Cell–ECM Graphs: A Graph-Based Method for Joint Analysis of Cells and the Extracellular Matrix

M.Ghafoor[a], JE.Parkinson[b,c], TE.Sutherland[d], M.Rattray[a]

a. Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PL, UK.

b. Lydia Becker Institute of Immunology and Inflammation, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PL, UK

c. Wellcome Centre for Cell-Matrix Research, Faculty of Biology, Medicine and Health, Manchester University of Manchester, Manchester, M13 9PL, UK

d. Institute of Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, AB25 2ZD, UK

Corresponding authors: magnus.rattray@manchester.ac.uk, mohamed.ghafoor@postgrad.manchester.ac.uk

Keywords: Cell graphs , Cell-ECM graphs, Cell-ECM Interactions, Extracellular Matrix

Abstract

The extracellular matrix (ECM) provides essential structural and biochemical support to tissues, yet its spatial organization is often underrepresented in computational analyses of spatial proteomics. We present Cell–ECM Graphs, a computational framework that integrates ECM and

cellular components into a unified graph structure, enabling joint modelling of cell–cell, ECM–ECM, and cell–ECM interactions. Using both simulated and real data, we evaluate performance across interaction profiling, node classification, spatial clustering, and explainable graph neural networks. Incorporating ECM enhances node classification and reveals biologically meaningful spatial patterns, including disrupted immune–ECM connectivity in allergen-exposed lungs. Cell–ECM Graphs offer a novel approach to analysing ECM architecture and its role in shaping the tissue microenvironment in health and disease.

Introduction

The extracellular matrix (ECM) is a three-dimensional, non-cellular scaffold present in all tissues. It is constructed from proteins, glycoproteins, glycosaminoglycans, and proteoglycans. At the cellular level, the ECM participates in cell adhesion, migration, proliferation, and differentiation. While at the organ level it influences biomechanical properties, such as tensile strength and elasticity[1].

The importance of the ECM is highlighted by the spectrum of severe genetic disorders caused by mutations in ECM related genes[2]. Furthermore, the lung ECM, which we demonstrate our novel method on, can be remodelled in diseases such as asthma[3], chronic obstructive pulmonary disorder[4] and lung cancer[5].

Cell graphs[6,7] utilize graph theory for the analysis of biological data and aim to learn structure-function relationships by modelling the structural organization of a tissue[8]. Generally, cells are represented as nodes and cell-cell interactions as edges, although the design may be modified by

the addition of other biological entities. Cell graphs have been applied to a broad range of tissues (e.g., brain[7], kidney[9] and bone[10]), experimental techniques (e.g., haematoxylin and eosin (H&E[11,12], imaging mass cytometry (IMC) and spatial transcriptomics[9]), and pathologies (e.g., cancer[7,12], diabetes[13] and fractures[10]). The graph representation enables the use of traditional graph analytics such as interaction tests [14] and deep learning methods such as graph neural networks (GNN)[15]. Furthermore, cell graphs and GNNs have been utilized with explainability techniques to identify important structures within the graph, such as nuclei importance in ductal carcinoma[16]. While many tools exist for analysing cells in biological tissues, few focus on the ECM, underscoring the need for novel methods to study the ECM.

Despite the vital nature of the ECM there has been little research attempting to incorporate it into cell graphs. ECM-aware cell graphs were created from H&E images whereby cells are nodes and coloured based on ECM surroundings[10]. However, this method is still cell-focused and does not have a separate representation of the ECM. Given the ECM's importance in the structure, function and pathology of organs, in addition to its inherent network-like structure, incorporating the ECM into cell graphs presents significant potential for analysis using graph theory and graph machine learning. Notably, by integrating ECM graphs with cell graphs, the application of graph analytical methods commonly used for cells-graphs can easily be adapted to focus on the ECM and cell-ECM interactions.

Our work presents a novel graph design called Cell-ECM graphs, which incorporates the spatial distribution of cells and the ECM into a single graph. Importantly, the graph design includes two types of nodes, cellular and extracellular. This enables the analysis of cell-cell, ECM-ECM and cell-ECM interactions, as well as alterations that may occur in disease. We test our method on both

simulated and real datasets and apply GNNs and explainability techniques to identify and visualize key cell-ECM and ECM-ECM interactions. Moreover, we show that cell-ECM graphs can improve node classification and spatial clustering compared to traditional methods. Finally, we show that down-stream graph analytics usually applied to cell graphs, such as interaction tests, can be adapted for use in Cell-ECM graphs.

Method

Simulated Spatial Proteomic Dataset with ECM Markers

To systematically evaluate cell-ECM Graphs, we generated a simulated IMC dataset with three ECM markers and controlled spatial organization of ECM and cell types. The ECM was simulated on a 500×500 grid by randomly generating 100 points and connecting them via Delaunay triangulation (ECM marker 1). Additionally, a hollow circular structure (ECM marker 2) containing a smaller circle (ECM marker 3) were added at random locations. Each ECM channel was assigned a distinct intensity range (500, 1000, and 1500, respectively) to simulate marker variability.

Cell types were then randomly overlaid onto the ECM: 4000 cells of type A on ECM marker 1, 200 cells of type B on ECM marker 2, and 50 cells each of types C and D on ECM marker 3. To test the ability of Cell-ECM Graphs to capture changes in ECM-ECM and cell-ECM interaction, two datasets were generated. The first altered ECM-ECM interactions by either removing or retaining connections between ECM markers 1 and 2. The second altered cell-ECM interactions by adding 50 type E cells to ECM marker 1. To mimic the scale of murine datasets, each dataset contained 40 regions of interest (ROIs), with 20 per condition.

Murine Imaging mass cytometry Dataset

The dataset is derived from the murine lung and the markers are outlined in **Table 1**. A total of 36 760 x 790µm regions of interest (ROI) were captured from 5 mice administered Phosphate-Buffered Saline (PBS), as controls, and 6 mice administered a multiallergen cocktail (Dust mite, Ragweed and Aspergillus fumigatus; DRA), both via the intranasal route [17]. Additional information regarding the IMC dataset can be found in the original paper [18].

Table 1. Cellular and ECM markers

Name	Target Alias	Metal Ion	Type
Hyaluronic acid binding protein	HABP	^{89}Y	ECM
Chitinase-like protein 4	Ym2	^{115}In	Cellular
Actin smooth muscle actin	aSMA	^{141}Pr	Cellular
Collagen VI	Col-VI	^{142}Nd	ECM
Heparan Sulfate	HepSul	^{143}Nd	ECM
Desmin	Desmin	^{144}Nd	Cellular
Protein S100-A4	s100a4	^{145}Nd	Cellular
Surfactant Protein C	SPC	^{146}Nd	Cellular
B220-PE	B220-PE	^{147}Sm	Cellular
Pan-cytokeratin	Pan-cyto	^{148}Nd	Cellular
CD11b	CD11b	^{149}Sm	Cellular
RELM alpha	RELMa	^{151}Eu	Cellular
Chitinase-like protein 3	Ym1	^{152}Sm	Cellular
Collagen IV	Col-IV	^{153}Eu	ECM
Chondroitin Sulphate	Chon-Sul	^{154}Sm	ECM
SCGB1A1	Scgb1a1	^{155}Gd	Cellular
Vimentin	Vimentin	^{156}Gd	Cellular
E-cadherin	E-Cad	^{158}Gd	Cellular

Fibrinogen	Fibrinogen	^{159}Tb	ECM
CD44	CD44	^{160}Gd	Cellular
Ly-6C	Ly6C	^{162}Dy	Cellular
Major Histocompatibility Complex class II	MHCII	^{163}Dy	Cellular
Von Willebrand Factor	vWF	^{164}Dy	Cellular
Collagen III	Col-III	^{166}Er	ECM
Fibronectin	Fibronectin	^{167}Er	ECM
Laminin γ1	lamg1	^{168}Er	ECM
Collagen I	Col-I	^{169}Tm	ECM
Podoplanin -FITC	Podo-FITC	^{170}Er	Cellular
Mucin 1	Muc1	^{172}Yb	Cellular
Tyrosine-protein kinase Mer	MerTK	^{174}Yb	Cellular
Platelet-derived growth factor receptor alpha	PDGFRa	^{175}Lu	Cellular
Lymphocyte antigen 6G	Ly6G	^{176}Lu	Cellular
Intercalator - (DNA)	DNA1	^{191}Ir	Cellular
Intercalator - (DNA)	DNA3	^{193}Ir	Cellular

Cell segmentation and annotation

Cell segmentation was performed using the Steinbock pipeline [21]. The DeepCell option with min-max normalization was selected. This approach uses a pre-trained model that predicts cell segmentations from cytoplasmic and nuclear markers from the raw image data. Cells were manually annotated using biological knowledge from known cell markers.

Cell-ECM graph generation

A schematic of steps required to build a Cell-ECM graph is outlined in **Fig 1**. The equation for a graph is $G = (V, E)$, where V represents a number of vertices and E a set of edges between vertices. The cell graph was constructed by treating each cell as a vertex (or node) and the cell type label assigned to each vertex. To define the edges (or links), representing cell-cell interactions, we find the $N_{edge}(G)$ nearest neighbouring cells based on spatial proximity. Subsequently, we pruned any edges with a distance greater than $D_{max}(CC)$ to refine the Cell graph. This "cell-cell distance" hyperparameter controls the length-scale of a cell interaction.

To represent the ECM as a graph, only the ECM markers were selected. The image was divided into P_{size} x P_{size} patches with N channels, where N is the number of ECM markers. K-means clustering, with K_{ECM} clusters, was then applied to these patches where patches were represented by a N-dimensional vector representing the mean of each marker within each patch. Patches assigned to a background cluster (little or no ECM expression and relabelled as ECM cluster 0) were removed and the remaining patches became vertices with a label determined by the cluster assignment. To create edges representing ECM-ECM interactions, an edge was added if two ECM nodes are neighbouring and no background nodes are between them.

The merging of the cell graph and ECM-graph through cell-ECM edges produced the final Cell-ECM graph. Cell-ECM interactions were added based on finding the $N_{edge}(G)$ nearest ECM nodes to each cell, If a cell is within $D_{max}(CE)$ of the ECM node, a link was added (this "cell-ECM distance" is the final hyperparameter which controls the length-scale of the cell-ECM interactions). This

integration created a comprehensive graph that captured cell-cell, ECM-ECM and cell-ECM interactions.

Several experiments were performed to test the effect of the Cell-ECM graph hyperparameters. The effects of modifying P_{size} and K_{ECM} are visualized in (**Fig. S1**). The silhouette score was used to select the optimal K_{ECM} as 4 and P_{size} as 5. The optimal $D_{max}(CC)$ and $D_{max}(CE)$ were selected based on the results of Bayesian optimization during node classification (see **Table 2** for optimized hyperparameters**)**.

