, Yuan Fang, Duyu Tang, Junzhou Huang, Hui Lu, and Jianhua Yao. scbert as a large-scale pretrained deep language model for cell type annotation of single-cell rna-seq data. *Nature Machine Intelligence*, 4(10):852–866, 2022.
- Seid Miad Zandavi, Derong Liu, Vera Chung, Ali Anaissi, and Fatemeh Vafaei. Fotomics: fourier transform-based omics imager for deep learning-based cell-identity mapping using single-cell omics profiles. *Artificial Intelligence Review*, 56(7):7263–7278, 2023.
- Yuansong Zeng, Jiancong Xie, Ningyuan Shanguan, Zhuoyi Wei, Wenbing Li, Yun Su, Shuangyu Yang, Chengyang Zhang, Jinbo Zhang, Nan Fang, et al. Cellfm: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. *Nature Communications*, 16(1):4679, 2025.
- Yitan Zhu, Thomas Brettin, Fangfang Xia, Alexander Partin, Maulik Shukla, Hyunseung Yoo, Yvonne A Evrard, James H Doroshov, and Rick L Stevens. Converting tabular data into images for deep learning with convolutional neural networks. *Scientific reports*, 11(1):11325, 2021.

A A Theoretical View of the Smoothness–Fidelity Trade-off

The empirical analysis suggests that spatial smoothness can facilitate learning, while gene mixing can reduce predictive information. We illustrate how these two effects can arise from the same pixel-aggregation operation.

Consider a binary classification problem with $Y \in \{-1, +1\}$. For the k_p genes assigned to pixel p , let

$$\mathbf{X}_p \mid Y = y = y\boldsymbol{\mu}_p + \boldsymbol{\varepsilon}_p, \quad \boldsymbol{\varepsilon}_p \sim \mathcal{N}(0, \sigma^2 I_{k_p}), \quad (7)$$

where noise is independent across genes and pixels. Mean aggregation produces

$$Z_p = \frac{1}{k_p} \mathbf{1}^\top \mathbf{X}_p. \quad (8)$$

Let $\bar{\mu}_p = k_p^{-1} \mathbf{1}^\top \boldsymbol{\mu}_p$.

Proposition A.1 (Smoothness–fidelity effect of pixel aggregation). Mean aggregation has two simultaneous effects.

First, for neighboring pixels p and q ,

$$\mathbb{E}[(Z_p - Z_q)^2 \mid Y] = (\bar{\mu}_p - \bar{\mu}_q)^2 + \sigma^2 \left(\frac{1}{k_p} + \frac{1}{k_q} \right). \quad (9)$$

Thus, holding the underlying spatial signal fixed, averaging more genes suppresses noise-induced local variation.

Second, let D_X^2 and D_Z^2 denote the squared Mahalanobis distances between the two classes before and after aggregation. Then

$$D_X^2 = \frac{4}{\sigma^2} \sum_p \|\boldsymbol{\mu}_p\|_2^2, \quad (10)$$

$$D_Z^2 = \frac{4}{\sigma^2} \sum_p k_p \bar{\mu}_p^2, \quad (11)$$

and therefore

$$D_X^2 - D_Z^2 = \frac{4}{\sigma^2} \sum_p \|\boldsymbol{\mu}_p - \bar{\mu}_p \mathbf{1}\|_2^2 \geq 0. \quad (12)$$

Aggregation therefore removes exactly the class-discriminative components that distinguish genes within the same pixel.

Proof. For each pixel,

$$Z_p = y\bar{\mu}_p + \bar{\varepsilon}_p, \quad \text{Var}(\bar{\varepsilon}_p) = \frac{\sigma^2}{k_p}.$$

Independence across pixels immediately gives Eq. (9).

Before aggregation, the two class means differ by $2\boldsymbol{\mu}_p$ within pixel p , and the covariance is $\sigma^2 I$, giving

$$D_X^2 = \frac{4}{\sigma^2} \sum_p \|\boldsymbol{\mu}_p\|_2^2.$$

After aggregation, the class means at pixel p differ by $2\bar{\mu}_p$, while its variance is σ^2/k_p . Hence

$$D_Z^2 = \frac{4}{\sigma^2} \sum_p k_p \bar{\mu}_p^2.$$

Finally, decompose

$$\boldsymbol{\mu}_p = \bar{\mu}_p \mathbf{1} + \boldsymbol{\mu}_p^\perp, \quad \mathbf{1}^\top \boldsymbol{\mu}_p^\perp = 0.$$

Orthogonality gives

$$\|\boldsymbol{\mu}_p\|_2^2 = k_p \bar{\mu}_p^2 + \|\boldsymbol{\mu}_p^\perp\|_2^2,$$

which yields Eq. (12). $\square$

The proposition provides a simple explanation for the empirical smoothness–fidelity trade-off. Averaging reduces gene-specific fluctuations, thereby lowering local spatial variation. On a regular grid, the sum of neighboring squared differences is the discrete Dirichlet energy, which penalizes rapidly varying spatial components and therefore connects this effect to the low-frequency structure measured in the main analysis. At the same time, averaging removes within-pixel gene contrasts that can carry class-discriminative information.

