GraphSC: An Exploration into the Applications of Graph Networks in Stem Cell Colony Analysis

Bumjin Joo

Advisor: Ritambhara Singh

Reader: Chen Sun

Submitted in partial fulfillment of the requirements for Undergraduate Honors in Computer Science at Brown University

Abstract

Recent increases in popularity for stem cell therapies have prompted interest in automated processes which could replace intensive characterization measurements. We introduce GraphSC, a fully automated pipeline which utilizes quantitative bright field micropscopy (QBAM) imagery and graph neural networks (GNNs) alongside publicly available tools like Cellpose and WIPP to predict tissue function. QBAM images of induced pluripotent stem cell–derived retinal pigment epithelial cells (iPSC-RPE) were used to train GraphSC's GNN and predict transepithelial electrical resistance (TER) and vascular endothelial growth factor (VEGF) secretion. We validate GraphSC's learned representations through multi-target prediction and transfer learning tasks. These results demonstrate that automated stem cell characterization can be improved with GNNs.

Acknowledgements

First and foremost, I would like to give my thanks to Dr. Ritambhara Singh. Thank you for having been such an amazing professor and PI throughout my various stages as an undergraduate. Thank you for having been so patient with me as I learned and grew through endless lectures, staff meetings, and lab meetings.

Likewise, Dr. Chen Sun has been foundational in my exploration through Deep Learning. Truly, this project would not have been possible without my endless learning from your seminars, courses, and staff meetings alike.

Atishay Jain and Matthew Murakami have also been excellent mentors since the start of this project. I would like to thank Atishay for consistently helping me build and guide this project through countless meetings; and Matthew for lending his expertise in countless essential tools like Cellpose and WIPP.

1 Introduction

Modern cell-based therapies have been shown to treat diseases like retinal degeneration, neurodegeneration, and certain cancers by replacing damaged native tissue with a new, functional implant. Moreover, modern cell therapy has only been accelerated by the development of stem cell reprogramming (Robinton and Daley, 2012). In particular, iPSCs (induced pluripotent stem cells) have further extended the potential of cell therapies by enabling the transplantation of iPSCs reprogrammed into specialized autologous tissue. However, a key constraint in stem cell treatment is the developmental process of these stem cells. The process of confirming the proper functional development "require[s] a trained user, are low throughput, expensive, endpoint, [and] are time consuming." Moreover, a scarcity of automated methods to validate stem cell development contributes to a difficulty in clinical applications of iPSC-based therapy (Schaub et al., 2020).

The introduction of Deep Learning (DL) techniques into clinical tasks has helped advance and accelerate our understanding of biological phenomena. As an extension on traditional machine learning (ML), DL takes advantage of stochastic calculus with deeply layered networks to model increasingly complex biology. In the context of iPSC characterization, previous works pioneer methodologies to automate this entire process, developing workflows that consist of microscopy, image processing, and finally Deep Learning techniques and entirely remove the necessity for labor intensive measurements.

These pipelines demonstrate how deep learning models, particularly Convolutional Neural Networks (CNNs), outperform various other models as part of the fully automated process. In particular, they demonstrate the ability for CNNs to outperform traditional machine learning methods while simultaneously requiring less data preprocessing steps to attain their predictions. However, these CNN models still underperform in certain critical aspects. Firstly, the published CNN mdoels still experience relatively high error on considerably consistent target values, indicating that a more sophisticated process might be necessary to fully exploit the depth of information in the microscopy images. Next, many papers take note of the inherent inability to interpret why and how predictions were made. Additionally, these projects report considerably high training costs, with large datasets and many train steps required for convergence.

To address these problems, we propose GraphSC, a graph neural network-based approach which reinterprets visual microscopy into graphs to learn stem cell colony-level features with minimal manual annotation. As a self-supervised network, we hope GraphSC yields generalizable representations of iPSC bright field microscopy to achieve state of the art results on down stream tasks that previously required entirely separate networks.

We evaluate our model on two key tasks: baseline prediction performance on stem cell developmental, or potency, metrics (TER, VEGF) and training complexity and stability. In particular, we reference paired microscopy-potency measurements to evaluate performance and compare our model to several baselines on the same task. For training complexity and stability, we perform hyperparameter tuning on both models by searching a predefined hyperparameter space and observing convergence patterns and train time on a consistent system.

