Inferring Spatially Resolved Transcriptomics Data from Whole Slide Images for the Assessment of Colorectal Tumor Metastasis: A Feasibility Study

Michael Fatemi[1], Eric Feng[2], Cyril Sharma[3], Zarif Azher[2], Tarushii Goel[4], Ojas Ramwala[5], Scott Palisoul[6], Rachael Barney[6], Laurent Perreard[7], Fred Kolling[7], Lucas A. Salas[8,9,10], Brock C. Christensen[8,9,11], Gregory Tsongalis[6], Louis Vaickus[6], Joshua J. Levy[6,8,12,13,*]

1. Department of Computer Science, University of Virginia, Charlottesville, VA
2. Thomas Jefferson High School for Science and Technology, Alexandria, VA
3. Department of Computer Science, Purdue University, West Lafayette, IN
4. Department of Computer Science, Massachusetts Institute of Technology, Cambridge, MA
5. Department of Computer Science, University of Washington, Seattle, WA
6. Emerging Diagnostic and Investigative Technologies, Department of Pathology and Laboratory Medicine, Dartmouth Health, Lebanon, NH
7. Dartmouth Cancer Center, Lebanon, NH
8. Department of Epidemiology, Dartmouth College Geisel School of Medicine, Hanover, NH
9. Department of Molecular and Systems Biology, Dartmouth College Geisel School of Medicine, Hanover, NH
10. Integrative Neuroscience at Dartmouth (IND) graduate program, Dartmouth College Geisel School of Medicine, Hanover, NH
11. Department of Community and Family Medicine, Dartmouth College Geisel School of Medicine, Hanover, NH
12. Department of Dermatology, Dartmouth Health, Lebanon, NH
13. Program in Quantitative Biomedical Sciences, Dartmouth College Geisel School of Medicine, Hanover, NH

* To whom correspondence should be addressed: joshua.j.levy@dartmouth.edu

Corresponding Author Contact Information:
Joshua J. Levy PhD
Assistant Professor of Pathology and Dermatology
Adjunct Assistant Professor of Epidemiology
Faculty, Quantitative Biomedical Sciences
Machine Learning Co-Director, Emerging Diagnostic and Investigative Technologies
Biostatistics and Bioinformatics Shared Resource, Dartmouth Cancer Center
Dartmouth-Hitchcock Medical Center
1 Medical Center Drive, Department of Pathology and Laboratory Medicine, Lebanon, NH 03756
Phone: (925) 457-5752 | Email: joshua.j.levy@dartmouth.edu

41   **Abstract**

42   Over 150,000 Americans are diagnosed with colorectal cancer (CRC) every year, and annually
43   over 50,000 individuals will die from CRC, necessitating improvements in screening,
44   prognostication, disease management, and therapeutic options. Tumor metastasis is the
45   primary factor related to the risk of recurrence and mortality. Yet, screening for nodal and distant
46   metastasis is costly, and invasive and incomplete resection may hamper adequate assessment.
47   Signatures of the tumor-immune microenvironment (TIME) at the primary site can provide
48   valuable insights into the aggressiveness of the tumor and the effectiveness of various
49   treatment options. Spatially-resolved transcriptomics technologies offer an unprecedented
50   characterization of TIME through high multiplexing, yet their scope is constrained by cost.
51   Meanwhile, it has long been suspected that histological, cytological and macroarchitectural
52   tissue characteristics correlate well with molecular information (e.g., gene expression). Thus, a
53   method for predicting transcriptomics data through inference of RNA patterns from whole slide
54   images (WSI) is a key step in studying metastasis at scale. In this work, we collected and
55   preprocessed Visium spatial transcriptomics data (17,943 genes at up to 5,000 spots per patient
56   sampled in a honeycomb pattern) from tissue across four stage-III matched colorectal cancer
57   patients. We compare and prototype several convolutional, Transformer, and graph
58   convolutional neural networks to predict spatial RNA patterns under the hypothesis that the
59   transformer and graph-based approaches better capture relevant spatial tissue architecture. We
60   further analyzed the model's ability to recapitulate spatial autocorrelation statistics using SPARK
61   and SpatialDE. Overall, results indicate that the transformer and graph-based approaches were
62   unable to outperform the convolutional neural network architecture, though they exhibited
63   optimal performance for relevant disease-associated genes. Initial findings suggest that different
64   neural networks that operate on different scales are relevant for capturing distinct disease
65   pathways (e.g., epithelial to mesenchymal transition). We add further evidence that deep
66   learning models can accurately predict gene expression in whole slide images and comment on
67   understudied factors which may increase its external applicability (e.g., tissue context). Our
68   preliminary work will motivate further investigation of inference for molecular patterns from
69   whole slide images as metastasis predictors and in other applications.
70
71   **Keywords:** Spatial transcriptomics, deep learning, graph neural network, transformers,
72   colorectal cancer, histomorphology

1. Introduction

Colorectal cancer is the third leading cause of cancer-related death in the United States, and there are disparities in screening and outcomes between age, race, and gender [1]. CRC incidence is rising among younger adults who are not typically incorporated into screening programs, illustrating the importance of developing timely and lower-cost prognostication methods to better assess the tumor's malignant potential. Currently, the Pathological TNM-staging system (pTNM), which determines staging based on the impact of local invasiveness– histological stage (T-stage), and metastatic potential– nodal (N-stage) and distant (M-stage) metastasis, is the most predictive factor for risk of recurrence and prognosis. Metastasis, in many cases, is challenging to assess at the time of primary tumor resection [2]. For instance, specimen inadequacy often hinders the complete inference of nodal involvement [3]. It is thus crucial to develop orthogonal, less invasive, but equally informative technologies that can shed light on disease pathology and prognosis. One promising direction is to study the Tumor Immune Microenvironment (TIME)– the amalgamation of immune cells, chemokines, cytokines, and other immune modulators, etc. that accrete at the invasive front and inside the tumor at the primary site [4–6]. Recent studies have demonstrated that monocyte/lymphocyte immune infiltrates and their spatial distribution, density, and relationships play an important role in providing a coordinated anti-tumoral response. Yet, the full importance of TIME has not been elucidated, as most clinical studies consider either a few canonical markers at a time (e.g., immunoscore, which assesses cytotoxicity at the primary site) or only study cell mixtures which lack a single-cell or spatial dimension [7].

Spatially-resolved transcriptomics (spatial omics), as enabled through technologies such as 10x Genomics Spatial Transcriptomics (ST, Pleasanton, CA) or Nanostring GeoMX Digital Spatial Profiling (DSP, Seattle, WA), is an actively growing area of research that provides rich information about how different areas of tissue interact by analyzing highly multiplexed gene expression at staggering spatial resolution. These technologies can be configured to study the distribution, density, and spread of tumor-infiltrating lymphocytes (TILs) as they may relate to concomitant somatic alterations [8,9]. Assay costs are currently exceedingly high, as profiling just four capture areas can cost tens of thousands of dollars, though costs are being driven downwards with new advances in chemistry and lower sequencing costs. Thus, sufficiently powering spatial transcriptomic association studies or extending their generalizability of the inferences to specific patient subgroups that lie outside of these small cohorts is challenging due to cost. In comparison, tissue slides stained with hematoxylin and eosin (H&E) to assess tissue histomorphology are routinely ordered at a very low cost, and there is ample evidence to suggest that many concurrent molecular alterations coincide with morphological features. Thus, the prediction of RNA expression using image data across a slide presents an opportunity to reveal critical prognostic information for patients at a lower cost, which can motivate relevant downstream analyses.

Deep learning approaches, which rely on using multi-layer artificial neural networks (ANN), have proven instrumental for image analyses in the context of digital pathology [10]. Of relevance for this study is the assessment of whole slide images (WSI), digitized tissue slides, from which machine learning applications can predict the primary site of a metastatic lesion, tumor stage, and the outcome of immunohistochemical stains. Convolutional neural networks (CNNs), a type of predictive machine learning model, are powerful tools for extracting dense information from high-dimensional image data. Prior works have employed these algorithms to extract morphological features from H&E-stained tissue to complement whole transcriptome analyses. As WSI can extend to hundreds of thousands of pixels along each spatial dimension, they are usually broken into subimages to enable efficient computation. Of relevance to our research

123   topic, He et al. (2020) used a DenseNet-101 model to regress on co-localized gene expression
124   levels[11], and Levy-Jurgensen et al. (2020) employed an InceptionV3 model to detect
125   dichotomized gene expression for given patches of tissue[12]. However, these techniques do not
126   analyze the potential for integrating spatial context outside the patch-level (i.e., spatially
127   correlated patches are assumed to be independent and identically distributed), and, therefore
128   may miss larger macroarchitectural contextual cues of aberrant expression.
129
130   Zeng et al. (2022) and Pang et al. (2021) investigated approaches to integrate broader spatial
131   context into their gene expression prediction models by using vision transformers and
132   demonstrated that it is possible to outperform convolutional architectures that make predictions
133   on individual patches (e.g., ST-Net) [13,14]. However, there remain several unknowns, such as 1)
134   the scale of tissue features relevant for inferring RNA, 2) how well these models preserve global
135   spatial expression characteristics (e.g., patterns of clustering), 3) whether relevant domain
136   knowledge can make inferences more informative, and 4) whether these effects depend on the
137   specific genes under study (i.e., what resolution/architectural context is optimal for specific
138   genes and what does it say about the tumor biology). Furthermore, many spatial omics studies
139   attempt to study one slide, where potential endogeneity is introduced by a matter of spatial
140   distance between collection sites.
141
142   Here, we compared a convolutional neural network and a graph attention network for inferring
143   spatially co-registered gene expression from WSI. We comment on the role the tissue
144   macroarchitecture may play in RNA inference as elucidated by changing subimage size and the
145   use of contextual models. We also assessed performance on a subset of immune-related genes
146   and how spatial characterization varies across these models/factors and slides. These
147   explorations will motivate future downstream work to characterize factors pertaining to tumor
148   nodal/distant metastasis.
149