This also clarifies why the two properties need not be fundamentally incompatible. If genes grouped within a pixel have the same class-dependent signal, i.e., $\mu_p = \bar{\mu}_p \mathbf{1}$, then Eq. (12) is zero even though averaging still reduces noise-induced spatial variation. More generally, spatial smoothness can also be obtained by placing similarly behaving genes nearby without merging their values.

The result corresponds directly to the two dominant effects observed in the joint regression analysis: greater low-frequency organization is associated with better prediction, whereas greater gene mixing is associated with worse prediction. The theory does not prescribe regression coefficients; rather, it shows how these opposite associations can arise from distinct consequences of the same image-construction process.

B Datasets and Evaluation Benchmarks

All benchmarks use data from the OpenProblems in Single-Cell Analysis platform with standardized preprocessing (Luecken et al., 2025).

Table 7: Overview of single-cell datasets. Imbalance denotes the ratio between the majority and minority class sizes. The NeurIPS 2021 dataset is used for cross-modality prediction (regression), where the target dimension corresponds to the number of surface proteins.

Dataset	Samples	Genes	Classes/Targets	Batches	Imbalance
Task 1: Cell Type Classification					
Diabetic Kidney Disease	39,176	27,980	13	11	54.7
GTEX v9	209,126	28,094	53	16	956.0
Immune Cell Atlas	329,762	29,335	35	12	964.4
HypoMap	384,925	41,642	13	24	5,350.2
Tabula Sapiens	483,152	55,904	160	29	11,734.7
Task 2: Cross-Modality Prediction (Gene-to-Protein)					
NeurIPS 2021 CITE-Seq	8,617	36,280	13 (Proteins)	12	N/A

B.1 Dataset Specifications

The Diabetic Kidney Disease (DKD) dataset (Wilson et al., 2022) profiles human kidney cortex from healthy and diabetic donors, including subtypes such as injured proximal tubule cells. GTEX v9 (Eraslan et al., 2022) contains 209,126 nuclei from eight tissues and 16 donors, with technical variability and donor-specific batch effects. The Immune Cell Atlas (ICA) (Domínguez Conde et al., 2022) covers immune cells across 16 tissues, including tissue-resident and circulating populations. HypoMap (Steuernagel et al., 2022) provides a murine hypothalamus reference of approximately 385,000 cells with diverse neuronal subtypes. Tabula Sapiens (TS) (Consortium* et al., 2022) contains nearly 500,000 cells from 24 organs. Its class-imbalance ratio exceeds $10,000\times$, making rare-cell performance particularly important. Table 7 reports the ratio of the largest to smallest class for each dataset.

The NeurIPS 2021 CITE-Seq (GEX2ADT) dataset (Stoeckius et al., 2017) contains bone marrow mononuclear cells from 12 healthy donors. The task is to predict 13 surface-protein abundances from gene expression. Multiple collection sites per donor create a nested batch structure, requiring models to distinguish proteomic signal from technical variation.

Figures 6 and 7 compare DKD images by cell type (rows) and mapping (columns), using the viridis colormap for single-channel images. They show the smooth clusters of DeepInsight and the finer texture of Random Mapping examined in the spectral analysis.

B.2 Evaluation Metrics

For cell-type annotation, we report accuracy and F1-Macro. We prioritize F1-Macro because it weights each cell type equally, whereas accuracy can be dominated by common classes in these imbalanced atlases.

For gene-to-protein regression, we report root mean squared error (RMSE), global Pearson correlation, and Spearman correlation. RMSE measures prediction error, Pearson correlation captures linear association, and Spearman correlation measures rank association. The rank-based metric complements Pearson correlation when expression scales are nonlinear or affected by extreme values.

C Detailed Experimental Settings

We detail the mapping implementations, baseline configurations, and training protocol below.

C.1 Gene Expression-to-Image (GE2I) Transformation Protocols

We adapt the official method implementations to a shared pipeline.

Preprocessing and Algorithm Hyperparameters.

Data preprocessing followed method-specific protocols to maintain consistency with original implementations. In general, gene expression counts were log-normalized ($\log(1 + x)$) and globally scaled to the $[0, 1]$ interval. For the DeepInsight family, we followed the original DeepInsight-style preprocessing and used a MinMax scaler, which demonstrated greater stability for manifold projections.

For algorithm configurations, DeepInsight variants used perplexity=30 for t-SNE, the default setting for PCA, and $n_neighbors=15$ with $min_dist=0.1$ for UMAP, employing cosine metrics and mean-aggregation for pixel collisions. MR-DeepInsight applied identical parameters across parallel views ($n_jobs=8$) with majority voting. Fotomics utilized a LogScaler and FFT-based assignment with mean-aggregation.

For adaptive approaches targeting square grids with side length $S = \lceil \sqrt{G} \rceil$, Genomap applied Gromov-Wasserstein Optimal Transport with entropic regularization ($\epsilon = 0.1$, $num_iter=200$) to generate weighted pixel superpositions. IGTD used the Error Reduction Repositioning (ERR) algorithm with $max_step=30,000$ and Pearson correlation to enforce a strictly one-to-one mapping. REFINED employed MDS initialization followed by hill-climbing with nearest-neighbor slotting to resolve conflicts. Random Mapping provided a baseline one-to-one assignment on the adaptive grid without semantic priors.

Resolution Normalization and Spectral Analysis.