2 Related Works

2.1 Stem Cells

When culturing iPSC colonies for the purpose of transplantation in therapy contexts, it is crucial to first observe the level of maturation to ensure that the cells have properly developed and will properly enable the therapeutic benefits of the treatment. In particular, the transepithelial electrical resistance (TEER or TER) values are most commonly used to assess the integrity of the stem cell colony (Srinivasan et al. (2015), Hori et al. (2024), Schaub et al. (2020)). To align the performance targets of this work with recent studies, these values will be noted as TER, though it is common to find works with either abbreviation. Additionally, the relative levels of secretion of vascular endothelial growth factor (VEGF) has been shown to indicate differing cell functionality and can indicate whether a stem cell colony has specialized into a unique function Nourse et al. (2010), Lee et al. (2018), Fan et al. (2011). It is important to note that both metrics are aggregate measures across the entire colony of cells, and that the entire colony can be characterized, at a given time, as sharing a certain TER or VEGF value.

2.2 DNNs on QBAM Microscopy

Throughout this work, I reference Schaub et al. (2020), whose methodology aims to predict TER and VEGF-ratio values from patient iPSC-RPE (Retinal Pigment Epithelium) cell images. In their work, the authors employ Deep Learning methods on QBAM, or quantitative bright-field absorbance microscopy, imagery of iPSC-RPE cells. This study paritcularly pioneers an automated, reproducible method of capturing QBAM images. Clinical-grade iPSC-RPE cells were gathered from both age-related macular degeneration (AMD) patients and healthy donors.

Schaub et al. use several ML and DL models in their findings. Firstly, a Convolutional Network (DNN-F) as well as a Multilayer Perceptron (MLP) are used to predict a single TER or VEGF potency value directly from QBAM microscopy. In particular, the VEGF value was presented as a ratio between the VEGF present on the basal and apical sides of the iPSC-RPE. Then, to validate their model against baseline performances, the authors train a separate convolutional network to segment the outlines of individual cells in a microscopy image, process the visual features of each cell (*e.g.* area, circumference, width) using the WIPP package from NIST, and then fit traditional machine learning models (TML) on those visual cell features to predict TER or VEGF values. Ultimately, the study demonstrates that their neural networks, and their DNN-F in particular, outperform the traditional machine learning models, as compared through the lens of root-mean-squared-error (RMSE).

Our GraphSC pipeline iterates upon the TML workflow by replacing DNN-S with a more modern package for cell segmentation–CellPose, and then using the per-cell information from WIPP to generate graphs to be processed by our graph network, which will be discussed more in the next section.

2.3 Graph Networks

Graph networks, or GNNs, have recently enjoyed extended popularity and interest under both pure computer science and biomedical applications alike (Li et al., 2021). As the name suggests, GNNs offer a general framework through which models can learn from data structured as graphs. In particular, GNNs are most applicable when the data can be split into a set of nodes and the prediction is expected to be based upon the relationships, or edges, between these nodes.

On a technical level, each node in a GNN updates its own representation by interacting with other nearby nodes. There are countless variants of GNNs which define different mechanisms by which the nodes interact with their neighbors.

Concretely, the input to a GNN layer will be a set of nodes (in the form of their current representations) $\{h_i \in \mathbb{R}^d | i \in \mathcal{V}\}$ and the set of edges between nodes $\mathcal{E}$. A GNN layer will then output transformed representations $\{h'_i \in \mathbb{R}^{d'} | i \in \mathcal{V}\}$, where some transformation function f_θ and aggregation function Aggregate will be applied to node i and its neighbors $\mathcal{N}_i = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}\}$:

$$\mathbf{h}'_i = f_\theta \left(\mathbf{h}_i, \text{Aggregate}(\{h_j | j \in \mathcal{N}_i\})\right)$$

2.3.1 Graph Attention Layers

Graph Attention Networks (GATs) are one such GNN variant. To perform an update, each node will attend to other nodes by using its own representation as the query. In this way, GATs generalize local-average neighbor aggregation used by GNNs by learning attention weights which enable a weighted average that highlights or hides neighboring nodes based on their learned importance.

