DGAT: A Dual-Graph Attention Network for Inferring Spatial Protein Landscapes from Transcriptomics

Haoyu Wang[1], Brittany Cody[2], Hatice Ulku Osmanbeyoglu[1*]

[1]Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[2]Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

*Correspondence should be addressed to H.U.O. (osmanbeyogluhu@pitt.edu), University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA; Phone: 412-623-7789; ORCID ID: 0000-0002-4972-4347

Abstract

Spatial transcriptomics (ST) technologies provide genome-wide mRNA profiles in tissue context but lack direct protein-level measurements, which are critical for interpreting cellular function and microenvironmental organization. We present DGAT (Dual-Graph Attention Network), a deep learning framework that imputes spatial protein expression from transcriptomics-only ST data by learning RNA–protein relationships from spatial CITE-seq datasets. DGAT constructs heterogeneous graphs integrating transcriptomic, proteomic, and spatial information, encoded using graph attention networks. Task-specific decoders reconstruct mRNA and predict protein abundance from a shared latent representation. Benchmarking across public and in-house datasets—including tonsil, breast cancer, glioblastoma, and malignant mesothelioma—demonstrates that DGAT outperforms existing methods in protein imputation accuracy. Applied to ST datasets lacking protein measurements, DGAT reveals spatially distinct cell states, immune phenotypes, and tissue architectures not evident from transcriptomics alone. DGAT enables proteome-level insights from transcriptomics-only data, bridging a critical gap in spatial omics and enhancing functional interpretation in cancer, immunology, and precision medicine.

Introduction

In multicellular systems, cell function is strongly influenced by spatial context. Understanding how cells are spatially arranged, interact, and respond within their microenvironment is essential to dissect complex developmental and disease processes. Spatial transcriptomics (ST) enables mRNA profiling while preserving tissue architecture and the cellular microenvironment. Next-generation sequencing (NGS)-based ST platforms—such as the original ST[1], 10x Visium, Slide-seq[2], Slide-seqV2[3], and high-definition spatial transcriptomics (HDST)[4]—use spatially barcoded probes to capture genome-wide mRNA from defined tissue regions ("spots") that typically encompass multiple cells arranged on a two-dimensional grid.

However, transcriptomic readouts alone do not capture functional protein-level information, which is often critical for interpreting cell types/states and predicting responses to targeted therapies. Research indicates a weak correlation at the level of individual cells between the abundance of most proteins and their associated transcripts[5]. This discrepancy may arise from errors in mRNA or protein measurement, the inherent stochasticity in RNA processing, translation, and protein transport, or confounding factors such as the time delay between transcription and translation. Consequently, relying solely on mRNA expression data often fails to provide a complete picture of cellular processes and tissue architecture. Recent technological advances provide simultaneous spatial profiling of genome-wide transcript and multiplexed protein expression for up to ~200 proteins with histopathological evaluation. These methods include Spatial PrOtein and Transcriptome Sequencing (SPOTS)[6], Spatial Multi-Omics (SM-Omics)[7], Spatial-CITE-seq (Cellular Indexing of Transcriptomes and Epitopes by sequencing)[8,9], and Stereo-CITE-seq[10] and the Visium CytAssist Spatial Gene and Protein Expression for FFPE platform by 10x Genomics. These are referred to here collectively as "spatial-CITE-seq". Yet, due to technological barriers and cost considerations, most studies quantify the transcriptome only and do not incorporate protein measurements.

Several machine learning (ML) strategies have been developed to predict missing modalities, such as proteins, from available data. In the imaging domain, deep learning methods based on image-to-image translation have enabled virtual staining of protein markers directly from hematoxylin and eosin (H&E) images[11]. More recent advances support the prediction of multiple markers simultaneously, and tools like Multi-V-Stain improve biological relevance by optimizing all markers jointly to produce multiplexed Imaging Mass Cytometry (IMC) stains[12]. While powerful, these image-based approaches are not applicable to transcriptomic data.

In parallel, multimodal single-cell datasets have inspired methods that learn RNA–protein relationships for downstream imputation tasks[13-16]. Methods like cTP-net (single-cell Transcriptome to Protein prediction)[17], STREAK (gene Set Testing-based Receptor abundance Estimation)[18], sciPENN (**s**ingle **c**ell **i**mputation **P**rotein **E**mbedding **N**eural **N**etwork)[19], scLinear[20] and SPIDER (surface protein prediction using deep ensembles from single-cell RNA sequencing)[21] have been tailored to learn mRNA-protein relationships from CITE-seq[5]/REAP-seq[22] —methods that profile mRNA and surface protein expression in single cells— to generate protein estimates for scRNA-seq datasets. While these tools have shown success in single-cell settings, they are not designed to incorporate spatial structure, limiting their utility for protein inference in ST data.

To overcome these challenges, we present DGAT (Dual-Graph Attention Network), a computational framework for imputing spatial protein abundance from transcriptomic measurements by transferring mRNA–protein relationships learned from spatial CITE-seq data. DGAT constructs heterogeneous graphs that integrate mRNA/protein expression with spatial coordinates and applies graph attention encoders to learn aligned representations of mRNA and protein features. A multi-branch decoder then predicts protein expression from these embeddings, enabling accurate inference in spot-based ST datasets (**Fig. 1**).

We benchmark DGAT using spatial CITE-seq datasets from human tonsil, breast cancer, glioblastoma, and malignant mesothelioma tissues, comprising over 26,000 spatial

spots, ~18,000 genes, and 31 shared proteins. Across leave-one-sample-out evaluations, DGAT consistently outperforms existing methods that do not utilize spatial information, demonstrating superior accuracy in predicting protein marker expression. Applying DGAT to diverse ST datasets, we impute 31 protein marker expression and show that the resulting profiles yield improved delineation of cellular states, immune phenotypes, and spatial domains such as germinal centers. By integrating protein-level insights into widely available ST datasets, DGAT enhances spatial domain detection, cell-type annotation, and biological interpretation, providing a powerful tool for tissue-level analysis in cancer, immunology, and precision medicine.

Results

Overview of the DGAT framework

An overview of DGAT is shown in **Fig. 1a**. The framework is designed to learn from spatial CITE-seq reference datasets with paired mRNA and protein measurements. Once trained, DGAT can predict protein expression from transcript-only spatial transcriptomics (ST) datasets.

DGAT receives as input paired gene and protein expression matrices with spatial coordinates (**Fig. 1b**). It constructs two types of graphs per sample: a spatial graph based on physical proximity, and feature similarity graphs based on transcriptomic and proteomic profiles. Each graph is processed by an independent Graph Attention Network (GAT)[23] with feature attention layers and residual connections[24-26] to create spatially-aware embeddings.

DGAT includes modality-specific encoders: the mRNA encoder integrates gene expression with spatial context, and the protein encoder models spatial co-expression patterns. Embeddings from both encoders are decoded in two ways: (i) a multilayer perceptron (MLP) reconstructs gene expression and (ii) a multi-branch protein decoder predicts protein levels, with each branch specialized for a distinct protein. To balance the alignment between gene and

protein modalities while maintaining their distinct feature representations, we applied three types of loss functions with cross-modality supervision mechanism during training. First, a latent alignment loss encourages convergence between the transcript and protein embeddings. Second, cross-prediction losses allow each encoder to guide the opposing decoder. Finally, reconstruction losses on both modalities further constrain learning. After training, spatial protein inference can be performed by passing mRNA profiles of ST data through the pretrained mRNA encoder and protein decoder (**Fig. 1c**).

Protein prediction benchmarking via sample holdout

We evaluated DGAT's predictive performance using a sample holdout strategy across six spatial CITE-seq samples derived from tonsil (n=2), estrogen receptor-positive breast cancer (BC), glioblastoma (GBM), and malignant mesothelioma tissues (n=2) (**Supplementary Table 1**). These datasets were chosen because they were among the few publicly or locally available spatial CITE-seq datasets, and they provided a diverse range of tissue types with shared protein panels, allowing systematic benchmarking of cross-sample protein prediction. Each dataset included spatially resolved transcriptomes and matched expression of ~30 proteins (**Supplementary Table 2**) measured via the 10x Genomics CytAssist gene and protein expression platform. For each experiment, DGAT was trained on all but one sample and used to predict protein expression in the held-out sample.

Performance was assessed using the Spearman correlation coefficient (ρ) to evaluate rank concordance between predicted and observed protein levels on holdout samples, and root mean square error (RMSE) used to quantify absolute prediction error across spatial spots. As a baseline comparison, we used normalized RNA expression values of the corresponding genes. Across all tissues, DGAT consistently outperformed the RNA baseline (**Fig. 2a**, **Supplementary Figure 1**). For instance, in the tonsil1 sample, the median protein-wise Spearman correlation increased from $\rho = 0.308 \pm 0.219$ (RNA baseline) to $\rho = 0.610 \pm 0.230$ (DGAT), while RMSE

decreased from 1.61 to 0.226 (**Fig. 2a**). Similar improvements were observed across other samples (**Fig. 2a**). Notably, improvements were observed for biologically and clinically relevant markers such as CCR7, CD14, CD19, CD68, CEACAM8, CXCR5, FCGR3A, HLA-DRA, ITGAM, ITGAX, KRT5, PDCD1, PTPRC, and SDC1 (e.g. **Fig. 2b** for CD68). Consistent trends were also evident in spot-wise analyses (**Supplementary Fig. 1**), highlighting DGAT's ability to generalize to unseen tissue samples and improve protein inference accuracy beyond transcript-based proxies.

The variation in predictive accuracy across tissues may be partly explained by the fact that our protein panel predominantly consists of immune markers, which are more abundantly and heterogeneously expressed in immune-rich tissues like tonsil and breast cancer. In contrast, glioblastoma and mesothelioma tissues typically exhibit lower immune infiltration or more complex tumor microenvironments, potentially making protein expression patterns harder to predict accurately.

We benchmarked DGAT against four state-of-the-art models developed based on paired single cell gene expression and protein expression datasets that do not consider spatial structure: Seurat v4[27], sciPENN[19], cTP-net[17], scLinear[20] (**Fig. 2c, Supplementary Fig. 1**). In all six held-out sets, DGAT achieved higher median correlation. To summarize performance across metrics, we computed an average rank score[28] (ARS), which integrates method rankings based on correlation and RMSE. DGAT ranked highest on ARS for protein-level predictions, demonstrating robust accuracy across datasets (**Fig. 2d**, **Supplementary Fig. 1b**), DGAT's advantage in spatial prediction highlights the benefits of incorporating spatial context. These results underscore DGAT's ability to infer biologically informative and tissue-relevant protein expression patterns.

