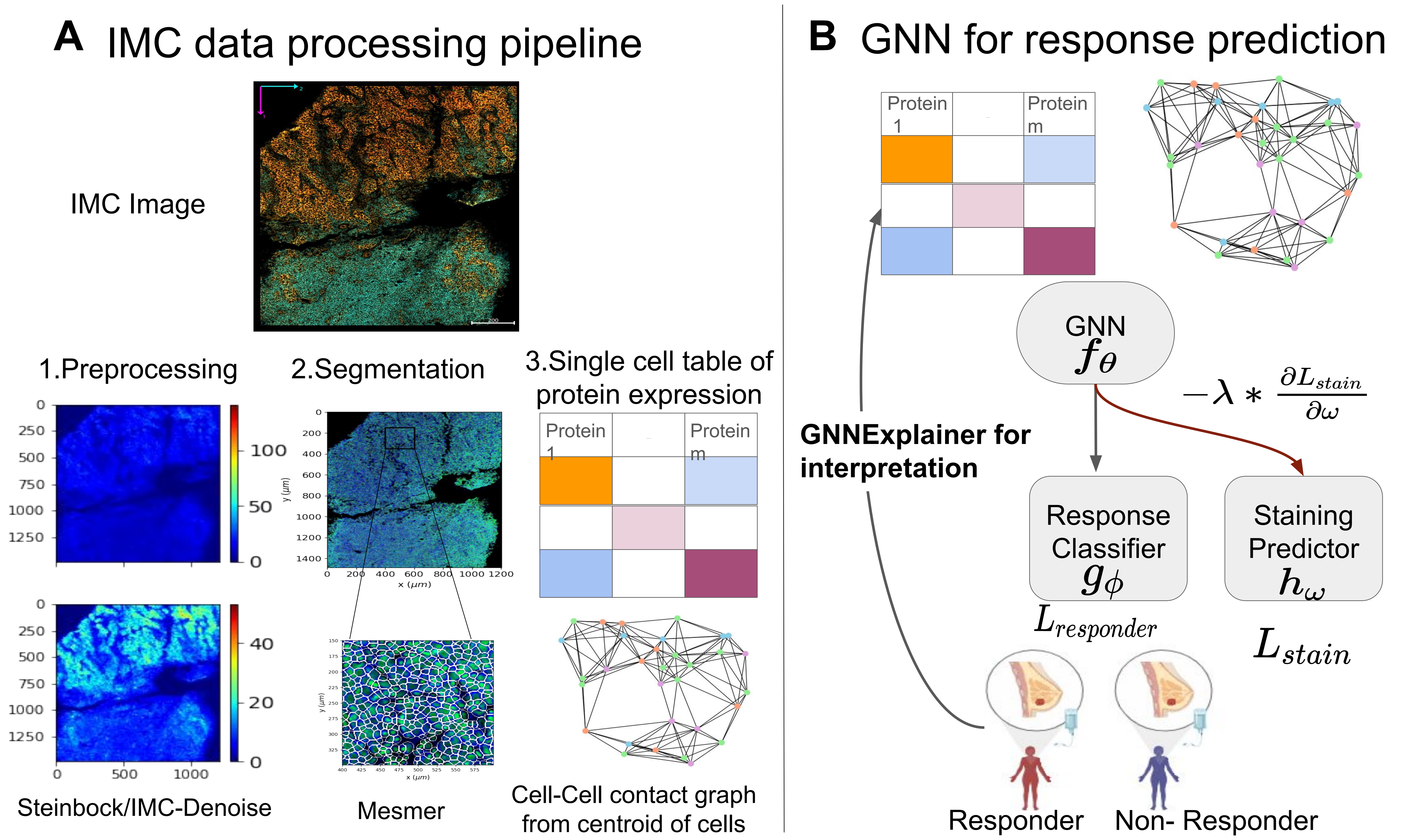
Batch-effect invariant graph neural networks for predicting chemotherapy response in triple-negative breast cancer patients

Asif Khan[1] Luciana Luque[2], Giuseppe Torrisi[2], Claudia Owczarek[3], Maddy Parsons[3], Chris Sander[1], Linus Schumacher[2]

1. Department of Systems Biology, Harvard Medical School, USA; 2. Centre for Regenerative Medicine, Institute for Regeneration and Repair, The University of Edinburgh, UK, 3, Molecular Biophysics, King's College London, UK.