150   **2. Material and Methods**
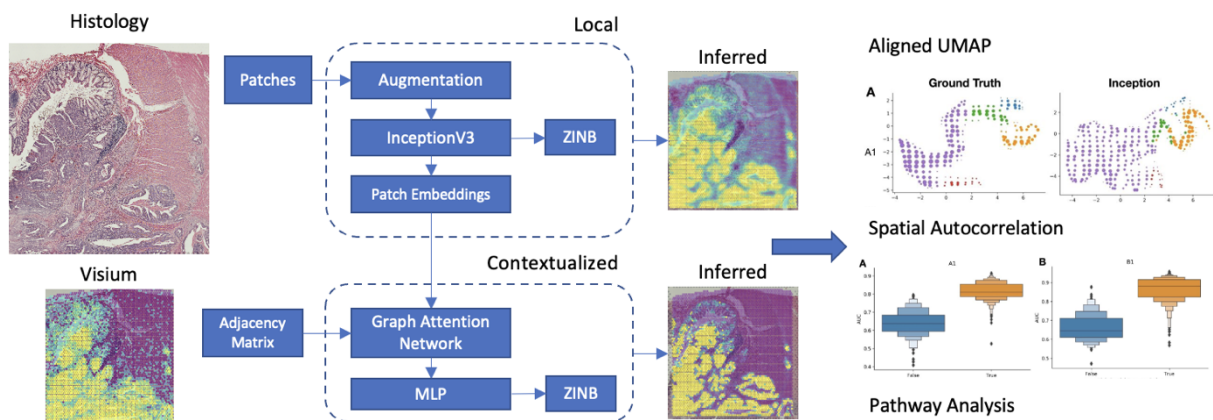
151
152
153   **Figure 1**. **Overview of tested model architectures:** Whole Slide Images are divided into
154   patches co-localized with the Visium spots and an Inception model is used to predict counts and
155   dichotomized expression for 1000 genes; Features derived from Inception model are
156   additionally fine-tuned using graph neural network for inference; inferred expression profiles are
157   compared to the ground truth through a cluster analysis, spatial autocorrelation tests and
158   pathway analysis

2.1. Overview

The primary goal of this work is to predict the gene expression detected by a Visium spot at any given location on the slide. Our method is as follows (**Figure 1**):

1. **Data Collection:** Acquire H&E whole slide images (WSI), and spatially-registered Visium assayed spatial transcriptomics slides from 4 stage-pT3 matched colorectal cancer patients at Dartmouth Hitchcock Medical Center, two patients without metastasis, one with nodal metastasis only and one with both nodal and distant metastasis.
2. **Preprocess**: Preprocess gene expression and WSI subarrays.
3. **Model Development**: Configure two modeling approaches: convolutional neural network and graph attention neural networks, the latter leveraging larger spatial context. These models will predict both binary (dichotomized expression) and count-based (continuous; zero-inflated negative binomial) objectives. We also configured additional approaches (e.g., Transformer) for comparison.
4. **Capture surrounding tissue context at different scales**: Ablation study over patch size to determine whether relevant biological information is encoded outside the Visium collection spot area.
5. **Leave one-patient-out cross-validation**: Evaluation on held-out slides/patients as a measure of external applicability.
6. **Recover Spatial Biology Inferences:** Spatial autocorrelation tests for the capacity of models to draw similar spatial inferences.

2.2. Data Collection

The primary dataset utilized in this study was acquired from four pathologic T Stage-III (pT3) matched (pTNM system) colorectal cancer patients at Dartmouth Hitchcock Medical Center, determined through a retrospective review of pathology reports from 2016 to 2019 following IRB review and approval. These four patients were drawn from a set of 36 patients included in a prior study [15]– half of the patients had concurrent tumor metastasis (slides A1 and B1 had tumor metastasis; slides C1 and D1 did not have tumor metastasis) and were otherwise matched on age, sex, tumor grade, tissue size, mismatch repair (MMR) status, and tumor site using iterative patient resampling with t-tests for continuous variables and fisher's exact tests for categorical variables. The four patients were subselected to restrict the tumor site (three in the right colon, one in the transverse colon), grade (three grade 1, one grade 2), node status (two metastasis cases with N-1a), and account for differences in sex (50% female within both the non-metastasis and metastasis groups). We restricted the cohort to patients without MMR deficiencies as determined through immunohistochemistry (IHC) to control for microsatellite instability status. Tissue blocks were sectioned into 10-micron thick layers, and specific capture areas that contained various distinct macroarchitectural regions (all containing epithelium, tumor-invasive front, intratumoral, lymphatics, etc.) were annotated by the practicing pathologist. A histotechnician carefully extracted / manually cut these capture areas from the tissue, and slides were sent to the Single Cell Genomics Core in the Center for Quantitative Biology for simultaneous H&E staining, imaging, and Visium profiling. After a deparaffinization step, spatial transcriptomics uses spatially-tagged oligo barcodes to 1) register spatial coordinates to collection spots, bound to the mRNA with a poly(A) tail for 2) reverse transcription into cDNA, after 3) permeabilization, and 4) sequencing for mRNA profiling. This allows for unbiased/gridded profiling of up to 5,000 spots (1-10 cells/spot) per 6.5mm by 6.5mm capture area.

2.3. Preprocessing

Spatial gene expression profiles contain information for 17,943 genes at almost 5,000 locations per slide (after filtering out non-tissue– total number of Visium dots: 4950, 4922, 4887, and 4169 per slide), sampled in a honeycomb formation. Each Visium spot covers a circular capture area with a diameter of 130 pixels at 20x magnification. After sequencing, we used the *SpaceRanger* package to preprocess the Visium reads into gene count matrices.

As whole slide images (WSIs) derived from the Visium capture areas (size of capture area– 6.5 x 6.5 mm) span thousands of pixels along each dimension, we subdivided the image into square patches centered on the Visium spot. We associated the gene expression of each patch based on the Visium spot at the center of the patch and ignored expression at other spots contained within the patch.

2.4. Inference Targets

We used the SpatialDE library to select the top 1000 genes based on their mean fraction of spatial variance (FSV) across all slides (i.e., selected genes which exhibited the greatest spatial variation across the four slides). We tested the capacity of our models to recover expression for all 1000 genes based on dichotomized expression (binary classification) and the original counts (regression) [16].

For binary classification, we classify tiles as having a "high" expression for a gene if its individual expression is greater than the median for that slide, as shown by Levy-Jurgensen et al.. For binary tasks, we used a weighted binary cross-entropy loss. The loss was independently taken for positive and negative Visium spots and summed together to account for unbalanced labels.

We model the gene expression distribution for regression tasks as a negative binomial distribution with zero inflation. The model predicts the parameters of the distribution (mean μ, dispersion factor σ, and inflation of zero count π) and is optimized with negative log-likelihood loss. The inferred value is equal to the expected value of the negative binomial distribution (accounting for zero inflation): $(1 - \pi)\mu + \pi(0) = (1 - \pi)\mu$.

2.5. Modeling Approaches

First, we compare the performance of the following models using the dichotomized labels and continuous regression objectives. For these models, we sought to establish whether increasing the spatial receptive field by varying the patch size (ablation study) and training graph attention networks (GATs) positively impacted our capacity to predict spatial gene expression. All models featured an output layer, which simultaneously predicted the expression of all selected genes (**Figure 1**). Details of these approaches can be found below:

2.5.1. Local Patch Prediction Model: We initialized our model using InceptionV3 weights (Szegedy et al., 2015) [17]. InceptionV3 was chosen because it has demonstrated high performance in gene imputation studies by Levy-Jurgensen et al. These models are trained for 25 epochs with a learning rate of 0.0001 and a batch size of 32. The patch size labeled Inception models they were configured to make predictions on (i.e., amount of incorporated surrounding information): **Inception-256** (256 pixels), **Inception-512** (512 pixels), and **Inception-768** (768 pixels) (model suffix indicates patch size).

2.5.2. Contextualized Patch Prediction Models: The Visium spots lie on a hexagonal array, each of which can be treated as a node in a graph, each connected to other Visium spots within 150

253 pixels. After training the InceptionV3 models from our local patch classification experiments, we
254 extracted patch image embeddings (e.g., n-dimensional descriptive vector) for each Visium spot
255 from the penultimate layer of the trained Inception-256 model. We test a graph attention network
256 (GAT) to see how iterative message-passing can improve model performance. We also
257 compared these results to that obtained using a Vision Transformer (ViT) [18(p16)]. GAT models
258 were labeled by the number of graph attention layers used to make predictions (i.e., the amount
259 of incorporated surrounding information; the number of layers dictates the size of the
260 neighborhood; eight attention heads per layer): **GAT-1** (1 layer), **GAT-2** (2 layers), and **GAT-4**
261 (4 layers) (model suffix indicates the number of layers). ViT models were labeled by the size of
262 the patch used to form contextual embeddings: **ViT-224** (224-pixel patch sizes) and **ViT-384**
263 (384-pixel patch sizes).

264 *2.5.3. Data Augmentation during training and hyperparameters*: To improve the robustness of
265 the model results to new tissue contexts, we transformed patch images by shifting and scaling
266 the pixel intensities by the mean and variance of ImageNet Then, we applied color jitter and
267 random rotations; between -15 and 15 degrees. Random horizontal and vertical flips also
268 augment patches if they are not used in the GAT. Through a coarse hyperparameter search,
269 learning rates were set to 1e-4, and models were trained for 25 epochs. Batch sizes for the
270 Inception model were set to 32, save for the regression models for patch sizes 512 and 768,
271 where batch sizes were set to 16 and 8, respectively.

272
273 **2.6. Comparison of Model Performance**

274
275 To compare model and patch-size performances, we performed leave-one-slide-out cross-
276 validation (CV), where three of the four slides were used for training/validation, and the final
277 slide was used for testing. This scheme was repeated four times to report on test performance
278 across all four slides in an unbiased manner using macro-averaged (across slides) median
279 (across genes) area under the receiver operating curve (AUROC) and Average Precision (AP)
280 statistics for the binary outcome and correlation coefficients (e.g., Spearman) to compare true
281 versus predicted counts. Non-parametric bootstrapping was used to assess statistical
282 significance through the calculation of a 95% confidence interval. During cross-validation, we
283 use the same set of *InceptionV3* embeddings to train the contextual models (i.e., GAT) that
284 corresponded to the same cross-validation fold. We acknowledge that these statistics could also
285 vary by models, patch size, and slides, so we plotted scatters of test AUCs of each gene to
286 compare these factors in a pairwise manner (e.g., for slide 1, comparing Inception to GAT, or
287 patch size 256 to 768) to draw additional inferences on suitable modeling approaches for a
288 subset of genes. Identifying a subset of genes that obtained optimal performance for one
289 approach versus another was as important as comparing overall performance. These
290 performance differences could suggest relevant tissue features at different scales. For instance,
291 the GAT could extract features from the larger macroarchitecture, indicating its relevance for a
292 gene that does not predict well from *InceptionV3*.

293
294 **2.7. Model Interpretation through Pathway Analysis and Gene Embeddings**

295
296 Due to fundamental limitations in tissue biology, it is unrealistic to expect that every gene can be
297 predicted from tissue histology. We sought to establish which types of genes could be inferred
298 from histology through pathway analysis. Using the *Elsevier Pathways* database available
299 through the Enrichr package, we performed a pathway analysis of the top 250 genes ranked by
300 AUROC averaged across the CV folds [19]. Enrichr reports overrepresented pathways using a
301 modified fisher's exact test. Detected pathways were filtered based on tissue specificity (i.e.,

302   could reasonably be involved with the colon). To determine whether different pathways could be
303   inferred from different tissue contexts, pathway analysis results were compared across models.