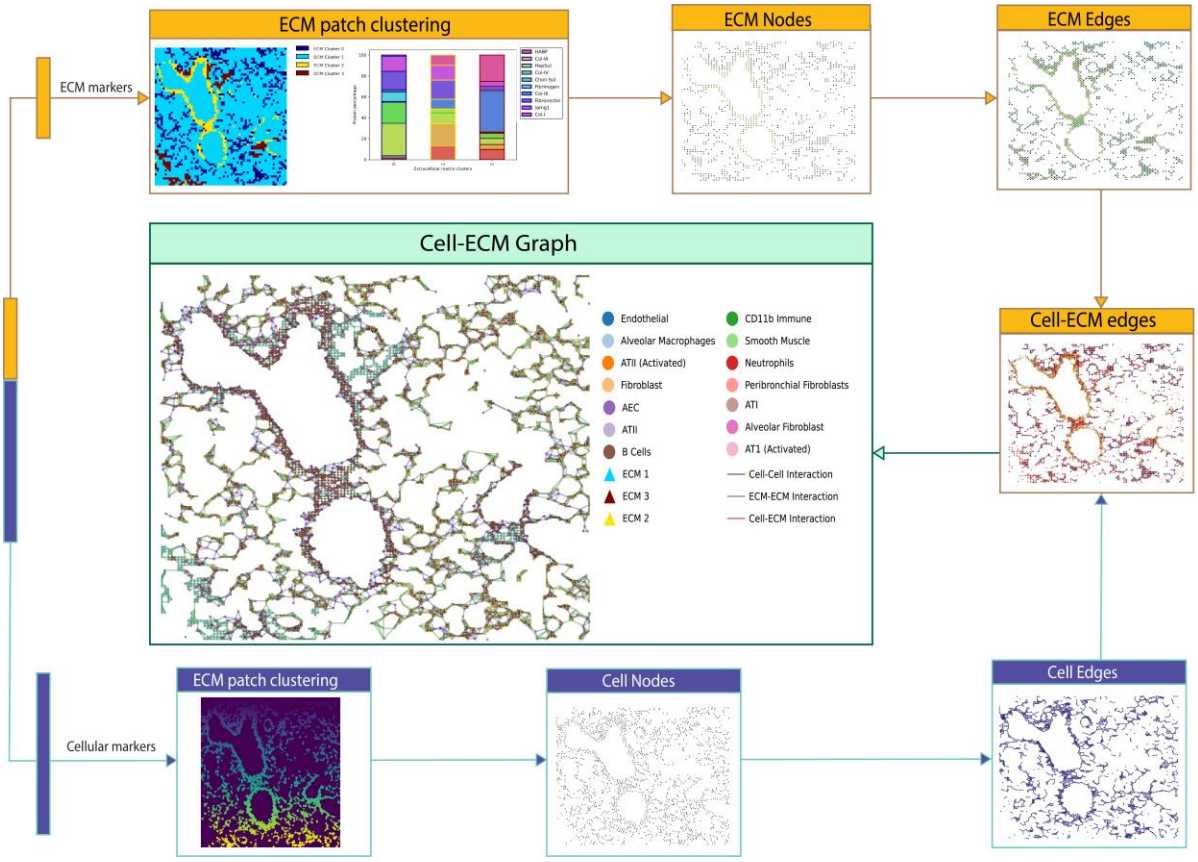


Figure 1: Cell-ECM graph Method. All the markers are divided into cellular (bottom, blue) and extracellular markers (top, orange). The cellular markers are first used to segment (DeepCell) and annotate cells, which subsequently become cellular nodes. Cell-cell interaction edges based on the 5 closest cells and pruned based on $D_{max}(CC)$ creating a cell graph. The image only using the ECM markers (top, orange) is then split into Nx5x5 patches, where N is the

number of ECM markers. Next, the mean is taken creating an average of ECM markers per patch and clustered into ECM clusters. After background removal, the remaining ECM patches become ECM nodes. ECM-ECM interaction edges are added if neighbouring. Next, cell-ECM interaction edges are added and pruned using $D_{max}(CE)$ to produce the final Cell-ECM graph (centre, green).

Table 2. Optimized hyperparameters of Cell-ECM Graphs

Hyperparameter	**Definition**	**Default**
$D_{max}(CC)$	Max cell-cell distance to be considered an interaction	17
$D_{max}(CE)$	Max cell-ECM distance to be considered an interaction	7
P_{size}	Patch size of ECM clusters	5
$N_{edge}(G)$	Number of nearest neighbour interactions	24
K_{ECM}	Number of ECM clusters for patch clustering	4

Interaction test

The interaction test followed that of imcRtools testInteractions function [21]. The 'classic' method was used to count and summarize interactions, whereby each count was divided by the total number of cell types or ECM clusters. The labels of these counts were randomly shuffled 1000 times to create a distribution of the interaction count under spatial randomness. The observed interaction count was compared against this Null distribution to derive empirical p-values (p=0.05). A significant value of +1 indicated significant interactions, -1 avoidance and 0 neither.

These values were also summed to compare the interactions of multiple ROIs under a specific condition.

Node Classification and Optimization

Once the graphs were built, we performed cellular node classification, using either the neighbouring cells (predicting cell types from direct cell-cell interactions), neighbouring ECM cluster (predicting cell types from direct cell-ECM interactions) and neighbouring cell and ECM (predicting cell types from interacting cells and ECM). This was performed using all the ROIs of the real murine dataset, a feature vector was created by counting the number of cellular, ECM or both interactions for a specific node. A K-nearest neighbour classifier was used to assign a label to each node. For Bayesian optimization, the data was split into 80% training, 10% validation and 10% testing split and optimized for 100 iterations using Mango [19] by maximizing the F1-score for cell graphs and cell-ECM graphs independently. The hyper parameter search space $D_{max}(CC)$and $D_{max}(CE)$ was from 0 to 50, while the search space for N_{edge}(G) and $K_{classifier}$ ranged from 3 to 30. Subsequently, with the optimal hyperparameters for each graph type, 5-fold cross validation was performed to generate more robust results.

Benchmarking Spatial Clustering Methods

To benchmark various spatial clustering methods, we first constructed a cell-graph and a cell-ECM graph from a simulated dataset containing four spatial regions. These regions differed in their combinations of cell-ECM interactions: two regions consisted of a single cell type on differing

ECMs, one featured a unique cell type on a specific ECM, and the last contained two cell types on the same ECM.

Cell neighborhood feature vectors were generated for each cell by counting the number of neighboring interactions with each cell type. Similarly, cell-ECM neighborhood feature vectors were generated by counting the number of interactions with both cell and ECM types. These feature vectors were clustered using K-means with K=4.

Local Indicators of Spatial Association Clustering (LisaClust) does not use a graph representation. It operates on spatial coordinates and cell type labels by performing local enrichment analysis using Ripley's K-function and clustering based on spatial autocorrelation. A key hyperparameter is the interaction radius, which was optimized (see **Supplementary Table S2**). Although originally designed for cell data, we adapted the method by treating ECM nodes analogously to cells.

Leiden clustering is a graph-based community detection algorithm that partitions the graph to optimize a quality function such as modularity. It does not use node features (i.e., cell and ECM labels). A key hyperparameter is the resolution, which controls cluster granularity—higher values yield more clusters and lower values yield fewer (see **Supplementary Table S3** for optimization).

Spectral clustering operates by constructing a similarity matrix from the graph, computing its Laplacian, and performing dimensionality reduction via its eigenvectors. K-means is then applied to the resulting low-dimensional representation. Like Leiden, it does not incorporate node features. The primary hyperparameter is the number of clusters K, set to 4 in this benchmark.

Graph Neural Network and Explainability

We implemented a graph neural network (GNN) for node classification on Cell–ECM graphs, node features were set as one-hot encoding of the node label (cell type or ECM cluster). The dataset was randomly split into 80% training, 10% validation, and 10% testing. The model consisted of three GCNConv layers followed by two fully connected layers. The hidden dimension was set to 366, and dropout was applied at a rate of 0.5. Message passing was followed by global add pooling to generate graph-level embeddings. The final output was passed through a log-softmax activation to predict graph class probabilities. The model hyperparameters were optimized using Bayesian optimization.

Adam optimizer with a learning rate of 1e-3 and negative log-likelihood loss function were used. Early stopping was employed based on validation loss to prevent overfitting.

To interpret model predictions, we applied Integrated Gradients[20], which attributes prediction outcomes to edge importance by integrating gradients from a baseline graph (edge weights set to 0) to the actual input graph (edge weights set to 1). For each edge, the attribution is approximate as:

$$Attribution_{ei} = \int_{\alpha=0}^{1} \frac{\partial F(x_a)}{\partial w_{ei}} d\alpha$$

where is $x_a$the graph with edge weights scaled by α and (x_a) is the model's output. The integral was estimated using a discrete Riemann sum implemented with the Captum library [21]. The GNN was built using PyTorch Geometric [22].

Code

The graphs were generated using Python, and the code for the Cell-ECM Graphs tool is publicly available at GitHub:https://github.com/moeghaf/CEGs. The fully executed notebooks to generate the main and supplementary results have been provided.

Data

The simulated datasets have been made publicly (https://zenodo.org/records/15523468), and the murine IMC dataset will be provided upon reasonable request.

Results

Overview of the Cell-ECM Graph Method

We developed a novel graph-based framework that integrates traditional cell-cell interactions with ECM components, enabling graph based-analysis of cell–cell, ECM–ECM, and cell–ECM interactions within a single graph structure. This method builds upon conventional cell graphs, which use only cellular nodes and cell–cell edges (**Fig. 2a**). In contrast, cell-ECM graphs incorporate both cell and ECM nodes, along with cell-ECM and ECM-ECM edges (**Fig. 2b**). From a single region of interest (ROI; **Fig. 2c**), we constructed both a cell graph (**Fig. 2d**) and a cell-ECM graph (**Fig. 2e**), illustrating the enhanced structural information captured by the latter.