Ablation and Interpretability Analyses

To systematically assess the contributions of DGAT's architectural components, we performed a series of ablation studies by selectively removing or modifying individual modules and design elements (**Supplementary Table 3**). First, we evaluated the impact of excluding the spatial graph, which caused a notable decline in median Spearman correlation (ρ) across datasets—from [Tonsil1: 0.610, Tonsil2: 0.658, BC: 0.579, GBM: 0.283, Meso1: 0.312, Meso2: 0.284] to [Tonsil1: 0.589, Tonsil2: 0.614, BC: 0.485, GBM: 0.153, Meso1: 0.266, Meso2: 0.178]—validating the critical role of spatial context in enhancing protein prediction. Similarly, removing the expression-based graph or relying on a single modality (e.g., RNA-only graph) further impaired performance (median ρ = [Tonsil1: 0.599, Tonsil2: 0.638, BC: 0.519, GBM: 0.228, Meso1: 0.310, Meso2: 0.291]), underscoring the necessity of multi-graph construction and modality-specific encoding.

Next, we investigated key architectural enhancements. Eliminating the feature-level attention mechanism reduced performance (median ρ = [Tonsil1: 0.585, Tonsil2: 0.661, BC: 0.533, GBM: 0.229, Meso1: 0.303, Meso2: 0.291]), highlighting its importance for prioritizing relevant molecular features. Substituting the protein-specific multi-branch decoder with a shared decoder similarly decreased accuracy (median ρ = [Tonsil1: 0.583, Tonsil2: 0.645, BC: 0.565, GBM: 0.251, Meso1: 0.305, Meso2: 0.280]), demonstrating the advantage of protein-specific mappings for flexible cross-modal learning. Additionally, ablating residual connections between GAT layers led to less stable convergence and further performance drops (median ρ = [Tonsil1: 0.571, Tonsil2: 0.622, BC: 0.460, GBM: 0.257, Meso1: 0.260, Meso2: 0.217]), suggesting these connections mitigate vanishing gradients and help preserve informative features across layers.

Finally, we examined the influence of loss functions. Removing the latent space alignment loss between RNA and protein embeddings impaired cross-modal consistency and prediction accuracy (median ρ = [Tonsil1: 0.585, Tonsil2: 0.656, BC: 0.477, GBM: 0.228, Meso1: 0.301, Meso2: 0.281]). Likewise, excluding the cross-prediction loss—where each

encoder supervises the decoder of the opposite modality—resulted in reduced generalization on unseen samples

To enhance model interpretability, we performed feature importance analysis using gradient-based saliency scores[29] on a representative tonsil sample (**Supplementary Table 4**). DGAT consistently prioritized genes with strong cell-type specificity and established protein-coding roles relevant to the predicted proteins. For instance, T cell proteins like CD3D and CD8 were associated with T cell–specific transcripts such as *GZMK*, *TRBC1*, *CXCL9*, and *SELL*, reflecting cytotoxic and memory functions. Myeloid markers CD68 and CD163 were linked to genes involved in innate immunity and myeloid differentiation, including *CXCL8*, *S100A8*, and *LY86*. Similarly, dendritic cell marker ITGAX (CD11c) correlated with *BATF*, *EBI3*, and *PLA2G7*, known mediators of dendritic cell activity. Stromal markers such as ACTA2 (α-SMA) and VIM (vimentin) were predicted using genes involved in extracellular matrix and cytoskeletal structure, including *COL14A1* and *TAGLN*. These findings demonstrate DGAT's capacity to capture biologically meaningful RNA–protein relationships that reflect cell identity and functional states.

DGAT Enables Protein Inference in Spatial Transcriptomics Datasets with Shared Tissue Features

To demonstrate DGAT's utility in ST datasets lacking protein measurements, we applied the pretrained model to publicly available human lymph node (LN) Visium data (**Fig. 3**) and triple-negative breast cancer (TNBC) tissue (**Fig. 4**) to predict the expression of 31 proteins. These tissues were selected because the DGAT training dataset includes closely related tissue types—tonsil and ER-positive (ER^+) breast cancer—which share key structural and molecular features with LN and TNBC, respectively. This setup allowed us to evaluate the model's capacity to generalize protein-level inference across similar yet distinct biological contexts.

In the LN dataset, clustering based on DGAT-inferred protein expression revealed spatially distinct regions that closely matched histologically annotated germinal centers (GCs)[30] (**Fig. 3a–b**). To quantitatively assess alignment with ground truth GC annotations, we framed the task as a binary classification problem, where spots assigned to clusters overlapping GCs were labeled as positive and all others as negative. This allowed us to compute classification metrics such as accuracy, precision, recall, and F1 score. DGAT-defined cluster 4 demonstrated strong alignment with GC regions, achieving an accuracy of 0.966, precision of 0.807, recall of 0.839, and an F1 score of 0.823. In contrast, transcriptome-based cluster 2 exhibited lower precision (0.413) and F1 score (0.577), despite high recall (0.954), indicating reduced specificity and greater noise in mRNA-only clustering. These results suggest that DGAT improves both spatial resolution and biological interpretability by filtering out transcriptomic noise.

We further benchmarked DGAT against several transcriptome-to-protein prediction models that do not incorporate spatial context, including CTP-net, scLinear, sciPENN, and Seurat. DGAT achieved the highest F1 score of 0.823, outperforming CTP-net (0.806), scLinear (0.782), sciPENN (0.291), Seurat (0.667), and transcriptome-only clustering (0.575) (**Fig. 3b**). These results emphasize the importance of integrating spatial context to accurately reconstruct protein-level features.

To assess the biological specificity of DGAT-inferred proteins, we evaluated their spatial enrichment in germinal center (GC) regions using the Wilcoxon rank-sum test with Benjamini–Hochberg correction. DGAT predicted protein and GC associations in the lymph node (LN) ST dataset showed strong concordance with protein–GC associations observed in two tonsil samples (Tonsil1 and Tonsil2; **Fig. 3c–d**), demonstrating the model's ability to generalize across tissues with shared immune architecture. Canonical GC markers—including PAX5, CD19, CD20, CXCR5, CR2, CD40, PDCD1 (PD-1), and PCNA—were significantly enriched in GC regions (**Fig. 3e-f**), consistent with their roles in B cell identity, follicular homing, immune signaling, and proliferation. Notably, enrichment signals were more statistically significant using

DGAT-inferred protein levels compared to mRNA expression, demonstrating DGAT's ability to recover biologically meaningful spatial patterns obscured at the transcript level.

We next applied DGAT to spatial transcriptomics data from TNBC tissue, a subtype not explicitly included in training but related through the presence of ER^+ breast cancer samples. Clustering based on DGAT-inferred proteins again outperformed transcriptome-only clustering, achieving a higher adjusted Rand index (ARI) of 0.504 compared to 0.447, relative to expert annotations of six histopathological regions (**Fig. 4a**). DGAT also correctly localized canonical protein markers to their expected compartments (**Fig. 4b**). In lymphocyte-rich regions, DGAT predicted elevated expression of immune markers such as CD163, CD68, CD14, ITGAX, and FCGR3A, consistent with macrophage infiltration (**Fig. 4c**). In normal stroma and lymphocyte-enriched areas, high predicted expression of KRT5 and BCL2 reflected basal epithelial identity and anti-apoptotic signaling. In ductal carcinoma in situ (DCIS) regions, DGAT predicted increased expression of PCNA, EPCAM, and SDC1, consistent with proliferative and epithelial features characteristic of early-stage tumor progression.

DGAT was also applied to a ST dataset (2,518 spots, 17,651 genes) from an ER-positive invasive ductal carcinoma sample profiled with 10x Visium. Pathologist-annotated H&E images defined five tissue regions for spatial domain evaluation. DGAT achieved higher ARI than baseline methods (**Supplementary Figure 2**), demonstrating improved spatial domain detection from transcriptomic data alone. Collectively, these results demonstrate that DGAT generalizes across related tissue types to infer spatially coherent, biologically meaningful protein expression landscapes. This enables enhanced spatial resolution and functional interpretation in transcriptomics-only datasets.

DGAT Accurately Infers Protein Expression in Unseen Tissue Types

To test DGAT's capacity for out-of-distribution generalization, we applied the pretrained model to spatial transcriptomics datasets from human melanoma and prostate cancer—tissue types not represented in the training set. Thrane[31] et al. manually annotated three distinct regions—melanoma, stroma, and lymphatic tissue—alongside an unannotated area (**Fig. 5a-b**). DGAT-inferred protein expression exhibited clear spatial specificity: canonical protein markers associated with each histopathological compartment were significantly enriched in their corresponding regions (**Figure 5c**). In the melanoma region, DGAT predicted elevated expression of epithelial and proliferation markers, including KRT5**,** PCNA**,** EPCAM, and SDC1. Stromal areas showed high levels of ACTA2, PECAM1, CEACAM8, and VIM, consistent with fibrovascular and mesenchymal signatures. In lymphatic tissue, DGAT identified increased expression of immune and B/T cell markers such as CXCR5, MS4A1, PAX5, HLA-DRA, CD8, CD3, and PTPRC, reflecting the expected immune cell infiltration.

DGAT also generalized effectively to prostate cancer spatial transcriptomics data, recovering spatially distinct protein expression patterns corresponding to epithelial, stromal, and immune-rich compartments (**Fig. 5d–e**). In this dataset, DGAT achieved the highest adjusted Rand index (ARI) of 0.683, outperforming all baseline methods in accurately recapitulating pathologist-defined tumor, stromal, and partially atrophic regions. The mRNA-only model performed moderately (ARI 0.638), while Seurat and sciPENN underperformed with ARIs of 0.269 and 0.243, respectively. Within prostate cancer tissue, tumor regions specifically enriched for CD63 and CD9, along with epithelial markers EPCAM and VIM (**Fig. 5f**). Stromal compartments were enriched for immune-related proteins CD14 and CD68, consistent with macrophage and monocyte infiltration. Partially atrophic areas exhibited intermediate expression profiles, reflecting a mixture of epithelial and immune signals indicative of a transitional microenvironment.

Across both datasets, DGAT-predicted protein landscapes closely mirrored known histological boundaries despite the domain shift from training tissues. These results suggest that DGAT captures broadly conserved RNA–protein relationships that remain informative across diverse tissue contexts. Importantly, DGAT's ability to recover biologically meaningful protein profiles in the absence of paired proteomic data underscores its potential as a powerful tool for retrospective spatial transcriptomics analysis, cross-tissue protein mapping, and hypothesis generation in under-characterized disease settings.

Discussion

The ability to infer spatial protein expression from transcriptomic data alone opens new avenues for understanding tissue organization, cellular heterogeneity, and microenvironmental dynamics in health and disease. Here, we introduced DGAT, a deep learning framework trained on spatial CITE-seq data that accurately imputes spatial protein expression in transcriptomics-only datasets. Our results demonstrate that DGAT not only recovers biologically meaningful protein distributions but also enhances key downstream analyses such as spatial clustering and spatial domain identification (e.g. GCs). Further, DGAT's application to in-house malignant pleural mesothelioma tissue, a rare and understudied cancer[32,33], demonstrates its value in expanding spatial proteomic profiling beyond well-characterized diseases. Hence, DGAT enables proteome-level insights in datasets where protein measurements are unavailable, offering a computational bridge between transcriptomic accessibility and proteomic relevance.