304   We also sought to assess how well the predicted gene expression profiles recapitulated
305   relationships/clustering between the Visium spots. This was accomplished through the
306   comparison of Uniform Manifold Approximation and Projection (UMAP) embeddings for the
307   ground truth and predicted expression profiles (on held-out slides) using *InceptionV3* and GAT
308   [20]. Ground truth and predicted gene expression profiles were projected to a lower dimensional
309   space using AlignedUMAP to preserve the orientation/alignment between the ground truth and
310   predicted expression profiles (i.e., positioning of Visium spots across projections is relatively
311   preserved) to enable comparison between the approaches. Visium spots corresponding to the
312   ground truth UMAP projections were clustered using Hierarchical density-based clustering
313   (HDBSCAN) [21]. HDBSCAN also identified outlier Visium spots removed from the cluster
314   analysis and scatterplots for all three datasets (ground truth, *InceptionV3*, GAT). Mapper, a
315   Topological Data Analysis method, was used to provide topological summaries of the
316   embeddings by reducing the number of points based on overlap and connectivity [22–24]. Mapper
317   embedding plots for *InceptionV3* and GAT were colored using the cluster information from the
318   ground truth expression to qualitatively assess topological consistency (i.e., were relationships
319   between Visium spots preserved). This procedure was repeated across the held-out test slides.

320   **2.8. Recapitulating Spatial Biology through Spatial Autocorrelation Tests**

321   In addition to tissue clustering, spatial variation is often used as a proxy for explaining the
322   diversity of cellular lineages interacting across the tissue slide. While this factor alone is not an
323   exhaustive assessment of spatial biology, this served as a target for our preliminary assessment
324   (an exhaustive exploration of spatial analyses is out-of-scope of this work, although it is a future
325   direction). After performing cross-validation, we sought to investigate the ability of our
326   algorithms to recapitulate known spatially variable genes on each of the held-out slides from the
327   cross-validation folds. We used two libraries to determine spatially variable (SV) genes:
328   SpatialDE and SPARK-X [25]. Gene expression counts were summed to create a total count
329   matrix. Genes with low overall expression across the slide (i.e., below a threshold of 30% slide
330   coverage) were removed. The data was then transformed to a normal distribution (by Anscombe
331   transform) to account for the negative binomial distribution of the gene expression. Since
332   SpatialDE's computation time increases cubically with each additional expression patch, we
333   reduced the resolution of the Visium data through 2x2 median pooling (i.e., taking median
334   expression for specific genes from 2x2 neighborhoods). The reduced memory requirements
335   allowed us to perform further histo-molecular assessments on the images, including clustering
336   based on spatial variability.

337
338   Using SpatialDE, we extracted the Fraction of Spatial Variance (FSV) and p-values for each
339   gene from the ground truth set. In addition, we used SpatialDE's Automatic Expression
340   Histology (AEH) to identify 5 groups of genes that were co-expressed spatially from the ground
341   truth data. Separately, we ran SPARK/SPARK-X on the 1000 genes from each held-out
342   validation slide, reporting projection and adjusted P-value (Bonferroni-corrected) statistics. We
343   separately ran this procedure for the ground truth expression and predicted expression profiles.
344   As an indication of performance, we expected p-values derived for the projection covariance
345   function for both the inferred and original expression data to correlate well with each other for
346   each slide. This was accomplished using the Fisher's exact test after dichotomizing spatial
347   autocorrelation statistics into "high autocorrelation" (low p-value) and "low autocorrelation" (high
348   p-value) and similar dichotomization for statistics extracted from the inferred expression.

349   Thresholds for dichotomization were chosen to maximize the magnitudes of the Fisher's exact
350   test statistics. Dichotomized spatial autocorrelation was cross-tabulated across the genes to
351   report odds ratios with p-values. An odds ratio and corresponding confidence interval of more
352   than one (i.e., statistically significant) would suggest the ability of the model to recapitulate
353   spatial autocorrelation from the slide. Separately, we sought to assess whether genes that were
354   predicted with high accuracy (AUC) were spatially autocorrelated using a similar methodology
355   (i.e., comparing dichotomized spatial autocorrelation on the ground truth expression with
356   dichotomized AUC for each modeling approach). We also compared model accuracies between
357   AEH groups of co-expressed genes identified from SpatialDE using Kruskal-Wallis ANOVA
358   statistical tests. Similar to the AUC comparison, we performed comparisons across models,
359   patch sizes, and slides. It should be noted that these analyses were done on the top 1000
360   spatially variable genes; dichotomized autocorrelation is for this reference group. For the
361   original and inferred expression, genes which exhibited spatial autocorrelation which differed
362   between patients with and without metastasis were selected for a pathway analysis using the
363   *MSigDB Hallmarks* gene sets via the Enrichr software.

364 **3. Results**
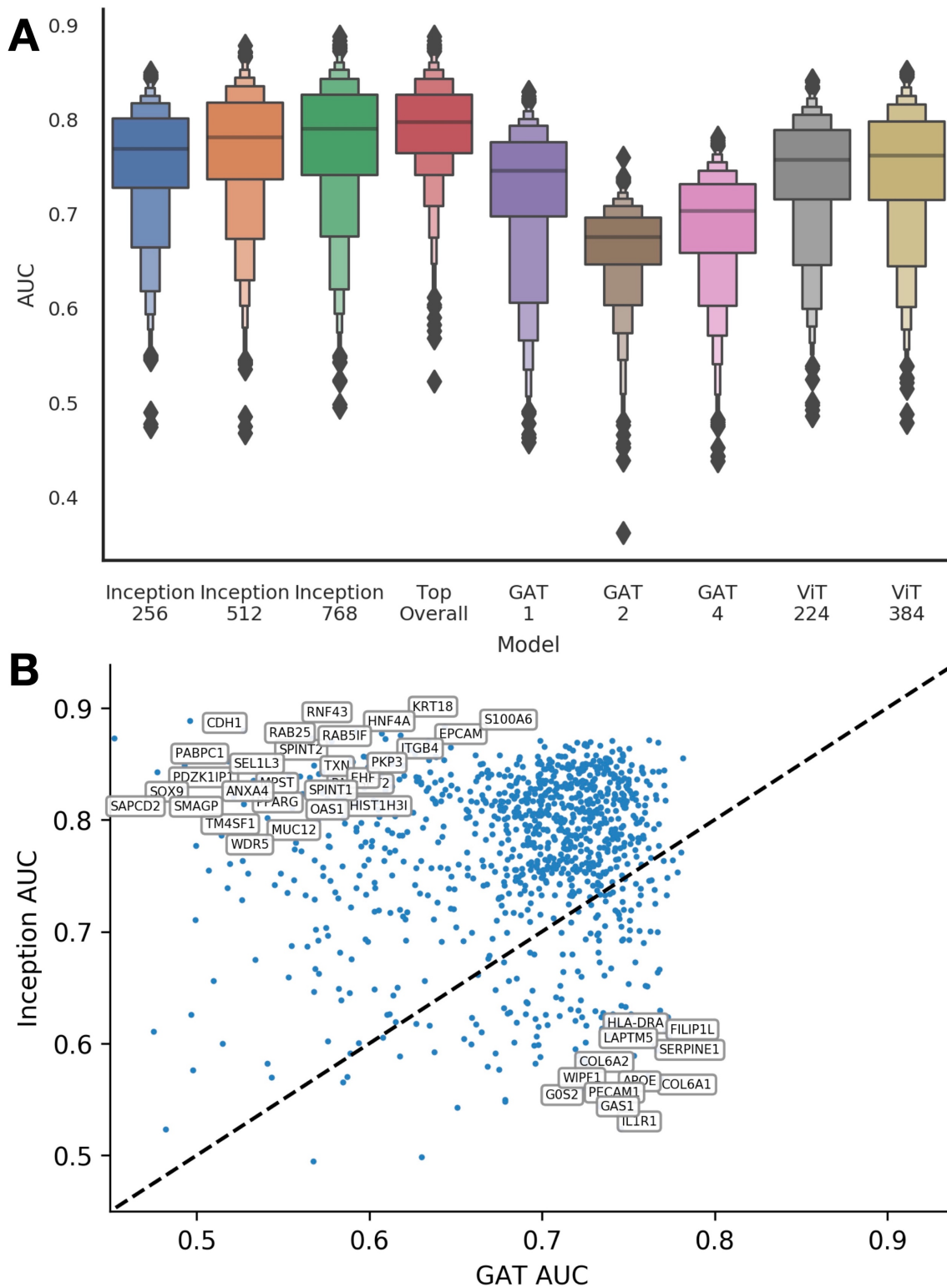
365
366 **Figure 2: Model Performance Comparison: A)** Boxenplots which depict the distribution of
367 AUC values for predicting dichotomized expression across all 1000 filtered genes; **B)** Scatter
368 plot of genes representing individual cross-validated AUC values for GAT-4 versus Inception-

369   768; while overall Inception outperforms GAT, there are several genes from which GAT
370   outperformed Inception

371   **3.1. Model Comparison**

372   First, we assessed the ability of the Inception model to predict gene expression without
373   considering the surrounding tissue macroarchitecture. Across the whole transcriptome, we
374   noted moderately strong concordance between the predicted and actual expression across the
375   held-out slide folds.
376
377   *3.1.1. Impact of patch size to leverage surrounding spatial context:* Importantly, the model
378   achieved optimal performance by increasing access to the surrounding macroarchitecture by
379   increasing the patch size.
380
381   *3.1.2. Impact of model architecture to leverage surrounding spatial context:* After optimizing
382   model architectures, the Inceptionv3 model appeared to outperform the GAT and Transformer
383   approaches for the whole transcriptome assessment. This was true for both classification and
384   regression modeling objectives. The GAT-1 model outperformed the GAT-4 model, which
385   incorporated more of the surrounding tissue context, for the classification task, though GAT-4
386   outperformed GAT-1 for regression and demonstrated a similar capacity to predict count
387   outcomes as Inception-512 (**Figure 2; Table 1**). A breakdown of these model performances for
388   individual held-out validation slides can be found in the appendix (**Supplementary Table 1;**
389   **Supplementary Figures 1-2**).
390
391   *3.1.3. GAT outperforms Inception on a subset of genes:* While overall, Inception outperformed
392   GAT, it is important to recognize that there are many genes for which GAT achieves superior
393   performance (**Figure 2B; Table 2; Supplementary Figure 1**). This was different than
394   comparing multiple Inception models with different patch sizes, where clearly Inception-768
395   outperformed Inception-256 on nearly all relevant genes (a scatter plot of the specific gene-level
396   AUC for specific genes for patch size 256 compared to 768 can be found in **Supplementary**
397   **Figure 2**). As different models demonstrate exemplary performance on different subsets of
398   genes, the combined accuracy (AUC) across the genes (i.e., selecting top performing model for
399   each gene based on CV-AUC) is 0.798 (95% CI [0.795-0.802]); the combined AP score is 0.67
400   (95% CI [0.659-0.677]) (**Figure 2**) (**Top Overall Model**).
401
402
403   **Table 1. Comparison of model performance for predicting dichotomized and log-**
404   **transformed expression for genes across the whole transcriptome**. These statistics are
405   created by taking the median of each performance statistic across all genes, averaged across
406   held-out slides.

