ME-VAE: **M**ulti-**E**ncoder **V**ariational **A**uto**E**ncoder for Controlling Multiple Transformational Features in Single Cell Image Analysis

Luke Ternes,[1] Mark Dane,[1] Marilyne Labrie,[2]

Gordon Mills,[2] Joe Gray,[1] Laura Heiser,[1] Young Hwan Chang[1]

1 Biomedical Engineering Department, Oregon Health & Science University

2 Cell, Developmental and Cancer Biology Department, Oregon Health & Science University

---

Abstract

Image-based cell phenotyping relies on quantitative measurements as encoded representations of cells; however, defining suitable representations that capture complex imaging features is challenging since there are many obstacles, including segmentation and identifying subcellular compartments for feature extraction. Variational autoencoder (VAE) approaches produce encouraging results by mapping from an image to a representative descriptor, and outperform classical hand-crafted features for morphology, intensity, and texture at differentiating data. Although VAEs show promising results for capturing morphological and organizational features in tissue, single cell image analyses based on VAEs often fail to identify biologically informative features due to the intrinsic amount of uninformative variability. Herein, we propose a multi-encoder VAE (ME-VAE) in single cell image analysis using transformed images as a self-supervised signal to extract transform-invariant biologically meaningful features. We show that the proposed architecture improves analysis by making distinct populations more separable compared to traditional VAEs and intensity measurements by enhancing phenotypic differences between cells and by improving correlations to other modalities.

Introduction

Understanding cellular changes and phenotypic pathways at the single cell level is becoming increasingly important and allows for a more comprehensive understanding of the cell state and single cell heterogeneity. Single cell RNA sequencing is a well-established methodology with many documented analytical tools to extract, quantify, normalize, and evaluate single cell differences at the transcriptional level.[1,2,3] Until recently, analyzing single cell imaging data in a similar fashion was limited to extracting mean intensity profiles, predefined shape, textural and morphological features, and images stained with only a few markers. Emerging multiplexed imaging technologies such as cyclic immunofluorence (CYCIF),[4] multiplexed immunohistochemistry (mIHC),[5] CO-Detection by indEXing (CODEX),[6] and Multiplexed ion beam imaging (MIBI)[7] can create images stained with a large number of markers, expanding the depth of features. Robust analysis methods for high dimensional multiplexed imaging data, however, are not as well developed as single cell RNAseq (scRNAseq) analysis tools. One major limitation with analyzing highly multiplexed single cell images is the ability to extract meaningful information on staining localization patterns that indicate divergent cell states, tumor progression and response to therapy. This problem is further confounded when relevant or novel features are obscured by irrelevant features that are dominant in a dataset. Imaging data has morpho-spatial information that is not captured using a simple mean intensity information of each marker, but successful quantification of these features can lead to improved analysis and understanding.[8]

The classical approach for image feature extraction is to manually create a list of desired features and design a metric or algorithm that quantifies those features within the image. This traditional method is biased to known and easily measured features and can miss subtle but important features. More robust image feature extraction has been addressed using deep learning architectures such as the Variational Autoencoder (VAE)[9] in other domains where feature representation can be automatically generated without any supervising information. However, the problem with VAE feature extraction in single cell imaging is that there are typically unimportant or uninformative features that drive differences between cells and skew the results in undesired ways.[10] In single cell imaging data, these unimportant features include any form of basic image transformation such as rotation,

transposition, affine/skew, and stretching. When these features are known and controllable transformations, they can be used for a self-supervised signal to remove the uninformative features and extract invariant features with respect to a set of transformations during model training. Despite having the same base information, the common uninformative features in transformed images distract deep learning architectures so that they ignore most of the relevant features.[11,12,13,14,15] This holds true in biological single cell imaging, where VAEs will ignore biologically meaningful features to focus on recreating the transformation features that have a high variance across the dataset. Whether in RNAseq or imaging data, untailored deep learning architectures are unable to overcome these uninformative features that skew analysis unless some modification is made to either their architecture or objective functions.[11,12,13,14,15] Use cases of autoencoders for scRNAseq data have documented the difficulty of correcting for these features and extracting everything that is relevant into a single latent space. Many recent works propose changing autoencoder architectures to coupled networks or using multiple latent dimensions to overcome this without the need for biased hyperparameter tuning and data normalization which are common practices in most scRNA analysis methods.[10,16,17,18] Similar methodologies have also been highly explored in the imaging domain as well, seeking to correct transformative features with coupled networks, direct latent space modifications, novel layer architectures, and training networks with combinations of corrected and uncorrected image data.[11,19,13,20,12,14,15] However, most of these corrected architectures are only targeted to one specified feature and are not able to generalize to other uninformative features without further modification.

Here we describe a novel method for single cell image feature extraction that removes specified uninformative features by making the features uniform and invariable across the reconstructions, using modified pairs of transformed input and output images by self-supervised transformation, utilizing multiple encoding blocks. Using this multi-encoder VAE to control for multiple transformational features, we highlight the ability to extract more biologically meaningful and transform-invariant single cell information with the proposed approach and better separate biologically distinct cell populations within a mixed dataset.

Results

A. Controlling for uninformative features.

When a single transformational feature varies across a single cell imaging dataset, Standard VAEs extract only the single dominant component to reconstruction. In a dataset where rotation varies from image to image, reconstructions along the principal component walk[21] only constitute the angle of the cell and downstream analysis will be heavily skewed by this extracted component (Fig 1a). In another dataset where polar orientation is the dominant feature, we observe the same behavior (Fig 1b); VAEs only extract the dominant uninformative features and ignore many other subtle features necessary for detailed reconstruction.

In order to overcome model hypersensitivity to dominant uninformative features, several architectures were proposed and tested using transformed image pairs to make deep learning models learn the latent space by ignoring uninformative features (Fig 1c,d,e). A Standard VAE with no control for any uninformative features was used as baseline and shows a high correlation between the embedding components and the respective feature metrics (Figure 1c and Supplemental Figure 1a). When a single factor is controlled (i.e. rotation), it becomes uncorrelated to the all VAE encodings, and even the max correlated component in the latent space is insignificant (Figure 1d and Supplemental Figure 1b). Controlling for one feature, however, does not significantly impact the other dominate transformation feature (i.e. polar orientation). With a double transformed output, attempting to correct for two features simultaneously, we saw decorrelation of both dominant features (Supplemental Figure 1c), but the reconstructed images of this architecture were poor (data not shown) reflecting a failure of the model to learn any relevant feature embedding. The VAE with transformed output was shown to work on simple transforms that the model can learn such as rotation, but more complex features or pairs of transformations like rotation combined with polar orientation proved too difficult for the model. Finally, when both uninformative features were controlled for using the proposed ME-VAE with transformed image pairs, we saw a decorrelation in both uninformative features that were controlled for, indicating that the VAE reconstructions learned to overcome the multiple dominant features and focus on underlying features that better separate cell populations (Figure 1e and Supplemental Figure 1d). Unlike with the corrected output VAE, the ME-VAE produced coherent

reconstructions (data not shown). Moreover, the proposed approach is both generalizable and scalable since it enables the control of many uninformative features together in a parallel combination by using a multi-encoder network where any number of encoders can be added in parallel and each encoder learns a single transformation.

B. Improving Biological Interpretation on Single Channel Images

To evaluate the models' ability to extract biologically relevant features with improved downstream usefulness, we analyzed a dataset (Methods A) of single cell CYCIF images taken of MCF10A non-malignant breast epithelium cell line, and we restricted our analysis to populations that had been treated with PBS (control) and TGFβ ligand to induce a change in cell state. We also limited the size of the images to the only utilize the channel stained for Epidermal Growth Factor Receptor (EGFR). The EGFR marker was chosen for analysis because it displayed all uninformative features such as rotation, polarity, and cell size that we were correcting for. Furthermore, the two ligand populations were chosen because they have similar distributions of mean EGFR intensity and cell size, making them difficult to naively separate (Fig. 2c), but qualitatively show phenotypic differences such as compartment localization and stain texture. The proposed ME-VAE was tested to see if it could better separate and extract biologically distinct features between the PBS and TGFβ cell populations compared to a Standard VAE within this labeled dataset. The separation was both quantitatively measured by k-means cluster purity and qualitatively via separation in UMAP embedding space. Within this dataset we show that the ME-VAE is able to better separate distinct cell populations between the PBS and TGFβ treated cell.

As can be observed in Figure 2a, the Standard VAE is incapable of separating the two cell populations, creating a mix of the labeled cell populations, both in k-means cluster space, as well as, in UMAP embedding space. The cells within UMAP regions also have an arbitrary range of phenotypes, and the only patterns that can be observed are the patterns in uninformative features such as rotation. By comparison, the ME-VAE shows a dramatic increase in k-means cluster purity, and a clear separation of labeled cell populations in UMAP (Fig 2b). All regional cell images within the multi-encoder's UMAP space show distinct phenotypic differences that separate the cell populations with biologically relevant features (stain localization, intensity, and subcellular pattern). In PBS dominated regions, we see very dim staining throughout the cell with the stain being most heavily

concentrated uniformly along the cellular membrane. TGFβ by comparison shows a cloudy diffuse concentration of stain throughout the cell with the heaviest concentration of stain localizing to one side of the nuclear membrane. These differences illustrate a clear difference in cellular regulation and compartmentalization of the EGFR protein induced by the presence of the TGFβ ligand. Several sub-clusters also form in UMAP space that subdivide the labeled populations, indicating features are being extracted that are able to identify different subpopulations within a treatment population, which can be used for more robust analysis. Based on observations from the multi-encoder output in Fig 2b, we inferred that the metric of radial slope would better capture the biological differences between the two populations (see Methods C, Supplemental Figure 3). Using this metric, we see improved separation and cluster purity compared to the naïve metrics (Fig 2c and 2d). The cluster purities from the inferred radial slope metric are still lower than the full ME-VAE cluster purity, indicating there is some other feature captured in the multi-encoder beyond the radial slope as well.

C. Use case with a Large Complex Dataset

The models were next trained on a larger portion of the CYCIF imaging dataset taken on MCF10A breast epithelium cell line. The expanded set was comprised of six ligands conditions treated for 48 hours and stained with choosen markers (see Methods A and Table 1). For this iteration, the ME-VAE was trained to control for shape and size in addition to the rotation and polar orientation performed previously as well as learned 23-channel images. This was done because the six cell populations had significantly different sizes and need to controlled to improve the model's ability to capture subcellular stain information.