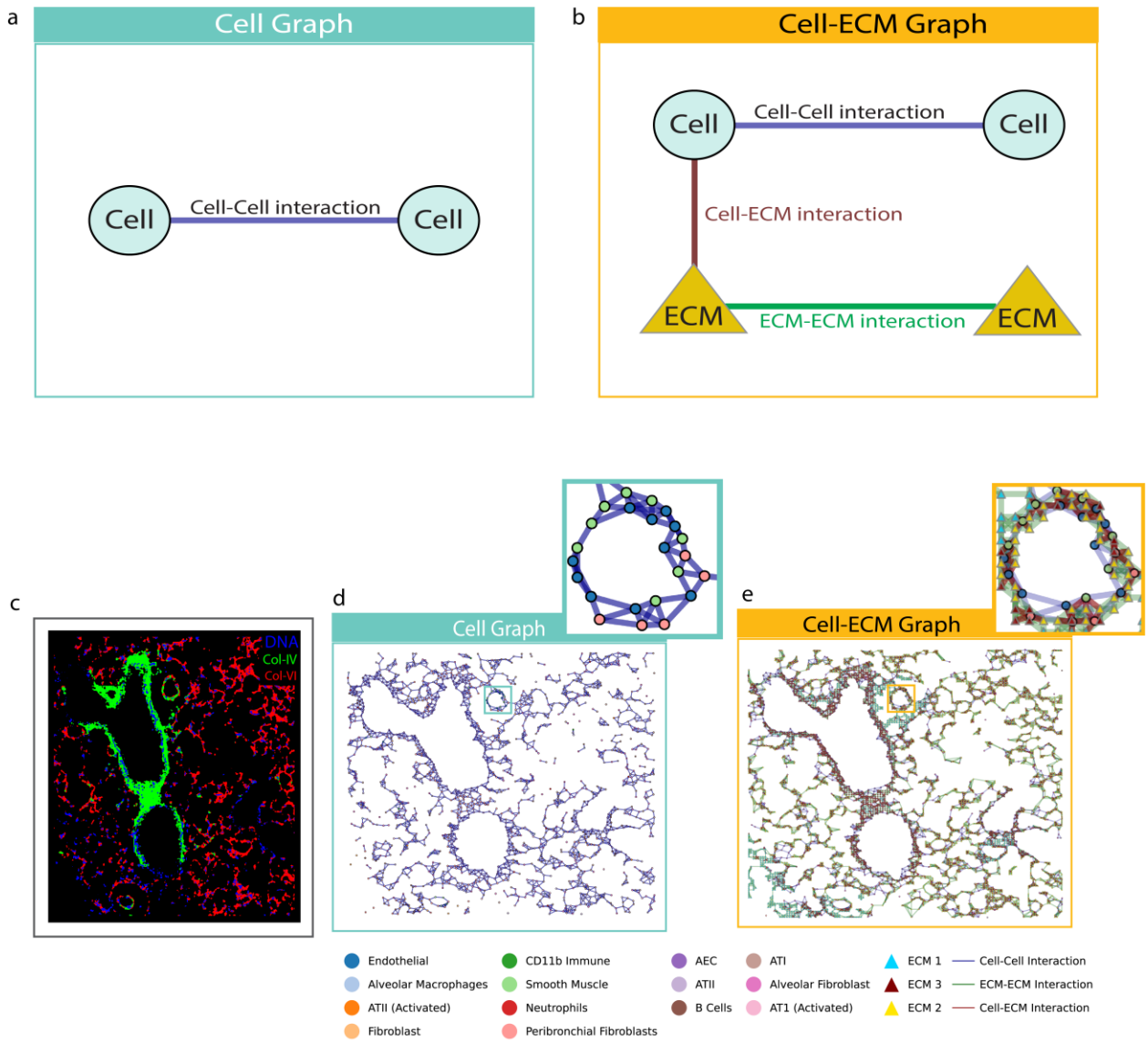


Figure 2: Overview of cell-ECM graph. **a)** Cell graph illustration showing cells as nodes and cell-cell interactions as edges. **b)** Cell-ECM Graph illustration, with cells (circle) and ECM (triangle) as nodes and cell-cell, ECM-ECM and cell-ECM interactions as edges. **c)** Visualization of single ROI with DNA (blue), collagen-IV (green) and collagen-VI (red) markers **d)** Cell graph and **e)** Cell-ECM Graph built from the same ROI with same region enlarged.

The cell-ECM Graph was constructed from a real murine IMC dataset. The first step involves generating a cell graph by representing segmented cells as nodes and their spatial interactions as edges (**Fig. 3a**). An ECM graph was then created by clustering ECM patches and connecting neighbouring clusters (**Fig. 3b**). Finally, edges were added between spatially interacting cell and

ECM nodes (**Fig. 3c**), resulting in the complete cell-ECM Graph (**Fig. 3d**). The entire murine IMC dataset ECM patch images (**Fig. S2**), cell graphs (**Fig. S3)**, ECM graphs **(Fig. S4)**, cell-ECM interactions (**Fig. S5)**, and cell-ECM graphs (**Fig. S6**) are provided.

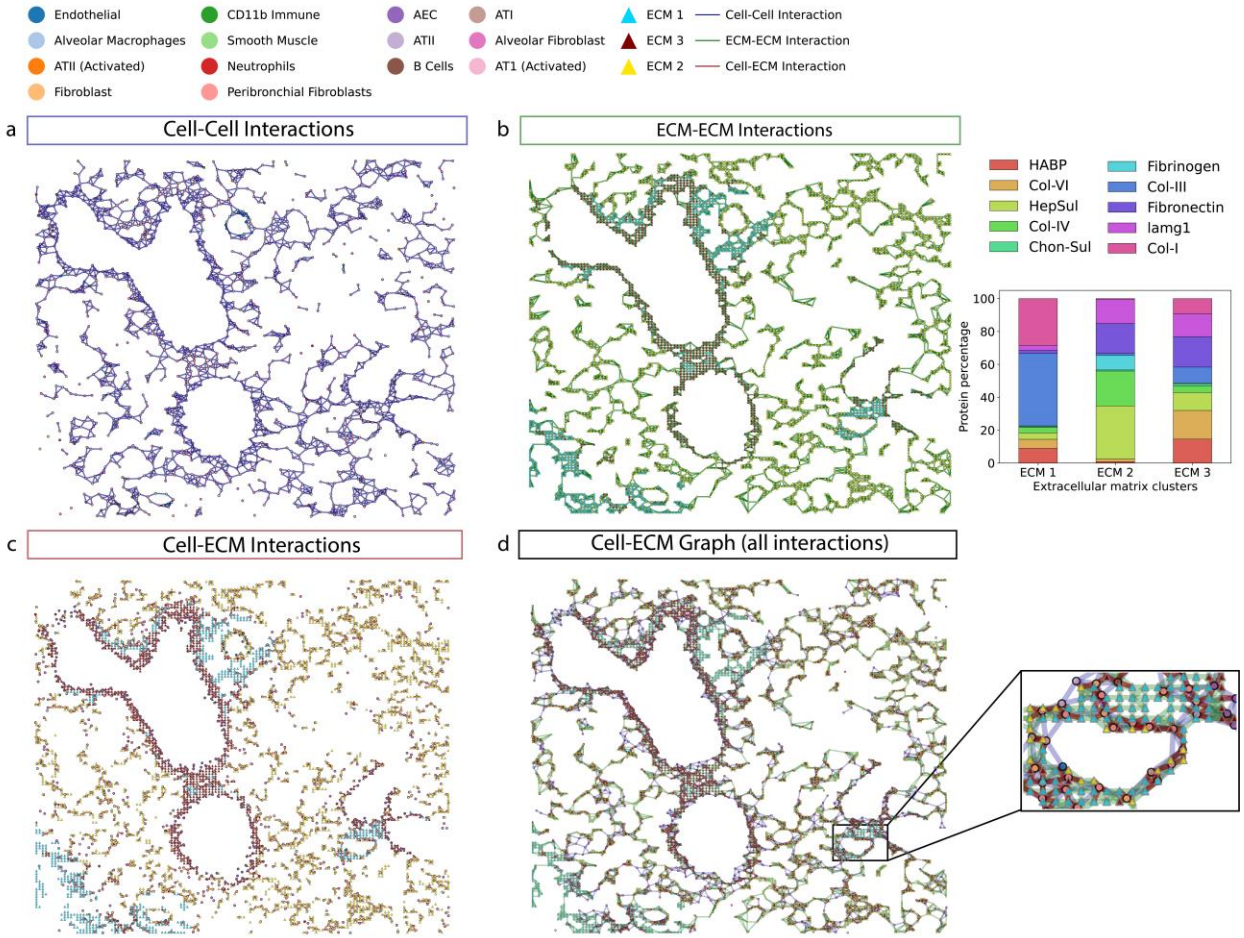


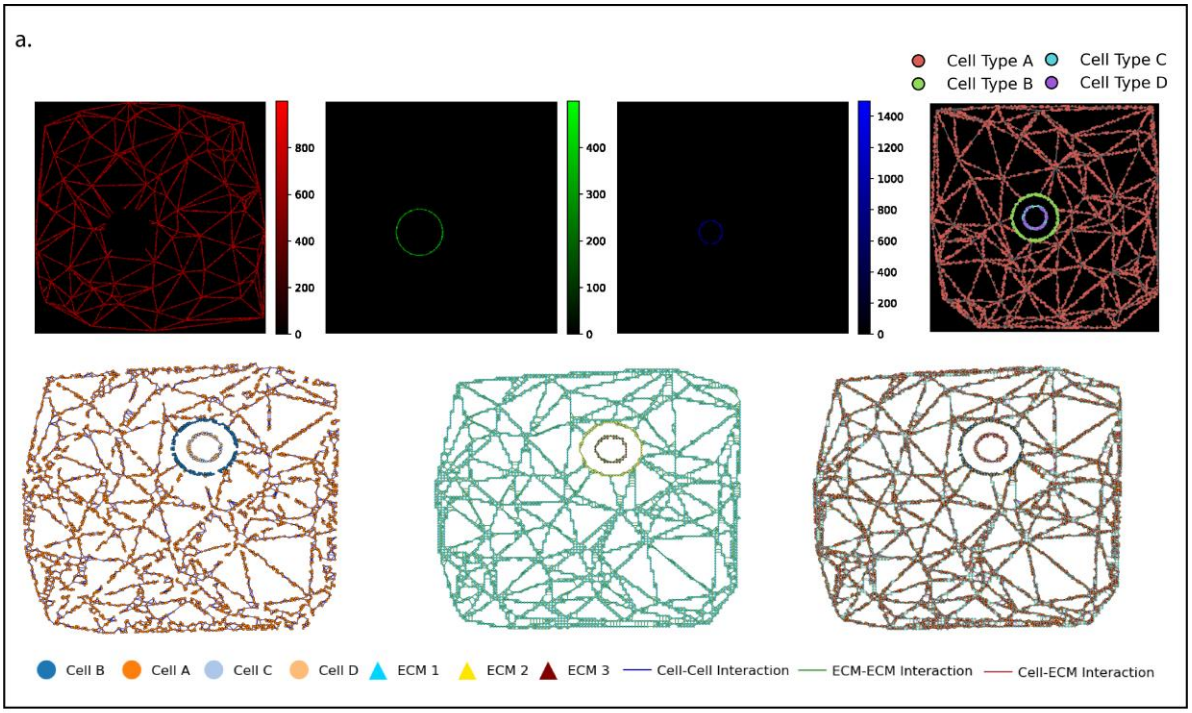
Figure 3: Visualization of Cell-ECM Graphs. **a)** Traditional cell graph representing cell-cell interactions as edges and cells depicted as nodes. **b)** ECM graph representing ECM-ECM interactions as edges, with ECM clusters as nodes and ECM molecule expression per ECM cluster. **c)** Both cells and ECM are nodes with cell-ECM interactions as edges**. d)** The full Cell-ECM graph with cells and ECM clusters as nodes and cell-cell, ECM-ECM and cell-ECM interactions as edges with a region enlarged.

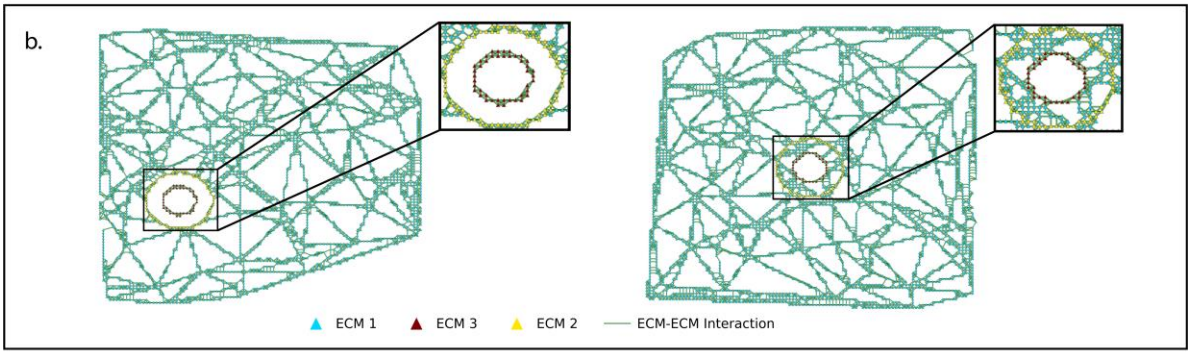
Simulated Spatial Proteomic Data for Assessment of Cell–ECM Graphs

To evaluate the utility of Cell–ECM Graphs under controlled conditions, we generated simulated spatial proteomic datasets comprising distinct ECM channels and annotated cell types. Specific cell types were localized to defined ECM clusters to establish ground truth cell–ECM relationships. As shown in Figure 4a, the simulation includes three ECM channels and four cell types, with overlaid cell–cell, ECM–ECM, and cell–ECM interactions visualized using the Cell–ECM Graph framework.