Specifically, we make use of GATv2 layers as introduced by Brody et al. (2021), which adjusts the original node attentioning mechanism introduced by Veličković et al. (2017) by performing the nonlinear transformation (most commonly a LeakyReLU) before applying a second linear transformation onto the already linearly transformed key / query node representations. We can represent this attentioning mechanism and re-express the general GNN aggregation function as follows. We can first define e, the attention score for any given pair of nodes h_i and h_j, as well as the normalized attention value for node i and all neighbors j $\alpha_{i,j}$:

$$e(\mathbf{h}_i, \mathbf{h}_j) = \mathbf{a}^\intercal \, \text{LeakyReLU}\left(\left[\mathbf{W}\mathbf{h_i} || \mathbf{W}\mathbf{h_j}\right]\right)$$

$$\alpha_{i,j} = \text{softmax}_j\left(e(\mathbf{h}_i, \mathbf{h}_j)\right) = \frac{\exp(e(\mathbf{h}_i, \mathbf{h}_j))}{\sum_{j' \in \mathcal{N}_)} \exp(e(\mathbf{h}_i, \mathbf{h}_{j'}))}$$

where both $\mathbf{W} \in \mathbb{R}^{d \times d'}$ and $\mathbf{a} \in \mathbb{R}^{2d'}$ are learned linear weights, and $||$ represents vector concatenation. Then, we can define the updated node representation $\mathbf{h}'_i$ as the attention-weighted sum across all neighbors, followed by a non-linearity function σ.

$$\mathbf{h}'_i = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \cdot \mathbf{W}\mathbf{h_j}\right)$$

2.3.2 Graph Interpretation of Stem Cell Colonies

Many studies have historically noted the ability for graph theory to better represent complex biological processes like molecular structures, DNA, and even cellular structure.

Moreover, recent publications have demonstrated, graph networks' ability to learn more detailed and complex models of cell colonies as a consequence of improved representations of cell interactions. In particular, Wang et al. (2021) note how graphs are able to better incorporate spatial structure and cell interactions to better represent both individual cells, as well as the entire sample sample of cells.

Despite the growing body of work being published for other biological tasks, like those mentioned above, we denote the surprising lack of studies on learning representations of stem cell colonies in particular.

2.4 Cellpose Segmentation

Cellpose is a generalist deep learning-based cell segmentation model which was trained on a dataset of of over 70,000 segmented objects. Moreover, Cellpose is periodically further trained on new segmentation objects to further improve performance. Cellpose employs a U-Net to predict the outlines of cells in a mask from an input microscope image of cell samples.

Oftentimes, iPSC segmentation models are built and trained for singular, specialized datasets, as was the case in Schaub et al. (2020) work. As a natural consequence, these models tend to be trained on smaller datasets, whose segmentation masks are often hand corrected, if not entirely created manually. Moreover, the cell image samples might also suffer from low diversity in factors like cell shape, cell type, imaging conditions, and more. In contrast, Cellpose releases model weights which were trained on a combination of several datasets from multiple sources in order to further promote data diversity and achieve state of the art cell segmentation performance.

3 Methods

Throughout this section, GraphSC's end-to-end process of converting microscopy data into cell outline segmentations, which is then converted into visual features and other such properties, and finally reinterpreted into graphs.

3.1 Constructing Graphs from Microscopy Images

Given the iPSC-RPE microscopy images from Schaub et al. (2020), we first utilize Cellpose to predict the segmentation masks of individual cells in the colony. Figure 2 illustrates a sample segmentation output from Cellpose.