In this larger dataset, the Standard VAE performs similarly to the smaller dataset, encoding cells based primarily on dominant features such as size and rotation while largely ignoring complex staining information (Fig 3a,b – left panel). Although visually there is some preferential localization (such as OSM on the left side and TGFβ on the right), it is clear that the ligand populations are thoroughly mixed with poor separability. When looking at the intensity profiles, we can see that size has a strong impact on this left/right embedding space (Fig 3b left). Most all other stains show little or no consistency within the embedding space, with the exception of DAPI and ki67. These stains, however, show the same left/right distribution as size. It is likely that this

distribution of staining intensities is simply a result of cell size since the whole cell mean intensity of a nuclear marker will decrease with larger cells and increase with smaller cells.

Despite the increased complexity of the multi-channel CYCIF images and diversity of the dataset, the ME-VAE still shows good separation of the labeled cell populations (Fig 3a - right). Subpopulations within a given label are even more distinct in this larger dataset, with multiple clusters for HGF, BMP2, and TGFβ. By analyzing intensity profiles and regional cell images of these subpopulations, we can see differences in expression (Fig 3b right and Supplemental Figure 2b - right). The UMAP intensity profiles also show clear patterns in stain intensity indicating that the VAE encoding space contains relevant biological information. The distribution pattern of size does show some presence in the UMAP space, but the effect is largely dulled in comparison to the Standard VAE.

There are many differences in stain expression that can be observed between and within the ligand treated cell populations. Here we discuss some of the most noticeable drivers of separation. In comparison with all other treatment groups, PBS shows a marked deacrease in ki67 expression. This indicates a relative decrease in proliferation compared to other markers. The TGFβ populations show a drastic increase in S6 expression relative to other ligand populations, indicating an increase in cell growth. This can be observed visually by looking at the regional cell images (Supplemental Figure 2b right); however, it is worth noting that this high expression of S6 is seen in both large and small cells treated with TGFβ. In both EGF and BMP2 treated populations, decreased expression of membrane adhesion proteins such Ecadherin and Cateninβ. This decreased expression presents visually as dim stain, but the marker is still localized to the membrane and is not missing or diffuse through the cell.

More interestingly, there are also differences in stain distribution that can be observed through the model. In both TGFβ and PBS, we see increased concentration of HES1 expression localized primarily to the nucleous, while in other populations the distribution is even throughout the cell. In the case of TGFβ this localization is accompanied by a large increase in expression intensity, but in the case of PBS, it is not significantly more intense than some of the other ligand populations. In a similar manner, Stat1a is observed to be primarily located in the

nucleous for TGFβ, BMP2, and OSM populations, but shows decentralized staining in cell images for other ligand populations. This is important because both HES1 and Stat1a are functional in the nucleus (Stat1 particularly as it translocates into the nucleous as part of its functional pathway).[22,23]

One final observation worth note is that p21 shows a unique ability to separate very distinct subpopulations in TGFβ, HGF, and BMP2. This would indicate that within these ligand populations, there are subsets of the population that are undergoing growth arrest via the inhibition of cell cycle progression via p21 regulation.

It is clear from these biological observations that the ME-VAE is able to capture relevant biological information and separate cell populations in a way that highlights important features without significant interference from the unformative features that were controlled for. By contrast, little to no biologically relevant information can be easily obtained from the latent encodings of the Standard VAE.

D. Correlation with Pathway activity profiles via Reverse Phase Protein Arrays (RPPA) and ME-VAE features using CYCIF

By reordering VAE features using hieachical clustering to form a feature spectra, we can extract broader patterns from the VAE and reduce the dimensionality of the feature set. While the Standard VAE shows very poor self correlation, with only a handful of features clustering together with significant correlation (Figure 4a left), we observe a very clear pattern of self-correlations between ME-VAE features indicating the model is successful at extracting several distinct expression patterns and that there are clear clusters of VAE features that encode for similar things (Fig 4a right). Using hierarchical clustering, we identify ten representative clusters from the ME-VAE latent space that illustrate different expression patterns. We can visually observe the patterns using a representative image for each cluster (Fig 4b). Using these feature clusters, we can create aggregated feature sets using the mean of all features in a cluster. Representative cell images were chosen by selecting cells that had high expression of their respective cluster to understand and interpret VAE features. Between clusters 0 and 1, we can visually see a difference in the ratio of nuclear size and cell size. Cluster 1 encodes for larger nuclei than is encoded in cells from cluster 0 (this pattern is reaffirmed in Fig. 5b where cluster 1 is correlated to DNA pathways and cluster 0 is not). Cluster 4 is a highly varied cluster, but contains large cells with more diffuse

intensity patterns. From these clusters of VAE features we can see that the ME-VAE architecture is extracting a combination of intensity and morpho-spatial profiles with at least 10 clear axes of variation. Using these reduced feature sets, we can analyze for biological meaning with fewer spurious correlations than comparing many to many.

A growing method for single cell analysis is to integrate multiple modalities. Multi-modal integration helps to validate where the two modalities overlap, expands the dataset with mutually exclusive features, and allows for cross-wise mapping of features. This is important for VAE based single cell image analysis because it frames inherently obscure encoding features in a biological context and validates that the extracted features are coherent. The increased feature range extracted by the ME-VAE allows for the cross-wise mapping and integration of complex CYCIF image features and other modalities, in this case reverse phase protein arrays (RPPA).

When correlating the 7 aggregated Standard VAE features with RPPA pathway activity, we notice two very distinct issues. First, there is only a single Standard VAE aggregated feature that shows any significant pattern of correlation, and that aggregated feature correlates to every RPPA pathway activity profiles (Fig 5a). The second is that there is a single RPPA pathway that correlates too every Standard VAE aggregated feature. When the datapoints of the Standard VAE are shuffled, we notice a similar overall magnitude of correlation values (data not shown), indicating that the correlations found between the Standard VAE and RPPA pathways are likely spurious. When correlating Standard VAE aggregated features to the extracted CYCIF metrics (Methods C, Supplemental Figure 3), the spearman correlations are small despite the increased sample size of n=73,134 (Fig 5a), with the largest correlations being restricted to nuclear markers such as DAPI and CyclinD1. As mentioned above this is likely an artifact of encoding for size as nuclear expressions can be a function of cell size.

By contrast the ME-VAE features show good variation and distinct patterning between all ligand treatments (Fig 5b, Supplemental Figure 4), resulting in more powerful spearman correlations with RPPA Pathways. Unlike with the Standard VAE, when the datapoints are shuffled, we see a significant decrease in correlation strength (data not shown), indicating that the found correlations are less likely to be spurious. The

improved correlations illustrate the multi-encoder's better applicability for multi-modal integration and comparison. Correlations between ME-VAE aggregated features and CYCIF metrics were also performed using spearman correlation using the full 71,314 sample size. For all 10 aggregated features, we see strong and consistent spearman correlations. These correlations illustrate that the aggregated features the ME-VAE extracts have strong biological significance in CYCIF and RPPA and builds confidence in the biological interpretation of the image encoding.

These biological correlations are validated by looking at representative images for each ligand treatment (Supplemental Figure 2 and 4). The stains shown were selected because they had high correlations to the aggregated ME-VAE features or showed distinct patterns visually. The same patterns observed in the correlation and expression matrices (Fig 5b), are also qualitatively confirmed by visual inspection. For example, S6 expression is high in BMP2, EGF, and TGFβ, and it is low in HGF, OSM, and PBS. Radial CyclinD1 slope, as shown in Supplemental Figure 3, is negative in BMP2, EGF, and TGFβ, whith high stain intensity in the inner compartment and rapid decrease toward the cell perimeter; conversely, HGF, OSM, and PBS show much dimmer CyclinD1 expression in the inner compartment. This pattern is even more clearly seen in the radial HES1 slope (Supplemental Figure 4), where HGF, OSM, and PBS show a more continuous stain abundance all the way to the cell membrane. Although the RPPA correlation sample size (n=6) is still too small to draw convincing biological meaning from this dataset, the correlations between proteins markers in CYCIF and RPPA pathways linked by VAE features, are supported by known literature. A simple validating correlation is seen with the correlation of DAPI expression (ME-VAE aggregated feature 1) to the DNA damage and repair (DDR) pathway. We can see that the ME-VAE aggregated feature is positively correlated to both DAPI in CYCIF and to the DDR pathway in RPPA, indicating that these two components are also transitively correlated. A more interesting example (ME-VAE aggregated feature 9) shows a strong correlation between the Stat3 radial slope of distribution and the epithelial-to-mesenchymal transition (EMT) pathway and hormone receptor pathway in RPPA. This is validated by prior literature showing that Stat3 distribution throughout the cell, its activation in the cytoplasm, and its translocation to the nucleus are important in the EGF induced epithelial-to-mesenchymal transition pathway.[24]

These findings based on the previous reseach support the relevance of ME-VAE features and demonstrate a potential application of multi-modal integration by using the proposed approach for single cell image analysis.

The ME-VAE isn't just able to extract biologically meaningful features, but is able to improve downstream analysis by increasing population separability (Fig 6) as measured by mean pairwise Tukey p-values and mean effect sizes to represent the separability of feature distributions between ligand pairs. For the given MCF10a dataset, the CYCIF markers were chosen with the known ligands and cell populations in mind to highlight differences between the populations and separate them. This results in already decent separability using just CYCIF mean intensity information (Fig 6 left). Within the CYCIF mean intensity features, the small mean Tukey pairwise p-value, indicates that for a given feature many of the populations are significantly different, and also have a decent effect size. Because VAE features are able to capture more detail and nuance than naïve intensity metrics, the multi-encoder is able to improve the separability. VAE features show lower mean Tukey pairwise p-values indicating a greater average significance in separability, and the effects sizes for those separations were larger. The exception to this, shown in the first example, is S6, where the CYCIF mean intensity shows better separability, but even in this example the multi-encoder's feature is still adequate. It is worth noting that ME-VAE latent space features will work in combination to represent even a single stain, and so separability can be expected to improve even further when utilizing the more than just one feature at a time.

