Unsupervised representation learning to interrogate cellular behaviours in 47-parameter single-cell resolution tissue imaging data

Salil Bhate, Julia Kennedy-Darling, Garry Nolan

Abstract

We consider here the problem of learning representations for high-parameter tissue imaging data, in order to visualize and understand variation in cellular behaviours and phenotypes. Our dataset contains 5 human tissue sections imaged with 47 antibody parameters (each a different color of the image), in which 2 million single-cells are visible in high-resolution, in their tissue context. We use convolutional neural networks to simultaneously learn image features and perform variational inference in a latent variable model for cellular imaging phenotypes, spatial contexts and the relationship between them. Using our image features, we find subclusters of cells by imaging phenotypes, despite having the same surface marker expression profiles. We visualize the relationship between spatial context and cell type. We utilize the latent representation and generative capacity of our model to describe axes of variation in imaging phenotypes and spatial contexts for specific cellular subsets. These axes recapitulate known biological phenomena. Our results highlight the role of representation learning approaches in overcoming the inherent complexity of high-parameter imaging data for deriving novel biological insights.

1 Introduction

Recent technological advances have enabled acquisition of tissue microscopy data, in which millions of individual single-cells can be resolved and many different mRNA and protein parameters as 'channels' or 'colors' can be simultaneously imaged. We consider here a dataset of 5 human immune tissues, imaged with 47 antibody parameters (each a different color of the image) obtained with the CODEX imaging platform [1]. This dataset contains 2 million cells, in tissue slices, with each cell occupying approximately a 30x30 pixel area in the image (Figures 1a and b). The antibody panel used targets cell-surface markers, intracellular markers as well as extracellular proteins.

Utilizing traditional cell-segmentation approaches, it is possible to obtain single-cell expression data (a cell's *expression* phenotype) and define its cell-type. However, there is strictly more information in the image of a cell than this. The rich patterns of protein expression across 47 parameters visible in an image (Figure 1a) (which we refer to as a cell's *imaging phenotype*) are not captured in its expression phenotype. Moreover, in high-parameter imaging data, we can see not only individual cells, but rather interacting cells, in their local microenvironment (a cell's *spatial context*). Methods are therefore warranted for interrogating cellular imaging phenotypes and spatial contexts in high-parameter tissue microscopy data, in order to summarize and understand biological phenomena governing the behaviour of cells in tissue.

Any such method for interrogating spatial contexts and cellular imaging phenotypes must solve the following problems:

Accepted as a poster at NeurIPS Computational Biology Workshop 2017

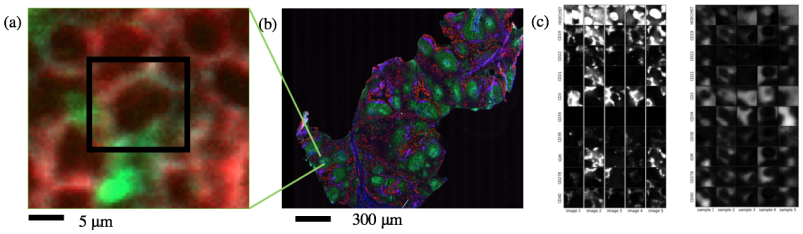

Figure 1: (a): Representative cellular image. 2 of 47 parameters shown. Area inside black box corresponds to an imaging phenotype, area outside is a spatial context (b) representative entire tissue image (contains 400,000 single cells), (c) Left: actual imaging phenotypes sampled from dataset. Right: generated samples of imaging phenotypes by sampling from trained latent variable model (20-d prior), 10 of 47 parameters shown

- Tissues (or even single cells) have not previously been imaged in 47 antibody parameters simultaneously. Thus, there are no prior image features which are biologically informative. Moreover, the patterns of antibody staining depend strongly on the tissue type and the specific protein parameter being imaged. Therefore, features, across combinations of parameters, must be *learned*
- Tissue images are a snapshot: cells with a given expression phenotype can be found with different imaging phenotypes and in different spatial contexts. Therefore, approaches must be able to represent uncertainty over the learned image features.
- Learned features must generalize to new tissue samples imaged with the same antibody parameters, and moreover must be useful to interrogate new tissues.

We sought to develop representation learning approaches to address the following biological questions.

- What are the key axes of variation a cell's imaging phenotype given its expression phenotype?
- How do we describe and interpret high-parameter spatial contexts in a given tissue?
- How does spatial context of a cell affect its imaging phenotype?

We therefore utilize convolutional neural networks to simultaneously learn image features and perform variational inference in a latent variable model for imaging phenotypes and spatial contexts. We explicitly model imaging phenotypes conditioned spatial contexts. We apply our model to a new tissue sample to extract image features. These features find, within cells that have the same expression phenotype, subclusters of cells by imaging phenotype. We interpret these clusters by sampling from our model, and contrast this to extracting actual images from the tissue. Using learned features for spatial contexts, we visualize how the type of a cell affects the spatial contexts it's found in.