Next, we extract visual cell features using WIPP, or the Web Image Processing Pipeline. WIPP processes both the originally microscopy image as well as the Cellpose segmentation, using both modalities to gather visual features for each cell which was segmented by

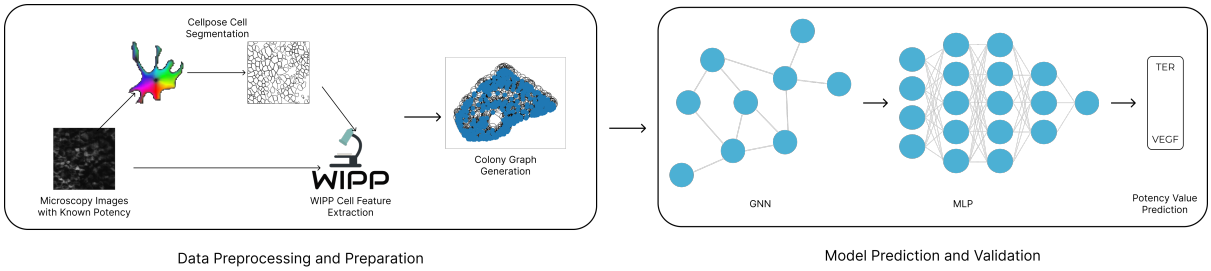

Figure 1: Methodology for stem cell colony analysis. **a)** An overview of the data preprocessing pipeline. Images are first segmented into cell outlines through Cellpose, and then have their features extracted through WIPP. The cell features for an entire image are then converted into a graph. **b)** Summary of the GNN architecture, which uses GATv2 layers to first encode the colony graph, and then uses a Multilayer Perceptron to predict out the associated potency values (TER, VEGF)

Cellpose. These visual features consist of observations like cell area, perimeter, height, and circularity, among others.

These WIPP visual features are then used to generate graphs to be used as input to our GNN network. To convert stem cell microscopy scans into learnable graphs, we first recognized that a colony of stem cells acts as a densely connected unit, with various colony factors being represented not just by any singular cell, but by observing connections between cells. In particular, we posit that the network's understanding of individual cells should be influenced by its physically-nearest neighbors, as cell interactions are most likely to happen with physically close cells. Thus, by using the distance between cell centroids as the criterion for edges between cell-nodes, and thresholding the number of neighbors any individual cell-node can have, we can construct graphs representing an entire microscopy sample of stem cells.

3.2 Graph Network Construction

Given some graph representation of a stem cell colony microscopy image, we aim to predict some number of continuous values which represent the TER and/or VEGF values for that sample.

We use a standard GAT architecture, with an encoder consisting of multiple GATv2 layers followed by a multilayer Dense prediction head to predict the regression values. In both the encoder and prediction heads, we utilize dropout layers, while the encoder receives additional batch normalization layers in order to minimize the impact of overfitting.

3.3 Fine Tuning

To further improve our model's performance, we performed hyperparameter and model tuning as a bayesian search using Optuna. Considering the large hyperparameter space with many different degrees of freedom, a standard grid search proved too computationally costly to be viable; therefore, a more efficient hyperparameter search, like the one implemented through Optuna, was highly valuable to the execution of these experiments.