A major advantage of DGAT lies in its ability to generalize across tissue types. We showed that a model trained on diverse spatial CITE-seq datasets could be effectively applied to unrelated transcriptomics-only datasets, including lymph node and breast cancer tissues, where it improved the concordance between predicted spatial clusters and expert pathological annotations. This suggests that the protein–transcript relationships learned by DGAT are

sufficiently robust to capture conserved biological patterns, even across different microenvironments.

While DGAT demonstrates strong generalization across the immune and cancer tissues profiled here, broader applicability will depend on the availability of spatial CITE-seq datasets spanning diverse organs and disease contexts. Currently, such datasets remain relatively limited, which may pose a barrier until more become publicly available. Moreover, because mRNA–protein relationships can vary substantially by tissue type and cell state, models trained on immune-rich or cancer samples may require retraining or fine-tuning to achieve optimal performance in tissues with distinct molecular landscapes, such as hepatic, neural, or fibrotic tissues. Similarly, while we evaluated DGAT on approximately 30 immune-related proteins, scaling to larger panels measured in some CITE-seq studies (200–300 proteins) should be feasible provided these markers are well represented in the training data, though this remains to be systematically benchmarked. Looking forward, incorporating diverse training data and leveraging spatial context will be critical for navigating tissue-specific regulatory differences and further enhancing DGAT's robustness.

Current spatial transcriptomics and spatial CITE-seq platforms are limited by their resolution and coverage. Capture spots often encompass multiple cells, leading to mixed signals that obscure cell-specific expression profiles. These technical constraints pose challenges for fine-grained spatial inference, particularly in complex tissues like tumors. However, emerging technologies—such as Visium HD and other high-resolution spatial omics platforms—promise to address these limitations by offering near-single-cell resolution and expanded molecular profiling. We anticipate that DGAT will be well suited for these next-generation datasets, where finer spatial granularity may further improve the fidelity of protein inference.

Looking forward, several extensions of this work are possible. First, expanding the

training corpus with additional spatial CITE-seq datasets from diverse tissues and disease contexts could improve the model's generalization capacity and resilience to dataset-specific biases. Second, integrating histopathological images (e.g., H&E or multiplexed immunofluorescence) may allow DGAT to incorporate morphological context, further enhancing protein prediction accuracy and interpretability. Lastly, beyond clustering, DGAT could support a broader range of spatial analyses, including ligand–receptor interaction mapping, cell–cell communication inference, and spatial trajectory reconstruction, offering a powerful toolset for dissecting tissue biology and informing therapeutic discovery.

In conclusion, DGAT provides a scalable, data-efficient solution for inferring protein-level spatial maps from transcriptomic inputs alone. As spatial omics technologies continue to evolve, DGAT stands to facilitate deeper biological insights in both basic and translational research settings—particularly where proteomic measurements are unavailable or impractical. DGAT represents a step toward democratizing spatial proteomics, enabling researchers to unlock functional insights from transcriptomic-only data without requiring expensive or technically demanding protein assays.

Methods

Malignant Mesothelioma Visium CytAssist v2 Library Preparation

Two formalin-fixed & paraffin-embedded (FFPE) malignant mesothelioma tissue blocks were obtained from National Meosthelioma Virtual Bank. Malignant mesothelioma samples were assessed for RNA quality, and all samples had a DV200 score above the minimum threshold ≥30%. Spatial transcriptomic and proteomic libraries were generated with Visium CytAssist for FFPE v2 gene expression kit (10x Genomics: 1000520) with the Human Immune Cell Profiling Panel (10x Genomics: PN-1000607). 5 micron FFPE sections were placed on Schott Nexterion Hydrogel Coated Slides (Schott North America: 1800434) or Fisherbrand Superfrost Plus Slides

(Fisher: 22-037-246) and processed through the 10x CytAssist for FFPE v2 protocol according to the manufacturer's instructions. H&E images were taken at 10x magnification using an EVOS 7000 Imaging System (Thermo Fisher Scientific).

Sequencing and Library Quality Control

Library QC was completed with an Agilent TapeStation 4150. Libraries were normalized and pooled to 2nM prior to loading on an Illumina NextSeq 2000, using a P3 100 flow cell with a target of 125 million reads per sample for transcriptomic libraries and 25 million reads per sample for proteomic libraries.

Data Preprocessing

We used datasets from three different platforms, including Visium CytAssist Gene and Protein Expression (tonsil1, tonsil2, breast cancer, glioblastoma from 10x Genomics and in-house malignant mesothelioma1, malignant mesothelioma2), 10x Visium platform (human breast cancer and human prostate cancer), the ST platform (human melanoma[31]). A detailed description and data sources are provided in **Supplementary Table 1**.

For all in house and public spatial CITE-seq (Visium CytAssist) and spatial transcriptomics datasets, low-quality spatial barcodes (spots) were removed if they contained fewer than 700 gene expressed or had >30% of UMIs derived from mitochondrial transcripts. Genes not expressed in at least 2.5% of spots were excluded to reduce technical noise and sparsity. Gene expression matrices were normalized to 10,000 UMIs per spot, log-transformed, and scaled to unit variance per gene. Values were clipped to a maximum of 10 to mitigate the influence of outliers.

Protein expression (ADT counts) was normalized using the centered log-ratio (CLR) method. After preprocessing, the dataset comprised 11,535 genes and 31 proteins across six

samples, which were integrated via shared features into a unified dataset for downstream modelling. The SCANPY toolkit[34] was utilized for all preprocessing steps.

Model Architecture: DGAT (Dual Graph Attention Transformer)

Graph Construction: We denoted the mRNA and protein expression matrices as $X_{mRNA} \in \mathbb{R}^{N\times G}$ and $X_{protein} \in \mathbb{R}^{N\times P}$, where N was the number of spots, and G and P were the number of genes and proteins, respectively. Each spot was modeled as a node in a graph with features defined by either mRNA or protein expression. Two types of adjacency matrices were constructed: spatial adjacency ($\mathbf{A}_{spatial}$) and modality-specific adjacency ($\mathbf{A}_{mRNA}$, $\mathbf{A}_{protein}$). The spatial graph connects each spot to its six nearest physical neighbors based on Euclidean distance of spot coordinates. Molecular graphs were built using k-nearest-neighbor (k=10) graphs in principle component (PCA) space (retaining 85% of variance) for each modality. The final adjacency matrices used for each modality were defined as:

$$A_{mRNA+s} = A_{mRNA} + A_{spatial}$$

$$A_{protein+s} = A_{protein} + A_{spatial}$$

Encoder design: While the mRNA and protein encoders differ in their input dimensionalities—reflecting the distinct numbers of mRNA and protein features—their internal architectures are otherwise identical to ensure consistent feature extraction across modalities. Each encoder consists of three Graph Attention (GAT) layers, interleaved with Layer Normalization and skip connections, enabling the learning of robust, generalized latent representations. The inputs to each encoder are (X_{mRNA}, A_{mRNA+s}) or $(X_{protein}, A_{protein+s})$. The encoding process is defined as: $H^1 = \sigma(\mathrm{LayerNorm}(\mathrm{GATConv}^1(X, A) + \mathrm{skip}^1_{en}(X)))$

$$H^{1\prime} = H^1 \odot \sigma(\mathrm{FeatureAttention}(H^1))$$

$$H^2 = \sigma(\mathrm{LayerNorm}(\mathrm{GATConv}^2(H^{1\prime}, A) + \mathrm{skip}^2_{en}(H^{1\prime})))$$

$$H^3 = \sigma(\mathrm{LayerNorm}(\mathrm{GATConv}^3(H^2, A) + \mathrm{skip}^3_{en}(H^2)))$$

where σ is the LeakyReLU activation, and ⊙ denotes element-wise multiplication. Multi-head feature attention is applied as:

$$A = \frac{1}{H}\sum_{h=1}^{H} \sigma'\,(W_a^{(h)} \cdot \mathrm{ReLU}(W_a^{(h)} + b_a^{(h)}) + b_a^{(h)})$$

where H denotes the total number of attention heads, $W_a^{(h)}$ and $b_a^{(h)}$ are the learnable weight matrices and bias terms for attention head *h*, respectively, and σ' is the sigmoid activation function used for gating.

The output of each encoder is a modality-specific latent representation computed as:

$$Z_{mRNA} = \mathrm{GATEncoder}_1(X_{mRNA}, A_{mRNA}) = H^3_{mRNA} \in \mathbb{R}^{N\times d}$$

$$Z_{protein} = \mathrm{GATEncoder}_2(X_{protein}, A_{protein}) = H^3_{protein} \in \mathbb{R}^{N\times d}$$

where d=1024 is the latent dimensionality and N is the number of spots or nodes in the graph. These latent embeddings are subsequently passed to the decoder modules for downstream tasks.

Decoder design: In the DGAT model, decoder modules are tasked with reconstructing mRNA and protein expression from the learned latent embeddings with distinct architecture design. The mRNA decoder is implemented as a fully connected feedforward neural network with residual connections, designed to map latent embeddings back to the original mRNA feature space:

$$D_{mRNA}(Z) = H^l_d = \sigma(W^l_{m_de} \cdot H^{l-1}_{m_de} + b^l_{m_de}) + \mathrm{skip}^l_{m_de}(H^{l-1}_{m_de})$$

where $W^l_{m_de}$ denotes learnable weights for l-th layer in mRNA decoder. $b^l_{m_de}$ is learnable bias term. σ represents LeakyReLU activation. $\mathrm{skip}^l_{m_de}$ is residual (skip connection) projection layer. $H^l_{m_de}$ is the layer output of the mRNA decoder.

The protein decoder consists of a shared two-layer trunk followed by multiple independent branches, one per protein, to model protein-specific reconstructions:

$$H'_{p_de} = \sigma(W^2_{p_de} \cdot \sigma(W^1_{p_de} \cdot Z + b^1_{p_de}) + b^2_{p_de})$$

$$D_p(H'_{p_de}) = \sigma(W^4_{p_de}\sigma(W^3_{p_de}H'_{p_de} + b^3_{p_de}) + b^4_{p_de})$$

$$D_{protein}(Z) = [D_1(H'_{p_de}), D_2(H'_{p_de}), \ldots, D_P(H'_{p_de})]$$

where H'_{p_de} denotes the shared feature extracted by the trunk in the protein decoder. $W^l_{p_de}$ represents the learnable weights for l-th layer. $b^l_{p_de}$ is learnable bias term. D_p is a protein branch which predicts one protein's expression. $D_{protein}$ is formed by concatenating the outputs of each protein-specific branch.