To assess the method's sensitivity to structural variation, we introduced two controlled modifications. Firstly, altered ECM-ECM interactions: a subset with ECM clusters 1 and 3 are disconnected; while in the second set, they are linked (**Fig. 4b**). Secondly, in terms of cell–ECM interactions: A fifth cell type (Cell Type E) was introduced, interacting with ECM cluster 1 alongside the dominant Cell Type A, simulating a rare but specific cell–ECM interaction (**Fig. 4c**).

These controlled variations allow us to benchmark the model's ability to detect both structural and compositional differences in the ECM microenvironment.


a.
Cell Type A
Cell Type B
Cell Type C
Cell Type D
Cell B
Cell A
Cell C
Cell D
ECM 1
ECM 2
ECM 3
Cell-Cell Interaction
ECM-ECM Interaction
Cell-ECM Interaction


b.
ECM 1
ECM 3
ECM 2
ECM-ECM Interaction

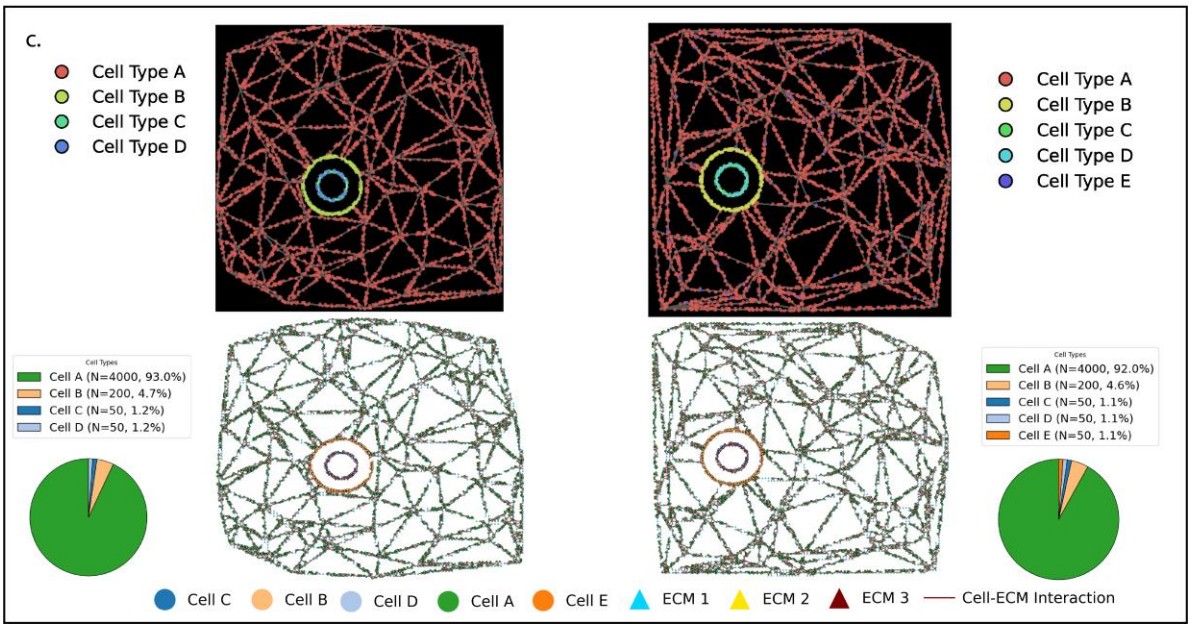

c.
Cell Type A
Cell Type B
Cell Type C
Cell Type D
Cell Type A
Cell Type B
Cell Type C
Cell Type D
Cell Type E
Cell Types
Cell A (N=4000, 93.0%)
Cell B (N=200, 4.7%)
Cell C (N=50, 1.2%)
Cell D (N=50, 1.2%)
Cell Types
Cell A (N=4000, 92.0%)
Cell B (N=200, 4.6%)
Cell C (N=50, 1.1%)
Cell D (N=50, 1.1%)
Cell E (N=50, 1.1%)
Cell C
Cell B
Cell D
Cell A
Cell E
ECM 1
ECM 2
ECM 3
Cell-ECM Interaction

Figure 4: Simulated dataset. **a)** Generation of a simulated dataset with 3 different ECM markers, and cell types placement. Creation of cell-ECM graph from the simulated dataset. **b)** Simulated ECM remodelling, with ECM clusters 1 and 3 interaction in one scenario, and not in the other. **c)** Simulated cell-ECM interactions, with a rare cell type (cell E) being introduced those interactions with ECM 1, along with cell type A.

Graph Neural Network with Explainability Highlights Key ECM-ECM and Cell–ECM Interactions

Graph neural networks (GNNs) were employed to classify entire graphs, distinguishing conditions such as control versus disease. To aid in identifying key interactions driving classification, we applied an explainability techniques to highlight the most influential graph edges (**Fig. 5**).

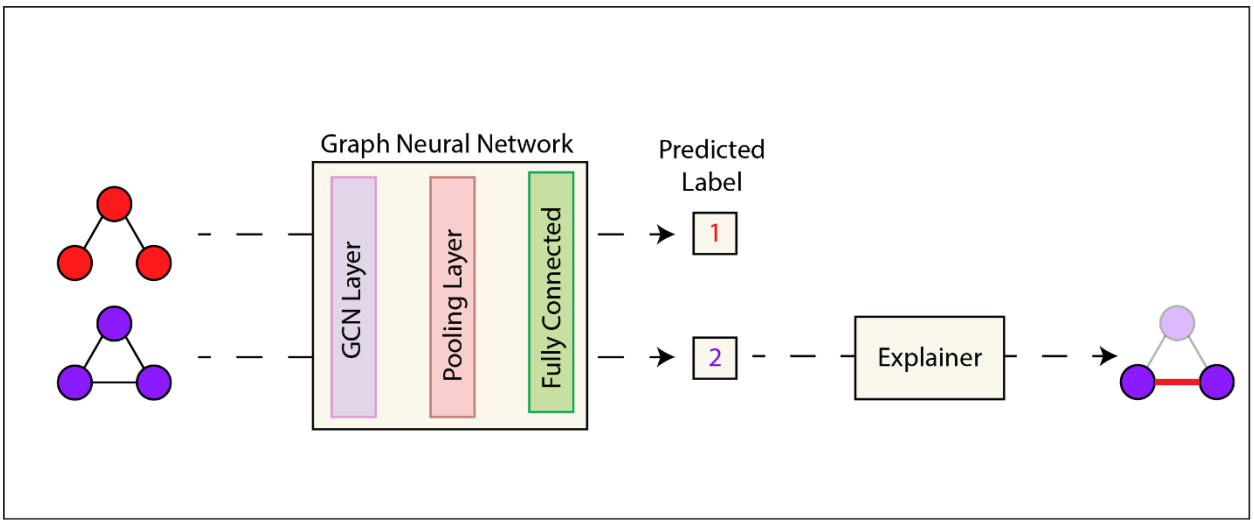


Figure 5. Explainable AI overview. A graph neural network is first trained to classify two groups of graphs with different interaction patterns. Subsequently a GNN explainer is used to highlight the important interactions used to distinguish the two groups.

We first trained the model on a simulated dataset of ECM graphs representing two distinct interaction patterns mimicking ECM remodelling. Each graph contained three ECM clusters. The key difference between groups was the interaction between ECM clusters 1 and 3: in Group 1, these clusters remained disconnected, whereas in Group 2, they were directly connected (**Fig.**

4b). The GNN reached 100% accuracy on the test set and the explainer successfully identified this distinction, with ECM 1–3 interactions ranking among the topmost important features in Group 2 (**Fig. 6a**) and ECM 1-2 interactions being the most important in Group 1. **Figure S7** shows the training and validation loss.

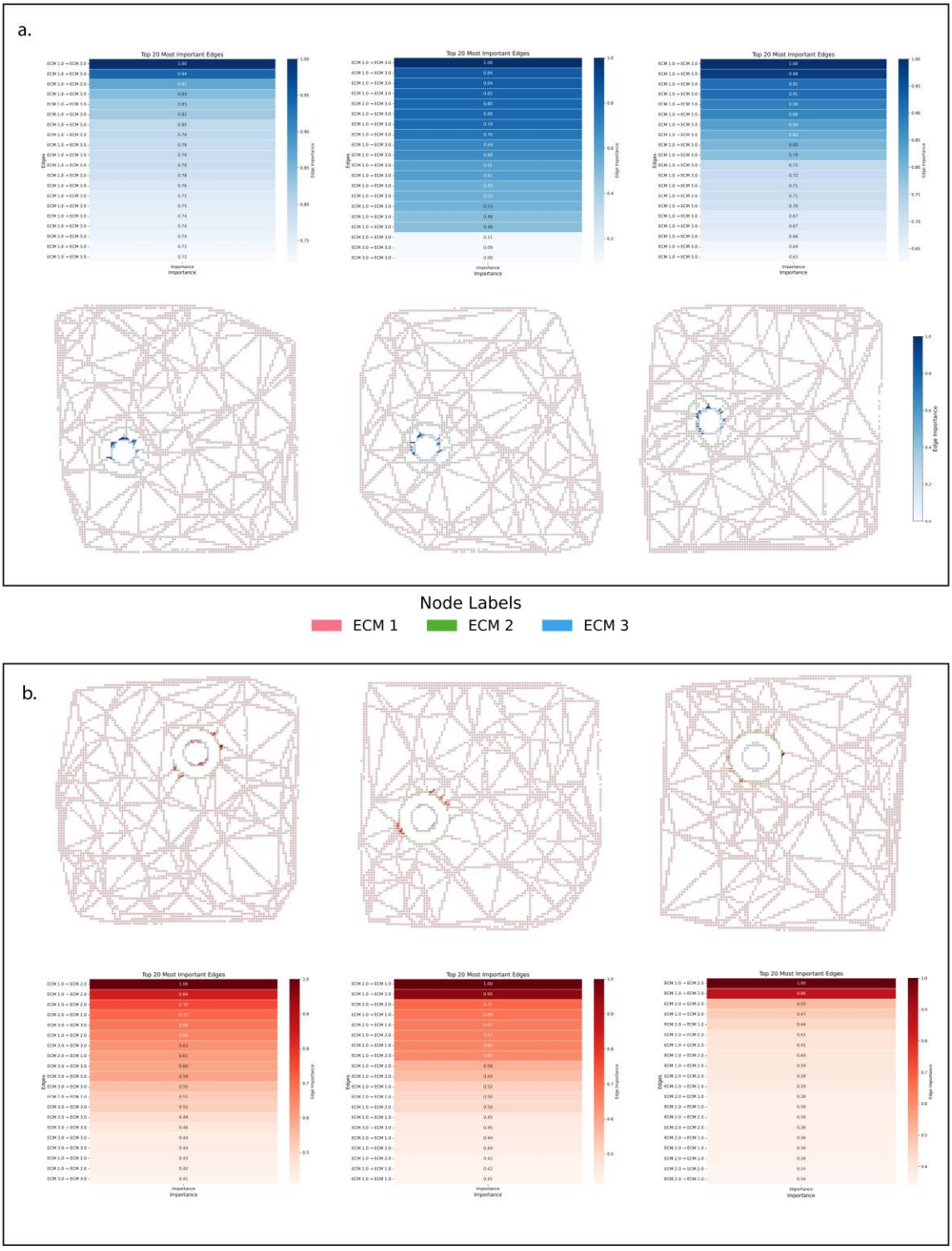


Figure 6. GNN Explainer on simulated ECM-Graphs. Visualization of GNN explainer results on simulated ECM graphs, highlighting the ECM remodelling: **a)** ECM 1 and 3 interactions present and **b)** absent.