We compute SE and LFER on raw generator outputs at their native resolution ($S \times S$ for adaptive grids), before resizing. This avoids suppressing high-frequency structure, particularly in Random Mapping, through interpolation.

For CNN training, we resize maps to 224×224 using bilinear interpolation with antialiasing, then normalize each channel using training-set means and standard deviations. Spectral descriptors therefore characterize the native maps rather than the resized CNN inputs.

C.2 Collision-Aware Unique Assignment

For mappings that produce continuous 2D coordinates, we replace conventional binning with a greedy assignment to distinct pixels. Let $\mathbf{z}_i \in \mathbb{R}^2$ be the scaled coordinate of gene i , and

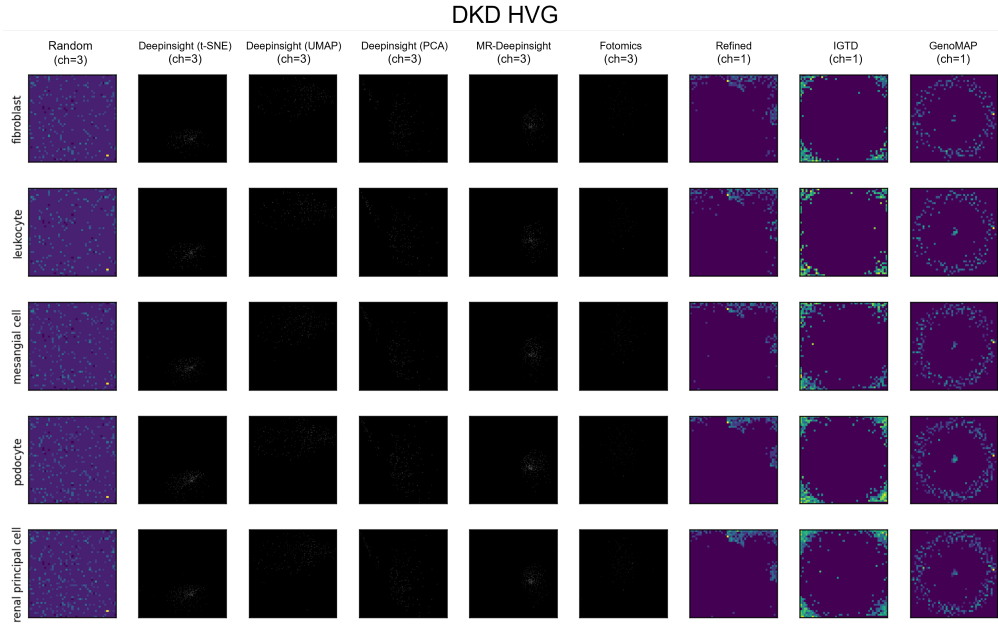


Figure 6: Visual comparison of gene expression-to-image mappings on DKD highly variable genes. Each row represents a specific cell type, while each column corresponds to a mapping strategy.

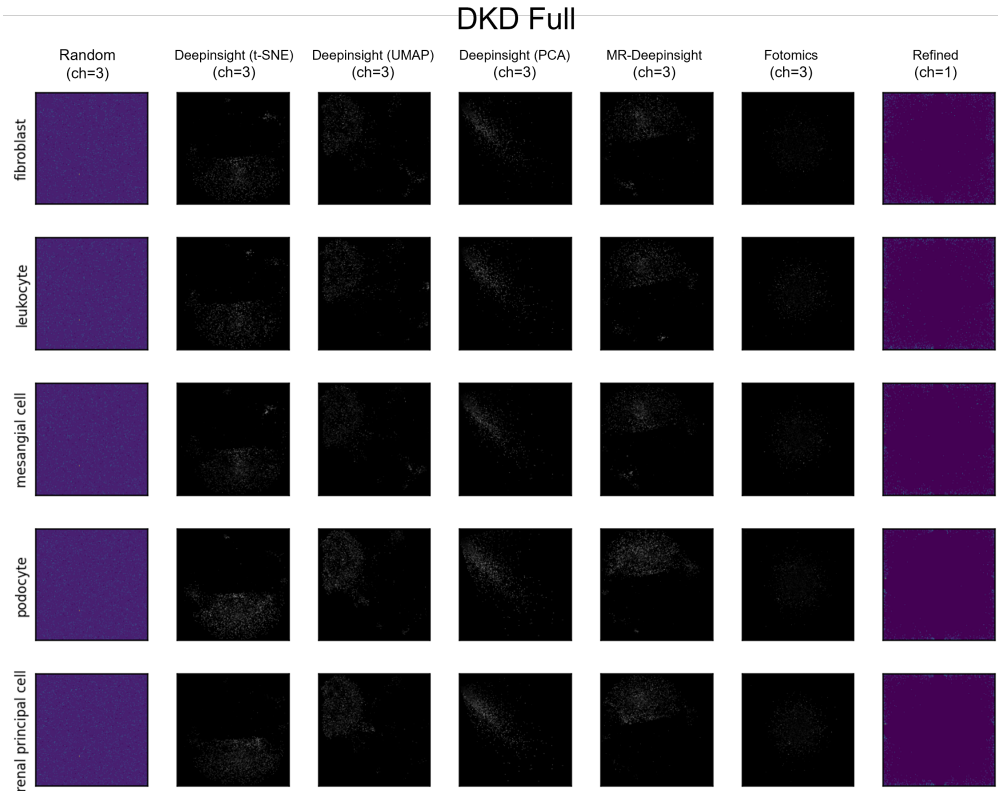


Figure 7: Visual comparison of gene expression-to-image mappings on DKD full genomes. Each row represents a specific cell type, while each column corresponds to a mapping strategy.