Discussion:

Just as it is necessary to pre-process, normalize, and remove unwanted features from scRNAseq or RPPA analysis, so too is it necessary to remove uninformative features from single cell imaging analysis in order to better extract features of interest. Without this guided feature alignment, VAE applications for single cell image analysis will only reconstruct dominant features while ignoring subtle more informative features (Fig 1a,b). Transform invariant autoencoders[13] for image analysis have been used before, but controlling for the multiple complex features present in single cell imaging data is difficult given a single encoder network. By making uninformative features invariable across a dataset, using pairs of transformed images in parallel encoding blocks (Fig 1e), VAE priority can be shifted to mutually shared biologically relevant information (Fig 2a, Fig 3). This

also results a more complex and meaningful latent space, as more encoding features are necessary to represent the complex features being reconstructed (Fig 4).

Feature extraction is important for all downstream analysis and interpretation, but often times naïve metrics are not enough to capture biological differences and separate cell populations, especially in datasets where labeled populations are not known beforehand. By separating populations with the ME-VAE, distinct populations and biologically meaningful metrics can be identified (Fig 2b, Fig 5, Supplemental Figure 2), with increased separability compared to naïve metrics using intensity information alone (Fig 2b, Fig. 6). This is because a well-trained encoder can capture complex patterns such as staining texture, pattern, and distribution that are not intuitively quantified by other metrics. More complex single cell analysis methods such as multimodal integration and comparison (Fig 5) require that a wide range of features are extracted that are reflected biologically (Supplemental Figure 4). By employing this architecture to extract a larger range of features and metrics from single cell images, the potential applications become available that were previously restricted due to inadequate cell representations.

The simplicity of the multi-encoder design, makes it easily incorporated into more complex deep-learning architectures, such as being augmented with a discriminator to improve reconstruction quality. This methodology is limited by the fact that the uninformative features being corrected for must meet two criteria: 1) being known beforehand so that it can be addressed with a new encoding block and transformed image pair; 2) be a known transform operation such as rotation, affine, or scale such that a respective randomly transformed image can be made using the operation. Despite this limitation, the majority of dominant uninformative features are based on known transformations, making the ME-VAE architecture widely applicable. Computationally the model is not significantly larger than a Standard VAE or other comparable architectures, as it only adds a single encoding block for each undesired feature. The majority of increased computation time comes from the image transformation steps.

Future applications of this architecture will allow complex features such as texture, pattern, and distribution to be extracted from single cell images free from the hassle of disentangling dominant uninteresting

transform features. Images contain morpho-spatial features not shared by their other single cell counterparts (scRNAseq and RPPA), and by implementing this architecture, the scientific community will be able to analyze these unique image features with the same robustness as algorithms made for the other well-established single cell modalities.

Methods

A. Datasets

MCF10A cell populations were treated with six ligands, PBS (control), HGF, OSM, EGF, EGF+BMP2, and EGF+TGFβ. After 48 hours, cells were fixed and subjected to cyclic immunofluorescence with 23 markers shown in Supplemental Table 1(data from the LINCS Consortium -- https://lincs.hms.harvard.edu/mcf10a/).[25] Cells were segmented using CellPose segmentation tool,[26] on the EGFR channel. Stains were normalized using histogram stretching to the 1st and 99th percentiles across intensities for the whole dataset. Image augmentations were applied for rotation, polar orientation, and size/shape (Supplemental Figure 5). Rotation is corrected by obtaining the major axis from the binary cell mask, then rotating the image using the python package OpenCV.[27] Polar orientation was corrected by calculating the angle toward the image's center of mass, then applying a flip/rotation to align the angle using the python Numpy package.[28] Size/shape was corrected simultaneously by registering the cell mask to a circle target image (code available here: https://github.com/GelatinFrogs/Cells2Circles). In total, 71,314 cells were processed through this pipeline.

Bulk Reverse Phase Protein Array (RPPA) was performed by the LINCS consortium[25] in parallel to the CYCIF imaging, on cell populations treated with the same ligands after 48 hours of exposure. The protein array incorporated 295 protein markers. As described by Akbani, R., et al.,[29] RPPA data were median-centered and normalized by standard deviation across all samples for each component to obtain the relative protein level. The pathway score is then the sum of the relative protein level of all positive regulatory components minus that of negative regulatory components in a particular pathway. Pathway members and weights were developed through literature review. Despite the available bulk RPPA dataset saving a smaller sample size than the single cell CYCIF dataset, it was chosen as the secondary modality because similar ligand separation and cluster patterns were

observed in both modalities, indicating an overlap in the information each contains (Supplemental Figure 6).

For correlation to CYCIF and RPPA pathways, the VAE latent space was restricted to smaller sets of aggregated features. These aggregated features were made using self-correlation of VAE features across individual cell metrics and averaging the VAE features for resulting hierarchical clusters (Figure 4). This was done to reduce the feature dimensionality and reduce spurious correlations in the biological findings. Representative images for each cluster were done by finding cells with a high average expression for all features within the cluster. For other analyses of VAE features, ME-VAE encoding features were restricted to 18 single features for each. The dimension of 18 was chosen because it is the number of mutual markers between the RPPA and CYCIF datasets. Explanatory features were chosen from the VAE encodings such that the inter-cluster variability was maximized and the intra-cluster variability was minimized using the following equation:

$$Feature\ Score = Var_{all} - \frac{\sum_{i=cluster} Var_{C_i}}{\#\ of\ Clusters}$$

B. VAE Models

To allow for fair comparison, the structure of the encoder and decoder blocks were kept consistent between networks, and the same latent dimension was used for all models for a given dataset (64 for the 1-channel dataset, 512 for the 23-channel dataset). Standard VAEs with matching pairs of single cell images were used to establish baseline performance (Fig 1c):

$$L_c = \|x - D(E(x))\| \rightarrow z = E(x)$$

By using an image randomly transformed w.r.t. a dominant feature as the input and controlling for the uninformative feature in the output image (Fig 1d), the model can self-supervise the transformation and will only encode novel features since the controlled features (such as rotation) no longer aid reconstruction:

$$L_d = \|x' - D(E(T^{-1}(x')))\| \rightarrow z = E(T^{-1}(x'))$$

where x' represents an image that has been transformed with a known transformation to remove one or more uninformative features and $T^{-1}(\cdot)$ represents a random transformation of the controlled image to create a dominant uninformative feature at a random degree.

The proposed multi-encoder architecture uses multiple transformed inputs with separate encoder blocks, where each block controls for a separate uninformative feature, and a single decoder block uses the shared latent space for reconstruction (Fig 1d):

$$L_e = \|x' - D(\,[E_1(T^{-1}(x')) \times E_2(T^{-1}(x')) \times \ldots \times E_i(T^{-1}(x'))]\,)\,\| \rightarrow$$

$$z_{shared} = [E_1(T^{-1}(x')) \times E_2(T^{-1}(x')) \times \ldots \times E_i(T^{-1}(x'))]$$

where each encoder's individual latent space is combined in an elementwise multiplication layer and $T_i^{-1}(\cdot)$ represents a different random transformation for individual uninformative features such as rotation, polar orientation, size, shape, etc. respectively. The shared latent space of the multi-encoder forces the deep learning model to encode features that are shared between each transformation, reinforcing the shared mutual information and eliminating the non-shared transformational information.

The multi-encoder architecture allows for image pairs to be randomly transformed, which can act as a balancing agent for imbalanced features. Furthermore, the corrected outputs serve as a weakly self-supervised signal for the model. With the extra information from the additional inputs, the model is able to overcome more complex transformations that failed in the corrected output architecture in Fig 1d, when multiple corrections are attempted. Paired images also serve one additional benefit of allowing for features to be retained in parallel encoders that might be lost due to artifacts in other corrections, i.e. artifacts within a polarity correction encoder will not be present in a rotation correction encoder.

C. Evaluations Metrics

In order to evaluate the model's ability to separate labeled cell populations, k-means clustering was applied to the encoding spaces using sklearn.[30] Cluster purity was then calculated by taking the percentage of the largest population for each cluster. UMAP embeddings were calculated using the UMAP python package.[31] Biological metrics were calculated to give VAE encoding features biological grounding (Supplemental Figure 3). Circularized cells were used for calculation because it made compartmentalization of the cell more consistent and uniform. Mean intensities were calculated for inner, middle, outer, and whole cell compartments. To calculate the radial slope, the mean intensity was taken from each radius of the circularized cell, then the linear regression of

the series was calculated using the scipy.stats package[32] in python. Self-correlations between VAE features were performed using spearman correlation and clustering was done in seaborn clustermap.[33] Representative cluster images were chosen based on high expression of the cluster's respective VAE features. RPPA pathways activity scores, VAE features, and biological metrics were all standardized prior to analyses using the sklearn StandardScaler function[30] in python. Correlations between RPPA pathway activities and VAE encodings and between CYCIF and VAE encodings were both calculated using the spearman correlation. To test for separability (Fig. 6), features were first tested using ANOVA, all of which proved significant. Subsequently, the post-hoc pairwise Tukey p-test was used to calculate the significance and effect size for each ligand pair. The mean p-value and effect size were reported to illustrate average separability.

Acknowledgements

This work was supported in part by the National Cancer Institute (U54HG008100, U54CA209988, U2CCA233280, U01 CA224012, R01 CA253860) and the OHSU Center for Spatial Systems Biomedicine. The resources of the Exacloud high performance computing environment developed jointly by OHSU and Intel and the technical support of the OHSU Advanced Computing Center are gratefully acknowledged.

Author contributions

Conceptualization, L.T. and YHC.; Methodology, L.T. and Y.H.C.; Software, L.T.; Formal Analysis, L.T.; RPPA Analysis, M.D. and M. L. Writing original draft, L.T.; Writing-Review and Editing, L.T. and Y.H.C.; Supervision, G.M, J.W.G, L.H and Y.H.C.; Funding Acquisition, J.W.G;

References

1 Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, **32**(4), 381–386. https://doi.org/10.1038/nbt.2859

2 Saelens, W., Cannoodt, R., Todorov, H. et al. A comparison of single-cell trajectory inference methods. *Nat Biotechnol*. **37**, 547–554 (2019). https://doi.org/10.1038/s41587-019-0071-9

3 Bergen, V., Lange, M., Peidli, S. et al. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* (2020). https://doi.org/10.1038/s41587-020-0591-3

4 Lin, J. R., Fallahi-Sichani, M., Chen, J. Y., & Sorger, P. K. (2016). Cyclic Immunofluorescence (CycIF), A Highly Multiplexed Method for Single-cell Imaging. *Current protocols in chemical biology*, 8(4), 251–264. doi:10.1002/cpch.14

5 Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods*. 2014;70(1):46–58. 10.1016/j.ymeth.2014.08.016.