Next, we evaluated cell–ECM graphs with a focus on cell–ECM interactions using a simulated dataset where ECM–ECM interactions were unchanged, but cell–ECM interactions varied. To mimic rare immune cell infiltration, Group 1 lacked Cell E, while Group 2 included a small number of Cell E nodes interacting with ECM 1 (**Fig. 4c**). The model achieved 100% accuracy on the test set and identified Cell E–ECM 1 edges as most important for classification (**Fig. 7**). **Figure S8** shows the training and validation loss.

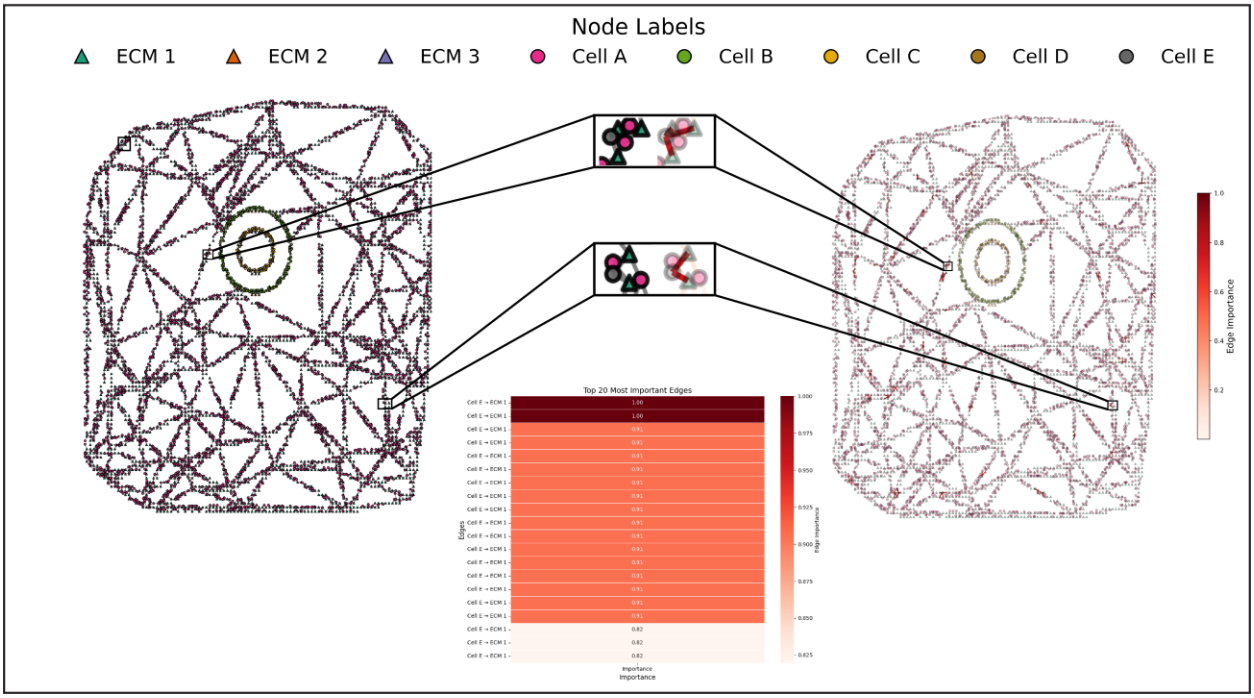
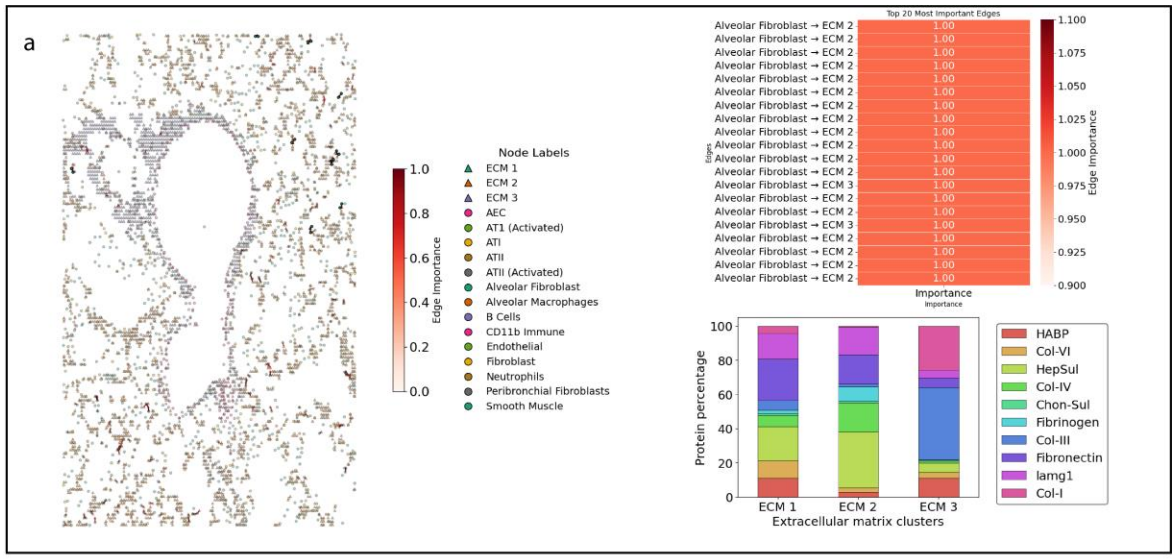
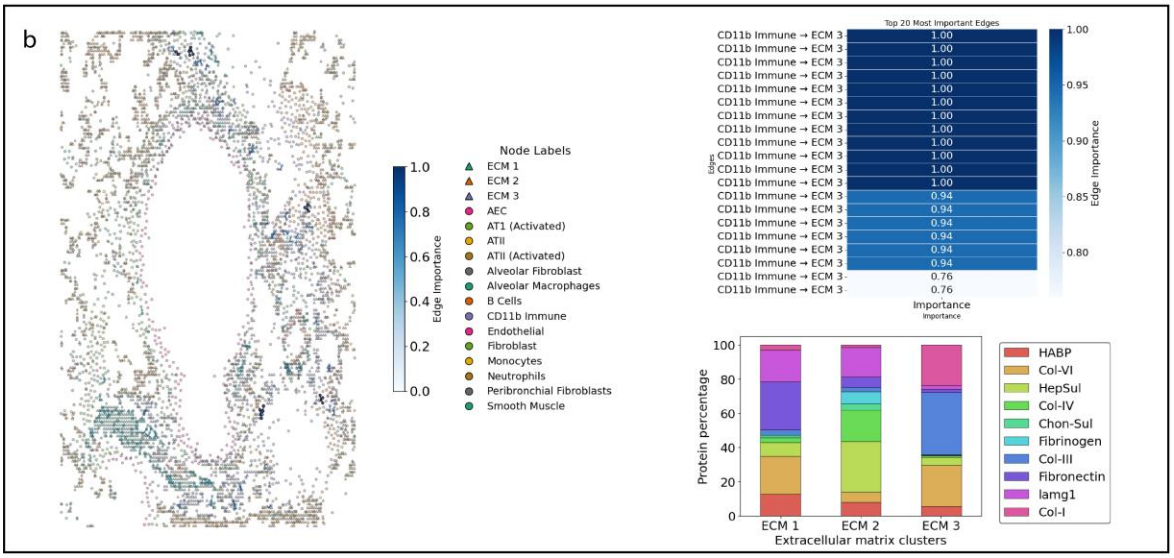


Figure 7. GNN Explainer on simulated cell-ECM Graphs. Visualization of GNN explainer results on simulated cell-ECM interactions, highlighting the ground truth modification of cell-ECM interactions, cell-E interacting with ECM cluster 1.

We then applied the same method to a real dataset, murine tissue treated with PBS or DRA. Focusing on cell-ECM interactions, an accuracy of 75% on the test set was achieved. In PBS tissue, the top-ranked interactions were between Alveolar fibroblasts cells and ECM 2 (**Fig. 8a**). While in

DRA-treated tissue, the most important interactions were between $CD11b^+$ immune cells with ECM 3 (**Fig. 8b**) and monocytes with ECM 1 (**Fig. 8c**).


a
Node Labels
ECM 1
ECM 2
ECM 3
AEC
AT1 (Activated)
ATI
ATII
ATII (Activated)
Alveolar Fibroblast
Alveolar Macrophages
B Cells
CD11b Immune
Endothelial
Fibroblast
Neutrophils
Peribronchial Fibroblasts
Smooth Muscle
Edge Importance
Top 20 Most Important Edges
Alveolar Fibroblast → ECM 2
Alveolar Fibroblast → ECM 3
Edges
Importance
Protein percentage
ECM 1
ECM 2
ECM 3
Extracellular matrix clusters
HABP
Col-VI
HepSul
Col-IV
Chon-Sul
Fibrinogen
Col-III
Fibronectin
Iamg1
Col-I


b
Node Labels
ECM 1
ECM 2
ECM 3
AEC
AT1 (Activated)
ATII
ATII (Activated)
Alveolar Fibroblast
Alveolar Macrophages
B Cells
CD11b Immune
Endothelial
Fibroblast
Monocytes
Neutrophils
Peribronchial Fibroblasts
Smooth Muscle
Edge Importance
Top 20 Most Important Edges
CD11b Immune → ECM 3
Edges
Importance
Protein percentage
ECM 1
ECM 2
ECM 3
Extracellular matrix clusters
HABP
Col-VI
HepSul
Col-IV
Chon-Sul
Fibrinogen
Col-III
Fibronectin
Iamg1
Col-I