The insights we derive from our model recapitulate studied biological phenomena. Our approach therefore highlights the feasibility of neural network based unsupervised representation learning approaches for high-parameter tissue imaging data and the utility of these approaches for biological discovery and hypothesis generation.

2 Approach

Our images are 2D, with 47 colors. Cell centers are identified using intensity peaks in the channel staining the nucleus. For each cell c, we consider the $90 \times 90 \times 47$ pixel image centered at the cell center and divide it into a 9x9 grid of $30 \times 30 \times 47$ images. We define a cell's spatial context, x_c^o, to be the (ordered) 8 outer images of the grid (excluding the center). The cell's imaging phenotype x_c^i is defined to be the center image (Fig 1a). Cells are segmented to compute single-cell expression phenotypes using standard watershed techniques.

Cell-*types* are identified by automated clustering and manual annotation of identified clusters with single-cell expression [2]. This information is used only for interpretation and not for training.

We follow the β-VAE approach [3], assuming the latent variable model:

$$p(x_c^i) = p(x_c^i|z_c^i)p(z_c^i) \text{ and } p(x_c^o) = p(x_c^o|z_c^o)p(z_c^o)$$

We use a 20 dimensional isotropic Gaussian prior, and assume $p(x_c^i|z_c^i)$ is Gaussian with means parametrized by a generative deep neural network. We make variational approximations to $p(z_c^i|x_c^i)$, $p(z_c^o|x_c^o)$ as Gaussians with diagonal covariance; these mean and covariance parameters are estimated by deep neural networks that share common convolutional layers. This is in order to learn image features with a multi-task training signal. We also make a similar approximation for $p(z_c^i|x_c^o)$ to model the relationship between spatial context and imaging phenotype. We use β = 3, which corresponds to increasing the regularization of the variational posterior towards the prior; this regularization has been shown to increase disentanglement of the latent space [3]

To condition our model to generate spatial context or imaging phenotype samples from a specific subset of cells (sections 4.1 and 4.2) , we sample from the distribution obtained by taking the average of the posteriors predicted by the neural network over this subset.

We train our model on 2 human tonsil tissue images, containing a total of 800,000 cells. We perform all analyses described below on a third tonsil tissue that is biologically independent from those used for training.

3 Previous work

Neural networks have been extensively applied to microscopy data for in numerous supervised contexts: classification of single-cell phenotypes by experimental condition [5] segmentation [4], histopathology classiication object segmentation metastases [6]. A model capable of generating samples of single-cell biological images has not been demonstrated, to our knowledge; in particular not in high-parameter data. Techniques for learning disentangled representations have been extensively studied in the context of natural image analysis [3,7]. Our method is based on the β-VAE described in [3]. This is the first application of deep neural networks to high-parameter single-cell resolution tissue images. Moreover, our focus is on learning unsupervised representations that assist in interrogation of the underlying biological phenomena, in contrast to supervised approaches.

4 Results

We successfully train and sample from our model (Figure 1c, right). The generated samples are blurry, since our model capacity is severely restricted, both by the small latent dimension and enhanced regularization towards the prior. However, these images are biologically plausible; qualitatively and in terms of the combinations of markers that cooccur (Figure 1c, left). We now describe several applications enabled by our network and learned features.

4.1 Image features for imaging phenotypes and spatial contexts

We first extracted our multi-task image features from the imaging phenotypes of all cells, in a tonsil tissue not used during training. These image features clearly separate cell-types (which were identified with segmentation and manual annotation) (Fig 2a). Despite the restricted expressive capacity of our model, training a linear logistic regression classifier on these image features to classify cell-types by imaging phenotype achieves a significantly higher F-1 score than a model trained on on average pixel intensity for the same images, even though the average pixel intensity is typically linearly related to the cell-type annotation (0.64 vs 0.54).

We identified sub-clusters of CD4 T cells by imaging phenotype, by clustering image features for each cell (k-means with 30 clusters). The images obtained by conditioning our model to generate images from these subclusters (Fig 2c) generally coincide with actual randomly extracted cellular imaging phenotypes (Fig 2b), demonstrating that we can use the generating network to reason about image-feature clusters. The clusters clearly correspond to to CD4 T cells actively interacting with B-cells.

By extracting features for spatial contexts from the test tissue sample, we assessed the relationship between a cell's type, and its imaging phenotype. Certain cell subsets (proliferating B cells and T follicular helper cells) are found only in specific contexts, and moreover they coincide. In contrast, other subsets, like CD40+ B cells, are found in numerous different contexts (Fig 3a). Thus, these spatial context image features provide a way to visualize structure of the tissue in terms of the spatial-contexts, as well as how the type of a cell affects which contexts it can be found in.