assume that the canvas has sufficient capacity, $|\Omega| = HW \geq G$. Rounding and boundary clipping define its target pixel:

$$\mathbf{t}_i = \Pi_{\Omega}(\text{round}(\mathbf{z}_i)), \quad (13)$$

where Π_{Ω} clips each coordinate to the valid grid range. The corresponding global assignment objective is

$$\min_{\mathbf{a}_1, \dots, \mathbf{a}_G \in \Omega} \sum_{i=1}^G \|\mathbf{a}_i - \mathbf{t}_i\|_2^2 \quad \text{subject to} \quad \mathbf{a}_i \neq \mathbf{a}_j \quad \text{for all } i \neq j. \quad (14)$$

Rather than solve this objective jointly, we process genes in a fixed order. Starting with $A_0 = \emptyset$, each gene is assigned to the nearest unoccupied pixel:

$$\mathbf{a}_i \in \arg \min_{p \in \Omega \setminus A_{i-1}} \|p - \mathbf{t}_i\|_2^2, \quad A_i = A_{i-1} \cup \{\mathbf{a}_i\}. \quad (15)$$

We build a cKDTree over the grid coordinates and maintain an occupancy mask. For each target, we query nearest neighbors with geometrically increasing candidate counts, $k = 1, 2, 4, \dots$, capped at $|\Omega|$. Occupied candidates are discarded; once free candidates are available, we select the closest one. If the adaptive search returns no valid candidate, an exhaustive search over the remaining free pixels provides a fallback. This avoids constructing a dense gene-pixel cost matrix, although searches can become expensive as the grid fills.

Provided $HW \geq G$, the procedure guarantees an injective assignment and hence $CR = 0$ and $APPG = 1$ on the native single-view grid. It minimizes displacement from the rounded target at each step, conditional on earlier assignments; it does not guarantee a global optimum of Eq. 14 or preservation of the original neighborhoods. The result depends on gene order and tie-breaking. We use this heuristic as a collision-removal intervention distinct from exact LSA, with its cross-dataset results reported in Table 5.

C.3 Baseline Model Configurations

We follow the hyperparameters and input protocols of OpenProblems (Luecken et al., 2025). XGBoost (Chen, 2016) uses precomputed PCA features, typically 50 components derived from HVGs, rather than raw counts. Seurat uses its default variable-selection, scaling, and classification workflow (Hao et al., 2021).

Generative and foundation models. For latent modeling, scANVI and its reference-mapping variant scANVI+scArches were configured with two hidden layers ($n_{\text{hidden}} = 128$), a 30-dimensional latent space, and a dropout rate of 0.2 (Lotfollahi et al., 2020). Training was conducted for approximately 100 to 200 epochs with batch covariates enabled. We evaluate scGPT (Cui et al., 2024) and CellFM (Zeng et al., 2025) using zero-shot embeddings. The evaluation uses the pretrained scGPT_human and scGPT_80M checkpoints with a weighted KNN classifier ($k = 10$) on the embeddings.

C.4 Standardized Training and Visualization

All GE2I representations use a shared ResNet-18 probe initialized with standard weights. We train with AdamW (initial learning rate 1×10^{-3} , weight decay 0.01), a 5-epoch linear warmup, and cosine annealing for up to 50 epochs. Early stopping uses a patience of 10 epochs.

D Cross-Modality Gene-to-Protein Prediction

We evaluate GE2I on CITE-Seq gene-to-protein prediction (Stoeckius et al., 2017) using full-transcriptome inputs. Each cell’s gene expression profile is represented as an image and used to predict the abundances of 13 surface proteins. Dataset specifications and image-construction procedures are provided in Appendices B.1 and C.1, respectively. We report root mean squared error (RMSE), overall Pearson correlation (PCC), and overall Spearman correlation to assess prediction error, linear association, and rank agreement with the measured protein abundances.

Table 8: Gene-to-protein prediction on CITE-Seq using full-transcriptome inputs. Results are reported as mean $\pm$ standard deviation. Lower RMSE and higher correlations indicate better prediction.

Method	RMSE $\downarrow$	Overall PCC $\uparrow$	Overall Spearman $\uparrow$
Random Mapping	0.3688 $\pm$ 0.00	0.7996 $\pm$ 0.00	0.7443 $\pm$ 0.01
DeepInsight (t-SNE)	0.3667 $\pm$ 0.00	0.8009 $\pm$ 0.00	0.7451 $\pm$ 0.01
DeepInsight (UMAP)	0.4397 $\pm$ 0.05	0.7497 $\pm$ 0.02	0.7266 $\pm$ 0.02
DeepInsight (PCA)	0.3741 $\pm$ 0.02	0.7995 $\pm$ 0.01	0.7647 $\pm$ 0.01
MR-DeepInsight	0.3479 $\pm$ 0.00	0.8237 $\pm$ 0.01	0.7797 $\pm$ 0.00
Fotomics	0.4125 $\pm$ 0.01	0.7404 $\pm$ 0.01	0.7076 $\pm$ 0.00
REFINED	0.3999 $\pm$ 0.02	0.7726 $\pm$ 0.01	0.7214 $\pm$ 0.01

Table 8 reports the results for the evaluated image mappings. Both random and structured layouts produce predictions correlated with measured protein abundances. DeepInsight (t-SNE) and Random Mapping achieve similar overall PCC values of 0.8009 and 0.7996, with RMSE values of 0.3667 and 0.3688, respectively. This evaluation extends the use of gene expression images from predicting categorical cell identities to estimating continuous protein measurements.