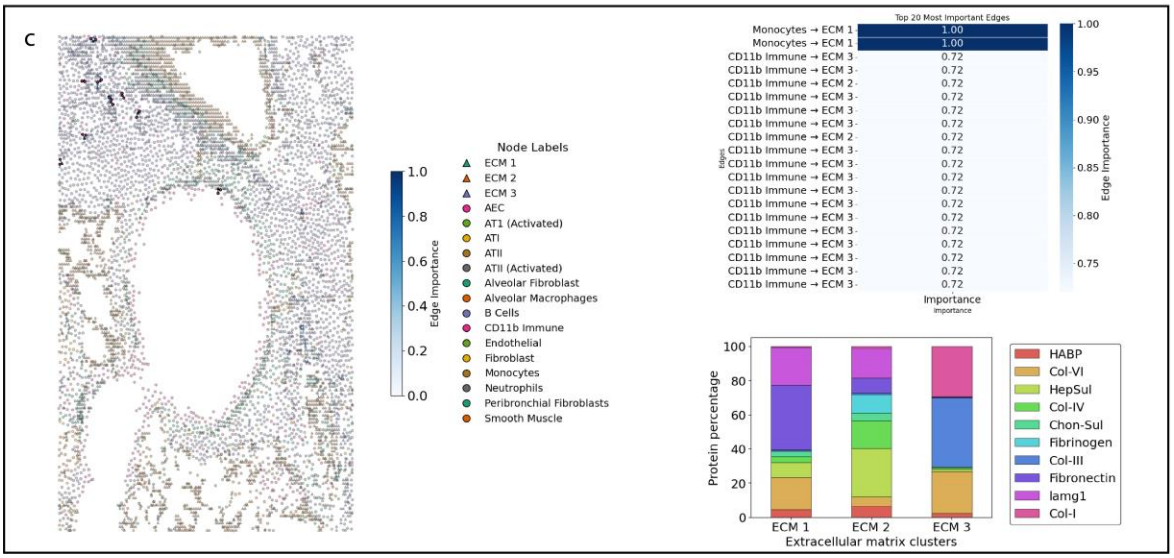
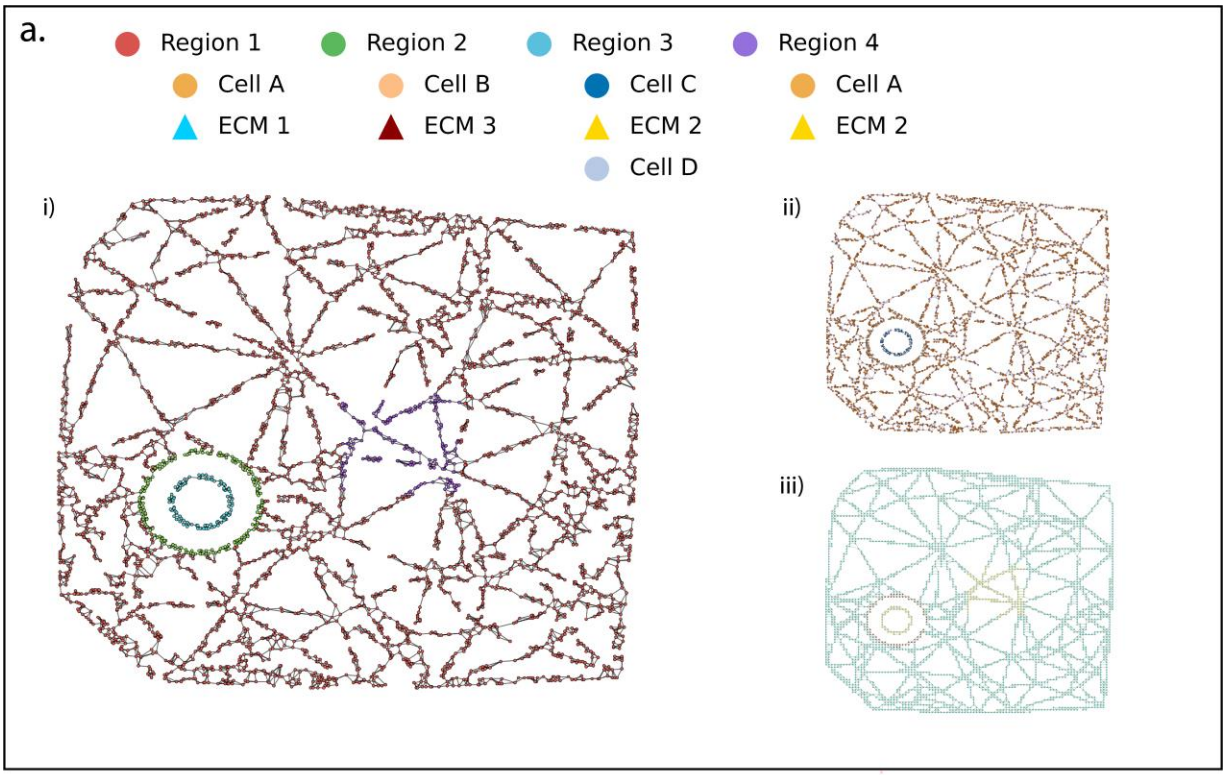

c
Node Labels
ECM 1
ECM 2
ECM 3
AEC
AT1 (Activated)
ATI
ATII
ATII (Activated)
Alveolar Fibroblast
Alveolar Macrophages
B Cells
CD11b Immune
Endothelial
Fibroblast
Monocytes
Neutrophils
Peribronchial Fibroblasts
Smooth Muscle
Edge Importance
Top 20 Most Important Edges
Monocytes → ECM 1
CD11b Immune → ECM 3
Edges
Importance
Protein percentage
ECM 1
ECM 2
ECM 3
Extracellular matrix clusters
HABP
Col-VI
HepSul
Col-IV
Chon-Sul
Fibrinogen
Col-III
Fibronectin
Iamg1
Col-I

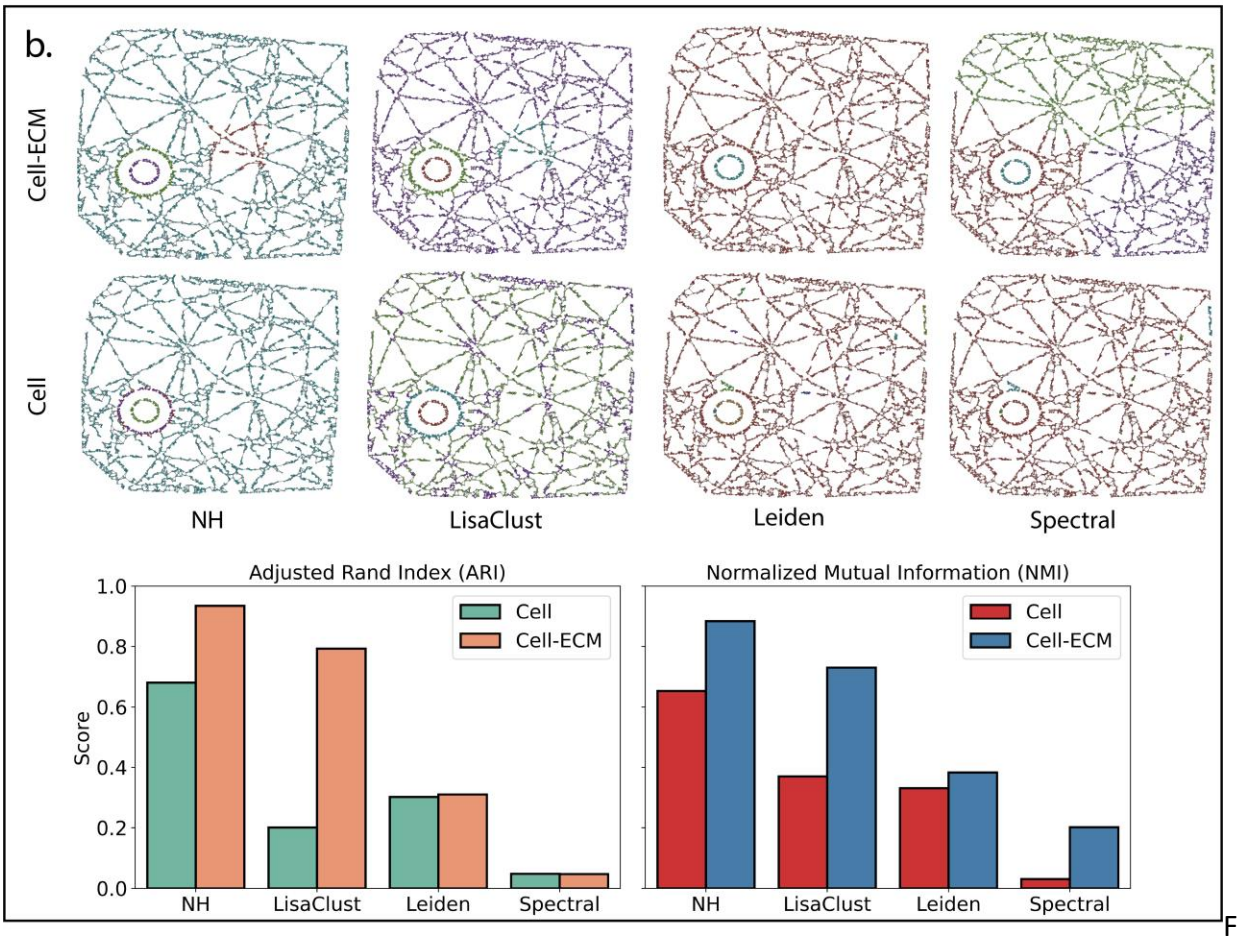
Figure 8. GNN Explainer on real cell-ECM Graphs. Visualization of GNN explainer results on real cell-ECM interactions in **a) PBS**, where alveolar fibroblasts interactions with ECM 2 were ranked important edges. Important cell-ECM interactions in **DRA** highlight **b)** CD11B immune cell interactions with ECM 3 and **c)** Monocyte interactions with ECM 1 as important interactions.

Benchmarking of cell-ECM graphs in terms of spatial clustering

To evaluate cell-ECM graphs ability to identify various spatial regions, we compared three clustering tools (neighbourhood analysis [23,24], LisaClust [25], spectral clustering [26] and Leiden clustering [27] using normalized mutual information (NMI) and adjusted rand index (ARI). Each method was adapted to incorporate the ECM information for comparison.

To assess each tool's ability to identify regions with different combinations of cell and ECM, we created a simulated dataset with four different spatial regions that differ in cellular and ECM combinations (**Fig. 9a**). Regions 1 and 4 have the same cell type but different ECM (cell type A on ECM 1 and 2 respectively). Region 2 contains a cell type on a unique ECM (cell type B on ECM 3) while region 3 contains two cell types on the same ECM (cell types C and D on ECM 2). Cell-ECM graphs delivered the best performance across all methods compared to using only the cellular information. In terms of methods, neighbourhood analysis achieved the highest performance.


a.
Region 1
Region 2
Region 3
Region 4
Cell A
Cell B
Cell C
Cell A
ECM 1
ECM 3
ECM 2
ECM 2
Cell D
i)
ii)
iii)


b.
Cell-ECM
Cell
NH
LisaClust
Leiden
Spectral
Adjusted Rand Index (ARI)
Normalized Mutual Information (NMI)
Score
1.0
0.8
0.6
0.4
0.2
0.0
Cell
Cell-ECM
NH
LisaClust
Leiden
Spectral

F

Figure 9. Benchmarking Spatial Clustering. **a)** Simulated dataset with i) 4 spatial regions that differ in cell and ECM combinations. Region 1 and 4 consists of the same cell type on different ECM, region 2 has a specific cell type on a unique ECM, while region 3 contains the two cell types on the same ECM. Visualize of the ii) cell-graph and iii) ECM graph. **b)** Visualization and results of benchmarking spatial clustering**.**

Cell–ECM Interactions Improves Node Classification

To further evaluate whether incorporating ECM information adds to the graph representation, we use node-level predictions to compare cell and cell-ECM graphs in a real dataset. To ensure fairness, all hyperparameters were optimized independently for each graph type. When using hyperparameters optimized for the cell graph, both the cell graph and cell–ECM Graph achieved a macro F1 score of 46.49%. Since removing cell-ECM edges from the cell–ECM graph reproduces the cell graph structure, the same result is expected. However, under cell–ECM graph optimized settings, the cell–ECM graph achieved a higher macro F1 score of 52.53%, while the cell graph scored 45.47% (**Table 3**). In addition, with the optimal hyperparameters for each graph type, we performed 5-fold cross validation, cell-ECM graphs outperformed cell graphs across all 5 folds (**Fig. 10**). These results indicate that incorporating ECM neighbours adds meaningful information for classifying cell types at the node level.