	Classification		Regression	
Model	**AUROC**	**AP**	**Spearman**	**Pearson R**
Inception-256	0.769±0.004	0.587±0.011	0.439±0.011	0.466±0.014
Inception-512	0.782±0.005	0.61±0.013	0.456±0.011	0.496±0.013
Inception-768	0.79±0.006	0.621±0.012	0.479±0.011	0.538±0.013
GAT-1	0.746±0.003	0.528±0.014	0.302±0.009	0.309±0.009
GAT-2	0.676±0.002	0.444±0.01	0.409±0.01	0.446±0.01
GAT-4	0.703±0.004	0.459±0.012	0.406±0.01	0.428±0.011

ViT-224	0.757±0.005	0.573±0.012	0.432±0.011	0.458±0.013
ViT-384	0.762±0.004	0.57±0.012	0.442±0.012	0.473±0.014

Table 2: Top 10 performing genes for Inception-768, Inception-256 and GAT-4, ranked by AUC

| Top-10 Inception				Top-10 GAT			
Name	**Inception-768**	**Inception-256**	**GAT**	**Name**	**Inception-768**	**Inception-256**	**GAT**
KRT8	0.889	0.850	0.491	*TRIP12*	0.753	0.745	0.829
S100AA10	0.884	0.846	0.714	*PGM3*	0.749	0.740	0.829
CDH1	0.880	0.848	0.510	*TMEM238*	0.855	0.825	0.825
RNF43	0.879	0.847	0.556	*SYNM*	0.700	0.674	0.823
KRT18	0.877	0.838	0.723	*SMAP1*	0.771	0.757	0.822
S100A6	0.875	0.842	0.606	*SERPINE1*	0.594	0.581	0.821
RAB25	0.873	0.844	0.545	*LSM4*	0.844	0.818	0.820
AXIN2	0.873	0.844	0.610	*GRIPAP1*	0.766	0.752	0.818
EPCAM	0.873	0.834	0.491	*TSPO*	0.822	0.796	0.818
TNS4	0.871	0.841	0.801	*HLA-DRA*	0.612	0.618	0.816
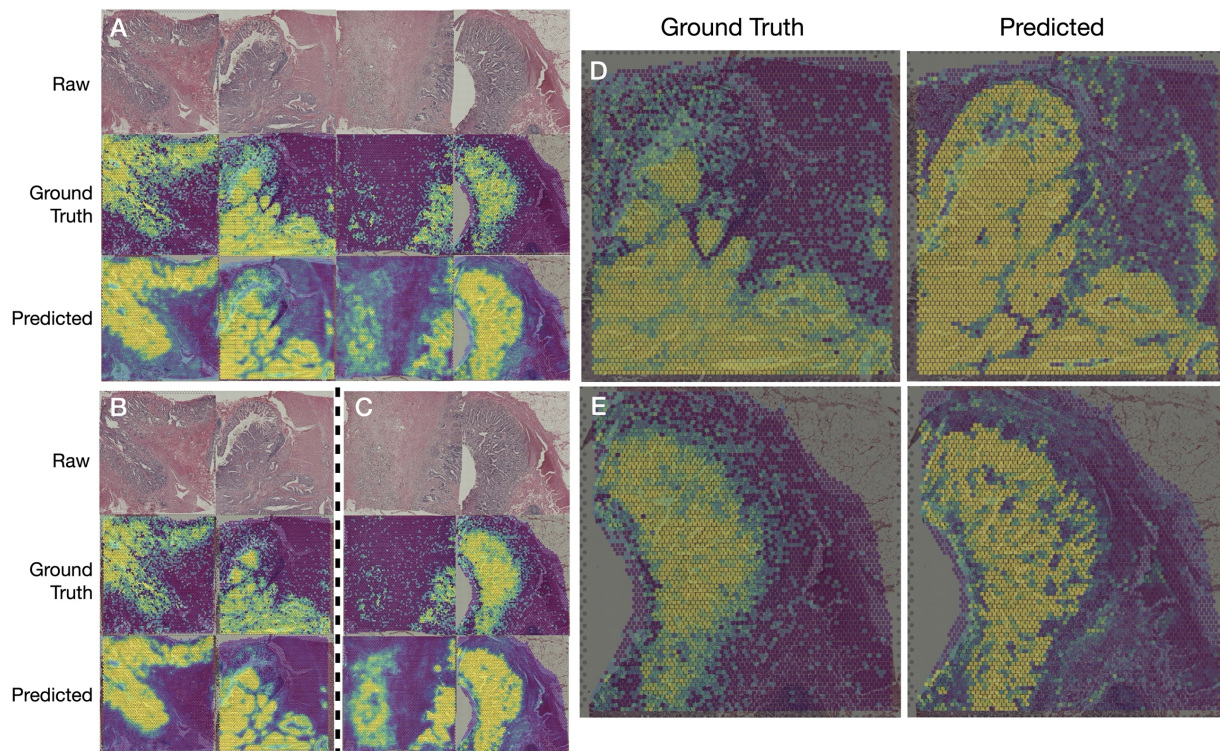
Figure 3. Comparison of log-scaled heatmaps of the ground truth gene expression (top) and output probabilities for above-median expression from the dichotomized classifiers: **A-C)** Inception-768 classifier's predictions, generating heatmaps for **A)** RAB25, **B)** TNS4, and **C)** AXIN2 genes; all predictions from the held-out test set. The corresponding median AUROCs for these genes were 0.873, 0.871, and 0.873. **D)** AXIN2 imputation on slide B1 with GAT. **E)** AXIN2 imputation on slide D1 with GAT.

3.2. Pathway Analysis

The top-performing genes by AUROC for both types of models were also highly related to tumor aggression and migration. For instance, *CDH1*, heavily implicated with tumor suppression, achieved AUROC of 0.880 [26]. *RAB25* (**Figure 3A**), which can serve as a tumor suppressor or oncogene depending on the context, obtained an AUROC of 0.879. *TNS4* has been heavily implicated for multiple prognostic outcomes following surgery and was predicted with an AUC of 0.871 [27]. *AXIN2* (**Figure 3C-E**), which inhibits the Wnt signaling pathway and serves to regulate immune cell infiltration [28], was detected with an AUROC of 0.873. The pathway analysis corroborated these findings and was highly relevant for metastasis formation. For instance, genes associated with cancer metastasis, cell motility and proliferation, glycolysis, and the epithelium to mesenchymal transition were identified by both models. Interestingly, the GAT model was able to identify genes related to anti-EGFR therapy resistance in colorectal cancer (**Supplementary Table 2**).

Inferred Spatial Expression is Topologically Consistent with Ground Truth
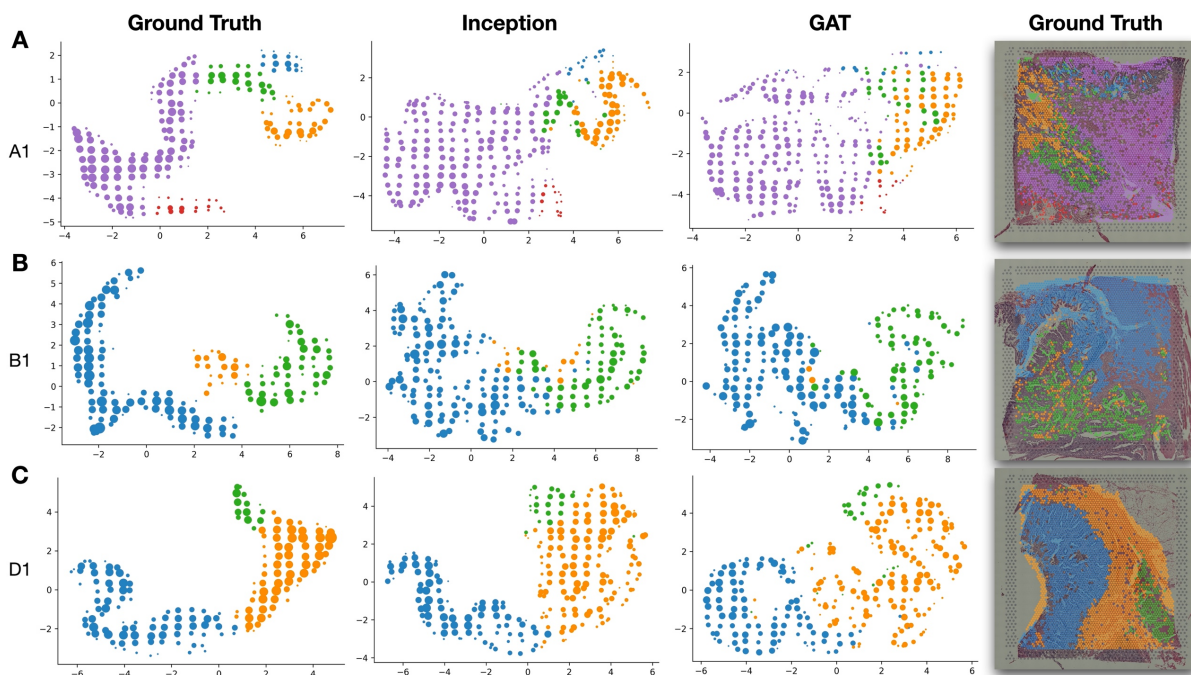
Figure 4: UMAP Embeddings of True and Predicted Gene Expression for slides **A)** A1, **B)** B1 and **C)** D1; embedding plots are summarized using Mapper, which flexibly clusters the expression data with overlapping clusters containing multiple Visium spots; Mapper nodes are sized by the number of associated spots and colored by the dominant cluster in the set of node-associated spots with cluster assignments determined using HDBSCAN; outliers were filtered from these embedding plots and the cluster assignments plotted over the WSI on the right

Visual inspection of mapper diagrams of clustered UMAP embeddings illustrates topological consistency (i.e., preserve relationships) between the predicted expression and the ground truth (**Figure 4**). Qualitatively, the gene expression embeddings produced from Inception appear to be more closely aligned with the ground truth embedding plots across the tissue types (e.g.,

449 **Figure 4 A, C**, where clusters are placed in the approximately same area for Ground Truth and
450 Inception in the Mapper diagrams).
451

3.3. Spatial Autocorrelation

453 We also compared the capabilities of each modeling approach for their ability to recapitulate
454 slide-level spatial autocorrelation parameters. Results indicate a large significant positive
455 association between predicted and actual spatial autocorrelation for both Inception and GAT
456 approaches. For half of the held-out slides, GAT demonstrated a larger effect estimate than
457 Inception (**Table 3**).
458

459 Separately, results indicate that highly spatially autocorrelated genes, as determined using the
460 actual gene expression, were predicted with higher accuracy using Inception and GAT versus
461 genes, which lacked spatial variation (**Table 3**). These models varied in their ability to associate
462 spatial variation with model accuracy. For Inception, there was a large statistically significant
463 effect (**Figure 5**), while spatial variation was not as associated with accuracy for the GAT– i.e.,
464 there was a statistically significant association for the first two slides and no statistically
465 significant effect for the final two. For the Inception model, groups of genes that tended to be co-
466 expressed, as determined through the AEH analysis on the raw expression data, were found to
467 have widely different accuracies (**Table 3; Supplementary Figures 3-5**). Similar to the spatial
468 variation analysis, GAT model accuracy did not vary substantially between AEH groups
469 determined from the raw expression data.
470

