Variational Autoencoding Tissue Response to Microenvironment Perturbation

Geoffrey F. Schau[a,b], **Guillaume Thibault**[a], **Mark A. Dane**[a], **Joe W. Gray**[a], **Laura M. Heiser**[a], **Young Hwan Chang**[a,b]

[a]Dept. of Biomedical Engineering, Center for Spatial Systems Biomedicine, Oregon Health & Science University, Portland, OR, USA

[b]Computational Biology Program, Oregon Health & Science University, Portland, OR, USA

Abstract

This work applies deep variational autoencoder learning architecture to study multi-cellular growth characteristics of human mammary epithelial cells in response to diverse microenvironment perturbations. Our approach introduces a novel method of visualizing learned feature spaces of trained variational autoencoding models that enables visualization of principal features in two dimensions. We find that unsupervised learned features more closely associate with expert annotation of cell colony organization than biologically-inspired hand-crafted features, demonstrating the utility of deep learning systems to meaningfully characterize features of multi-cellular growth characteristics in a fully unsupervised and data-driven manner.

1. INTRODUCTION

The presence of constituent components within the cellular microenvironment and their effect on growth, differentiation, and therapeutic response of tissue is of paramount importance in the field of spatial systems biology.[1, 2] Recent advances in high-throughput systematic screening technologies enable quantification of phenotypic differences among a variety of cell populations in response to diverse chemical and genetic treatments.[3, 4] The microenvironment microarray (MEMA) platform[5, 6] is designed to generate images that capture diverse phenotypic changes of cellular populations exposed to soluble ligands and insoluble extracellular matrix (ECM) proteins. High-throughput generation of these types of data require powerful and sophisticated algorithms to capture features of interest to better form and validate biological hypotheses. Presently, image-based cell profiling methods utilize classical image quantification approaches to extract hundreds of features from high content images to quantify the perturbagens' effects on feature gradients. However, defining and characterizing microenvironment-dependent multi-cellular spatial organization has remained an unmet computational challenge. Although popular techniques extract features such as cell counts, cellular spatial relationships (i.e., neighborhood information), or distance to cells in a specific sub-cellular structure,[4] these features are limited to

characterizing spatial organization of individual cells and often require significant biological expertise to design.

In biomedical imaging analysis, deep learning techniques that employ convolutional neural networks (CNNs) to extract deep hierarchical spatial features directly from raw pixel image data have been shown to outperform classical methods that analyze hand-crafted features.[7] Applications of deep convolutional neural networks in cellular imaging have shown promising utility for classification, segmentation, and dimensionality reduction in diverse biomedical contexts.[8, 9] Multi-agent learning models, including generative adversarial networks (GANs)[10] and variational autoencoders (VAEs),[11] have recently been shown to be capable of learning salient features of high-throughput imaging screens at cellular and sub-cellular resolution.[7, 12] Although powerful, GAN architecture has been shown to struggle in capturing multiple modes of input data, which limits the interpretability of their learned features.[13] Unlike GANs, VAE latent features conform to expected prior distributions, which enables elegant interpretation and visualization of what these models learn. To characterize features of multi-cellular growth patterns associated with microenvironment perturbation, this work applies convolutional variational autoencoding architecture to analyze images of normal human mammary epithelial cells grown on the MEMA platform. The main advantages of our approach are:

- *Multi-cellular spatial organization characterization*: Unlike current image-based cell profiling methods that focus on single-cell analysis, our approach is designed to learn biologically meaningful spatial organization of multi-cellular populations.

- *Principal Feature Manifold*: We introduce a novel method to visually interpret meaningful high-dimensional learned features of a VAE model by generating synthetic samples within the principal component plane of the model's learned feature space.

2. METHODS

2.1 Deep Variational Autoencoding Networks

The variational autoencoder (VAE) architecture introduced by Kingma and Welling[11] is designed to elucidate salient features of data in a data-driven and unsupervised manner. A VAE model seeks to train a pair of complementary networks: an *encoder* network θ that seeks to model an input x_i as a hidden latent representation z, and a *decoder* network ϕ that seeks to reconstitute x_i from its latent representation z. The VAE loss function shown in Equation 1 regularizes model training with an additional Kullback-Leibler (KL) divergence term that penalizes the distribution of z with respect to a given prior, which in our case is the standard normal Gaussian distribution, $p(z) = N(0,1)$. By specifying a latent dimension z less than the input dimension of x_i, a VAE model learns optimized encoding and decoding functions that enable reconstruction of an input sample subject to capacity constrains of the latent feature space within the model.

$$\mathcal{L}_i(x_i, \theta, \phi) = -\mathbb{E}_{z \sim q_\theta(z|x_i)}\Big[\log p_\phi(x_i|z)\Big] + \mathrm{KL}\Big(q_\theta(z|x_i)\big\|p(z)\Big)$$

The VAE model trained in this study incorporates two-dimensional convolutional layers to encode spatial information of multi-cellular organization of cells grown in diverse microenvironments. Specifying a limiting bottleneck on the latent feature space forces the model to learn salient features of the dataset and reduce the dimensionality of input features for further downstream analyses.

2.1.1 Learning Model Design—The encoder and decoder models used in this study are congruent and composed of five 2D convolutional layers each containing 64 filters with same padding and rectified linear unit activations on all layers except for the final sigmoidal decoder layer. The outer two convolutional layers have a 3×3 kernel, the inner two layers have a 2×2 kernel, and the latent layer is composed of 16 hidden features, which illustrated good trade-off between model capacity and training loss. Both the encoder and decoder are optimized with the RMSProp optimizer against the custom variational loss function that penalizes the binary cross entropy between input and reconstruction as well as the KL divergence between the latent space sample and standard normal distribution. The models designed for this study were written in Python using Keras[14] with Google's Tensorflow backend,[15] and trained using Nvidia Tesla V100 GPUs mounted on the Exacloud high performance computing environment at OHSU. The code used to train and evaluate the models used in this study is publicly available at https://www.github.com/schaugf/ImageVAE.