6 Dakshinamoorthy G, Singh J, Kim J, Nikulina N, Bashier R, Mistry S, et al. Abstract 490: Highly multiplexed single-cell spatial analysis of tissue specimens using CODEX. *Cancer Res*. 2019;79(13 Supplement):490 10.1158/1538-7445.Am2019-490.

7 Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, et al. Multiplexed ion beam imaging of human breast tumors. Nat Med. 2014;20(4):436–42. 10.1038/nm.3488.

8 Way, G.P., et al., Predicting cell health phenotypes using image-based morphology profiling. *Molecular Biology of the Cell*. (2021). https://doi.org/10.1091/mbc.E20-12-0784

9 Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

10 Geddes, T., Kim, T., Nan, L. *et al*. Autoencoder-based cluster ensembles for single-cell RNA-seq data analysis. *BMC Bioinformatics* **20,** 660 (2019). https://doi.org/10.1186/s12859-019-3179-5

11 Hinton G.E., Krizhevsky A., Wang S.D. (2011) Transforming Auto-Encoders. In: Honkela T., Duch W., Girolami M., Kaski S. (eds) Artificial Neural Networks and Machine Learning – ICANN 2011. ICANN 2011. Lecture Notes in Computer Science, vol 6791. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21735-7_6

12 Bai, Y., & Duan, L. L. (2019). Tuning-Free Disentanglement via Projection. *arXiv preprint arXiv:1906.11732*.

13 Bepler, T., Zhong, E., Kelley, K., Brignole, E., & Berger, B. (2019). Explicitly disentangling image content from translation and rotation with spatial-VAE. In *Advances in Neural Information Processing Systems* (pp. 15435-15445).

14 Matsuo, T., Fukuhara, H., & Shimada, N. (2017, September). Transform invariant auto-encoder. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2359-2364). IEEE.

15 Guo, X., Zhu, E., Liu, X., & Yin, J. (2019, August). Affine Equivariant Autoencoder. In *IJCAI* (pp. 2413-2419). https://doi.org/10.24963/ijcai.2019/335

16 Way, G.P., Zietz, M., Rubinetti, V., et al. Compressing gene expression data using multiple latent space dimensionalities learns complementary biological representations. Genome Biol 21, 109 (2020). https://doi.org/10.1186/s13059-020-02021-3

17 Christopher Heje Grønbech, Maximillian Fornitz Vording, Pascal Timshel, Casper Kaae Sønderby, Tune Hannes Pers, Ole Winther, scVAE: Variational auto-encoders for single-cell gene expression data, Bioinformatics, btaa293, https://doi.org/10.1093/bioinformatics/btaa293

18 Rohan Gala, Agata Budzillo, Fahimeh Baftizadeh, Jeremy Miller, Nathan Gouwens, Anton Arkhipov, Bosiljka Tasic, Gabe Murphy, Hongkui Zeng, Michael Hawrylycz, Uygar Sümbül. bioRxiv 2020.06.30.181065; doi: https://doi.org/10.1101/2020.06.30.18106

19 Rybkin, O., Daniilidis, K., & Levine, S. (2020). Simple and Effective VAE Training with Calibrated Decoders. *arXiv preprint arXiv:2006.13202*.

20 Park, T., Zhu, J. Y., Wang, O., Lu, J., Shechtman, E., Efros, A. A., & Zhang, R. (2020). Swapping Autoencoder for Deep Image Manipulation. *arXiv preprint arXiv:2007.00653*.

21 Geoffrey Schau, *et al*. Variational Autoencoding Tissue Response to Microenvironment Perturbation" *SPIE Medical Imaging Conference*, San Diego, Feb 16-21, 2019; https://doi.org/10.1117/12.2512660

22 Kaoru Murata, et al. Hes1 Directly Controls Cell Proliferation through the Transcriptional Repression of p27Kip1. *Molecular and Cellular Biology*. **25** (10) 4262-4271, (2005). doi: 10.1128/MCB.25.10.4262-4271.2005

23 Nikolai N. Khodarev, Bernard Roizman and Ralph R. Weichselbaum. Molecular Pathways: Interferon/Stat1 Pathway: Role in the Tumor Resistance to Genotoxic Stress and Aggressive *Growth*. *Clin Cancer Res*. **18** (11) 3015-3021 (2012). doi: 10.1158/1078-0432.CCR-11-3225

24 Wang T, Yuan J, Zhang J, et al. Anxa2 binds to STAT3 and promotes epithelial to mesenchymal transition in breast cancer cells. *Oncotarget*. 2015;6(31):30975-30992. doi:10.18632/oncotarget.5199

25 The LINCS Consortium: http://www.lincsproject.org

26 Stringer, C., Wang, T., Michaelos, M. *et al*. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18,** 100–106 (2021). https://doi.org/10.1038/s41592-020-

27 Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. (2000).

28 Harris, C.R., *et al*. Array programming with NumPy. *Nature* **585,** 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2

29 Akbani, R., et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun*. **5**, 3887 (2014). https://doi.org/10.1038/ncomms4887

30 Fabian Pedregosa, *et al*. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830 (2011).

31 McInnes, L, Healy, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*. 1802.03426, 2018

32 Pauli Virtanen, *et al*. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**(3), 261-272 (2020).

33 Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, **6**(60), 3021 (2021). https://doi.org/10.21105/joss.03021

Figures

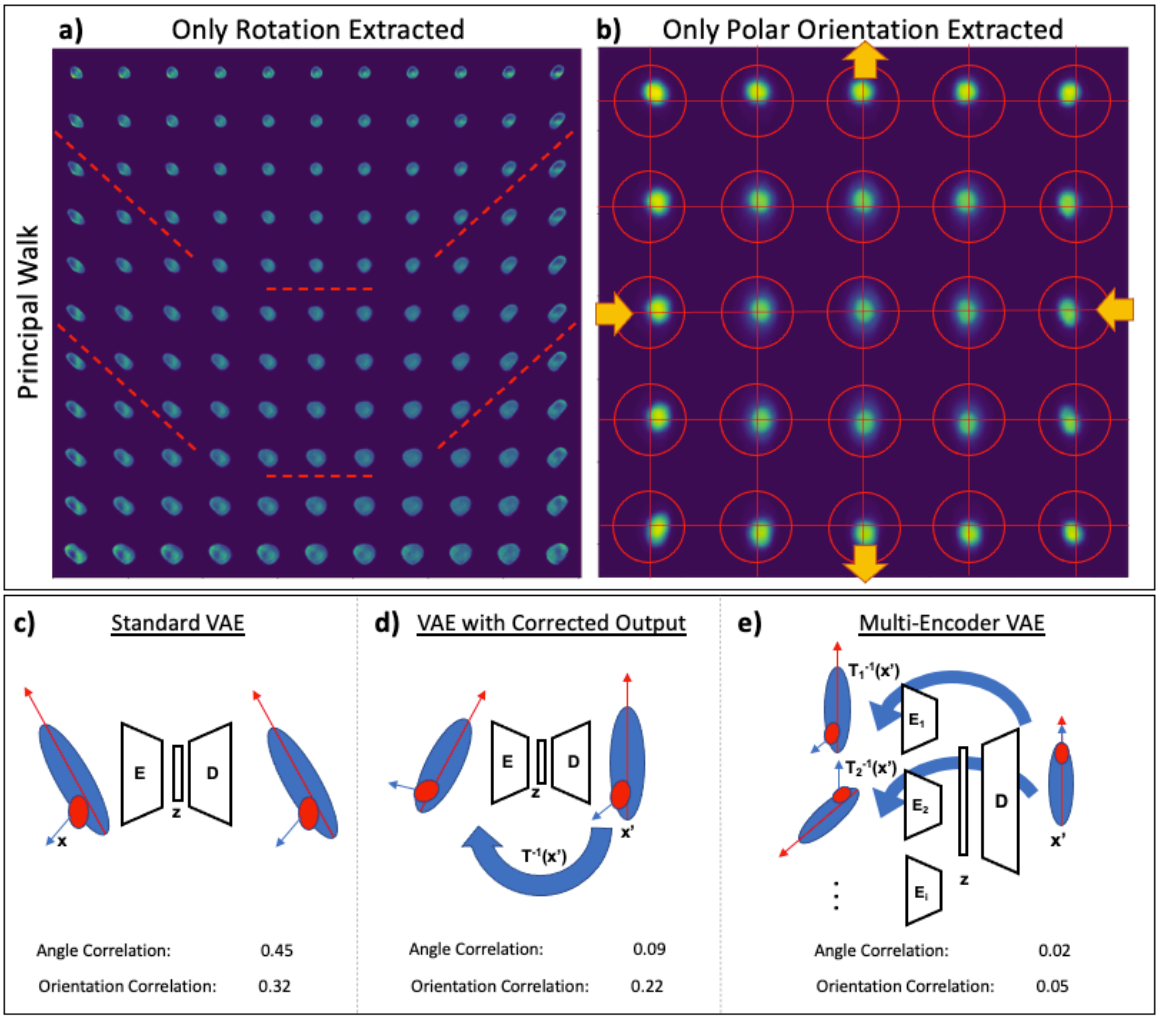


Figure 1: VAE hypersensitivity and proposed model architecture. VAE analysis of two separate datasets are shown, each governed by a single biologically uninformative feature **a)** rotation and **b)** polar orientation. Principal walk reconstructions[21] show the VAEs' governing features across the latent space through a range of image reconstructions. To correct this model hypersensitivity several architectures were tested: **c)** Standard VAE with matched raw images. **d)** VAE with paired randomly transformed input and controlled output images **e)** the proposed multi-encoder VAE: VAE with corrections for multiple features (rotation, polar orientation, size, shape, etc.) using parallel encoder models, a shared latent space, and a single decoder model. In c)-e), a correlation between the embedding components and the respective feature (angle and orientation) is measured to quantify how effectively the model removes uninformative features.