***Loss functions*:** We designed three types of loss functions, each serving a distinct purpose to facilitate mRNA-to-protein prediction. The alignment loss defined as $\mathcal{L}_{align} = MSE(Z_{mRNA}, Z_{protein})$, where MSE is mean square error, aims to maximize the concordance between the mRNA and protein latent embedding spaces. The reconstruction loss, defined as

$$\mathcal{L}^{mRNA}_{recon} = RMSE(X_{mRNA}, D_{mRNA}(Z_{mRNA}))$$

$$\mathcal{L}^{protein}_{recon} = RMSE(X_{protein}, D_{protein}(Z_{protein}))$$

Where RMSE represents the root mean square error, ensures that the latent representations produced by each encoder retain its distinct modality-specific features. Additionally, we introduce a prediction loss by cross-feeding the latent embeddings into the decoder of the opposite modality, defined as

$$\mathcal{L}^{mRNA}_{predict} = RMSE(X_{mRNA}, D_{mRNA}(Z_{protein}))$$

$$\mathcal{L}^{protein}_{predict} = RMSE(X_{protein}, D_{protein}(Z_{mRNA}))$$

supervises and reinforces the model's ability to perform cross-modality inference.

Finally, the total loss function is defined as:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}^{mRNA}_{recon} + \beta \cdot \mathcal{L}^{protein}_{recon} + \gamma \cdot \mathcal{L}_{align} + \delta \cdot \mathcal{L}^{mRNA}_{predict} + \eta \cdot \mathcal{L}^{protein}_{predict}$$

where $\alpha, \beta, \gamma, \delta, \eta$ are hyperparameters which control the relative importance of each loss component.

Model training: The training and evaluation process for the DGAT model is designed to ensure the accurate cross-modality prediction for protein expression from spatial transcriptomics data. The key steps are as follows: The weights $\alpha, \beta, \gamma, \delta, \eta$ were set based on preliminary experiments to $\alpha = 5, \beta = 1, \gamma = 1, \delta = 3, \eta = 5$. The alignment loss directly couples the mRNA and protein latent spaces, which can lead to over-regularization if overemphasized. To enhance the model's generalization, we applied dropout rate of 0.3 for both mRNA encoder and protein encoder. Moreover, to stabilize the training and avoid over-optimization for the alignment loss, the weight of it is set to 0 if the loss is lower than 0.015. Limiting it dynamically retains modality-specific nuances, reducing the risk of both embeddings collapsing into a single, less informative and less biological meaningful space. Separate Adam optimizers are used for each encoder and decoder. The learning rates for encoders are 5×10^{-4}, while they are 1×10^{-4} for decoders. In addition, gradient norms are clipped to 1.0 to prevent exploding. In addition, A small weight decay of 2×10^{-5} is applied to the encoders to improve generalization. The learning rate is reduced by 20% every 10 epochs to encourage convergence. Finally, training is stopped if the total loss does not improve for 10 consecutive epochs.

We adopt a noise-based early stopping criterion ("EB-criterion") that does not require held-out validation data[35]. The key idea is to monitor when the model's parameter updates become dominated by noise rather than useful learning signals. To do this, we estimate the signal-to-noise ratio of parameter gradients during training. When this ratio becomes low—indicating that continued training is unlikely to yield meaningful improvement—we stop training. To ensure robustness, we delay the stopping check until after 50 epochs and smooth the evidence signal by averaging over the most recent 10 epochs. Training stops when this

averaged signal crosses a predefined threshold (0.96), indicating that learning has effectively plateaued.

***Model performance*:** Leave-One-Out Testing (LOOT) is used to evaluate the model's generalization. Each sample in the dataset is treated as a test set while the remaining samples are used for training. This ensures robust performance evaluation across diverse tissue types. To evaluate the accuracy of the protein imputation methods, we calculated the following metrics: (a) Spearman correlation coefficient (ρ): This quantified the rank correlation between the predicted and measured protein expression levels, helping us assess how well the imputed data preserves the ordering of the original data. (b) Root Mean Square Error (RMSE): This metric measured the overall error between imputed and actual protein levels, providing a measure of model accuracy. (c) Average Rank Score (ARS): We computed ARS as a combined metric (ref), incorporating both Spearman correlation and RMSE. This gave us an overall performance indicator, with a higher ARS value representing better accuracy across the evaluation metrics. These metrics were used to assess the accuracy of protein imputation at both the protein level (quantifying the accuracy of protein expression values) and the spot level (assessing accuracy for individual spots).

Benchmarking procedure

We compared DGAT against cTP-net[17], sciPENN[19], scLinear[20] and Seurat:

- cTP-net: Protein abundances were inferred from gene expression matrices using the official cTP-net training script with default parameters. SAVER-X denoising was omitted due to runtime and software dependencies.
- sciPENN: Preprocessing and training model were conducted following default parameters provided in the GitHub repository (https://github.com/jlakkis/sciPENN).

- Seurat v4: Data were processed using SCTransform and dimensionality reduction was performed with PCA. Transfer anchors and protein data imputation were computed using FindTransferAnchors and TransferData with default settings **(https://satijalab.org/seurat/).**
- scLinear: We followed the tutorial provided on the Zenodo https://zenodo.org/records/10602824 with default settings.

Feature importance analysis

To interpret the model's protein prediction behavior and identify key transcriptomic contributors, we performed a feature importance analysis based on gradient attribution. For each protein-specific decoder branch, we computed the gene-level importance scores by aggregating the input-gradient product ($\partial y/\partial x \times x$) across all cells, normalized by z-scoring. This method highlights genes whose expression changes have the strongest influence on protein prediction, thereby capturing both sensitivity and directionality.

Hardware & Reproducibility

All experiments were conducted on NVIDIA l40s GPUs with 40GB memory. Training a single DGAT model took ~1 hour for 6 samples. Code and pretrained models are available at: https://github.com/osmanbeyoglulab/DGAT/. Software dependencies are Python 3.10.8, PyTorch 2.0.1, Scanpy 1.11.1, NumPy 1.24.1, SciPy 1.15.3, scikit-learn 1.2.

Data Availability

The published human tonsil, breast cancer and glioblastoma Visium CytAssist Gene and Protein Expression samples that supports the finding of this study can be downloaded from the 10× Genomics website (https://www.10xgenomics.com/datasets/visium-cytassist-gene-and-protein-expression-library-of-human-tonsil-with-add-on-antibodies-h-e-6-5-mm-ffpe-2-standard;

https://www.10xgenomics.com/datasets/gene-and-protein-expression-library-of-human-glioblastoma-cytassist-ffpe-2-standard; https://www.10xgenomics.com/datasets/gene-and-protein-expression-library-of-human-breast-cancer-cytassist-ffpe-2-standard; https://www.10xgenomics.com/datasets/gene-protein-expression-library-of-human-tonsil-cytassist-ffpe-2-standard). The in-house malignant mesothelioma Visium CytAssist Gene and Protein Expression samples are available in GEO GSE300851. The ST TNBC breast cancer data from Wu et al.[36], including pathological annotations, were accessed through the Zenodo data repository (https://doi.org/10.5281/zenodo.4739739). The published melanoma ST data that supports the finding of this study can be downloaded from http://www.spatialtranscriptomicsresearch.org/. Additionally, raw count matrices, images, and spatial data for prostate, breast cancer, and lymph node 10x Visium datasets are accessible at https://support.10xgenomics.com/spatial-gene-expression/datasets.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HW performed computational experiments and analyses, helped to develop the algorithmic approaches and helped to write the paper. BC conducted the pathological review of mesothelioma samples and performed germinal center annotations for tonsil samples. HUO supervised the project, helped to develop the algorithmic approaches and wrote the paper.

Acknowledgments

We thank Xiaojun Ma and Annalisa Barrata for their valuable feedback on the manuscript. This work was supported by the National Institutes of Health (R35GM146989). This work was also

supported by the Centers for Disease Control and Prevention in association with the National Institute for Occupational Safety and Health and the National Mesothelioma Virtual Bank (U24OH009077). Data analyses were supported by the University of Pittsburgh Center for Research Computing and the Extreme Science and Engineering Discovery Environment (XSEDE), supported by the National Science Foundation (OCI-1053575), through the Bridges-2 system at the Pittsburgh Supercomputing Center (ACI-1445606). Histology sectioning was performed by the Pitt Biospecimen Core (RRID:SCR_025229). 10x Visium library preparation and Illumina sequencing were conducted by the Health Sciences Sequencing Core (RRID:SCR_023116) at UPMC Children's Hospital of Pittsburgh, Rangos Research Center. Additional support for services and instrumentation used in this project was provided, in part, by the University of Pittsburgh, the Office of the Senior Vice Chancellor for Health Sciences, the Department of Pediatrics, and the Richard King Mellon Foundation for Pediatric Research.

References

1 Stahl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).
2 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).
3 Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* **39**, 313-319, doi:10.1038/s41587-020-0739-1 (2021).
4 Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* **16**, 987-990, doi:10.1038/s41592-019-0548-y (2019).
5 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868, doi:10.1038/nmeth.4380 (2017).
6 Ben-Chetrit, N. *et al.* Integration of whole transcriptome spatial profiling with protein markers. *Nat Biotechnol* **41**, 788-793, doi:10.1038/s41587-022-01536-3 (2023).
7 Vickovic, S. *et al.* SM-Omics is an automated platform for high-throughput spatial multi-omics. *Nat Commun* **13**, 795, doi:10.1038/s41467-022-28445-y (2022).
8 Liu, Y. *et al.* High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat Biotechnol* **41**, 1405-1409, doi:10.1038/s41587-023-01676-0 (2023).
9 Liu, Y. *et al.* Spatial-CITE-seq: spatially resolved high-plex protein and whole transcriptome co-mapping. *Res Sq*, doi:10.21203/rs.3.rs-1499315/v1 (2022).