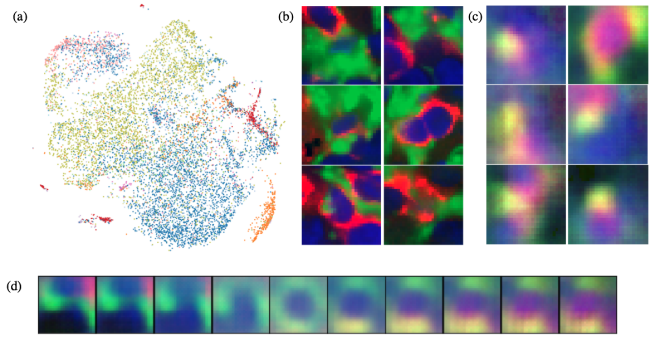

Figure 2: (a) t-stochastic neighbor embedding of imaging phenotype features, color corresponds to cell-type. (b) Actual imaging phenotypes extracted randomly from a cluster of extracted imaging phenotype features. Red: CD3 (T cells), Green: CD19 (B cells), Blue: Hoechst (Nucleus). (c) Imaging phenotypes sampled by our model conditioning on the same cluster of features as (b); colors the same. (d) Samples obtained by varying a proliferating B cell along its most variable latent axis; colors the same.

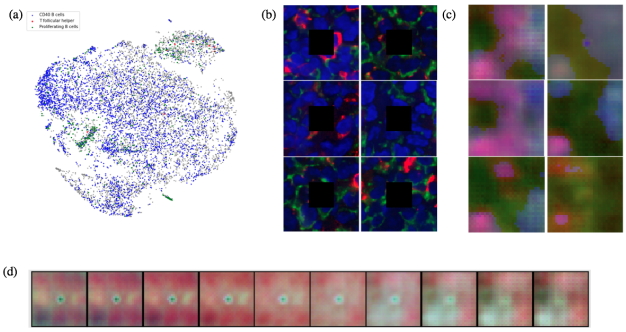

Figure 3: (a) t-stochastic neighbor embedding of spatial context features (b) Actual spatial contexts extracted randomly from a cluster of extracted spatial context features. Red: CD3 (T cells), Green: CD19 (B cells), Blue: Hoechst (Nucleus). (c) Spatial contexts sampled by our model conditioning on the same cluster of features as (b); colors the same. (d) Samples obtained by varying the average T follicular helper cells along the most variable latent axis for that cell type; colors the same.

We clustered spatial contexts by their image features, and extracted images as with imaging phenotype features (Fig 3b,c). The highlighted cluster clearly captures an interface between regions of T cells and B cells. Our features are informative on a tissue image sample not used during training, demonstrating their generalisability.

4.2 Latent axes of variation for proliferating B and TFH cells

We next used the latent space and the generative capacity of our model to investigate the primary axes of variation within a subset of B cells, proliferating B cells. For every cell, we computed the posterior mean estimated by our inference network. Comparing the posterior means restricted to this cellular subset, to those across all cells, we found one axis in the latent space that had significantly higher variation in proliferating B cells than overall. To interpret this variation, we generated images, varying only this axis [Fig 2d]. These images show a range from dividing B cell (with no T cell markers present), to a single B cell (with T cell markers present). This recapitulates the dynamics of B cell maturation in the germinal center response: alternating from interacting with T cells to rapidly proliferating.

We next identified axes of variation for the spatial contexts of T follicular helper (TFH) cells. We found one axis in the latent space that had significantly higher variation in TFH cells relative to all T cell subsets. By generating images varying only this axis, we see that it corresponds to a transition from the edge of a B cell follicle - surrounded by CD22+ B cells, to the boundary of B-cells and T-cells. This reflects the movements of TFH cells in the tissue.

Conclusion

We demonstrated that a combination of generative modelling and feature extraction can be used for extracting biological insights from high-parameter single-cell imaging data about imaging phenotypes and spatial contexts. Our model generalized well to tissue sample not seen during training. We anticipate that development of further explicit generative image modelling [9] and semi-supervised approaches [8] that leverage the representation learning capacity of neural networks will provide increasingly complex biological insights into cellular behaviour and organization in tissue.

References

[1] Y. Goltsev et al. (2017) *Deep profiling of mouse splenic architecture with CODEX multiplexed imaging*, Cell Sneak Peek

[2] N. Samusik et al. (2016) *Automated mapping of phenotype space with single-cell data*, Nature Methods

[3] I. Higgins et al. (2016) *beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework*, ICLR

[4] S. Sadanandan et al. (2017) *Automated Training of Deep Convolutional Neural Networks for Cell Segmentation*, Scientific Reports

[5] O. Kraus. et al. (2017) *Automated analysis of high-content microscopy data with deep learning*, Molecular Systems Biology

[6] Y. Liu et al. (2017) *Detecting Cancer Metastases on Gigapixel Pathology Images*

[7] X. Chen et al. (2016) *InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets*, Advances in Neural Information Processing Systems 2016

[8] N. Siddharth et al. (2017) *Learning Disentangled Representations with Semi-Supervised Deep Generative Models*, Advances in Neural Information Processing Systems

[9] M. Johnson et al. (2016) *Composing graphical models with neural networks for structured representations and fast inference*, Advances in Neural Information Processing Systems