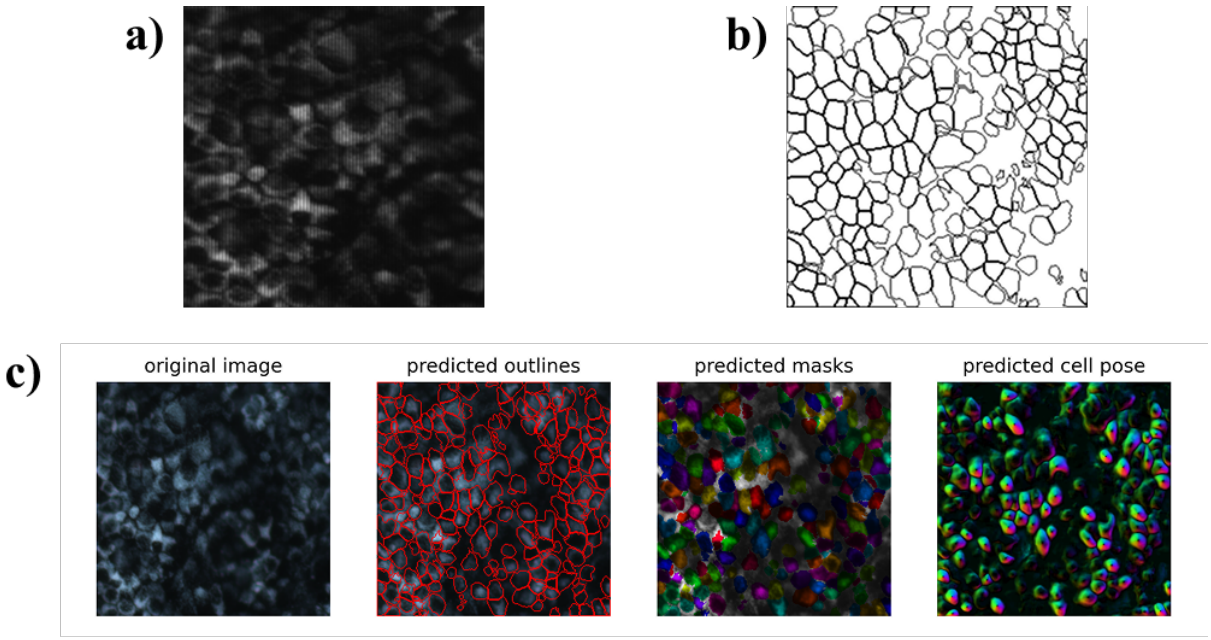

Figure 2: Visualization of Cellpose inputs and outputs. **a)** An input microscopy image is provided as a tiff file. **b)** The output segmentation for the sample image in a). **c)** A visualization of Cellpose predictions, including the input, predicted output cell segmentations, and masks / flows which illustrate the directional gradient-based mechanism used by Cellpose to predict cell regions.

3.4 Cell Colony Characterization

Our model consists of Graph Convolutional Layers applied sequentially, followed by a Dense Network which acts as a prediction head. During the standard training process, we train this GCN model on TER values and both the numerator and denominator for the VEGF ratio, a direct comparison to Schaub et al. (2020)'s DNN-F. As the standard task, We optimize models which predict TER and VEGF values separately, though this process is experimented with as discussed below. We predict that the best GCN models for each task will still outperform DNN-F, due to the ability for GCN models to better represent the biological phenomena underlying the cell colony.

3.5 Guillotine Regularization / Projection Heads

As explored in Chen et al. (2020) and Bordes et al. (2022), many DL models have been able to achieve and demonstrate self supervised learning and broader generalization capabilities through the use of *projection heads*. These projection heads are small neural networks which are trained on top of a *trunk* model during early training iterations, but then discarded or replaced with new networks later into the training regime. Both studies indicate that models which have stronger generalization capabilities tend to perform stronger when undergoing this Guillotine Regularization process (projection head removal).

4 Experiments

We perform experiments on the standard VEGF and TER regression task. In particular, we hope to directly compare the performance of our GNN in order to evaluate whether GraphSC can achieve SOTA on these tasks.

4.1 Implementation and Metrics

We train our GNN for 300 epochs, with an Adam optimizer on learning rate 0.004, with exponential learning rate decay and L2 normalization. We evaluate our root mean squared error (RMSE) metric as the square root of the sum of squared error divided by the total number of predicted values across the entire split.

4.2 Data

Throughout this project, we utilize the AMD patient stem cell microscopy images which were first released by Schaub et al.. In particular, this dataset provides 3 unique clones per each of 3 unique donors, with imaging and potency metrics (TER, VEGF) being provided at consistent time steps through the culturing process. As detailed in earlier sections, stem cell microscopy images are converted into individual cell features through WIPP, and then subsequently converted into graphs.

In contrast with Schaub et al.'s work, in which a model is both trained and evaluated on data entirely from a single clone and donor, we hold out a validation and test split as being a specific clone of a specific donor (*e.g.* clone C of donor 3), and use all other remaining data in order to evaluate each model's performance on a broader scale. We posit that train and evaluation data being sourced from the same clone-donor subset might significantly reduce the variance in the distribution underlying the input images, and that our broader formulation of the task better tests a model's ability to generalize.

4.3 VEGF Target Re-expression

In their formulation of the VEGF task, Schaub et al. (2020) attempt to predict the ratio of the presence of VEGF on the apical versus basal side of the stem cell colony.