10 Liao, S. *et al.* Integrated Spatial Transcriptomic and Proteomic Analysis of Fresh Frozen Tissue Based on Stereo-seq. *bioRxiv*, 2023.2004.2028.538364, doi:10.1101/2023.04.28.538364 (2023).
11 Burlingame, E. A. *et al.* SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning. *Sci Rep* **10**, 17507, doi:10.1038/s41598-020-74500-3 (2020).
12 Berijanian, M. *et al.* Unsupervised many-to-many stain translation for histological image augmentation to improve classification accuracy. *J Pathol Inform* **14**, 100195, doi:10.1016/j.jpi.2023.100195 (2023).
13 Wu, K. E., Yost, K. E., Chang, H. Y. & Zou, J. BABEL enables cross-modality translation between multiomic profiles at single-cell resolution. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2023070118 (2021).
14 Alatkar, S. A. & Wang, D. CMOT: Cross-Modality Optimal Transport for multimodal inference. *Genome Biol* **24**, 163, doi:10.1186/s13059-023-02989-8 (2023).
15 Cohen Kalafut, N., Huang, X. & Wang, D. Joint variational autoencoders for multimodal imputation and embedding. *Nature Machine Intelligence* **5**, 631-642, doi:https://doi.org/10.1038/s42256-023-00663-z (2023).
16 Lan, M., Zhang, S. & Gao, L. Efficient Generation of Paired Single-Cell Multiomics Profiles by Deep Learning. *Adv Sci (Weinh)* **10**, e2301169, doi:10.1002/advs.202301169 (2023).
17 Zhou, Z., Ye, C., Wang, J. & Zhang, N. R. Surface protein imputation from single cell transcriptomes by deep neural networks. *Nat Commun* **11**, 651, doi:10.1038/s41467-020-14391-0 (2020).
18 Javaid, A. & Frost, H. R. STREAK: A supervised cell surface receptor abundance estimation strategy for single cell RNA-sequencing data using feature selection and thresholded gene set scoring. *PLoS Comput Biol* **19**, e1011413, doi:10.1371/journal.pcbi.1011413 (2023).
19 Lakkis, J. *et al.* A multi-use deep learning method for CITE-seq and single-cell RNA-seq data integration with cell surface protein prediction and imputation. *Nat Mach Intell* **4**, 940-952, doi:10.1038/s42256-022-00545-w (2022).
20 Hanhart, D., Gossi, F., Rapsomaniki, M. A., Kruithof-de Julio, M. & Chouvardas, P. ScLinear predicts protein abundance at single-cell resolution. *Commun Biol* **7**, 267, doi:10.1038/s42003-024-05958-4 (2024).
21 Chen, R., Zhou, J. & Chen, B. Imputing abundance of over 2,500 surface proteins from single-cell transcriptomes with context-agnostic zero-shot deep ensembles. *Cell Syst* **15**, 869-884 e866, doi:10.1016/j.cels.2024.08.006 (2024).
22 Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* **35**, 936-939, doi:10.1038/nbt.3973 (2017).
23 Veličković, P. *et al.* Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
24 Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. in *Proceedings of the AAAI conference on artificial intelligence.*
25 Bresson, X. & Laurent, T. Residual gated graph convnets. *arXiv preprint arXiv:1711.07553* (2017).
26 Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).
27 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
28 Li, C. Y., Hong, Y. J., Li, B. & Zhang, X. F. Benchmarking single-cell cross-omics imputation methods for surface protein expression. *Genome Biol* **26**, 46, doi:10.1186/s13059-025-03514-9 (2025).

29 Koo, P. K., Majdandzic, A., Ploenzke, M., Anand, P. & Paul, S. B. Global importance analysis: An interpretability method to quantify importance of genomic features in deep neural networks. *PLoS Comput Biol* **17**, e1008925, doi:10.1371/journal.pcbi.1008925 (2021).
30 Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* **40**, 661-671, doi:10.1038/s41587-021-01139-4 (2022).
31 Thrane, K., Eriksson, H., Maaskola, J., Hansson, J. & Lundeberg, J. Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. *Cancer Res* **78**, 5970-5979, doi:10.1158/0008-5472.CAN-18-0747 (2018).
32 Ma, X. *et al.* Spatial landscape of malignant pleural and peritoneal mesothelioma tumor immune microenvironment. *bioRxiv*, 2023.2009.2006.556559, doi:10.1101/2023.09.06.556559 (2023).
33 Hmeljak, J. *et al.* Integrative Molecular Characterization of Malignant Pleural Mesothelioma. *Cancer Discov* **8**, 1548-1565, doi:10.1158/2159-8290.CD-18-0804 (2018).
34 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).
35 Mahsereci, M., Balles, L., Lassner, C. & Hennig, P. Early stopping without a validation set. *arXiv preprint arXiv:1703.09580* (2017).
36 Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334-1347, doi:10.1038/s41588-021-00911-1 (2021).
37 Jiang, X. *et al.* iIMPACT: integrating image and molecular profiles for spatial transcriptomics analysis. *Genome Biol* **25**, 147, doi:10.1186/s13059-024-03289-5 (2024).

Figures Legends

Figure 1: **Schematic of DGAT framework. (a)** DGAT integrates spatial transcriptomics and proteomics by constructing modality-specific heterogenous graphs based on molecular similarity and spatial proximity. Graph Attention Networks generate spatially-aware embeddings, which are aligned via latent space loss and cross-model supervision. These embeddings are decoded through a multilayer perceptron for mRNA reconstruction and a multi-branch decoder for protein prediction, enabling protein imputation from transcript-only data **(b)**.

Figure 2: **Protein prediction performance of DGAT in held-out spatial CITE-seq samples. (a)** The heatmap shows correlations between measured protein abundances and DGAT-predicted protein abundance across different samples compared to normalized RNA expression. The highest correlation value for each protein is highlighted with a pink box. **(b)** Spatial maps of samples are shown, colored by predicted protein abundance, normalized mRNA expression, or measured protein abundance of CD68 across all samples. Correlation values between predicted protein abundance (or normalized mRNA expression) and measured protein abundance are displayed in the top-left corner of each plot. All protein and RNA expression levels (counts) are shown on a log scale. **(c)** Performance of DGAT on held-out spatial CITE-seq samples compared to mRNA, cTP-net, scLinear, sciPENN and Seurat v4. Boxplots show mean Spearman correlations between predicted and actual protein expression across spots. **(d)** Barplot for ARS value of each method at the protein levels. Data are presented as the average of the six ARS values (from six datasets, six repetitions) across all experiments. Methods are ordered according to their performance.

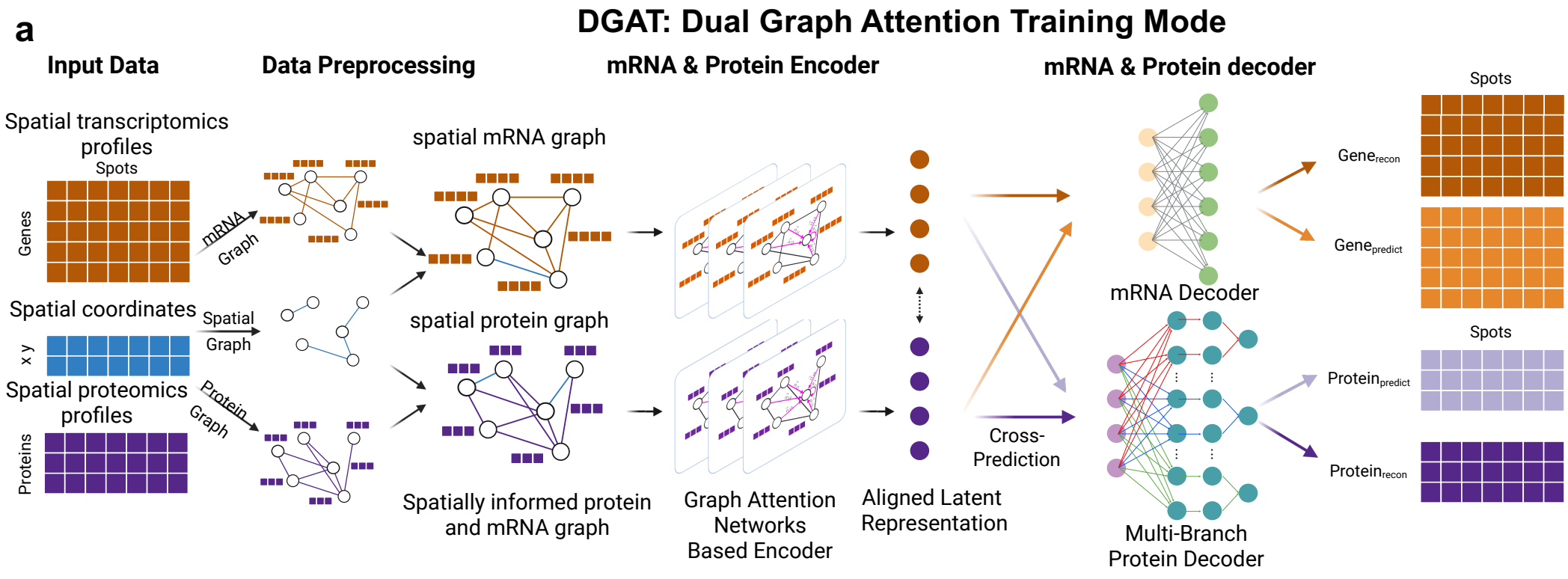
Figure 3: DGAT infers spatial protein expression and identifies germinal center-associated proteins in human lymph node tissue. (a) H&E-stained lymph node section with

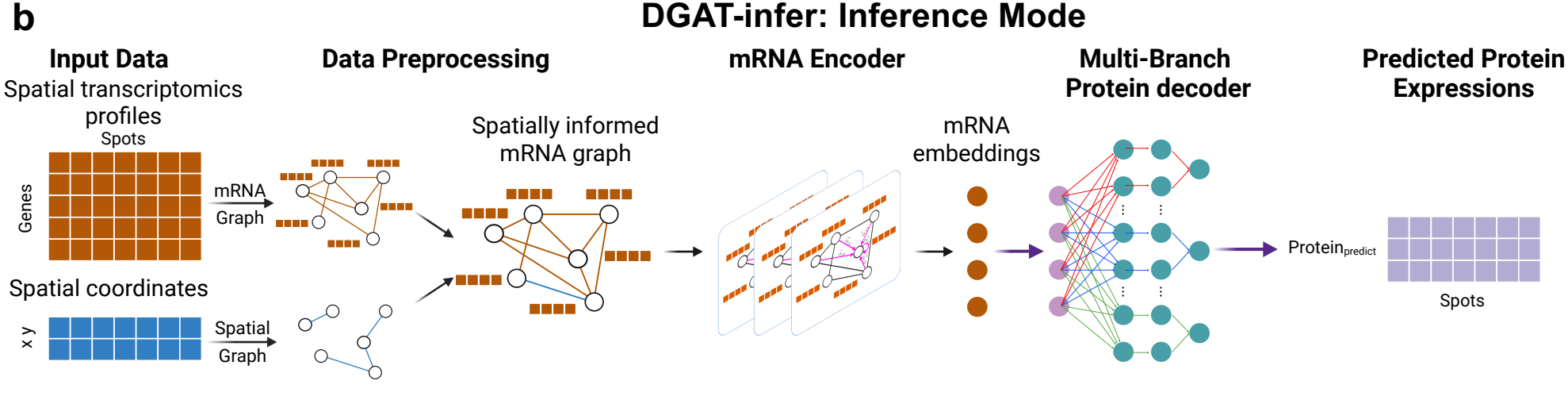
pathologist-annotated germinal centers (GCs). **(b)** Spatial clustering of tissue spots based on predicted protein expression from DGAT, cTP-net, scLinear, sciPENN, Seurat (anchor transfer), and raw mRNA expression, using the Leiden algorithm. DGAT achieves the highest F1 score (0.823), indicating superior spatial resolution in distinguishing GC from non-GC regions. **(c-d)** H&E-stained tonsil sections with pathologist-defined GCs used for benchmarking. **e)** Differential protein expression (GC vs. other regions) across DGAT-predicted proteins (lymph node), experimentally measured proteins (two tonsil replicates), and GC-enriched mRNA signals (mRNA_GC_LN). Circle size reflects statistical significance ($-\log_{10}$ adjusted p); color indicates direction of enrichment. **(f)** Spatial expression of CD19, CXCR5, MS4A1, PAX5, and PD1 in lymph node (top, DGAT-predicted) and tonsil tissues (middle and bottom, measured). DGAT recapitulates GC-localized protein expression patterns consistent with experimental data. All comparisons are statistically significant (adjusted $p < 1\times10^{-15}$).