471 Using the ground truth, Inception, and GAT results, we identified the set of genes, which
472 exhibited different spatial variation for primary sites with (A1, B1) and without metastasis (C1,
473 D1) based on the dichotomized thresholds. Through a pathway analysis (gene set testing of
474 Cancer Hallmark genes), genes that were differently autocorrelated were related to the epithelial
475 to mesenchymal transition (**Supplementary Table 3**).
476
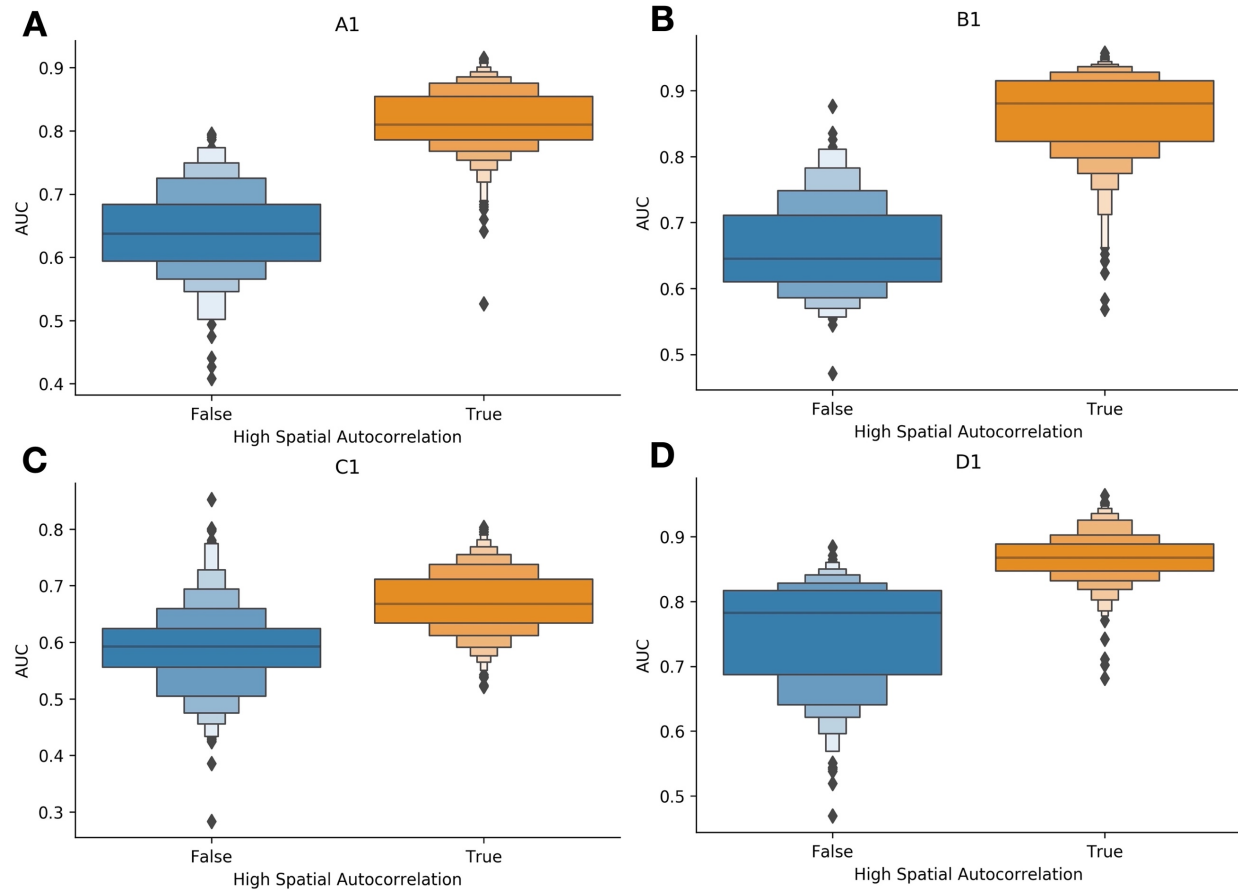
Figure 5. Boxenplots illustrating the predictive accuracy of Inceptionv3 (AUC, y-axis) across genes, separated by whether highly significant spatial variation was reported (blue versus orange, x-axis); gathered from validation slides held-out of the training/validation set. The positive association demonstrates higher accuracy for genes with significant spatial variability (i.e., not distributed randomly); these genes were more accurately predicted by our RNA inference model. **A-D)** correspond to slides A1-D1, respectively

Table 3: Statistical testing results from spatial autocorrelation analysis, comparing: A) Spatial autocorrelation from raw expression (high/low) with inferred spatial autocorrelation from predicted expression values (high/low); **B)** Spatial autocorrelation from raw expression (low/high) with model AUC (high/low); **C)** Whether model accuracy changed depending on the AEH group; **A)** and **B)** were determined through Fisher's exact tests while **C)** utilized Kruskal-Wallis ANOVA testing

Slide	Model	Recover Spatial Autocorrelation		Spatial Autocorrelation vs AUC		AEH group vs AUC
		OR	P-Value	OR	P-Value	P-Value
A1	**Inception**	722.27	6.55e-65	0.01	4.05e-99	1.78e-39
B1		40.09	1.30e-34	0.01	8.29e-85	3.93e-162
C1		68.82	1.25e-16	0.07	1.05e-57	2.00e-119
D1		92.18	1.30e-20	0.02	5.41e-96	2.46e-166
A1	**GAT**	101.64	1.46e-61	0.45	2.02e-05	9.52e-02

B1	51.12	1.16e-24	0.58	0.01	1.37e-01
C1	24.80	3.10e-06	0.93	0.66	8.54e-01
D1	99.15	5.06e-22	0.99	1.00	6.68e-01

492

4. Discussion

493

494
495 Assessment of Colon cancer tumor recurrence risk depends on the ability to assess lymph node
496 status. For cases where lymph nodes are unable to be completely assessed, leveraging
497 information found in the tumor immune microenvironment at the primary resection site
498 presents a viable alternative. Yet most technologies to assess the primary site lack a spatial
499 component (e.g., bulk expression), which does not enable a comprehensive characterization of
500 the tissue. Spatial transcriptomic technologies enable high multiplexing at incredible spatial
501 resolution. Due to both fiscal and sampling (i.e., placement of capture area) constraints,
502 findings are not likely to be clinically actionable or reproducible through a low-cost test.
503 Inferring a spatial digital biomarker from a routine histological slide (WSI) has the potential to
504 enable high-throughput tissue characterization for the creation of nascent decision-making aids
505 which can complement existing specimen findings. In this study, we explored the potential for
506 deep learning models to predict spatial gene expression from formalin-fixed specimens, the
507 ability to recapitulate well known spatial findings, and comment on the degree to which spatial
508 information at higher-order contexts (i.e., macroarchitecture) plays a role in modeling spatial
509 expression.
510
511 We add further evidence that deep learning models can accurately predict gene expression in
512 whole slide images. Furthermore, we demonstrate that increasing the receptive field can
513 improve the performance of certain subsets of genes. Interestingly, although InceptionV3
514 outperformed other modeling approaches overall, this did not apply to all genes. We noted that
515 certain genes were predicted more effectively at local spatial resolutions (Inception) while
516 others benefited from considering a broader architectural context (GAT). For instance, we
517 noticed that *COL6A1* and *COL6A2* were consistently predicted better by the GAT model as
518 compared to Inceptionv3 [29–31]. *COL6A1* is a crucial component of collagen secretion,
519 extracellular matrix maintenance, and mesenchymal phenotype promotion. While both models
520 demonstrated the capacity to recapitulate well-known cancer markers, it was clear that certain
521 genes can be better predicted by considering receptive field size and choosing a model that
522 best incorporates this spatial information.
523
524 Spatial autocorrelation was recapitulated for these slides from the deep learning model
525 predictions. We noticed that increases in modeling accuracy were associated with spatial gene
526 expression variation. Genes that were differentially autocorrelated between metastatic and non-
527 metastatic tumors corroborated with well-established oncogenic pathways. This supports our
528 overarching modeling approach to characterize spatial heterogeneity. As some genes were
529 better predicted using the broader spatial architecture, it would make sense to decide which
530 modeling approach to utilize on a gene-by-gene basis and whether the local versus spatial
531 context is preferred. This allows for the selection of optimal modeling approaches in a gene-
532 specific manner to extend the broad applicability of our framework across all disease-relevant
533 genes.
534

535   Importantly, the techniques featured in this work will prove useful for inferring gene expression
536   on slides from external cohorts. This is necessary due to the prohibitive costs of spatial
537   transcriptomics. If validated properly on a slightly larger set of slides while carefully controlling
538   for potential confounders (e.g., site, MMR status, etc.), these methods may overcome sources
539   of patient-specific batch effects to allow the report of less biased effect estimates for large-
540   scale cohort studies. Validation efforts on these held-out slides can feature validation
541   assessments such as Spark-X and SpatialDE to ensure spatial heterogeneity from the inferred
542   expression data is similar to the initial internal validation cohort. To this end, we achieved
543   remarkable performance for the ability to predict spatial heterogeneity, which will potentially
544   help power future studies elucidating spatial transcriptomic predictors of colon metastasis.
545
546   Our results did not support the hypotheses that, overall, spatial gene expression estimation
547   would benefit from message-passing between patch locations via transformer models and
548   graph convolutional networks. However, these neural networks outperform local patch
549   prediction across a large set of genes, suggesting the relevance of these methods for genes
550   which leverage the spatial context. For GAT, underperforming genes can potentially be
551   explained by the fact that graph convolutional networks tend to smoothen features (Li et al.,
552   2018), which may weaken the model's predictions if the gene expression data is not as smooth.
553   While these findings could alternatively suggest that neighboring patches may not be as
554   correlated as suspected, this may only hold true for a subset of genes. Increasing graph
555   convolution layers may result in optimal performance for genes, which rely on higher-order
556   dependencies between tissue regions; however, such genes may be outnumbered by genes
557   better suited for the Inception approach.
558
559   There were a few study limitations of note. Our validation scheme featured the use of held-out
560   slides. Generally, as two cases had tumor metastasis, whereas the other two did not have
561   concurrent metastasis, we believed held-out slides would have similar heterogeneity in
562   expression and morphology as the training/validation slides. Indeed, it is possible that tissue
563   expression/morphology existed outside of this range, differentially impacting the
564   GAT/transformer models designed to capture long-range spatial dependencies. As we had
565   manually selected capture areas, we did not expect slides to have exactly matching/analogous
566   architectural features; thus, the GAT/transformer model could have been unable to generalize
567   as it aims to integrate this higher-level information. There were four whole slide images for both
568   training and cross-validation, and while these provided over ten thousand patches in the training
569   set when taken individually, they provided limited opportunities for models to learn diverse
570   global contexts that transformer models would have benefitted from. These challenges could
571   have been ameliorated by pretraining the GAT/Transformer on a variety of colorectal cancer
572   tissue contexts, which does not explicitly require spatial omics data. We also acknowledge the
573   impact of evolving workflows for spatial resolution of transcriptomics information. Many
574   workflows rely on the prediction of tissue features from fresh or fresh frozen tissue. We utilized
575   formalin-fixed paraffin-embedded (FFPE) tissue slides. Only recently have deparaffinization and
576   assay workflows been developed for FFPE. As many of these specimen processing workflows
577   are still under development, manual staining and imaging could have both introduced batch
578   effects and impacted tissue quality. For instance, while a plethora of data preprocessing
579   workflows offer capabilities to combat batch effects through normalization, a robust
580   comparison of preprocessing workflows was outside of the study scope. Our group is keen to
581   adopt future iterations of the FFPE workflows. We expect that protocol updates for the FFPE
582   workflow and data collection will provide major improvements in specimen processing, data
583   quality, and resolution, improving prediction models. Enlarging the capture area and evaluating
584   complementary molecular assays (e.g., spatial proteomics) will improve the resolution and