E Additional Ablations

This section collects supplementary result tables for the expression-space baseline and the visual training choices analyzed in the main text.

E.1 Full-Genome MLP Ablation

Table 9: Tuned full-genome Simple MLP baseline on raw expression vectors. Results use the same splits as the GE2I classification benchmarks.

Metric	DKD	GTE _x v9	HypoMap	ICA	TS
Acc	0.9660 $\pm$ 0.0011	0.9078 $\pm$ 0.0025	0.9927 $\pm$ 0.0027	0.9002 $\pm$ 0.0017	0.4941 $\pm$ 0.1406
F1-M	0.9312 $\pm$ 0.0161	0.3624 $\pm$ 0.0239	0.6903 $\pm$ 0.2036	0.8285 $\pm$ 0.0165	0.3044 $\pm$ 0.0524

The tuned Simple MLP uses raw expression vectors and the same splits as GE2I. This comparison tests whether an image representation is necessary to benefit from the larger gene set. The MLP is competitive on DKD, ICA, and TS, but its F1-Macro is lower than that of the best structured GE2I methods on GTE_x v9 and HypoMap. Raw information can therefore explain some full-genome gains, while projection provides useful inductive bias in other settings. The distinction between metrics is important: on GTE_x v9, the MLP attains high accuracy despite weaker F1-Macro, so its aggregate correctness does not imply equally strong performance across cell types. This ablation supports a dataset-dependent role for projection rather than a general superiority of either representation family.

E.2 Supplementary Visual-Training Results

The following tables provide the complete backbone and augmentation results discussed in Section 5.4.

E.3 Interpretability Trade-off

DeepInsight (t-SNE) yields more concentrated Grad-CAM responses than Random Mapping: CAM entropy is lower (0.907 vs. 0.950), center mass higher (0.521 vs. 0.427), and mean radius smaller (0.201 vs. 0.227). Canonical-marker hit ratio increases from 0.25 to 0.50, and mean marker CAM percentile from 0.559 to 0.742. The structured layout recovers more

Table 10: Collision-removal results for DeepInsight (t-SNE) across DKD backbones. Superscripts report relative performance changes.

Backbone	Acc	F1-M
ResNet-18	0.9640 ^{+1.2%} ± 0.0055	0.9094 ^{+3.4%} ± 0.0410
ResNet-50	0.9623 ^{+1.8%} ± 0.0065	0.9071 ^{+4.6%} ± 0.0358
ConvNeXt-Tiny	0.9573 ^{+0.3%} ± 0.0044	0.9022 ^{+4.0%} ± 0.0371
EfficientNet-B0	0.9654 ^{+1.1%} ± 0.0016	0.9202 ^{+7.0%} ± 0.0369
ViT-B/16	0.9512 ^{+1.2%} ± 0.0084	0.8618 ^{+4.4%} ± 0.0074

Table 11: Random mapping performance across backbones on DKD.

Backbone	Acc	F1-M
ResNet-18	0.9609 ± 0.00	0.8635 ± 0.02
ResNet-50	0.9605 ± 0.00	0.8984 ± 0.01
ConvNeXt-Tiny	0.9512 ± 0.01	0.8576 ± 0.02
EfficientNet-B0	0.9603 ± 0.00	0.8587 ± 0.01
ViT-B/16	0.9271 ± 0.00	0.8080 ± 0.01

Table 12: Image augmentation on DKD. Each augmentation is evaluated relative to the same mapping without augmentation. All tested configurations, including their combination, are reported.

Method	Augmentation	ACC	F1-Macro
DeepInsight (t-SNE)	None	0.9537 ± 0.00	0.8868 ± 0.02
	Geometric	0.9532 ± 0.01	0.8819 ± 0.02
	Bio. Cutout	0.9513 ± 0.01	0.8540 ± 0.04
	Bio. Noise	0.9482 ± 0.01	0.8771 ± 0.01
	Bio. Intensity	0.9566 ± 0.00	0.8938 ± 0.01
	All	0.4689 ± 0.31	0.3227 ± 0.22
Random	None	0.9421 ± 0.00	0.8053 ± 0.02
	Geometric	0.9550 ± 0.01	0.8416 ± 0.04
	Bio. Cutout	0.9532 ± 0.01	0.8319 ± 0.03
	Bio. Noise	0.9578 ± 0.01	0.8668 ± 0.01
	Bio. Intensity	0.9563 ± 0.01	0.8652 ± 0.03
	All	0.9516 ± 0.01	0.8534 ± 0.01
Refined	None	0.9566 ± 0.00	0.8704 ± 0.03
	Geometric	0.9497 ± 0.01	0.8263 ± 0.01
	Bio. Cutout	0.9552 ± 0.00	0.8698 ± 0.01
	Bio. Noise	0.2926 ± 0.22	0.0932 ± 0.12
	Bio. Intensity	0.9584 ± 0.00	0.8837 ± 0.02
	All	0.5444 ± 0.25	0.1941 ± 0.15

Table 13: Grad-CAM interpretability comparison on DKD. Canonical-marker coverage is measured as the number of established markers falling inside the salient region. Higher center mass and marker hit ratio are better; lower entropy and radius are better.