Table 3. Cell vs Cell-ECM Node Classification HP after optimization

Optimized	$D_{max}CC$	$D_{max}CE$	N_{graph}	$K_{classifier}$	CEG F1	CG F1
Cell graph	50	0	3	29	0.4649	0.4649
Cell-ECM graph	17	7	24	15	**0.5253**	0.4547

- CG = Cell graph
- CEG = Cell-ECM graph

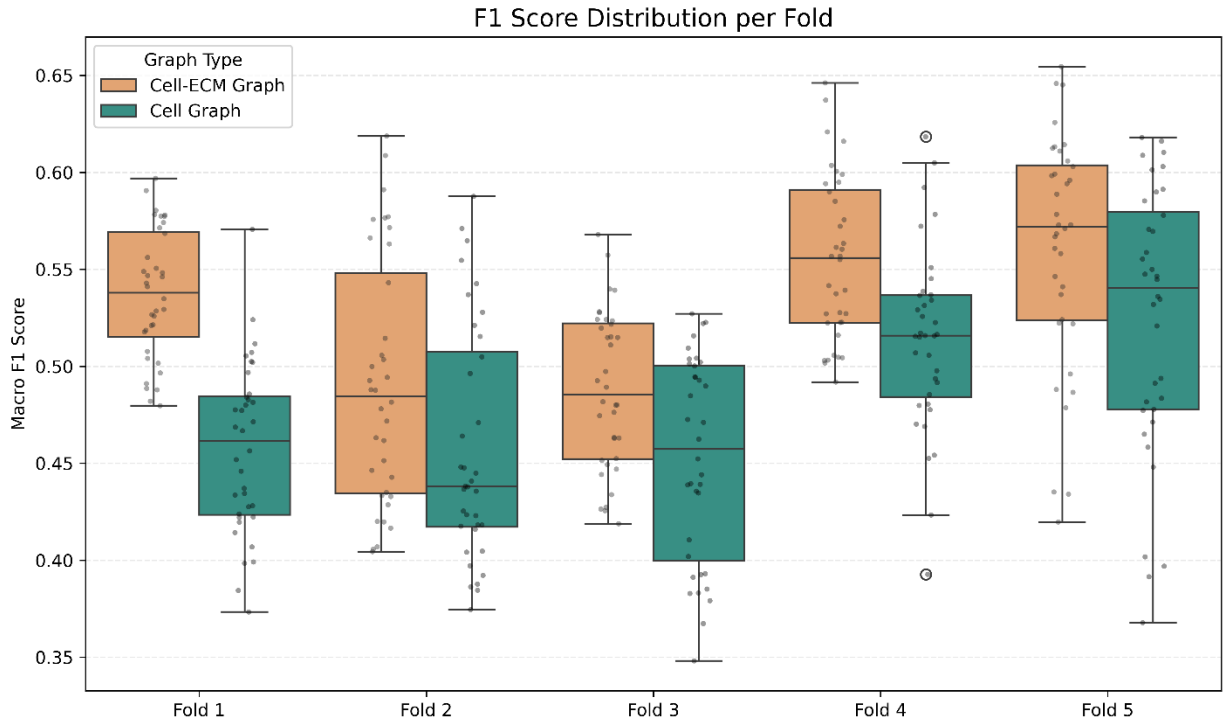
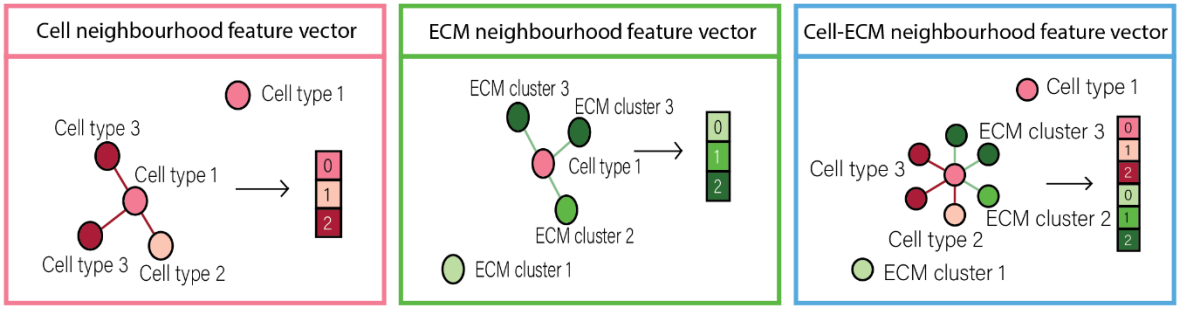


Figure 10. Five-fold cross validation of Node Classification. Using all available ROIs in the real murine dataset, five-fold cross validation was performed to compare performance of cell and cell-ECM graphs and node classification with the optimized hyperparameters for each graph type.

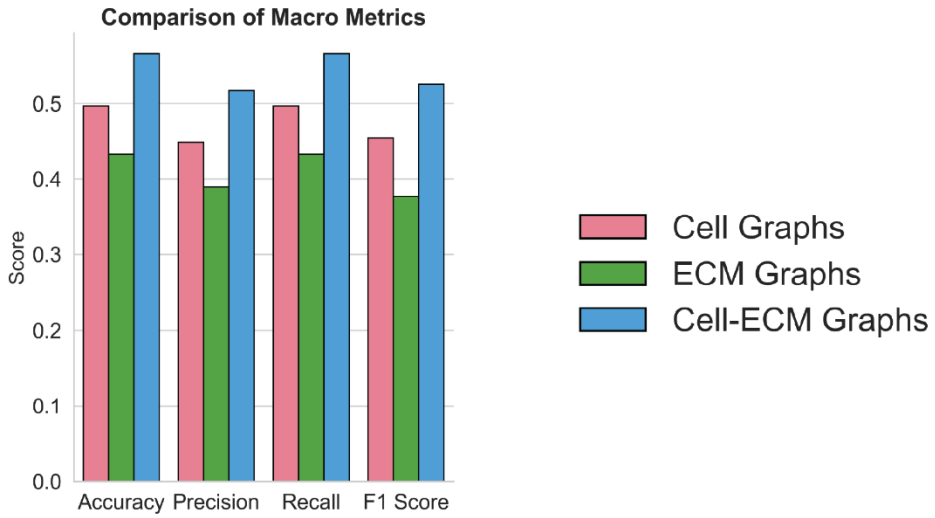
Using the optimal cell-ECM parameters, we further evaluated performance across individual cell types using either neighbouring cell interactions, ECM interactions or combined cell and ECM interactions (**Fig. 11a**). ECM interactions alone yielded the lowest scores but were still predictive.

Using cell-cell interactions improved performance compared to the ECM alone, while combining both cell and ECM interactions performed best across all metrics (**Fig. 11b**). Cell-type-specific results are shown in **Fig. 11c**.

a


Cell neighbourhood feature vector
Cell type 1
Cell type 3
Cell type 1
Cell type 3
Cell type 2
0
1
2
ECM neighbourhood feature vector
ECM cluster 3
ECM cluster 3
Cell type 1
ECM cluster 2
ECM cluster 1
0
1
2
Cell-ECM neighbourhood feature vector
Cell type 1
ECM cluster 3
Cell type 3
ECM cluster 2
Cell type 2
ECM cluster 1
0
1
2
0
1
2

b


Comparison of Macro Metrics
Score
0.0
0.1
0.2
0.3
0.4
0.5
Accuracy
Precision
Recall
F1 Score
Cell Graphs
ECM Graphs
Cell-ECM Graphs

c

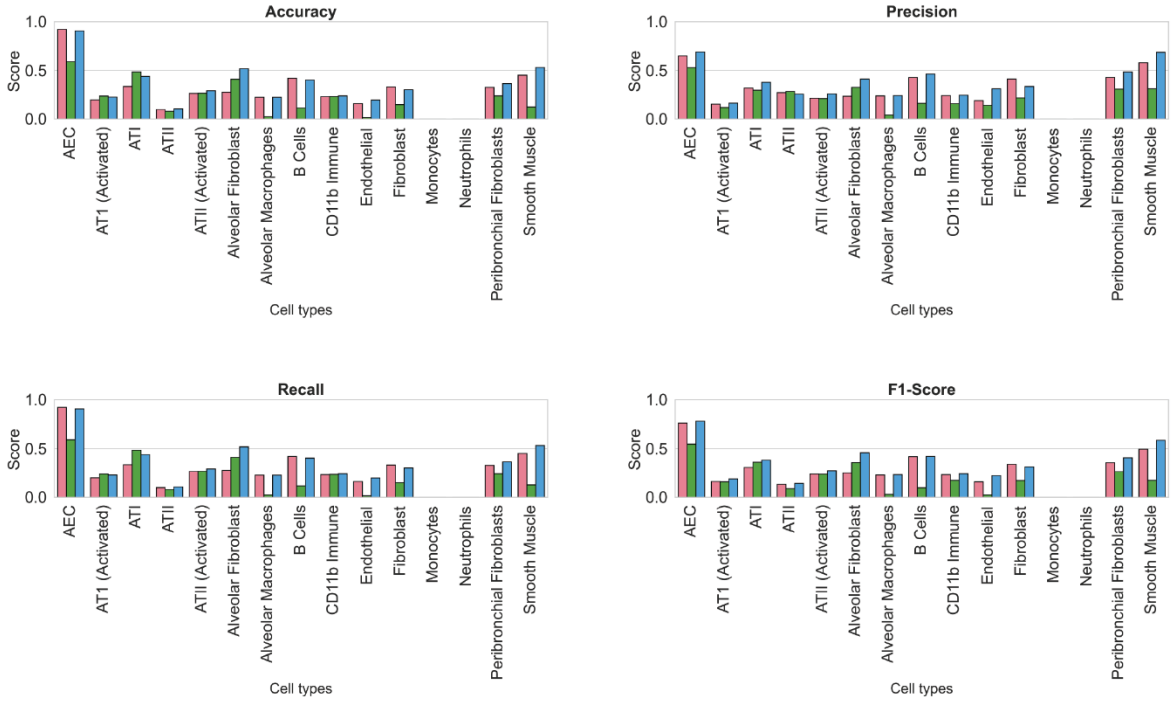

Accuracy
Precision
Recall
F1-Score
Score
0.0
0.5
1.0
AEC
AT1 (Activated)
ATI
ATII
ATII (Activated)
Alveolar Fibroblast
Alveolar Macrophages
B Cells
CD11b Immune
Endothelial
Fibroblast
Monocytes
Neutrophils
Peribronchial Fibroblasts
Smooth Muscle
Cell types

Figure 11. Node Classification through cell, ECM and cell-ECM neighbourhoods. **a)** Schematic showing how cell, ECM and cell-ECM neighbourhood features are calculated. **b)** Macro performance of each graph design and **c)** cell-specific performance.

Interaction Analysis of Cell-ECM and ECM-ECM interactions with Cell-ECM Graphs

Interaction tests are commonly used to identify significant interactions in cell graphs, the same approach can be extended to cell–ECM graphs. This enables the detection of significant ECM–ECM and cell–ECM interactions. When applied to the simulated ECM graphs, the ground truth ECM remodelling was evident through changes in edge counts but was not captured as significant (**Fig. S10**).

We then applied the permutation test to the real murine dataset, comparing PBS and DRA conditions (**Fig. 12**). No significant changes in ECM–ECM connectivity were observed. However, altered cell–ECM interactions were detected. DRA exposure increased interactions between ATI, ATII, and fibroblasts with ECM cluster 2. Immune–ECM interactions were also affected: B cells showed reduced interactions with ECM cluster 3, while $CD11b^{+}$ immune cells exhibited a significant increase in interactions with ECM cluster 3. Monocytes, absent in PBS, interacted with ECM cluster 1 in DRA.