585 scope of our findings. Although this work does not explicitly assess the tumor immune
586 microenvironment, the methodology explored here can help facilitate spatial analyses and will
587 be used to motivate future clinical findings. Future work will clearly demarcate these regions
588 (e.g., intratumoral, invasive margin, away from the tumor) for a more granular analysis that is
589 well-adjusted for potential confounders.
590
591
592 **5. Conclusions**
593 Tumor metastasis is heavily tied to poor prognostic outcomes and risk of recurrence.
594 Assessment of key niches at the primary site (e.g., tumor immune microenvironment) may
595 reveal histomorphological or biological factors relating to nodal involvement or distant
596 metastasis. In this work, we investigated the potential to infer spatially resolved transcriptomics
597 from the tissue histology of colorectal cancer patients through several sophisticated neural
598 network approaches. Our findings indicate that neural networks can be effectively employed in
599 this capacity and that selection of a neural network model could be informed by its relevance to
600 a molecular pathway resolved from histological features at different scales. Findings reaffirm
601 the role of the epithelial-to-mesenchymal transition as an important metastasis-related process.
602 In the future, with additional algorithmic fine-tuning, data curation, and standardization, there
603 are opportunities to generalize these findings to perform large-scale spatial molecular
604 assessments.
605
606
607 **Competing Interests**
608 None to disclose.
609
610 **Authors' Contributions**
611 The conception and design of the study were contributed by JL and LV. MF, EF, and CS
612 spearheaded study design, analysis, interpretation and initial manuscript drafting. All authors
613 contributed to writing and editing of the manuscript and all authors read and approved the final
614 manuscript.
615
624

References

1. Wong MC, Huang J, Lok V, et al. Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location. *Clin Gastroenterol Hepatol*. 2021;19(5):955-966.

2. Amin MB, Edge S, Greene F, et al., eds. *AJCC Cancer Staging Manual*. 8th ed. Springer International Publishing; 2017. Accessed April 28, 2021. https://www.springer.com/gp/book/9783319406176

3. Senthil M, Trisal V, Paz IB, Lai LL. Prediction of the Adequacy of Lymph Node Retrieval in Colon Cancer by Hospital Type. *Arch Surg*. 2010;145(9):840-843. doi:10.1001/archsurg.2010.182

4. Nearchou IP, Gwyther BM, Georgiakakis ECT, et al. Spatial immune profiling of the colorectal tumor microenvironment predicts good outcome in stage II patients. *Npj Digit Med*. 2020;3(1):1-10. doi:10.1038/s41746-020-0275-x

5. Uttam S, Stern AM, Sevinsky CJ, et al. Spatial domain analysis predicts risk of colorectal cancer recurrence and infers associated tumor microenvironment networks. *Nat Commun*. 2020;11(1):3515. doi:10.1038/s41467-020-17083-x

6. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018;24(5):541-550. doi:10.1038/s41591-018-0014-x

7. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer*. 2020;20(11):662-680. doi:10.1038/s41568-020-0285-7

8. Wu Y, Cheng Y, Wang X, Fan J, Gao Q. Spatial omics: Navigating to the golden era of cancer research. *Clin Transl Med*. 2022;12(1):e696. doi:10.1002/ctm2.696

9. Suwalska A, Zientek L, Polanska J, Marczyk M. Quantifying Spatial Heterogeneity of Tumor-Infiltrating Lymphocytes to Predict Survival of Individual Cancer Patients. *J Pers Med*. 2022;12(7):1113. doi:10.3390/jpm12071113

10. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539

11. He B, Bergenstråhle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng*. Published online June 22, 2020:1-8. doi:10.1038/s41551-020-0578-x

12. Levy-Jurgenson A, Tekpli X, Kristensen VN, Yakhini Z. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. *Sci Rep*. 2020;10(1):18802. doi:10.1038/s41598-020-75708-z

13. Zeng Y, Wei Z, Yu W, et al. Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. *Brief Bioinform*. 2022;23(5):bbac297. doi:10.1093/bib/bbac297

14. Pang M, Su K, Li M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. Published online November 28, 2021:2021.11.28.470212. doi:10.1101/2021.11.28.470212

15. Levy JJ, Bobak CA, Nasir-Moin M, et al. Mixed Effects Machine Learning Models for Colon Cancer Metastasis Prediction using Spatially Localized Immuno-Oncology Markers. *Pac Symp Biocomput Pac Symp Biocomput*. 2022;27:175-186.

16. Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. *Nat Methods*. 2018;15(5):343-346. doi:10.1038/nmeth.4636

17. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. Published online December 11, 2015. doi:10.48550/arXiv.1512.00567

18. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: ; 2020. Accessed August 28, 2021. https://openreview.net/forum?id=YicbFdNTTy

19. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14(1):128. doi:10.1186/1471-2105-14-128

20. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw*. 2018;3(29):861. doi:10.21105/joss.00861

21. McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. *J Open Source Softw*. 2017;2(11):205. doi:10.21105/joss.00205

22. Veen HJ van, Saul N, Eargle D, Mangham SW. Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *J Open Source Softw*. 2019;4(42):1315. doi:10.21105/joss.01315

23. Tauzin G, Lupo U, Tunstall L, et al. giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration. *ArXiv200402551 Cs Math Stat*. Published online April 6, 2020. Accessed July 23, 2020. http://arxiv.org/abs/2004.02551

24. Singh G, Mémoli F, Carlsson G. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. :10.

25. Zhu J, Sun S, Zhou X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol*. 2021;22(1):184. doi:10.1186/s13059-021-02404-0

694 26. Richards FM, McKee SA, Rajpar MH, et al. Germline E-cadherin Gene (CDH1) Mutations
695   Predispose to Familial Gastric Cancer and Colorectal Cancer. *Hum Mol Genet*.
696   1999;8(4):607-610. doi:10.1093/hmg/8.4.607

697 27. SAWAZAKI S, OSHIMA T, SAKAMAKI K, et al. Clinical Significance of Tensin 4 Gene
698   Expression in Patients with Gastric Cancer. *In Vivo*. 2017;31(6):1065-1071.
699   doi:10.21873/invivo.11171

700 28. Pai SG, Carneiro BA, Mota JM, et al. Wnt/beta-catenin pathway: modulating anticancer
701   immune response. *J Hematol OncolJ Hematol Oncol*. 2017;10(1):101. doi:10.1186/s13045-
702   017-0471-6

703 29. Huang MS, Fu LH, Yan HC, et al. Proteomics and liquid biopsy characterization of human
704   EMT-related metastasis in colorectal cancer. *Front Oncol*. 2022;12:790096.
705   doi:10.3389/fonc.2022.790096

706 30. Li X, Li Z, Gu S, Zhao X. A pan-cancer analysis of collagen VI family on prognosis, tumor
707   microenvironment, and its potential therapeutic effect. *BMC Bioinformatics*. 2022;23:390.
708   doi:10.1186/s12859-022-04951-0

709 31. van Huizen NA, Coebergh van den Braak RRJ, Doukas M, Dekker LJM, IJzermans JNM,
710   Luider TM. Up-regulation of collagen proteins in colorectal liver metastasis compared with
711   normal liver tissue. *J Biol Chem*. 2019;294(1):281-289. doi:10.1074/jbc.RA118.005087

712
713
714

Appendix

Supplementary Table 1. Comparison of model performance for predicting dichotomized and log-transformed expression for genes across the whole transcriptome. These statistics are created by taking the median across all genes, reported for each held-out slide.

Model	Slide	Classification		Regression	
		AUROC	AP	Spearman R	Pearson R
Inception-256	Overall	0.768	0.603	0.444	0.477
	A1	0.78	0.675	0.494	0.511
	B1	0.834	0.796	0.612	0.605
	C1	0.625	0.432	0.214	0.301
	D1	0.831	0.51	0.454	0.491
Inception-512	Overall	0.781	0.63	0.461	0.498
	A1	0.79	0.688	0.515	0.537
	B1	0.856	0.816	0.639	0.647
	C1	0.629	0.459	0.22	0.299
	D1	0.847	0.555	0.469	0.51
Inception-768	Overall	0.787	0.639	0.486	0.54
	A1	0.802	0.705	0.551	0.578
	B1	0.864	0.823	0.671	0.655
	C1	0.636	0.474	0.239	0.4
	D1	0.845	0.553	0.481	0.525
GAT-1	Overall	0.743	0.539	0.413	0.43
	A1	0.785	0.669	0.364	0.376
	B1	0.811	0.73	0.603	0.582
	C1	0.606	0.409	0.239	0.335
	D1	0.77	0.349	0.444	0.428
GAT-2	Overall	0.677	0.444	0.407	0.45
	A1	0.769	0.645	0.511	0.545
	B1	0.772	0.612	0.574	0.568
	C1	0.589	0.333	0.146	0.265
	D1	0.578	0.184	0.421	0.428
GAT-4	Overall	0.702	0.462	0.453	0.499
	A1	0.679	0.571	0.517	0.551
	B1	0.764	0.614	0.622	0.643
	C1	0.604	0.379	0.228	0.356
	D1	0.759	0.283	0.445	0.454
GAT-4-AXIN2	Overall	0.702	0.462	0.545	0.578
	A1	0.679	0.57	0.511	0.446
	B1	0.764	0.614	0.745	0.771
	C1	0.604	0.379	0.286	0.46
	D1	0.759	0.283	0.637	0.633
GAT-4-CDH1	Overall	0.800	0.707	0.546	0.571
	A1	0.826	0.747	0.468	0.478
	B1	0.858	0.791	0.753	0.741
	C1	0.648	0.462	0.31	0.425