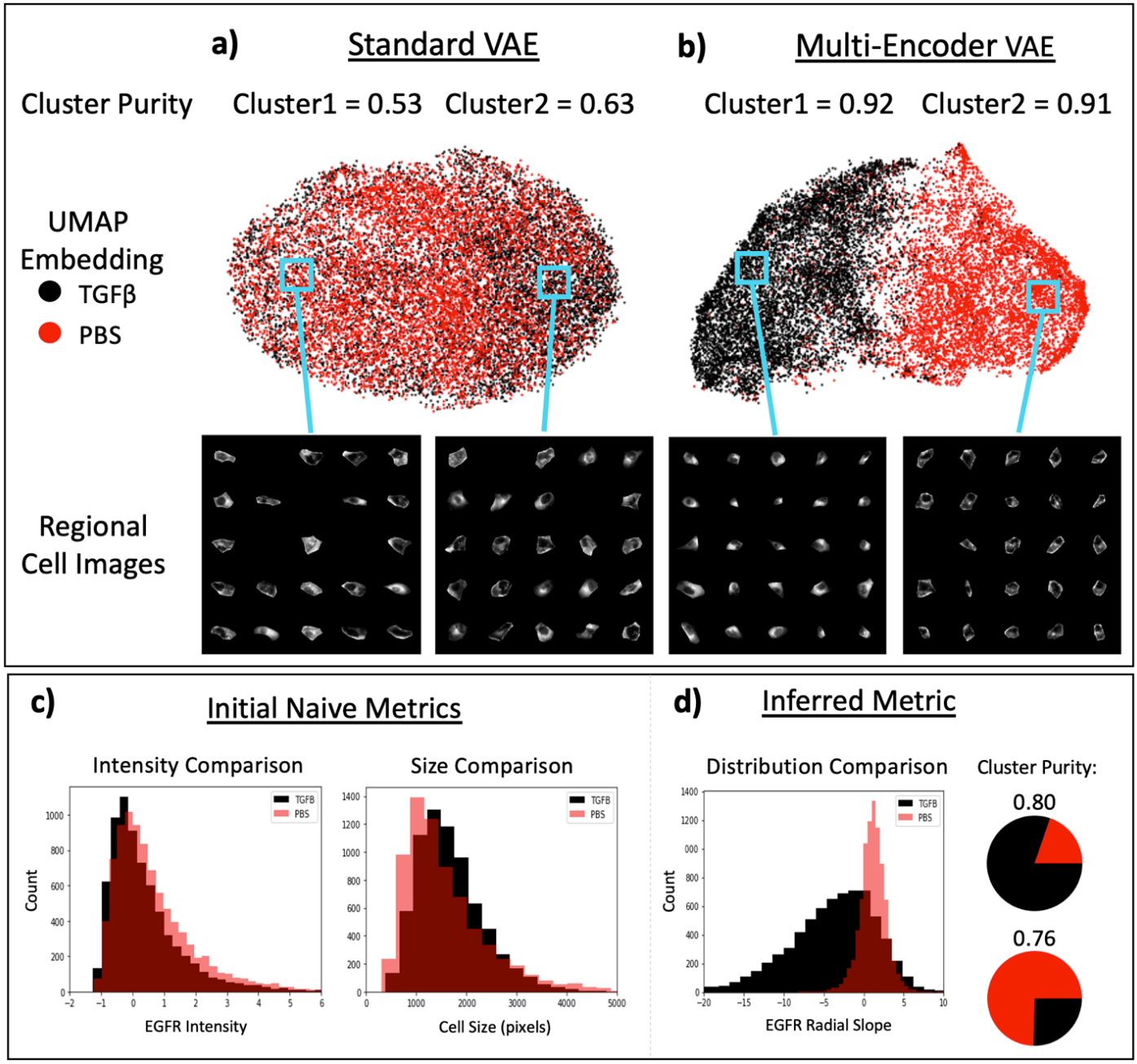


Figure 2: Separation of biologically distinct cell populations. a,b) The models are quantitatively evaluated using cluster purity (k-means, n=2). Qualitative comparison is made using visual separation of two labeled cell populations in UMAP embedding space and visual analysis of cells from UMAP regions to identify biologically distinct factors. **c)** Cells are compared using initial naïve metrics such as mean EGFR intensity and cell size to show the difficulty separating the cell populations. **d)** Cells are compared using radial slope (a metric inferred from visual analyzing the regional cell images in **b**).

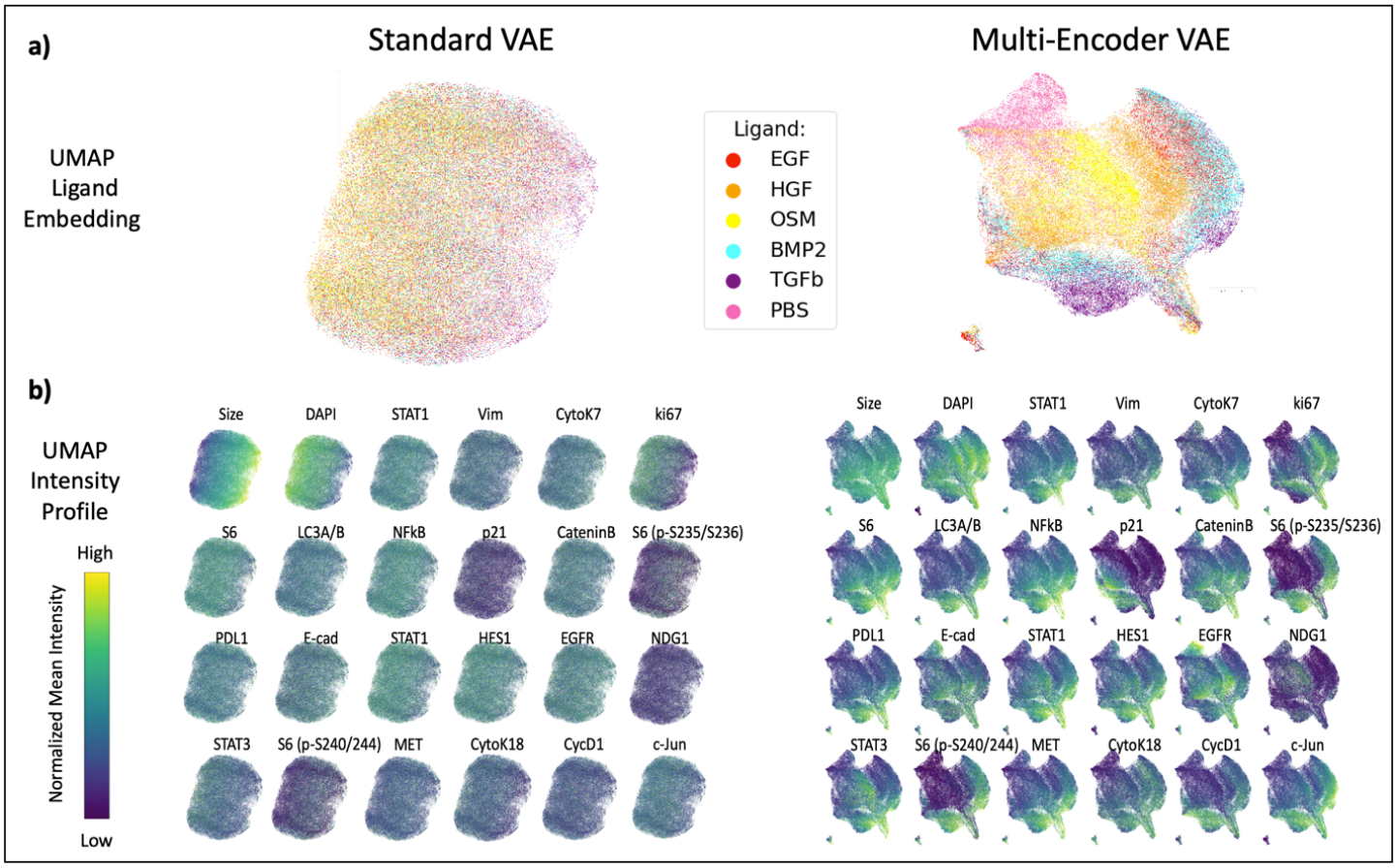


Figure 3: Ligand separation and feature distribution. a) UMAP embeddings for respective VAE encodings, allowing for qualitative visual evaluation of ligand separability**. b)** Distribution of stain features across UMAP space, colored by intensity.

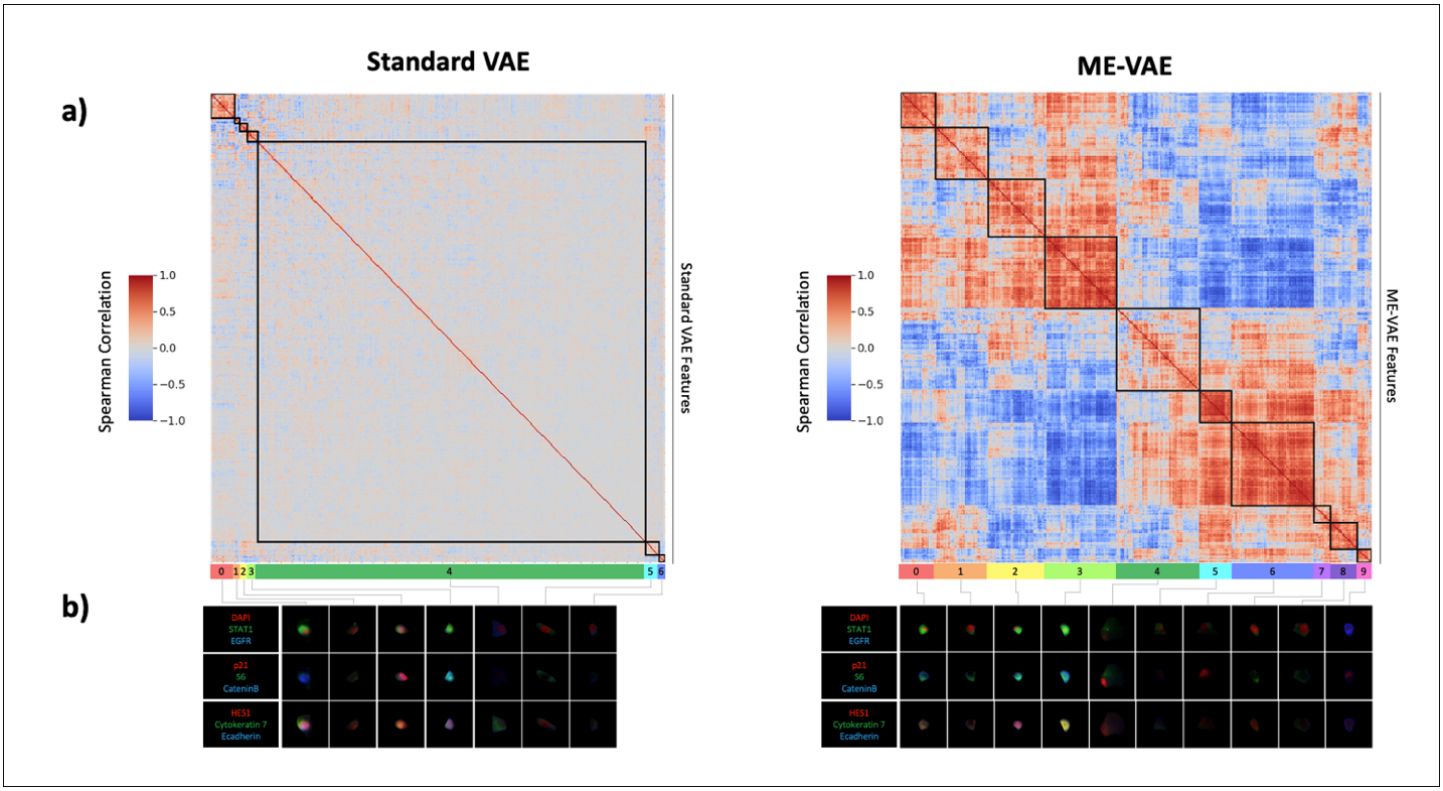


Figure 4: VAE feature self-correlation and clustering of aggregated feature sets.
a) Using the single cell observations as features, correlations are drawn between sets of individual VAE feature pairs. These features are then hierarchically clustered to observe patterns and reduce VAE features to aggregated feature sets.
b) Cell images were assigned aggregated feature scores using the mean expression of each feature in a cluster. Shown are representative cells that are highly expressing for each respective cluster.