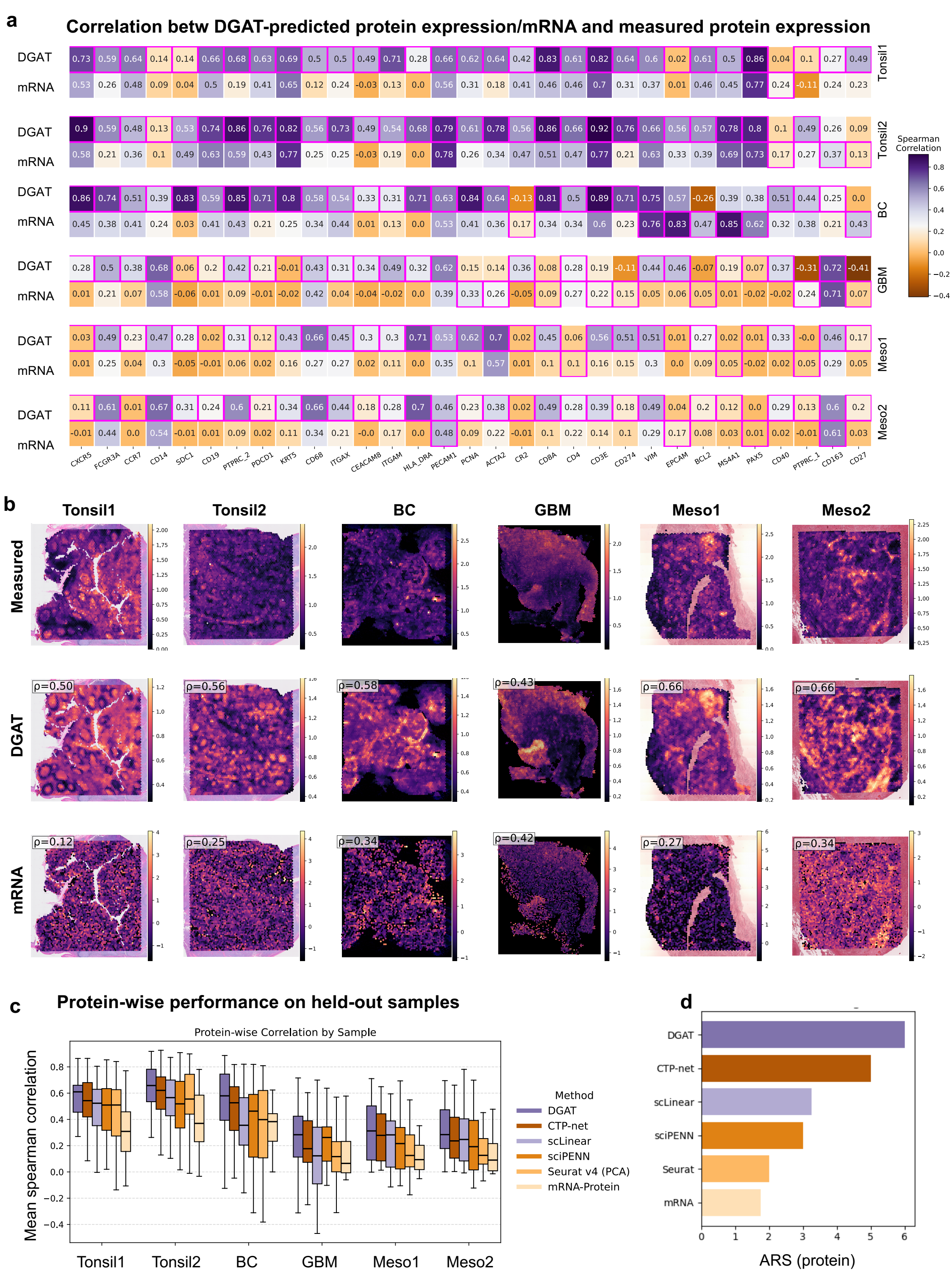
Figure 4: Application of DGAT to a human triple-negative breast cancer spatial transcriptomics dataset. (a) H&E-stained image, pathologist annotations, and **(b)** spatial clustering of a triple-negative breast cancer (TNBC) sample from Wu *et al.* based on DGAT-predicted protein expression and raw mRNA expression. Clustering performance in recovering annotated pathological regions is quantified using the adjusted Rand index (ARI). **(c)** Dot plot showing mean DGAT-predicted protein expression differences between each pathological region and all others in sample TNBC. Dot size represents statistical significance ($-\log_{10}$ FDR) from the Wilcoxon rank-sum test. Rows and columns are hierarchically clustered. Displayed proteins are the top-ranked (lowest *P*-value) for each pathological region.

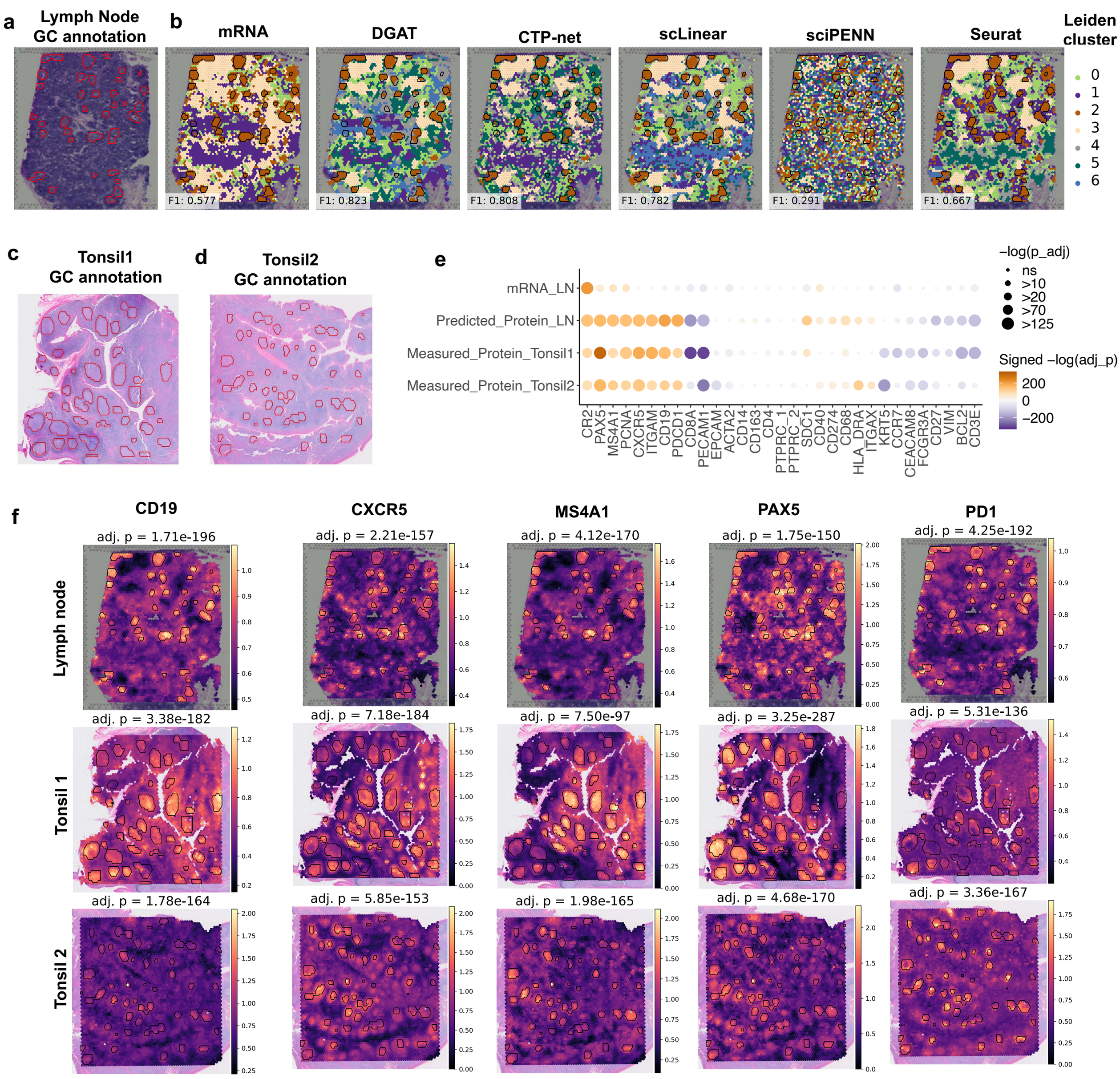
Figure 5: DGAT accurately infers spatial protein expression in previously unseen tissue types. (a) Histological annotations of the melanoma samples in the original study[31]**. (b)** Spatial clustering of melanoma tissue spots based on predicted protein expression from DGAT, cTP-