	D1	0.867	0.827	0.652	0.64
GAT-1-ZINB	Overall	n/a	n/a	0.393	0.408
	A1	n/a	n/a	0.441	0.429
	B1	n/a	n/a	0.551	0.562
	C1	n/a	n/a	0.15	0.19
	D1	n/a	n/a	0.429	0.449
GAT-4-ZINB	Overall	n/a	n/a	0.41	0.432
	A1	n/a	n/a	0.485	0.542
	B1	n/a	n/a	0.602	0.621
	C1	n/a	n/a	0.128	0.2
	D1	n/a	n/a	0.426	0.364
ViT-224	Overall	0.756	0.583	0.436	0.454
	A1	0.768	0.644	0.511	0.506
	B1	0.826	0.762	0.601	0.567
	C1	0.604	0.413	0.177	0.255
	D1	0.827	0.513	0.455	0.487
ViT-384	Overall	0.765	0.601	0.446	0.468
	A1	0.794	0.678	0.513	0.521
	B1	0.836	0.788	0.612	0.583
	C1	0.606	0.406	0.213	0.307
	D1	0.823	0.530	0.446	0.462

722
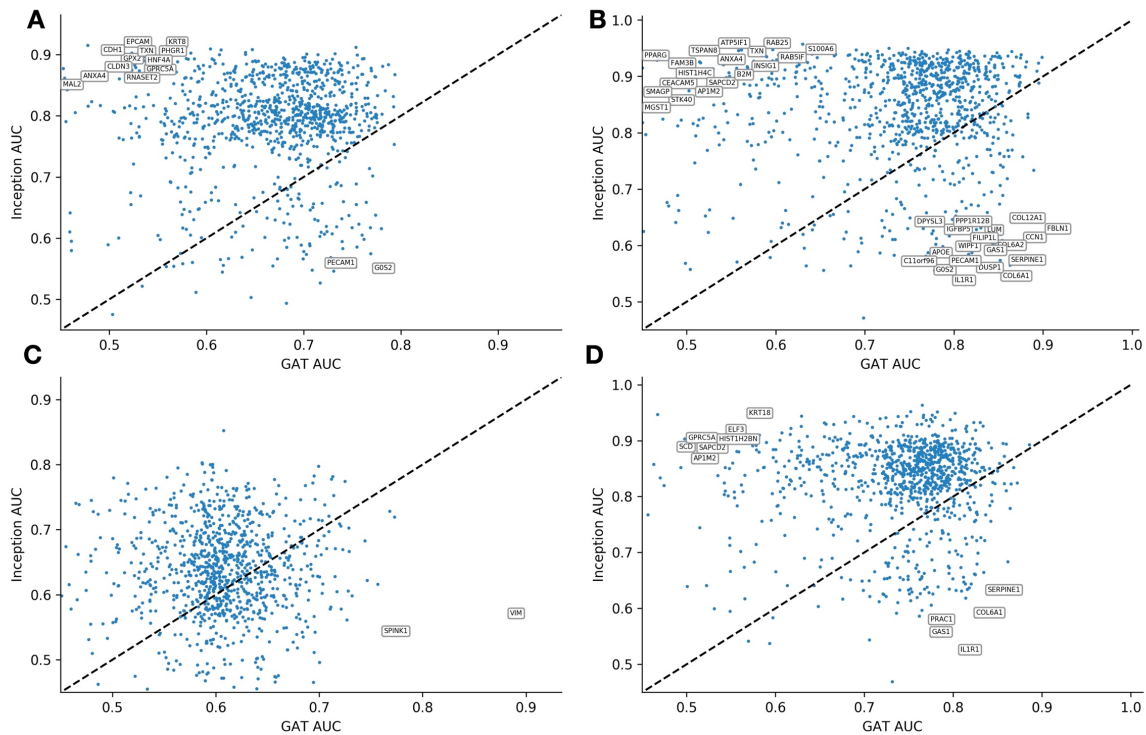
723
Supplementary Figure 1: Scatterplot depicting gene-specific AUCs for GAT and Inception-768 for slides: A) A1; **B)** B1; **C)** C1; **D)** D1

726
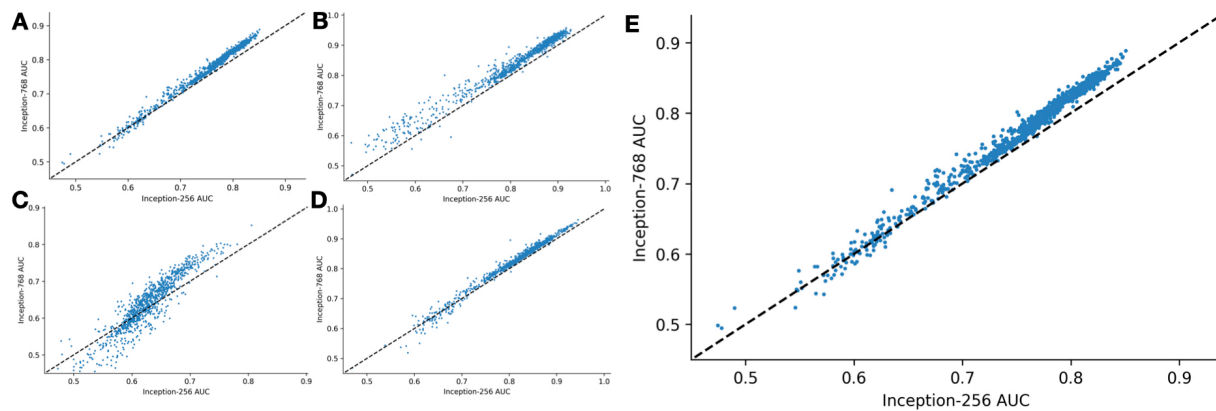
Supplementary Figure 2: Scatterplot depicting gene-specific AUCs for Inception-256 and Inception-768 for slides: A) A1; **B)** B1; **C)** C1; **D)** D1, **E)** Overall

Supplementary Table 2: Enrichr pathway results for genes found to be accurately predicted from the tissue histology via the Inception and GAT approaches; pathways were filtered based on tissue specificity
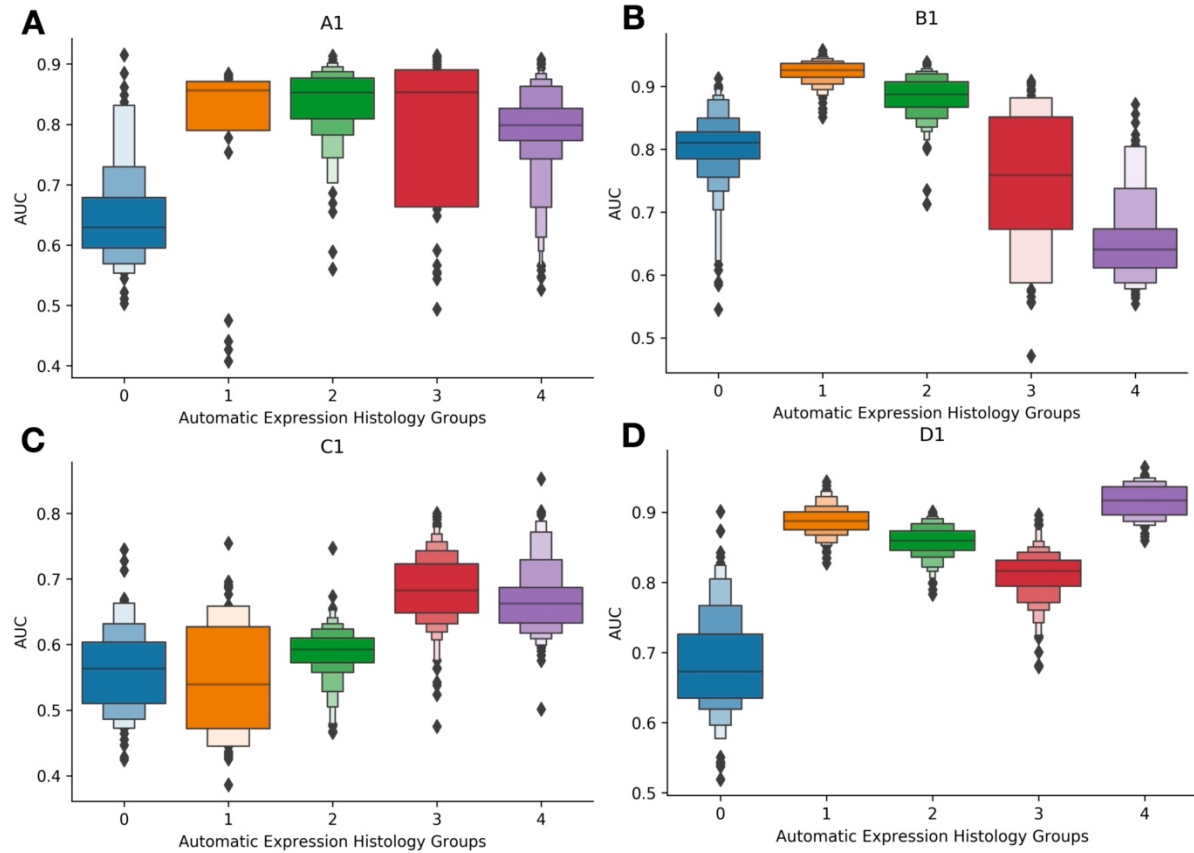
	Inception		GAT		
Term	**Overlap**	**Adjusted P-value**	**Term**	**Overlap**	**Adjusted P-value**
EPCAM in Cancer Cell Motility and Proliferation	9/36	2.79e-07	Proteins with Altered Expression in Cancer Metastases	9/106	1.22e-03
Proteins with Altered Expression in Cancer Metastases	11/106	2.82e-05	Desmosome Assembly	4/18	4.56e-03
Epithelial Cell in the Saliva Formation	5/13	7.15e-05	Corneodesmosomes in Atopic Dermatitis	4/19	5.15e-03
Androgens in Sebocyte Maturation	8/65	1.69e-04	Proteins with Altered Expression in Cancer Metabolic Reprogramming	7/85	5.50e-03
Proteins Involved in Ulcerative Colitis	11/141	1.69e-04	WNT in Epithelial to Mesenchymal Transition in Cancer	5/43	8.09e-03
Desmosome Assembly	5/18	1.70e-04	Metabolic Effects of Oncogenes and Tumor Suppressor in Cancer Cells	6/68	8.09e-03
Corneodesmosomes in Atopic Dermatitis	5/19	2.03e-04	Proteins Involved in Myocardial Ischemia	11/252	1.18e-02
Proteins Involved in Colorectal Neoplasms	9/99	2.35e-04	Desmosome Dysfunction in Cardiomyocyte	3/12	1.26e-02
WNT in Epithelial to Mesenchymal Transition in Cancer	6/43	5.66e-04	RAGE/AGER and S100 Proteins in Cardiovascular Injury	4/30	1.56e-02
Acinar Cells in the Saliva Formation	4/13	5.66e-04	Desmosomes Role in Dilated Cardiomyopathy	3/15	2.31e-02
Epithelial to Mesenchymal Transition in Cancer: Overview	8/90	5.85e-04	Epithelial to Mesenchymal Transition in Cancer: Overview	6/90	2.58e-02
Metastatic Colorectal Cancer	9/121	7.01e-04	Proteins Involved in Atherosclerosis	9/200	2.58e-02
Proteins with Altered Expression in Cancer Metabolic Reprogramming	7/85	2.43e-03	Cetuximab Resistance in Colorectal Cancer	5/64	3.05e-02
Telogen Maintenance in Androgenic Alopecia	4/23	4.15e-03	Proteins Involved in Colorectal Neoplasms	6/99	3.41e-02
Cancer Cells Inhibit Adipocyte Differentiation	4/27	6.43e-03	Glycolysis Activation in Cancer (Warburg Effect)	4/40	3.41e-02
TGFB Family in Epithelial to Mesenchymal Transition in Cancer	6/80	8.55e-03			