However, we posit that the conversion of the raw, perceptible data of VEGF presence into the aforementioned ratio creates an additional layer of abstraction that might cause worsened performance. To explore the validity of this claim, we additionally explore if our GNN performs better when given supervisory feedback of the directly tangible VEGF presence values, rather than the ratio of the values. RMSE for this task is reported in the same manner as other experiments.

4.4 Multi-target Prediction

We further explore the necessity, or potential lack thereof, in training entirely unique models to predict values for individual targets (TER, VEGF). We reformulate the regression task as predicting 2 or 3 different values (depending on the presentation of the VEGF

targets) per input sample, and evaluate model performance through RMSE. We predict that a model which can perform well on the joint task will be able to create a higher quality representation of the cell colony present in the input microscopy image, as the different values require specific features relevant to characterizing the colony.

4.5 Guillotine Regularization

Considering the limited size of our dataset, we are unable to properly execute Guillotine Regularization by definition. However, we still were interested in exploring the transfer learning and, more broadly speaking, generalization capabilities of out model. To explore the generalizability of our GNN's learned graph representations of the stem cell colony, we additionally perform projection head replacement during training to change the potency value target in a manner similar to Guillotine Regularization. In particular, we pretrain the GNN for 200 epochs on one prediction target (VEGF or TER). Then, we freeze all trainable parameters in the model except for the last linear layer, which is replaced by a newly instantiated layer with the correct number of output units for the relevant prediction task. Afterwards, the newly added linear probe is trained for 100 epochs on the new target task, and RMSE is reported in the same manner as other experiments.

5 Results

5.1 GraphSC Learns Better, Faster

	GraphSC	DNN-F	MLP
TER	**21.7**	70.6	84.7
VEGF	**84.9**	121.3	

Table 1: Model performance on TER and VEGF prediction tasks, quantified in Root Mean Squared Error (RMSE)

	GraphSC	DNN-F
TER	**21.7**	70.6
VEGF	**84.9**	121.3

Table 2: Model performance on TER and VEGF prediction tasks, quantified in Root Mean Squared Error (RMSE)

Table 2 summarizes the results of our baseline TER and VEGF prediction tasks, where we compare GraphSC against the state of the art CNN model DNN-F. The results show that our model achieved significantly lower RMSE values. Moreover, figure 3 demonstrates the remarkably faster training time of GraphSC as compared to DNN-F; where training a single instance of DNN-F took about as long as training 60 different GNNs during our hyperparameter tuning processes, even despite running on the exact same hardware.

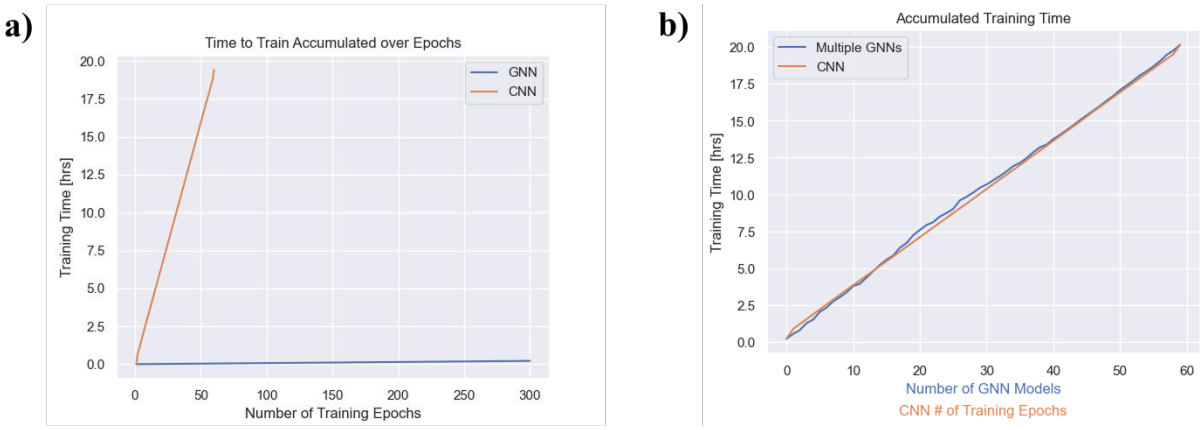

Figure 3: The time to train the CNN (DNN-F) and GNN (DNN-F) on stem cell microscopy data. **a)** Plotted training time across epochs for a singular model. Our GNN trained significantly faster than the CNN, taking only about 15 minutes for 300 epochs, as compared to almost 20 hours for only 60 epochs. **b)** Visualized training time across epochs for a single CNN instance versus training time for multiple GNN instances.The time to train DNN-F for a single epoch was roughly as costly as training a single GNN model.