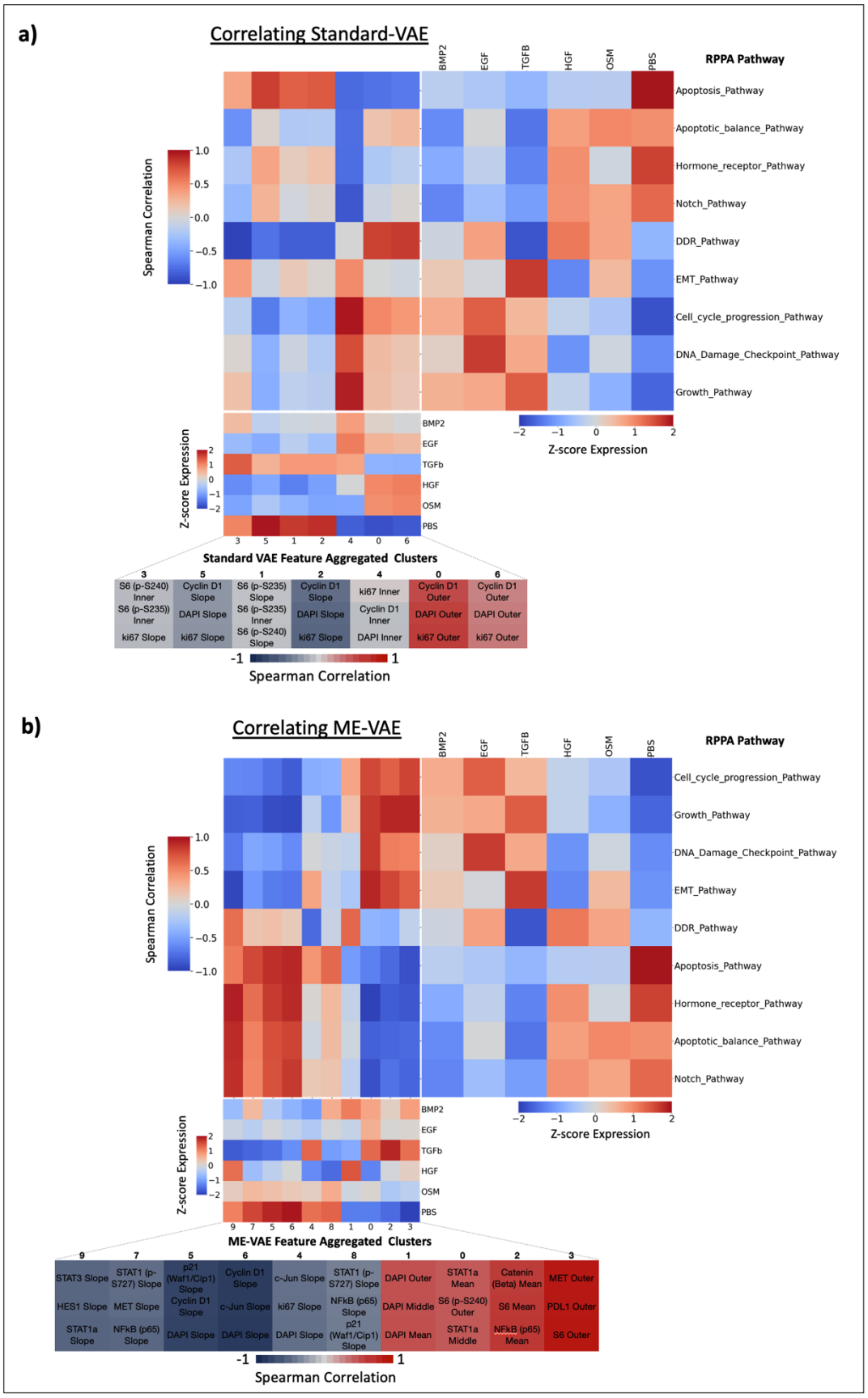


Figure 5: Correlating VAE features to RPPA pathway activity profiles and CYCIF. Correlation matrices between RPPA pathway activity scores and **a)** Standard VAE and **b)** ME-VAE aggregated features. Samples from the two modalities were paired by their ligand treatments, resulting in a sample size of n=6. RPPA pathways and VAE features were hierarchically clustered to show prominent patterns in correlation. VAE features were also correlated to several metrics of CYCIF expression (mean inner, mean middle, whole cell mean, and radial slope) for all 23 stains. This was done using the full dataset with single cell images (sample size n=73,134). The table of CYCIF correlations shows the top three correlations for each VAE feature.