Setting	Random	DeepInsight (t-SNE)	Key evidence
CAM entropy ↓	0.950	0.907	more compact saliency
Center mass ↑	0.427	0.521	stronger spatial focus
Marker hit ratio ↑	0.25	0.50	better biological alignment
DCT marker coverage ↑	2/5	5/5	SLC12A3, PVALB, TRPM6, CALB1, FXYD2
TAL marker coverage ↑	2/5	4/5	UMOD, SLC12A1, KCNJ1, CLDN16
Podocyte marker coverage ↑	2/5	5/5	NPHS1, NPHS2, PODXL, WT1, SYNPO

canonical markers within hotspots for distal convoluted tubule cells, thick ascending limb cells, and podocytes.

Spatial concentration alone would not establish biological relevance, since the mapping itself affects the geometry of saliency regions. Marker recovery provides complementary evidence that the structured hotspots align with known cell-type features. These results concern the interpretability of the evaluated Grad-CAM maps, not causal validation of the identified genes, and should be read separately from the classification comparison.

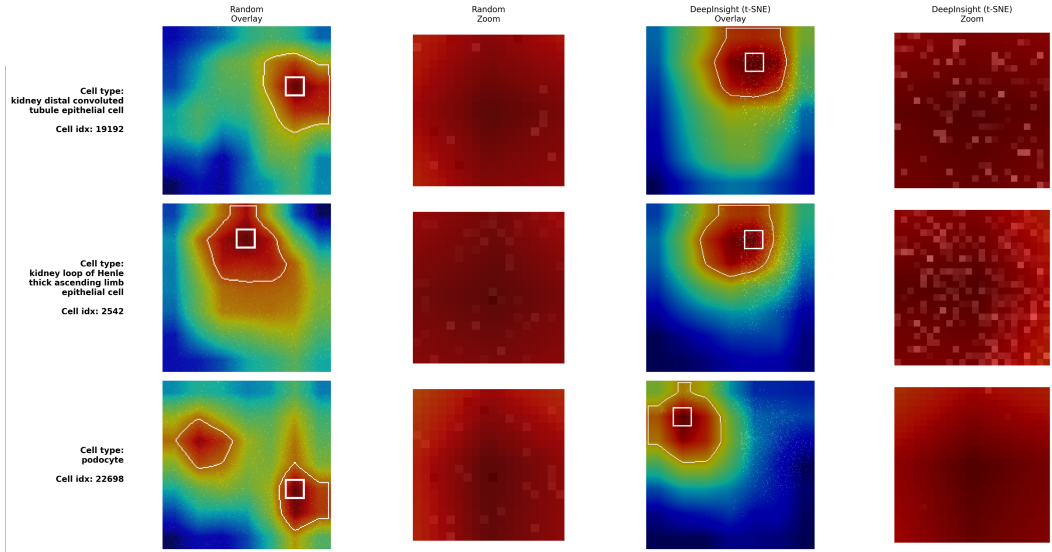


Figure 8: Representative Grad-CAM examples for three cell types where DeepInsight-tSNE shows stronger biological alignment than the random mapping. In distal convoluted tubule, thick ascending limb, and podocytes, the structured mapping yields tighter saliency regions and substantially higher canonical-marker recovery within the hotspot.

F Comprehensive Metric Breakdown

Table 14 reports spectral descriptors and performance for each method–dataset pair used in Section 5.2.

Full-transcriptome manifold layouts create more gene collisions than their HVG counterparts. CR increases with full-genome inputs; for DeepInsight (t-SNE) on HypoMap, it rises from 0.198 to 0.641. Compact gene clusters can thus reduce the number of expression values represented separately.

IGTD and REFINED achieve near-maximal smoothness without gene collisions. Both methods have $CR = 0$ and LFER near 1.0. Their layouts preserve separate gene assignments while concentrating spectral power at low frequencies. This distinguishes collision avoidance from texture preservation and motivates treating both as potential contributors to performance.

Table 14: Detailed metrics for spectral properties, information density, and performance across all datasets.