|		
Fatty Acid Synthase (FASN) Signaling	3/16	1.52e-02
WNT Signaling Activation by Blocking of Tumor Suppressors	4/36	1.57e-02
HPV Infection and Cancer	5/65	1.80e-02
Androgens in Adipocyte Activation	3/18	1.80e-02
Glycolysis Activation in Cancer (Warburg Effect)	4/40	1.97e-02
Metabolic Effects of Oncogenes and Tumor Suppressor in Cancer Cells	5/68	1.98e-02
Proteins Involved in HPV Infection	3/19	1.98e-02
CDH1 Down regulation Promotes Cancer Cell Migration and Metastases	4/44	2.45e-02
Androgen Deficiency in Male Obesity	3/24	3.31e-02
Sialophorin -> CTNNB/MYC/TP53 Signaling	3/25	3.60e-02
mRNA Degradation	3/26	3.91e-02
WNT Canonical Signaling Activation in Cancer	3/27	4.16e-02
Proteins Involved in Arterial Hypertension	9/255	4.30e-02
Hyaluronic Acid, CD44 and HMMR in Cancer Cell Invasion and Survival	4/56	4.35e-02
Estrogen Deficiency in Female Obesity	2/9	4.35e-02
Adipocyte Hypertrophy and Hyperplasia	3/29	4.50e-02
Proteins Involved in Hepatocellular Carcinoma	6/130	4.74e-02
Adherens Junction Assembly (Cadherins)	3/30	4.83e-02
Proteins with Altered Expression in Cancer-Associated Sustaining of Proliferative Signaling	7/175	5.00e-02

735
736
737

Supplementary Figure 3: Boxenplots illustrating the predictive accuracy of Inceptionv3 (AUC, y-axis) across genes, separated by the genes' Automatic Expression Histology groups (colors, x-axis); gathered from validation slides held-out of the training/validation set. Different groups were assigned for each slide. **A-D)** correspond to slides A1-D1 respectively
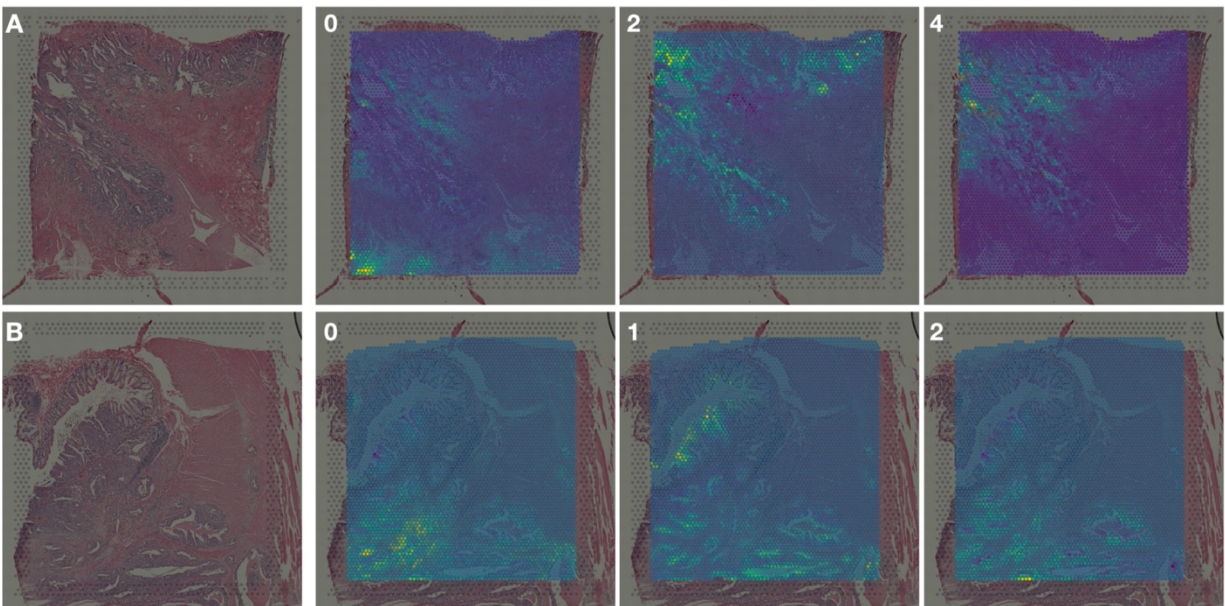
745 **Supplementary Figure 4: Spatial expression patterns for select AEH groups for slides:**
746 **A)** A1; **B)** B1; genes from the first featured AEH group for each slide were predicted with
747 low accuracy; genes from the final two AEH groups were predicted with high accuracy
748
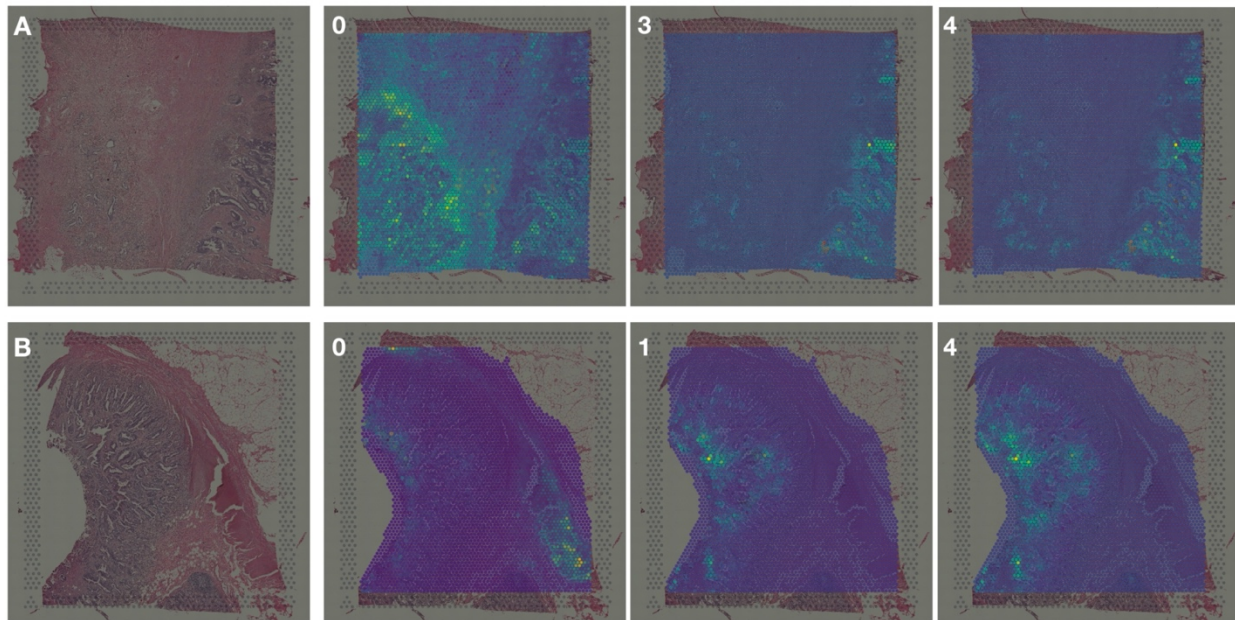
749
750 **Supplementary Figure 5: Spatial expression patterns for select AEH groups for slides:**
751 **A)** C1; **B)** D1; genes from the first featured AEH group for each slide were predicted with
752 low accuracy; genes from the final two AEH groups were predicted with high accuracy
753
754 **Supplementary Table 3: Genes differentially spatially autocorrelated between**
755 **METS/No-METS**

	Ground Truth		
Randomly Selected Genes	**Pathways**	**Overlap**	**Adj. P-Value**
STXBP3	Epithelial Mesenchymal Transition	17/200	2.53e-09
IMPAD1	Coagulation	11/138	0.000006432
ING1	TGF-beta Signaling	11/200	0.002024
YY1AP1	Myogenesis	11/200	0.0001214
IGFBP4	Apical Junction	10/200	0.0001214
ACO1	Hypoxia	5/54	0.0005478
ID3	Cholesterol Homeostasis	6/144	0.02979
PTAR1	UV Response Dn	7/199	0.02582
SYNM	UV Response Up	7/200	0.0286
SPRY2	IL-2/STAT5 Signaling	7/200	0.02582
RABL3	Estrogen Response Late	6/158	0.02582
N4BP2L2	p53 Pathway	4/74	0.02582
	Inception		

Randomly Selected Genes	Pathways	Overlap	Adj. P-Value
DPYSL3	Epithelial Mesenchymal Transition	9/200	3.41e-13
CALD1	TGF-beta Signaling	3/138	0.004267
G0S2	Coagulation	3/200	0.001833
IGHG1	TNF-alpha Signaling via NF-kB	2/54	0.003624
LIMS2	Angiogenesis	2/200	0.05432
COL6A2	Xenobiotic Metabolism	2/200	0.02452
C1S	Hypoxia	2/200	0.02452
SERPINF1	Myogenesis	2/200	0.02452
FOSB	Complement	2/200	0.02452
SORBS1	Inflammatory Response	1/36	0.02452
WIPF1	IL-6/JAK/STAT3 Signaling	1/87	0.1169
PCOLCE	Interferon Alpha Response	1/97	0.119

GAT

Randomly Selected Genes	Pathways	Overlap	Adj. P-Value
ACTA2	Epithelial Mesenchymal Transition	18/200	1.85e-27
COL5A2	Angiogenesis	4/36	0.000002955
COL3A1	Myogenesis	5/200	0.00009592
COL12A1	Glycolysis	3/200	0.01585
GREM1	Apical Junction	3/200	0.01585
CDH11	Hedgehog Signaling	2/138	0.1196
SPARCL1	Apical Surface	2/144	0.1321
CRYAB	Coagulation	2/200	0.06564
POSTN	UV Response Dn	2/200	0.06564
TPM2	TGF-beta Signaling	1/36	0.1475
DES	Hypoxia	1/44	0.09354
DCN	Adipogenesis	1/54	0.09354

756
757
758