5.2 GraphSC is Capable of Predicting Multiple Targets

	GraphSC	DNN-F
TER	21.7	70.6
VEGF, Ratio	84.9	121.3
VEGF, 2 Values	57.3	96.7
Both	110.5	–

Table 3: GraphSC GNN's predictive performance on VEGF, measured in RMSE.

We report the results of our VEGF Target Re-expression experiments in Table 3. We demonstrate how our graph network achieves lower RMSE values when predicting the directly tangible VEGF values rather than the ratio. Additionally, though the model was trained for the full 300 epochs to ensure convergence, we also note that the GNN reached these performance metrics in just over 100 epochs as well.

We also include performance on 3 target-prediction: TER, VEGF apical, and VEGF basal. Although the GNN trained on all 3 targets underperforms in predictive capability as compared to the specialized models, GraphSC still demonstrates performance similar to specialized DNN-F models.

Overall, these results demonstrate GraphSC's ability to learn latent representations relevant to multiple, minimally related target values within a unified model.

5.3 GraphSC Learns Higher Quality Latent Representations

Table 4 indicates that our approach was successful in the transfer learning task as well, achieving competitive performance when pretrained on either target and fine tuned for a

	GraphSC (pretrain)	GraphSC (standard)	DNN-F (standard)
TER	65.6	**21.7**	70.6
VEGF	**41.2**	57.3	96.7

Table 4: GraphSC GNN's pretrain predictive performance v.s. standard GraphSC GNN training, DNN-F standard training. GraphSC (pretrained) was first pretrained on the other target type for 200 epochs before finetuning for 100 epochs. GraphSC (standard) and DNN-F were trained for 300 epochs as in earlier experiments.

limited number of epochs on the other. Considering that all weights are frozen except for a newly instantiated linear probe, we claim that this demonstrates the model's propensity to learn a generalizable representation of the stem cell colony.

Due to time and resource constraints, more work and novel tasks could not be developed for this approach. However, this approach is simple to apply to novel tasks which we hope to explore in future works.

5.4 Graph Network Training Instability

Our GAT model trained remarkably unstably during hyperparameter fine tuning. This was most pronounced when observing two similar sets of significantly similar hyperparameters that yielded incredibly differing qualities of output predictions. Moreover, this occurred despite training to an excessive number of epochs (500 iterations) and happened regardless of the dataset or target type.

We believe that this training instability might be a consequence of the GNN architecture as well as the complex nature of the biological task. To this end, graph neural networks have been documented to suffer from training and prediction instability. In particular, Klabunde and Lemmerich (2022) detail this instability, and how it manifests in multiple instantiations of identical models, across both standard Graph Convolutional Networks as well as Graph Attention Networks. Moreover, the authors similarly report that this issue persisted even despite large numbers of training iterations.

In an attempt to combat this, we attempted several regularization strategies, including L2-Loss, learning rate warmup and decay, as well as batch normalization and dropout. However, despite including all of these techniques, while additionally searching over their hyperparameters, we still found unstable prediction quality between markedly similar looking models.

To further evaluate this claim, we hope to continue validating GraphSC's predictions by exploring the diversity in range of values predicted by the model. Likewise, we hope to further explore the interpretability of the GAT by visualizing and potentially validating important cell features for TER and VEGF values, as well as exploring cell interactions which might further influence these target values. However, we must also recognize that, withstanding pre-exsting literature which already explores the influence of such factors on our focused target values, it is difficult to validate these factors as correctly being attended to.

Moreover, we must recognize that there may also be instability and stochasticity underlying our data as well. This noise may come in several forms such as microscopy image quality or naturally-occurring biological randomness which can't fully be accounted for through just image modalities.