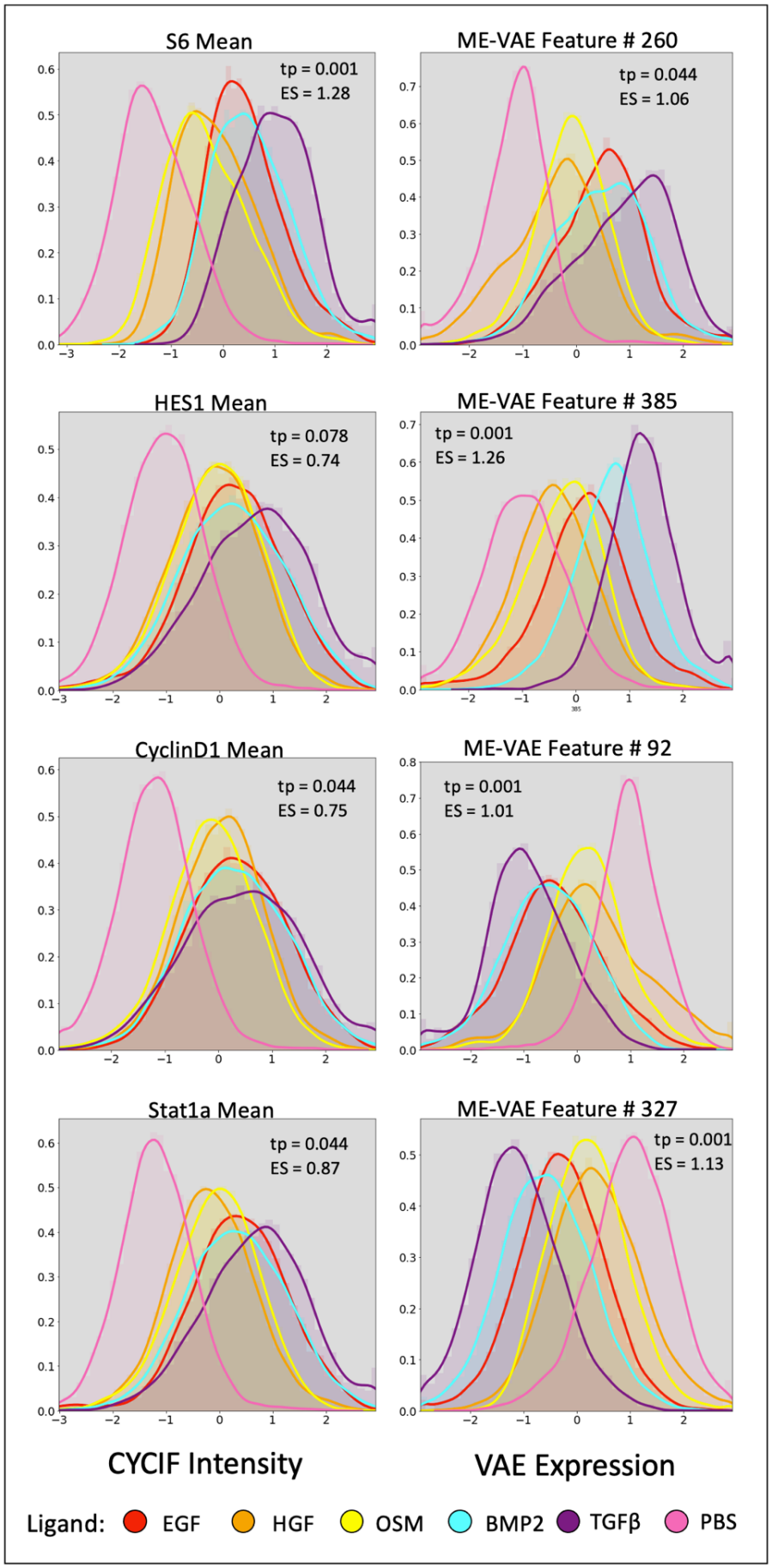
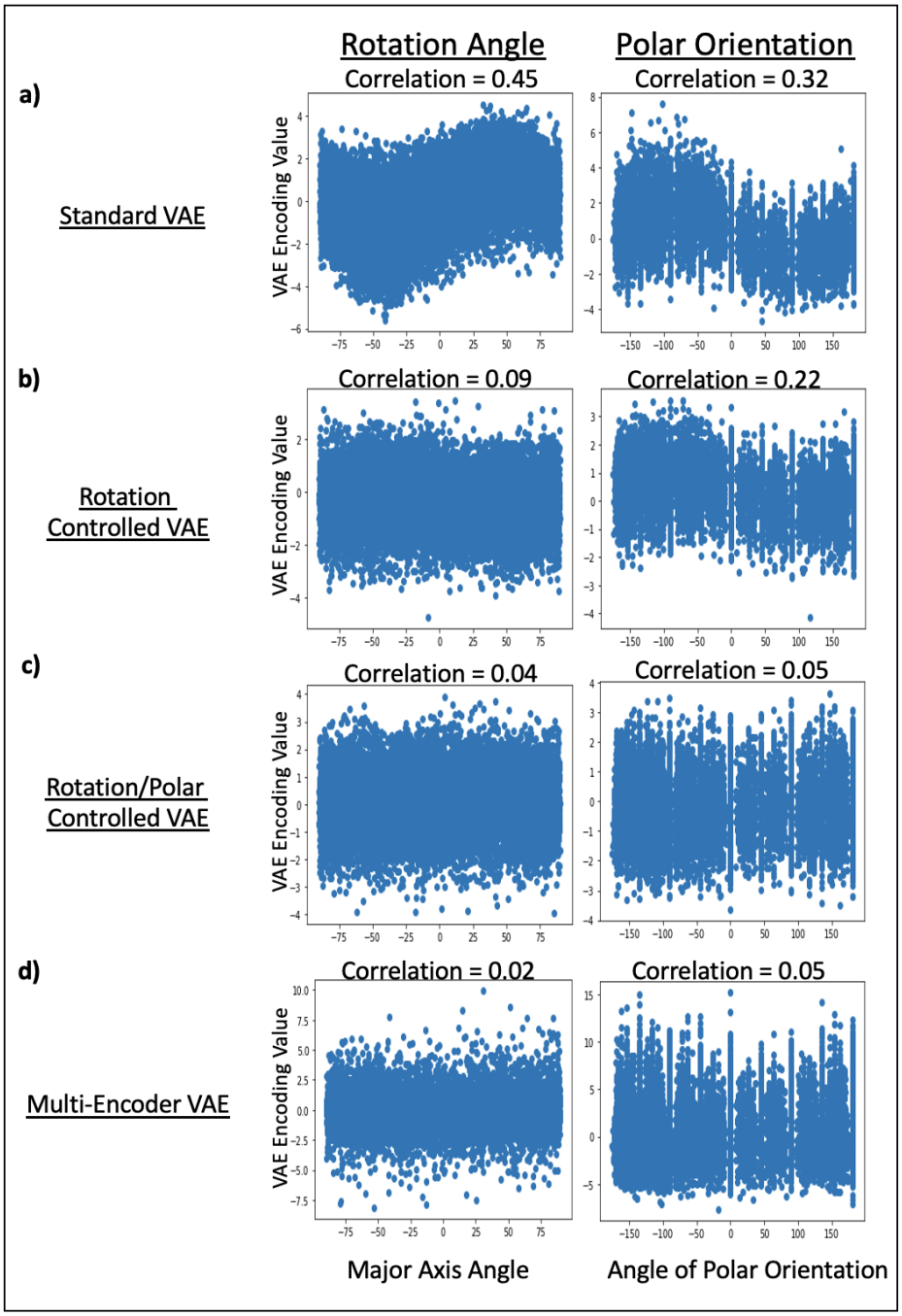
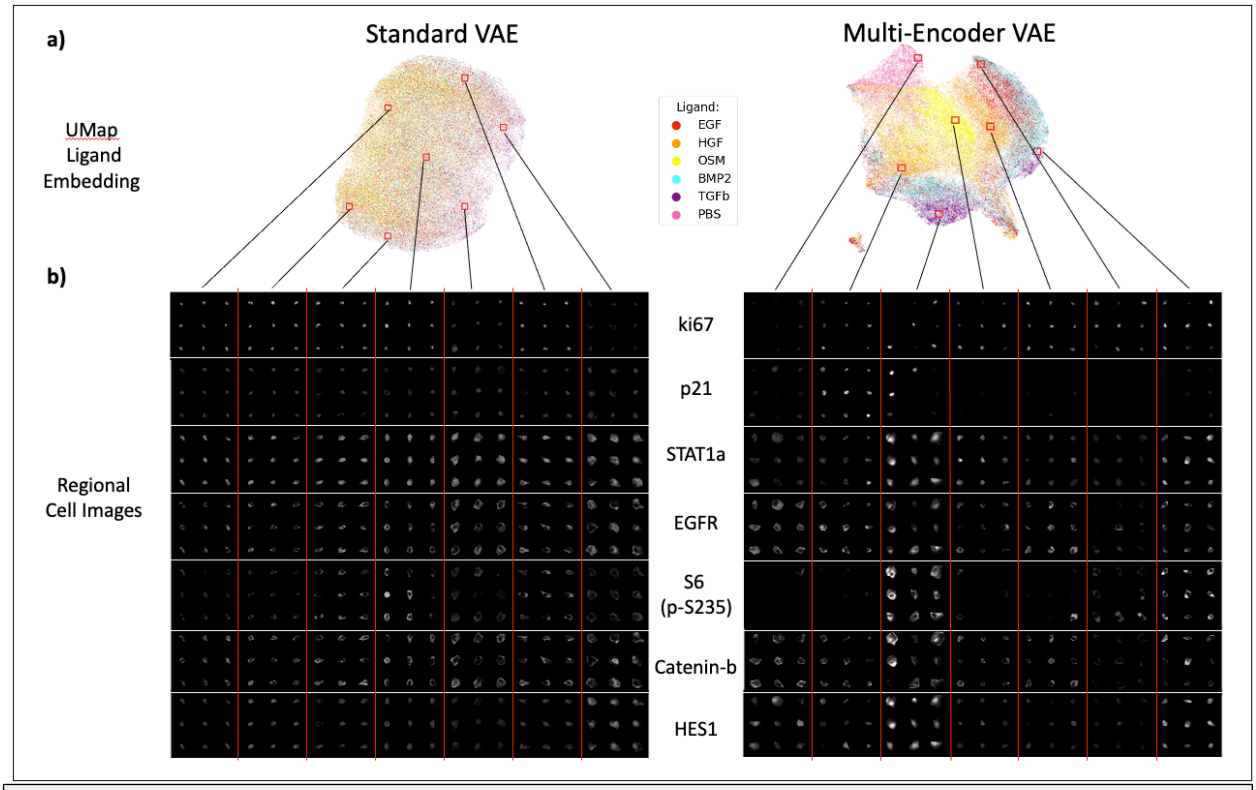


Figure 6: Separability of ligands. Density function and histogram for several CYCIF and ME-VAE feature pairs. The mean Tukey-pairwise p-value (tp) and mean effect size (ES) shown for each feature.

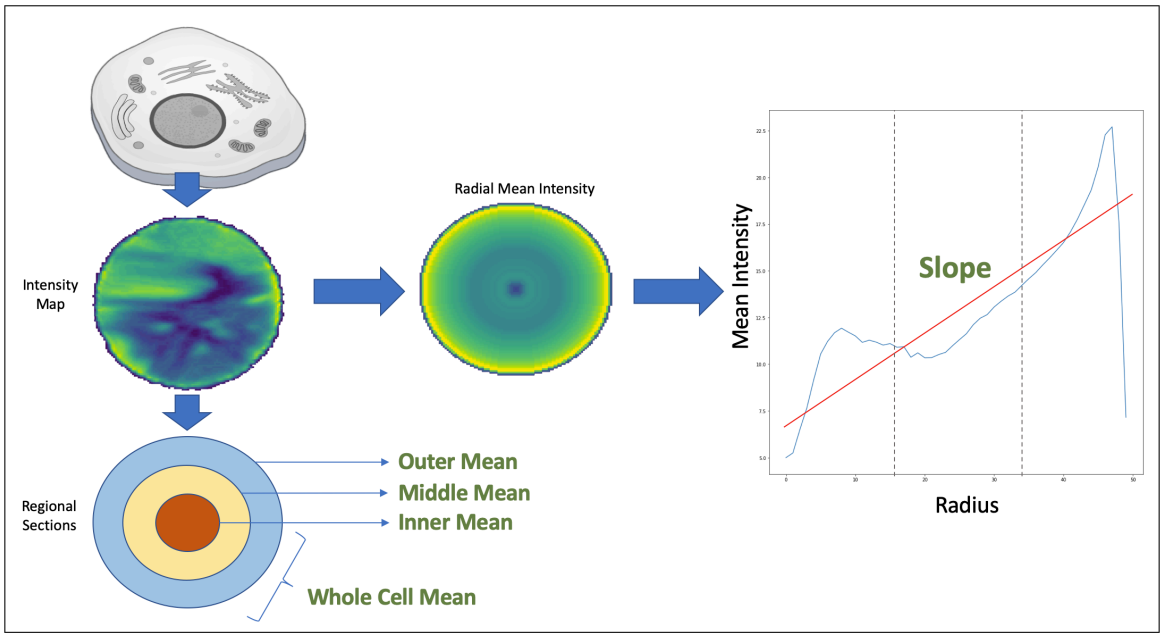
Supplemental Material



Supplemental Figure 1: Mitigation of uninformative features for tested architectures. Encoding spaces for each VAE method were analyzed for correlation with uninformative features. Scatter plot and correlation is shown for the latent space component that had the highest correlation to given the metric. **a)** Standard VAE used as baseline to show high correlation between encoded features and undesired features. **b)** Output corrected transform invariant VAE controlling for rotation only. **c)** Output corrected transform invariant VAE controlling for rotation and polar orientation. **d)** Proposed multi-encoder VAE correcting for both rotation and polar orientation.



Supplemental Figure 2: Regional cell images across UMAP clusters. Regional cell images were sampled from locations throughout UMAP space to highlight the differences in expression pattern. Stains shown were selected based on a combination of being correlated to important VAE features and hand-selection for known variance.



Supplemental Figure 3: Extracted biological metrics from CYCIF. Cell intensity maps were circularized to allow easy compartmentalization. The inner, middle and outer mean intensities were extracted by dividing the cell into thirds radially. The mean intensity of the whole cell was also taken. The radial mean intensity map was created by taking the average intensity for each radius across the circularized cell. The slope of the radial mean intensity map was then taken to create a single metric for stain distribution.