Method	Dataset	APPG	CR	LFER	SE	Acc	F1-M
Highly Variable Genes (HVG)							
Random	DKD	1.000	0.000	0.177	3.029	0.900	0.773
Random	GTE _x	1.000	0.000	0.198	3.033	0.653	0.211

Continued on next page

Table 14 – Continued from previous page

Method	Dataset	APPG	CR	LFER	SE	Acc	F1-M
Random	HypoMAP	1.000	0.000	0.191	3.030	0.980	0.557
Random	ICA	1.000	0.000	0.207	3.034	0.683	0.484
Random	Tabula Sapiens	1.000	0.000	0.221	3.026	0.384	0.215
DeepInsight (t-SNE)	DKD	0.869	0.127	0.900	2.846	0.892	0.742
DeepInsight (t-SNE)	GTEEx	0.831	0.170	0.823	3.520	0.626	0.194
DeepInsight (t-SNE)	HypoMAP	0.832	0.198	0.933	3.088	0.983	0.572
DeepInsight (t-SNE)	ICA	0.832	0.179	0.954	2.887	0.693	0.504
DeepInsight (t-SNE)	Tabula Sapiens	0.797	0.196	0.945	2.769	0.371	0.184
DeepInsight (UMAP)	DKD	0.976	0.025	0.988	0.582	0.900	0.759
DeepInsight (UMAP)	GTEEx	0.757	0.235	0.871	2.043	0.677	0.221
DeepInsight (UMAP)	HypoMAP	0.807	0.183	0.996	0.455	0.981	0.578
DeepInsight (UMAP)	ICA	0.778	0.205	0.953	1.254	0.705	0.510
DeepInsight (UMAP)	Tabula Sapiens	0.610	0.345	0.873	2.587	0.392	0.197
DeepInsight (PCA)	DKD	0.230	0.246	0.511	3.997	0.903	0.781
DeepInsight (PCA)	GTEEx	0.159	0.293	0.401	4.408	0.623	0.176
DeepInsight (PCA)	HypoMAP	0.171	0.109	0.442	4.253	0.983	0.572
DeepInsight (PCA)	ICA	0.157	0.185	0.314	4.614	0.686	0.481
DeepInsight (PCA)	Tabula Sapiens	0.173	0.127	0.387	4.460	0.374	0.202
MR-DeepInsight	DKD	1.784	0.103	0.895	1.768	0.908	0.797
MR-DeepInsight	GTEEx	1.656	0.165	0.832	2.558	0.684	0.199
MR-DeepInsight	HypoMAP	1.641	0.165	0.919	1.686	0.992	0.593
MR-DeepInsight	ICA	1.614	0.178	0.909	2.104	0.748	0.639
MR-DeepInsight	Tabula Sapiens	1.565	0.186	0.917	1.894	0.387	0.210
FOTOMICS	DKD	0.863	0.140	0.620	3.510	0.902	0.742
FOTOMICS	GTEEx	0.919	0.086	0.450	4.189	0.658	0.201
FOTOMICS	HypoMAP	0.885	0.116	0.569	3.738	0.981	0.520
FOTOMICS	ICA	0.930	0.074	0.611	3.444	0.707	0.510
FOTOMICS	Tabula Sapiens	0.927	0.077	0.636	3.375	0.383	0.209
Refined	DKD	1.000	0.000	0.995	0.555	0.902	0.753
Refined	GTEEx	1.000	0.000	0.991	0.545	0.637	0.194
Refined	HypoMAP	1.000	0.000	0.997	0.522	0.981	0.560
Refined	ICA	1.000	0.000	0.995	0.608	0.663	0.484
Refined	Tabula Sapiens	1.000	0.000	0.993	0.646	0.373	0.215
IGTD	DKD	1.000	0.000	0.996	0.467	0.902	0.733
IGTD	GTEEx	1.000	0.000	0.994	0.481	0.666	0.211
IGTD	HypoMAP	1.000	0.000	1.000	0.422	0.984	0.539
IGTD	ICA	1.000	0.000	0.999	0.429	0.690	0.500
IGTD	Tabula Sapiens	1.000	0.000	0.996	0.463	0.393	0.207
Genomap	DKD	1.000	1.000	0.776	1.877	0.881	0.772
Genomap	GTEEx	1.000	1.000	0.411	2.815	0.644	0.207
Genomap	HypoMAP	1.000	1.000	0.610	2.400	0.977	0.542
Genomap	ICA	1.000	1.000	0.712	2.221	0.668	0.474
Genomap	Tabula Sapiens	1.000	1.000	0.833	1.608	0.361	0.210
Full-Transcriptome							
Random	DKD	1.000	0.000	0.252	4.409	0.942	0.805
Random	GTEEx	1.000	0.000	0.244	4.409	0.849	0.460
Random	HypoMAP	1.000	0.000	0.241	4.607	0.990	0.655
Random	ICA	1.000	0.000	0.233	4.433	0.879	0.731
Random	Tabula Sapiens	1.000	0.000	0.241	4.755	0.457	0.345
DeepInsight (t-SNE)	DKD	0.489	0.482	0.943	1.958	0.954	0.887
DeepInsight (t-SNE)	GTEEx	0.446	0.504	0.867	2.643	0.855	0.506
DeepInsight (t-SNE)	HypoMAP	0.293	0.641	0.950	2.205	0.992	0.749
DeepInsight (t-SNE)	ICA	0.402	0.581	0.954	2.036	0.877	0.748
DeepInsight (t-SNE)	Tabula Sapiens	0.255	0.644	0.946	2.336	0.427	0.247
DeepInsight (UMAP)	DKD	0.549	0.440	0.904	2.743	0.957	0.892
DeepInsight (UMAP)	GTEEx	0.262	0.655	0.802	3.204	0.846	0.489
DeepInsight (UMAP)	HypoMAP	0.190	0.766	0.916	2.878	0.994	0.771
DeepInsight (UMAP)	ICA	0.358	0.635	0.818	3.172	0.882	0.763
DeepInsight (UMAP)	Tabula Sapiens	0.070	0.795	0.917	3.008	0.428	0.267
DeepInsight (PCA)	DKD	0.116	0.422	0.765	2.791	0.958	0.846