6 Conclusion

In this paper, we proposed GraphSC, an automated microscopy-to-regression prediction pipeline which achieves state of the art performance by taking advantage of Graph Networks to learn deeper representations of the input colonies. Our approach moreover demonstrates the ability for our model to learn these representations with low computational cost. Additionally, we validate our model's learned representations by performing transfer learning tasks constrained to a small number of train iterations.

However, we also hope to continue exploring the interpretability and quality of learning of our graph networks in an attempt to both validate training instability as well as explore the importance of our learned feature importance values. More extensive microscopy image sets along with associated potency experimental values might help stabilize training instability, while we may also find further improved performance with manually annotated cell segmentation masks, as this would enable us to fine tune a local instance of Cellpose for higher quality segmentation masks. Finally, our model may prove capable in more diverse stem cell colony characterization tasks; however, we could not evaluate this claim due to a lack of available data.

References

Bordes, F., Balestriero, R., Garrido, Q., Bardes, A., and Vincent, P. (2022). Guillotine regularization: Why removing layers is needed to improve generalization in self-supervised learning. *arXiv preprint arXiv:2206.13378*.

Brody, S., Alon, U., and Yahav, E. (2021). How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR.

Fan, W., Crawford, R., and Xiao, Y. (2011). The ratio of vegf/pedf expression in bone marrow mesenchymal stem cells regulates neovascularization. *Differentiation*, 81(3):181–191. Mini-Series on Stem Cells.

Hori, T., Okae, H., Shibata, S., Kobayashi, N., Kobayashi, E. H., Oike, A., Sekiya, A., Arima, T., and Kaji, H. (2024). Trophoblast stem cell-based organoid models of the human placental barrier. *Nat. Commun.*, 15(1):962.

Klabunde, M. and Lemmerich, F. (2022). On the prediction instability of graph neural networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 187–202. Springer.

Lee, H., Lee, S.-I., Ko, Y., and Park, J.-B. (2018). Evaluation of the secretion and release of vascular endothelial growth factor from two-dimensional culture and three-dimensional cell spheroids formed with stem cells and osteoprecursor cells. *Adv. Clin. Exp. Med.*, 27(7):971–977.

Li, R., Yuan, X., Radfar, M., Marendy, P., Ni, W., O'Brien, T. J., and Casillas-Espinosa, P. M. (2021). Graph signal processing, graph neural network and graph learning on biological data: a systematic review. *IEEE Reviews in Biomedical Engineering*, 16:109–135.

Nourse, M. B., Halpin, D. E., Scatena, M., Mortisen, D. J., Tulloch, N. L., Hauch, K. D., Torok-Storb, B., Ratner, B. D., Pabon, L., and Murry, C. E. (2010). Vegf induces differentiation of functional endothelium from human embryonic stem cells. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 30(1):80–89.

Robinton, D. A. and Daley, G. Q. (2012). The promise of induced pluripotent stem cells in research and therapy. *Nature*.

Schaub, N. J., Hotaling, N. A., Manescu, P., Padi, S., Wan, Q., Sharma, R., George, A., Chalfoun, J., Simon, M., Ouladi, M., Simon, Jr, C. G., Bajcsy, P., and Bharti, K. (2020). Deep learning predicts function of live retinal pigment epithelium from quantitative microscopy. *J. Clin. Invest.*, 130(2):1010–1023.

Srinivasan, B., Kolli, A. R., Esch, M. B., Abaci, H. E., Shuler, M. L., and Hickman, J. J.

(2015). TEER measurement techniques for in vitro barrier model systems. *J. Lab. Autom.*, 20(2):107–126.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wang, Y., Wang, Y. G., Hu, C., Li, M., Fan, Y., Otter, N., Sam, I., Gou, H., Hu, Y., Kwok, T., Zalcberg, J., Boussioutas, A., Daly, R. J., Montúfar, G., Liò, P., Xu, D., Webb, G. I., and Song, J. (2021). Cell graph neural networks enable the digital staging of tumor microenvironment and precise prediction of patient survival in gastric cancer.