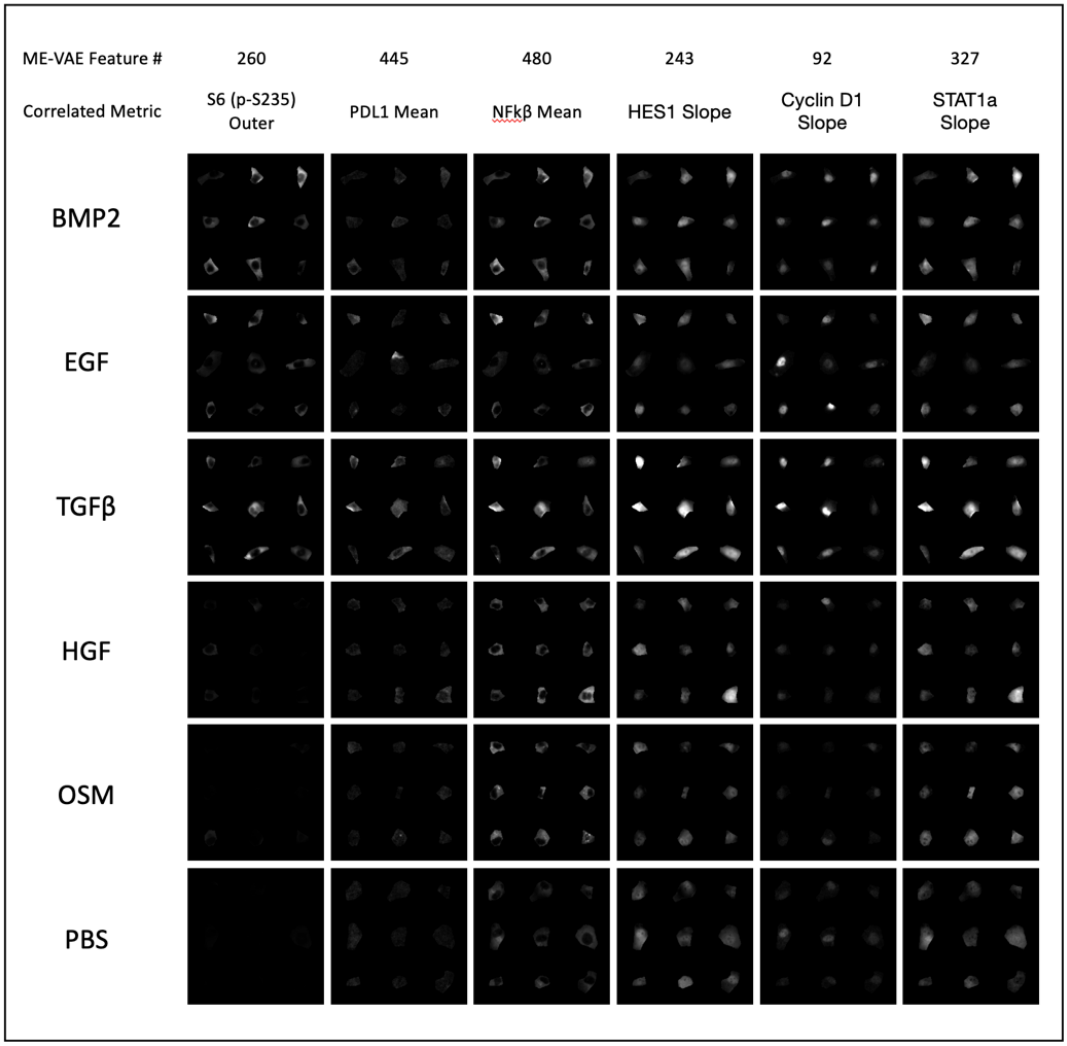
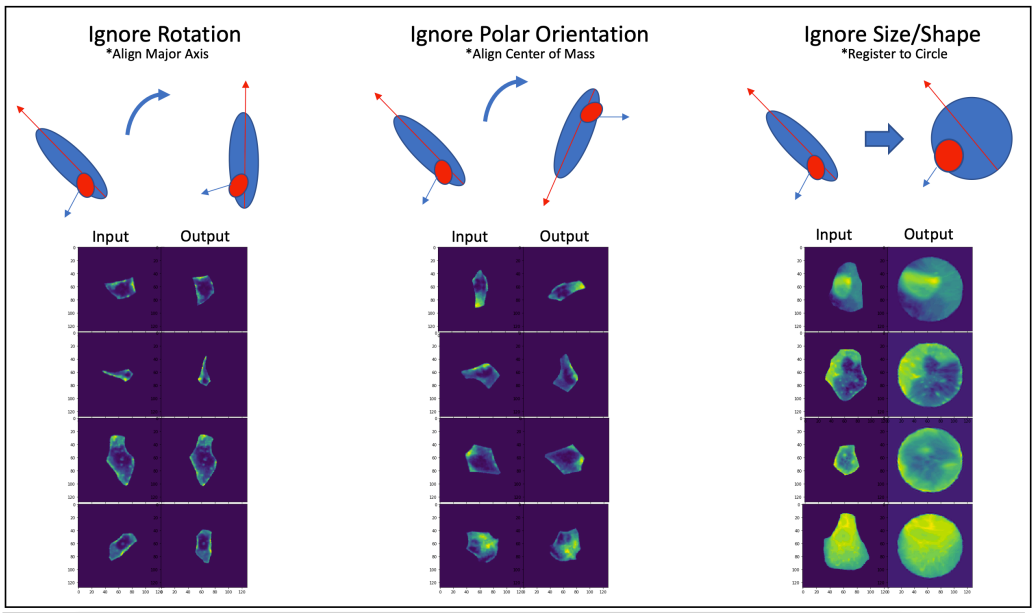
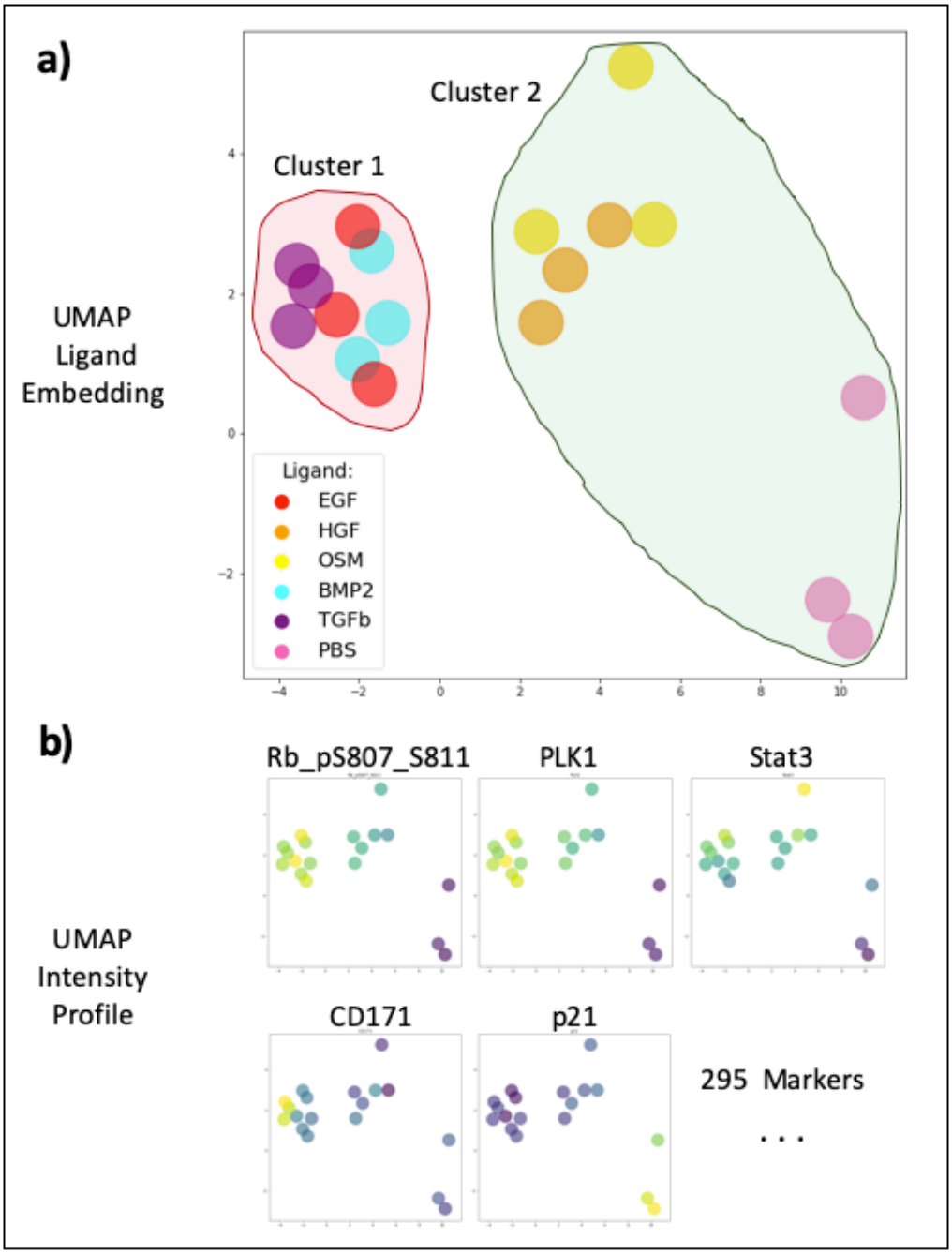


Figure 4: Representative cell images for each ligand treatment. Representative cell images are shown for each ligand treatment (rows) and are shown using several stains (columns). Each column also includes a VAE # that ties back to the multi-encoder feature that is highly correlated.



Supplemental Figure 5: Cell image augmentation and correction. Examples of image corrections for rotation, polar orientation, and size/shape, shown using EFGR channel of randomly selected images.



Supplemental Figure 6: Bulk RPPA analysis and clustering.
a) Independent analysis of the Bulk RPPA dataset shows distinct clustering of ligand populations in UMAP embeddings space where **b)** selected markers show clear patterns of distribution between the clusters.

Supplemental Table 1: CYCIF Markers

Channel	Marker
1	DAPI
2	STAT1 (p-S727)
3	Vimentin
4	Cytokeratin 7
5	ki67
6	S6
7	LC3A/B
8	NFkB (p65)
9	p21 (Waf1/Cip1)
10	Catenin (Beta)
11	S6 (p-S235/S236)
12	PDL1
13	E-cadherin
14	STAT1 (alpha-isoform)
15	HES1
16	EGFR
17	NDG1 (p-T346)
18	STAT3
19	S6 (p-S240/244)
20	MET
21	Cytokeratin 18
22	Cyclin D1
23	c-Jun