Continued on next page

Table 14 – Continued from previous page

Method	Dataset	APPG	CR	LFER	SE	Acc	F1-M
DeepInsight (PCA)	GTEEx	0.080	0.480	0.650	3.228	0.842	0.442
DeepInsight (PCA)	HypoMAP	0.084	0.460	0.734	2.677	0.993	0.728
DeepInsight (PCA)	ICA	0.100	0.405	0.644	3.342	0.872	0.748
DeepInsight (PCA)	Tabula Sapiens	0.048	0.447	0.753	2.715	0.442	0.305
MR-DeepInsight	DKD	0.876	0.572	0.891	2.436	0.970	0.927
MR-DeepInsight	GTEEx	0.714	0.548	0.785	3.186	0.871	0.366
MR-DeepInsight	HypoMAP	0.557	0.648	0.896	2.574	0.999	0.802
MR-DeepInsight	ICA	0.625	0.679	0.901	2.481	0.899	0.822
MR-DeepInsight	Tabula Sapiens	0.284	0.726	0.904	2.693	0.410	0.253
FOTOMICS	DKD	0.288	0.636	0.900	1.767	0.954	0.852
FOTOMICS	GTEEx	0.195	0.696	0.798	2.547	0.833	0.418
FOTOMICS	HypoMAP	0.249	0.672	0.871	1.882	0.991	0.719
FOTOMICS	ICA	0.287	0.577	0.866	2.009	0.861	0.720
FOTOMICS	Tabula Sapiens	0.196	0.610	0.810	2.296	0.460	0.295
Refined	DKD	1.000	0.000	0.999	0.456	0.957	0.870
Refined	GTEEx	1.000	0.000	0.995	0.523	0.847	0.488
Refined	HypoMAP	1.000	0.000	1.000	0.427	0.992	0.727
Refined	ICA	1.000	0.000	1.000	0.432	0.883	0.730
Genomap	DKD	1.000	1.000	1.000	1.004	0.408	0.092

G Statistical Analysis Diagnostics

Clustered inference accounts for repeated seeds, but full-transcriptome coverage is uneven. A cluster is one dataset–method pair, and G denotes the number of such clusters. We compute CR1 cluster-robust standard errors across repeated seeds. The reported z^* is the coefficient divided by its clustered standard error. Two-sided p -values use a t distribution with $G-1$ degrees of freedom, written t_{G-1} : 34 degrees for the 35 Full clusters and 44 degrees for the 45 HVG clusters. The q -values are Benjamini–Hochberg-adjusted p -values over the 16 descriptor tests across both outcomes and input scopes. This adjustment accounts for testing many coefficients simultaneously; $q < 0.05$ marks results that remain significant at a 5% false-discovery-rate threshold.

The Full and HVG models use 105 and 135 runs, respectively. Of 270 potential runs, 243 contain both outcomes and 240 also contain all four descriptors. Thirty Full and 40 HVG runs use APPG/CR values imputed by the observed-seed mean within the same dataset–method–scope group. Full coverage is uneven: Genomap contributes three DKD runs, REFINED contributes 12 runs, and IGTG has no complete records. The reported standard errors treat the supplied imputed values as fixed.

HVG conclusion: no descriptor reaches $q < 0.05$ for F1-Macro; for accuracy, APPG is positive and CR is negative. LFER and SE do not reach the adjusted threshold (Table 15). Compared with the Full results, the HVG pattern indicates that the observed descriptor–performance relationships depend on how many genes are represented.

Table 15: Beta regression in the HVG setting. Descriptors enter jointly with dataset-by-seed fixed effects, and SE is clustered by dataset–method. The p -values use a t distribution with 44 degrees of freedom; q is the Benjamini–Hochberg-adjusted p -value.

Target	Descriptor	Coef.	Cluster SE	z^*	p	q
Accuracy	APPG	0.146	0.044	3.34	0.002	0.004
	CR	-0.097	0.030	-3.26	0.002	0.004
	LFER	0.061	0.061	1.00	0.320	0.394
	SE	0.022	0.016	1.40	0.167	0.223
F1-Macro	APPG	0.172	0.091	1.88	0.066	0.096
	CR	-0.020	0.053	-0.39	0.701	0.747
	LFER	0.014	0.076	0.19	0.852	0.852
	SE	0.016	0.022	0.72	0.473	0.541

Full-transcriptome spectral descriptors are collinear, but their positive associations survive key exclusions. After accounting for fixed effects, the Full VIFs for APPG, CR, LFER, and SE are 1.96, 3.41, 11.57, and 8.43; the HVG values are 1.27, 1.01, 2.22, and 2.47. The higher Full LFER and SE values show that the two spectral coefficients are best interpreted together. Their positive Full associations remain when using only directly measured descriptors or excluding Genomap. Allowing dataset-specific precision reduces the APPG and CR associations, while method fixed effects attenuate the spectral associations by focusing the comparison on variation within each method.