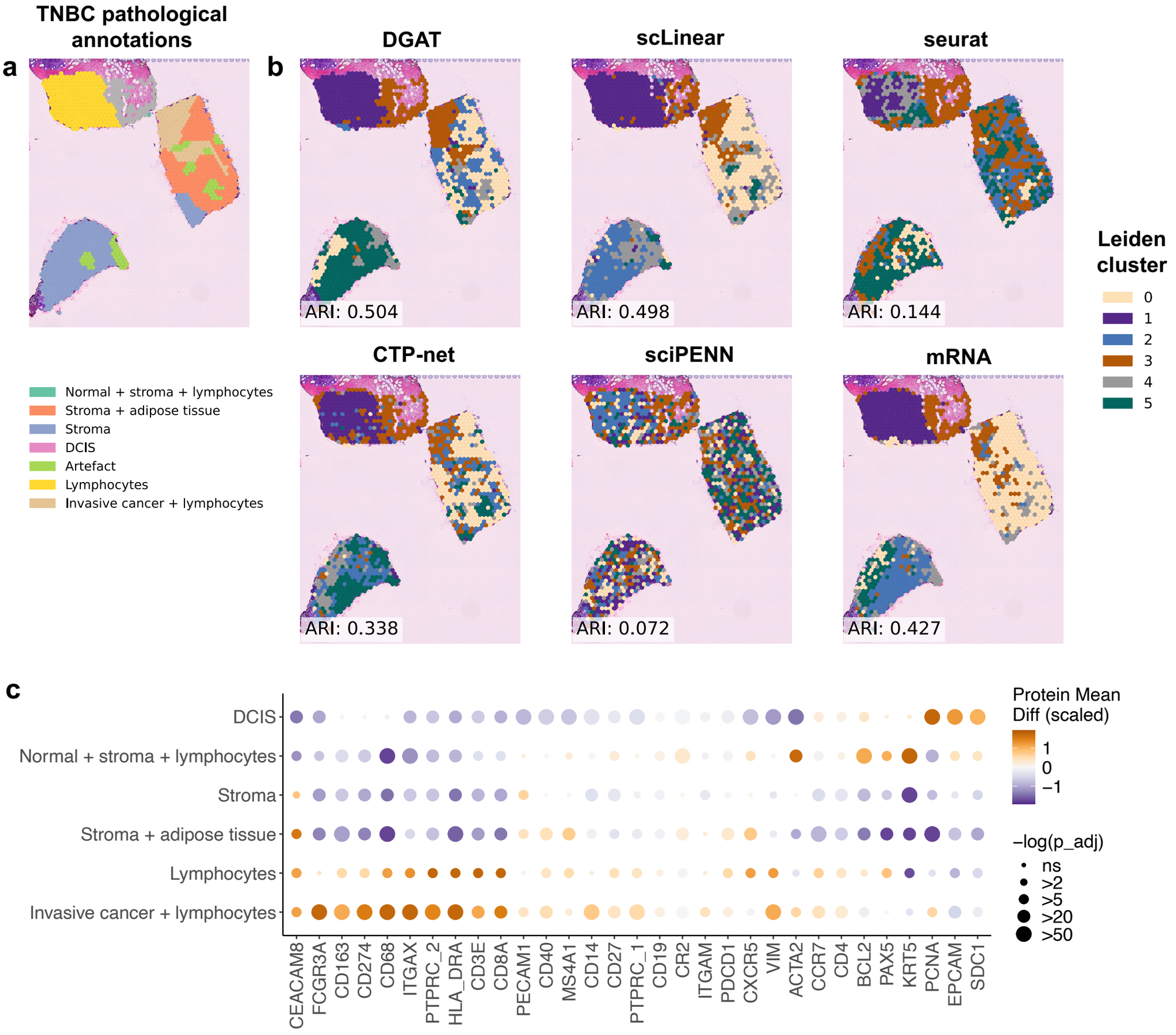
net, scLinear, sciPENN, Seurat (anchor transfer), and raw mRNA expression using the Leiden algorithm. Clustering accuracy in recovering annotated pathological regions is assessed using the adjusted Rand index (ARI**). (c)** Dot plot showing mean DGAT-predicted protein expression differences between each pathological region and all others in melanoma. Dot size represents statistical significance ($-\log_{10}$ FDR) from the Wilcoxon rank-sum test; rows and columns are hierarchically clustered. **(d)** Histological annotations of the human prostate cancer samples in the original study[37]. **(e)** Spatial clustering of prostate cancer tissue spots based on protein predictions from DGAT and baseline methods, as in **b**. **(f)** Dot plot of DGAT-predicted protein expression differences between pathological regions in prostate cancer, with statistical significance and clustering visualized as in **f**.

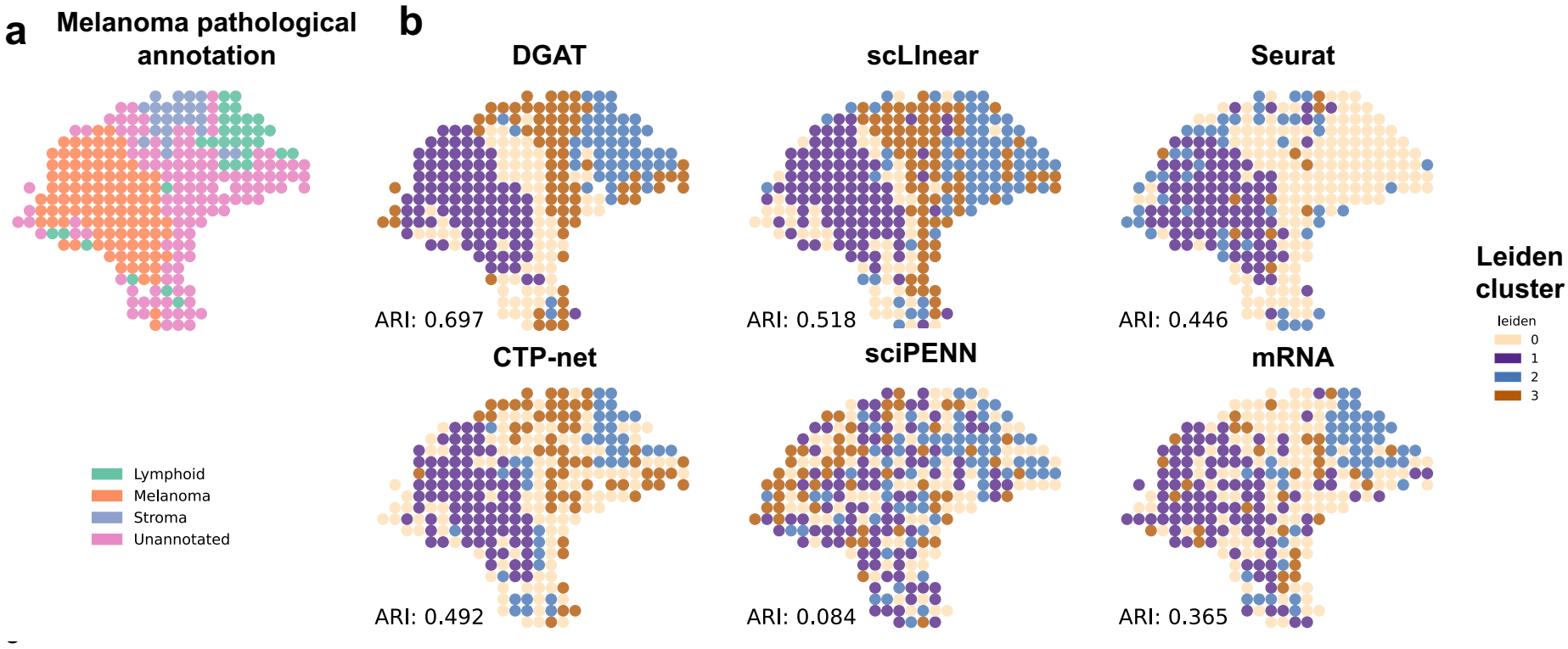

a
DGAT: Dual Graph Attention Training Mode
Input Data
Data Preprocessing
mRNA & Protein Encoder
mRNA & Protein decoder
Spatial transcriptomics profiles
Spots
Genes
mRNA Graph
spatial mRNA graph
Spatial coordinates
x y
Spatial Graph
spatial protein graph
Spatial proteomics profiles
Proteins
Protein Graph
Spatially informed protein and mRNA graph
Graph Attention Networks Based Encoder
Aligned Latent Representation
Cross-Prediction
mRNA Decoder
Multi-Branch Protein Decoder
Gene recon
Gene predict
Spots
Protein predict
Protein recon
Spots

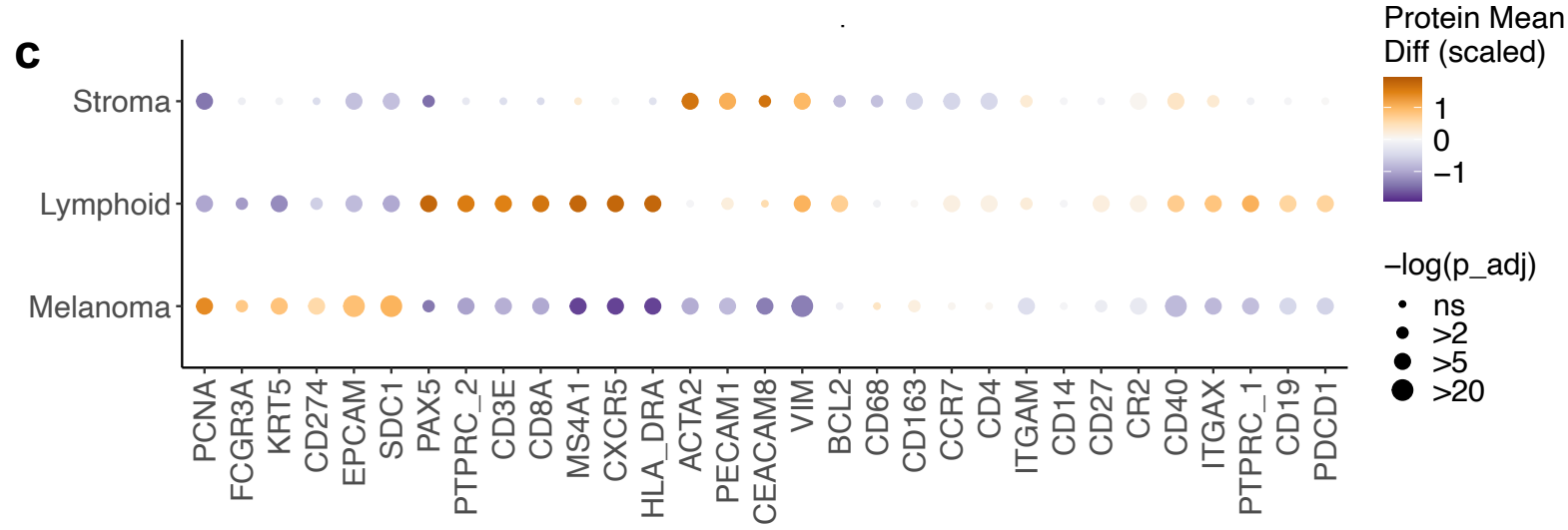

b
DGAT-infer: Inference Mode
Input Data
Data Preprocessing
mRNA Encoder
Multi-Branch Protein decoder
Predicted Protein Expressions
Spatial transcriptomics profiles
Spots
Genes
mRNA Graph
Spatially informed mRNA graph
Spatial coordinates
x y
Spatial Graph
mRNA embeddings
Protein predict
Spots