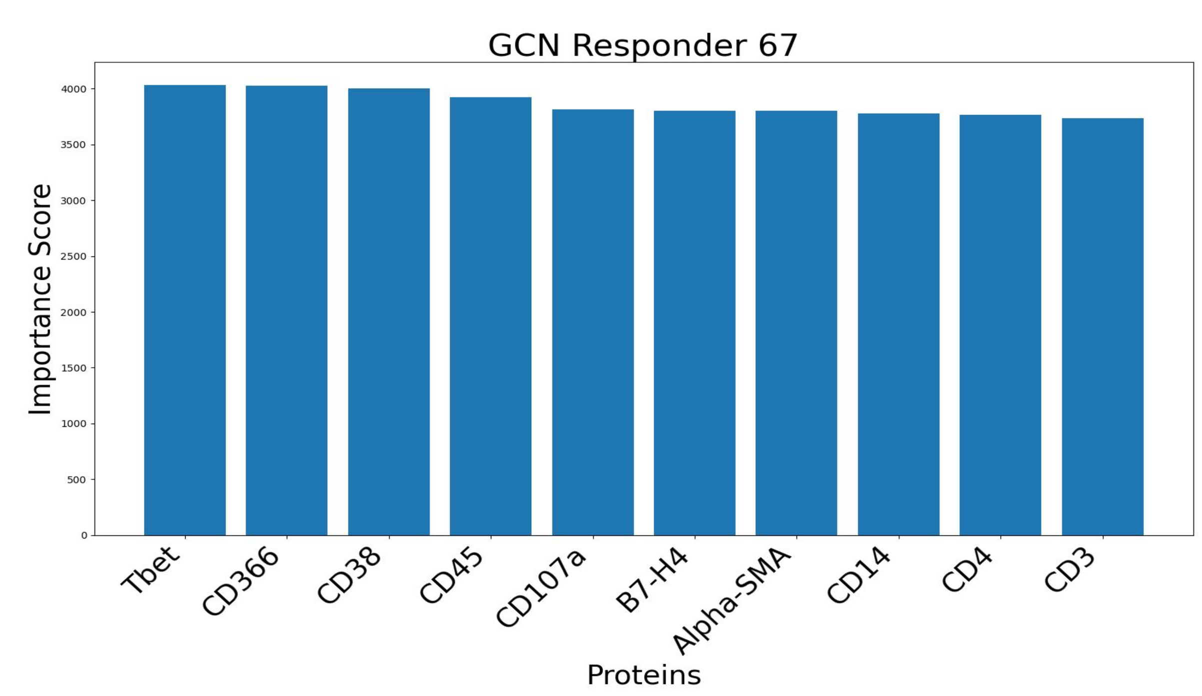
Overview of feature extraction from raw IMC images and building a response prediction model. (A) feature extraction using Steinbock and IMC-Denoise tools to mitigate artefacts such as hot pixels and shot noise. Cell segmentation is performed using Mesmer, where the centroids of identified cells denote their spatial coordinates. These centroids are then used to construct a graph, wherein the mean pixel intensities per cell are used to quantify its protein expression. (B) This protein expression data is integrated for training a GNN to predict treatment response. Following model training, GNNExplainer is used to find the most predictive protein profiles on a held-out patient data.

Methodology

- Extract protein expression profiles and centroids of individual cells from IMC images.
- Use a GNN to predict whether patients will respond to chemotherapy, while also removing batch effects using a surrogate optimization objective.

Results

- Compare the performance of the model trained with and without the staining predictor.
- used the GNNExplainer (Ying et al. 2019) algorithm to assess the importance of protein markers in the prediction of responders.



Feature Attribution. Top ten protein markers ranked by their importance score determined using GNNExplainer algorithm.

	Batch Correction		No Batch Correction	
	ROI	Patient	ROI	Patient
AUC	**0.797**	**0.928**	0.766	0.857
Accuracy	**0.751**	**0.947**	0.766	0.894
F1 Score	**0.752**	**0.923**	0.739	0.833

Batch effect correction with staining loss leads to improved performance, encouraging the GNN to learn representations invariant of batch effects.