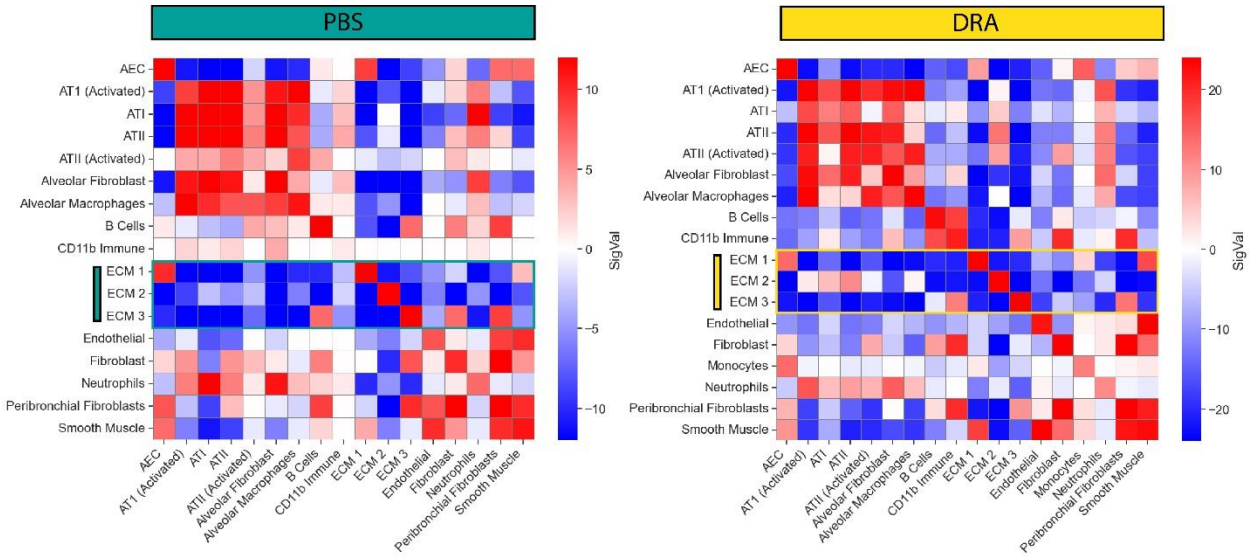


Figure 12. Interaction test on cell-ECM graphs. Interaction test was performed on cell-ECM graphs which enable significant ECM-ECM and cell-ECM interactions to be found in PBS vs DRA.

Discussion

Cell-ECM graphs is a graph representation capable of incorporating the cellular and ECM components into a unified graph enabling a more complete tissue representation. To model a tissue, a graph is constructed comprising nodes of cells and ECM clusters, along with cell-cell, ECM-ECM and cell-ECM edges.

Several computational methods have been developed for the spatial analysis of cells in spatial proteomics, such as cell segmentation [28], single-cell clustering and visualization [29], interaction analysis [14], spatial community analysis [30], and cellular neighborhood analysis [31]. In contrast, there is a lack of computational tools to aid ECM analysis. Studies that have included ECM markers have tended to conduct comparatively limited ECM analysis compared to cellular analysis, such as visualization of co localization [32] and measurement of intensity by expansion around biological structures[34].

Methods commonly used for spatial analysis of cells can be adapted for use in cell-ECM graphs. We illustrate this by using an interaction test on cell-ECM graphs, which allowed the visualization of significant cell-cell, ECM-ECM and cell-ECM interactions. We also extend cellular neighbourhood analysis to incorporate ECM, resulting in cell–ECM and ECM neighbourhood analyses. Among all methods benchmarked, cell–ECM neighbourhood analysis performed best. Both cell–ECM neighbourhood analysis and LisaClust (adapted to include ECM) successfully identified all four spatial regions. Although LisaClust is not graph-based, it effectively captures local spatial associations and improved with incorporating the ECM. In contrast, Leiden and spectral clustering are graph-based but do not incorporate cell type or ECM cluster labels, likely contributing to their lower performance. Even with these methods cell-ECM graphs outperformed the cell only representations.

Graph deep learning methods with explainability were adapted to focus on the ECM and cell-ECM interactions. GNN explainers have been applied to a range of biological datasets, such as the mutagenicity of molecules [35], drug-disease therapeutic mechanisms [36], EEG data of schizophrenia [37] and breast cancer cell graphs [38]. However, to the best of our knowledge, we are the first to apply it to the ECM. Our GNN explainer successfully identified ECM remodeling and rare cell-ECM interactions in the simulated dataset. In terms of the real murine lung dataset, alterations in monocyte/CD11b+immune cell ECM interactions were highlighted. The role of monocytes and CD11b+ immune cells in DRA exposed mice has been previously documented [39].

In summary, cell-ECM graphs represent a tool for the analysis of ECM and cell-ECM interactions in spatial proteomic datasets. It is the first graph design that incorporates both cells and the ECM

into a single graph in high dimensional spatial proteomics. Alternative graph designs should be considered in future work, with modified node and edge features.

Acknowledgements

MG was supported by the Wellcome Trust Immunomatrix in Complex Disease PhD programme (WT ref. 218491/Z/19/Z).

References

1. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200 (2010).
2. Lamandé, S. R. & Bateman, J. F. Genetic Disorders of the Extracellular Matrix. *Anat. Rec. Hoboken Nj 2007* **303**, 1527–1542 (2020).
3. The role of the extracellular matrix and specific growth factors in the regulation of inflammation and remodelling in asthma - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0163725809000023?casa_token=irz5kRf-B_wAAAAA:mOOqIVtFfqnByFYqninq1Izua8-LAXtzZPR7B3UaCPbLoV5YdzVEg87jOrHXSo8HJdIBoztRs1U.
4. Karakioulaki, M., Papakonstantinou, E. & Stolz, D. Extracellular matrix remodelling in COPD. *Eur. Respir. Rev.* **29**, (2020).
5. Götte, M. & Kovalszky, I. Extracellular matrix functions in lung cancer. *Matrix Biol.* **73**, 105–121 (2018).

6. Gunduz, C., Yener, B. & Gultekin, S. H. The cell graphs of cancer. in *ISMB/ECCB (Supplement of Bioinformatics)* 145–151 (2004).
7. Demir, C., Gultekin, S. H. & Yener, B. Augmented cell-graphs for automated cancer diagnosis. *Bioinformatics* **21**, ii7–ii12 (2005).
8. Cell-graphs: image-driven modeling of structure-function relationship: Communications of the ACM: Vol 60, No 1. https://dl.acm.org/doi/abs/10.1145/2960404.
9. Shao, X. *et al.* Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat. Commun.* **13**, 4429 (2022).
10. Bilgin, C. C., Bullough, P., Plopper, G. E. & Yener, B. ECM-aware cell-graph mining for bone tissue modeling and classification. *Data Min. Knowl. Discov.* **20**, 416–438 (2010).
11. Studer, L., Wallau, J., Dawson, H., Zlobec, I. & Fischer, A. Classification of intestinal gland cell-graphs using graph neural networks. in *2020 25th International conference on pattern recognition (ICPR)* 3636–3643 (IEEE, 2021).
12. Bilgin, C., Demir, C., Nagi, C. & Yener, B. Cell-Graph Mining for Breast Tissue Modeling and Classification. in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 5311–5314 (2007). doi:10.1109/IEMBS.2007.4353540.
13. Striegel, D. A., Hara, M. & Periwal, V. The Beta Cell in Its Cluster: Stochastic Graphs of Beta Cell Connectivity in the Islets of Langerhans. *PLOS Comput. Biol.* **11**, e1004423 (2015).
14. Schapiro, D. *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).
15. Corso, G., Stark, H., Jegelka, S., Jaakkola, T. & Barzilay, R. Graph neural networks. *Nat. Rev. Methods Primer* **4**, 1–13 (2024).

16. Jaume, G. *et al.* Quantifying explainers of graph neural networks in computational pathology. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 8106–8116 (2021).
17. Combined sensitization of mice to extracts of dust mite, ragweed, and Aspergillus species breaks through tolerance and establishes chronic features of asthma - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0091674909002358?casa_token=z1CuV2pt6Q8AAAAA:v5nJVIKAX05O-xKpj5I1_iBQlEaelSAwh65TARMA9Vj0pHO1ESZfZPMg_dY8jJ11KB_KrrMo8io.
18. Parkinson, J. E. *et al.* Extracellular matrix phenotyping by imaging mass cytometry defines distinct cellular matrix environments associated with allergic airway inflammation. *bioRxiv* 2024–11 (2024).
19. Sandha, S. S., Aggarwal, M., Fedorov, I. & Srivastava, M. Mango: A Python Library for Parallel Hyperparameter Tuning. in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 3987–3991 (2020). doi:10.1109/ICASSP40776.2020.9054609.
20. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. Preprint at https://doi.org/10.48550/arXiv.1703.01365 (2017).
21. Kokhlikyan, N. *et al.* Captum: A unified and generic model interpretability library for PyTorch. Preprint at https://doi.org/10.48550/arXiv.2009.07896 (2020).
22. Fey, M. & Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric. Preprint at https://doi.org/10.48550/arXiv.1903.02428 (2019).

23. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).

24. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

25. Patrick, E. *et al.* Spatial analysis for highly multiplexed imaging data to identify tissue microenvironments. *Cytometry A* **103**, 593–599 (2023).

26. Ng, A., Jordan, M. & Weiss, Y. On Spectral Clustering: Analysis and an algorithm. in *Advances in Neural Information Processing Systems* vol. 14 (MIT Press, 2001).

27. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

28. Bannon, D. *et al.* DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes. *Nat. Methods* **18**, 43–45 (2021).

29. Levine, J. H. *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).

30. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).

31. Goltsev, Y. *et al.* Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981 (2018).

32. Elaldi, R. *et al.* High Dimensional Imaging Mass Cytometry Panel to Visualize the Tumor Immune Microenvironment Contexture. *Front. Immunol.* **12**, (2021).

33. Ramaglia, V. *et al.* Multiplexed imaging of immune cells in staged multiple sclerosis lesions by mass cytometry. *eLife* **8**, e48051 (2019).

34. Wang, Y. J. *et al.* Multiplexed In Situ Imaging Mass Cytometry Analysis of the Human Endocrine Pancreas and Immune System in Type 1 Diabetes. *Cell Metab.* **29**, 769-783.e4 (2019).
35. Ying, Z., Bourgeois, D., You, J., Zitnik, M. & Leskovec, J. GNNExplainer: Generating Explanations for Graph Neural Networks. in *Advances in Neural Information Processing Systems* vol. 32 (Curran Associates, Inc., 2019).
36. Zhang, Y. *et al.* DDTExplainer: Mining Drug-Disease Therapeutic Mechanisms based on GNN Explainability. in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 1350–1355 (2024). doi:10.1109/BIBM62325.2024.10822060.
37. Zhdanov, M., Steinmann, S. & Hoffmann, N. Investigating Brain Connectivity with Graph Neural Networks and GNNExplainer. in *2022 26th International Conference on Pattern Recognition (ICPR)* 5155–5161 (2022). doi:10.1109/ICPR56361.2022.9956201.
38. Jaume, G. *et al.* Quantifying Explainers of Graph Neural Networks in Computational Pathology. in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 8102–8112 (IEEE, Nashville, TN, USA, 2021). doi:10.1109/CVPR46437.2021.00801.
39. Plantinga: Conventional and monocyte-derived CD11b+... - Google Scholar. https://scholar.google.com/scholar_lookup?doi=10.1016%2Fj.immuni.2012.10.016&pmid=23352232.