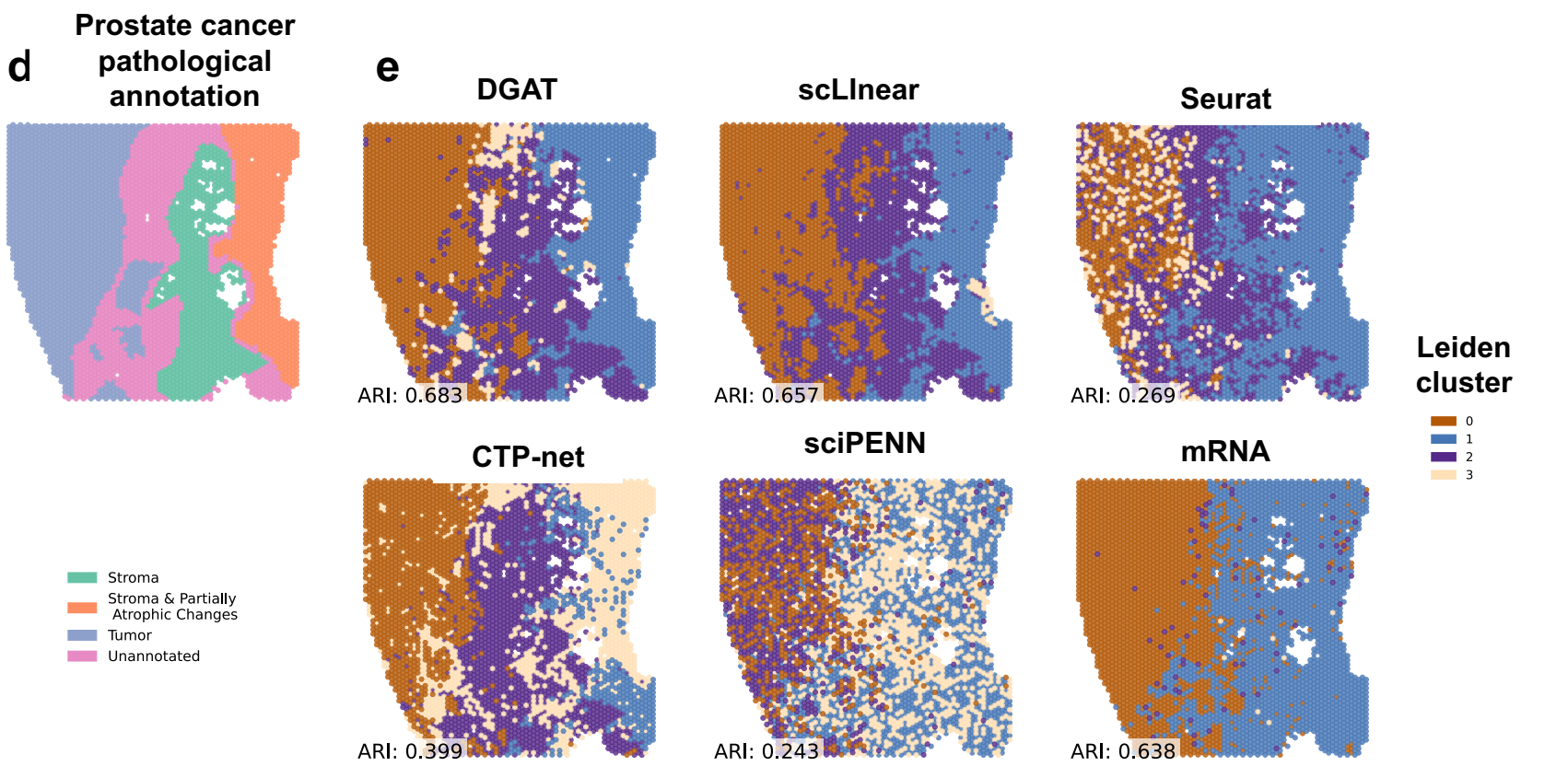

a
Correlation betw DGAT-predicted protein expression/mRNA and measured protein expression
DGAT
mRNA
Tonsil1
Tonsil2
BC
GBM
Meso1
Meso2
CXCR5 FCGR3A CCR7 CD14 SDC1 CD19 PTPRC_2 PDCD1 KRT5 CD68 ITGAX CEACAM8 ITGAM HLA_DRA PECAM1 PCNA ACTA2 CR2 CD8A CD4 CD3E CD274 VIM EPCAM BCL2 MS4A1 PAX5 CD40 PTPRC_1 CD163 CD27
Spearman Correlation
b
Tonsil1
Tonsil2
BC
GBM
Meso1
Meso2
Measured
DGAT
mRNA
ρ=0.50
ρ=0.56
ρ=0.58
ρ=0.43
ρ=0.66
ρ=0.66
ρ=0.12
ρ=0.25
ρ=0.34
ρ=0.42
ρ=0.27
ρ=0.34
c
Protein-wise performance on held-out samples
Protein-wise Correlation by Sample
Mean spearman correlation
Tonsil1
Tonsil2
BC
GBM
Meso1
Meso2
Method
DGAT
CTP-net
scLinear
sciPENN
Seurat v4 (PCA)
mRNA-Protein
d
DGAT
CTP-net
scLinear
sciPENN
Seurat
mRNA
ARS (protein)

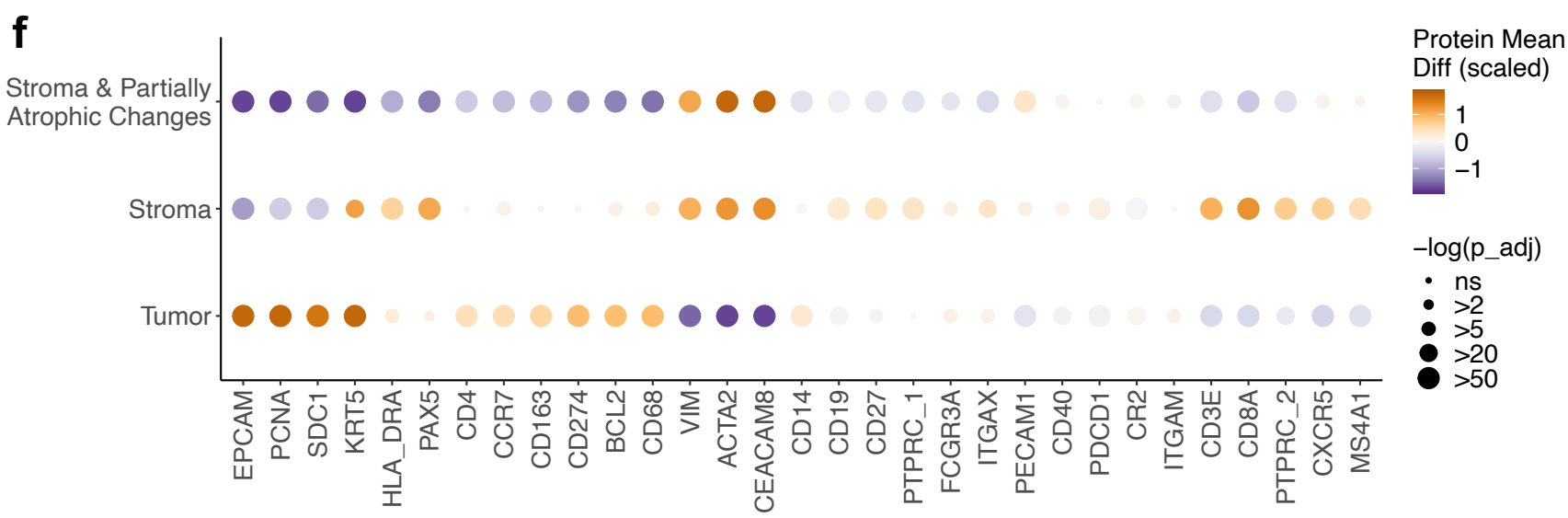

a
Lymph Node GC annotation
b
mRNA
DGAT
CTP-net
scLinear
sciPENN
Seurat
Leiden cluster
0
1
2
3
4
5
6
F1: 0.577
F1: 0.823
F1: 0.808
F1: 0.782
F1: 0.291
F1: 0.667
c
Tonsil1 GC annotation
d
Tonsil2 GC annotation
e
mRNA_LN
Predicted_Protein_LN
Measured_Protein_Tonsil1
Measured_Protein_Tonsil2
CR2 PAX5 MS4A1 PCNA CXCR5 ITGAM CD19 PDCD1 CD8A PECAM1 EPCAM ACTA2 CD14 CD163 CD4 PTPRC_1 PTPRC_2 SDC1 CD40 CD274 CD68 HLA_DRA ITGAX KRT5 CCR7 CEACAM8 FCGR3A CD27 VIM BCL2 CD3E
–log(p_adj)
ns
>10
>20
>70
>125
Signed –log(adj_p)
200
0
–200
f
CD19
CXCR5
MS4A1
PAX5
PD1
Lymph node
Tonsil 1
Tonsil 2
adj. p = 1.71e-196
adj. p = 2.21e-157
adj. p = 4.12e-170
adj. p = 1.75e-150
adj. p = 4.25e-192
adj. p = 3.38e-182
adj. p = 7.18e-184
adj. p = 7.50e-97
adj. p = 3.25e-287
adj. p = 5.31e-136
adj. p = 1.78e-164
adj. p = 5.85e-153
adj. p = 1.98e-165
adj. p = 4.68e-170
adj. p = 3.36e-167


TNBC pathological annotations
a
b
DGAT
scLinear
seurat
CTP-net
sciPENN
mRNA
ARI: 0.504
ARI: 0.498
ARI: 0.144
ARI: 0.338
ARI: 0.072
ARI: 0.427
Normal + stroma + lymphocytes
Stroma + adipose tissue
Stroma
DCIS
Artefact
Lymphocytes
Invasive cancer + lymphocytes
Leiden cluster
0
1
2
3
4
5
c
DCIS
Normal + stroma + lymphocytes
Stroma
Stroma + adipose tissue
Lymphocytes
Invasive cancer + lymphocytes
CEACAM8
FCGR3A
CD163
CD274
CD68
ITGAX
PTPRC_2
HLA_DRA
CD3E
CD8A
PECAM1
CD40
MS4A1
CD14
CD27
PTPRC_1
CD19
CR2
ITGAM
PDCD1
CXCR5
VIM
ACTA2
CCR7
CD4
BCL2
PAX5
KRT5
PCNA
EPCAM
SDC1
Protein Mean Diff (scaled)
1
0
−1
−log(p_adj)
ns
>2
>5
>20
>50


a
Melanoma pathological annotation
Lymphoid
Melanoma
Stroma
Unannotated
b
DGAT
ARI: 0.697
scLInear
ARI: 0.518
Seurat
ARI: 0.446
CTP-net
ARI: 0.492
sciPENN
ARI: 0.084
mRNA
ARI: 0.365
Leiden cluster
leiden
0
1
2
3


c
Stroma
Lymphoid
Melanoma
PCNA
FCGR3A
KRT5
CD274
EPCAM
SDC1
PAX5
PTPRC_2
CD3E
CD8A
MS4A1
CXCR5
HLA_DRA
ACTA2
PECAM1
CEACAM8
VIM
BCL2
CD68
CD163
CCR7
CD4
ITGAM
CD14
CD27
CR2
CD40
ITGAX
PTPRC_1
CD19
PDCD1
Protein Mean Diff (scaled)
1
0
−1
−log(p_adj)
ns
>2
>5
>20


Prostate cancer
d
pathological annotation
Stroma
Stroma & Partially Atrophic Changes
Tumor
Unannotated
e
DGAT
ARI: 0.683
scLInear
ARI: 0.657
Seurat
ARI: 0.269
CTP-net
ARI: 0.399
sciPENN
ARI: 0.243
mRNA
ARI: 0.638
Leiden cluster
0
1
2
3


f
Stroma & Partially Atrophic Changes
Stroma
Tumor
EPCAM
PCNA
SDC1
KRT5
HLA_DRA
PAX5
CD4
CCR7
CD163
CD274
BCL2
CD68
VIM
ACTA2
CEACAM8
CD14
CD19
CD27
PTPRC_1
FCGR3A
ITGAX
PECAM1
CD40
PDCD1
CR2
ITGAM
CD3E
CD8A
PTPRC_2
CXCR5
MS4A1
Protein Mean Diff (scaled)
1
0
−1
−log(p_adj)
ns
>2
>5
>20
>50