2.2 MEMA Dataset

This study seeks to uncover the role of microenvironmental perturbations in the growth of normal human mammary epithelial cells (HMECs) by evaluating phenotypic response to 57 ligands and 47 extracellular matrix (ECM) components using the microenvironment microarray platform.[5] In this assay, ECM proteins are robotically printed into micro-well plates to form 300 μm spots upon which cells bind and grow. Additionally, soluble ligands are added to each well, thereby creating a combinatorial microenvironmental perturbation comprised of one ECM and one ligand per spot. After three days of growth, cells are fixed and stained for Keratin 19 (luminal marker in the red channel), Keratin 5 (basal marker in the green channel), and DAPI (nuclear marker in the blue channel). Input data from this study are 37,269 images of individual MEMA spots down-sampled from full-resolution (1200×1200) to 256×256 pixels. Detailed experimental description, data, and meta-data of the data-generating process are available at the MEP-LINCS Synapse wiki: https://www.synapse.org/mep_lincs.

3. RESULTS AND DISCUSSION

3.1 VAE Analysis

A VAE model was trained for 100 epochs on the 37,269 MEMA spot images evaluated in this study. Input image reconstructions shown in Figure 1 illustrate that the trained model

learns sufficient spatial features of spot organization to reconstruct an input image from 16 learned latent features. Although the reconstructions are clearly lossy, they suggest that organization, intensity, and distribution of signal within the spots is learned. Notably, despite the clear heterogeneity in the dataset, the learned reconstructions are generated from a set of 16 learned features that conform to the expected standard normal prior placed on the learning loss function. Because the prior places no constraint on relationships between learned features, correlations between learned features exist. Interestingly, both the number of cells on each spot and localized abundance of the KRT19 luminal marker, both of which are typically used to characterize spot organization, appear to associate within the learned VAE feature space, which is visualized in two dimensions with the t-SNE algorithm.[16]

Local sub-regions of the learned VAE feature space are further visualized in the two-dimensional t-SNE projection by superimposing the input images onto the t-SNE coordinates as illustrated in Figure 2. By examining the embedding space in this manner, local regions of the learned feature space appear to group MEMA spots by similar features such as shape, color, and morphology.

3.2 Latent Space Walking

To provide a qualitative assessment of the learned VAE features, we employ a latent space walking procedure that holds all but one learned feature fixed at the latent dimension's expected value (zero) while the feature of interest is swept through the inverse cumulative distribution function (CDF) of the standard normal Gaussian, as in Kingma, *et al.*[11] By passing synthetic latent feature samples through the trained decoder network, the VAE generates samples that correspond to changes in a single feature of interest while holding the rest constant. At left in Figure 3 illustrates the effect each learned VAE feature (shown in columns) has on the decoded synthetic sample by sweeping it through the CDF of the standard normal distribution (shown in rows). Although this representation can provide a qualitative assessment of each of the independent learned features, this established analysis does not consider recurring correlations between independent features. The nature of neural computing and the covariance matrix shown in Figure 1 suggest that learned features interact in complex, non-linear ways that cannot be visualized with this class of latent space walking techniques.

3.2.1 Principal Feature Manifold—To improve the interpretability of the learned latent feature space, we introduce a novel principal feature manifold (PFM) visualization approach. Our technique is based off principal component analysis (PCA), which computes a set of principal components that capture sources of significant variation within a dataset. In brief, PCA transforms an input dataset into projection matrix $\mathbf{T}$ by rotating the input data $\mathbf{X}$ by a computed weight matrix $\mathbf{W}$, which is derived from the eigen decomposition of the data's covariance matrix.

$$\mathbf{T} = \mathbf{X}\mathbf{W}$$

We leverage the dimensionality reduction properties of PCA to visualize a principal feature space. To do so, we first reduce the learned VAE feature space to the first two principal components using PCA. Because the variance of each principal component is known, we then sample a bivariate percentile distribution $\widehat{\mathbf{X}}$ that is scaled to the variance of the first and second principal components to span the sampling space we wish to visualize. We next multiply the sampled percentile grid by the inverse of the principal component matrix $\mathbf{W}^{-1}$ to rotate the uniform grid back into VAE feature space. The trained decoder network ϕ then transforms the resulting VAE space samples into synthetic input images $\mathbf{S}$ which can be visualized in two dimensions, as shown at right in Figure 3.

$$\mathbf{S} = \phi(\widehat{\mathbf{X}}) = \phi\!\left(\mathbf{TW}^{-1}\right)$$

This approach illustrates variability in the learned feature set by decoding higher-dimensional feature interactions presented in the first principal component plane. Although the information contained in a classical latent space walk and the introduced principal feature manifold are similar, the PCA formulation enables evaluation of the entire latent feature space in a simple two-dimensional image. While similar to the t-SNE space embedding shown in Figure 2, the PFM approach is uniquely capable of generating arbitrary synthetic samples using the trained decoder model.

3.3 Measuring Organization with Human Annotation

Presently, measuring organization of MEMA spots requires single-cell segmentation and feature extraction to first classify every cell on the spot as either basal or luminal based on expression of keratin markers. Spot organization is then computed as hand-crafted metric that measures relative abundance of keratin 19 (KRT19), a structural component of epithelial cells, within the central core region of the spot with respect to the outer region. Although reasonably effective in this experiment, similar types of hand-crafted features require sophisticated pre-processing steps and special knowledge of the biological phenomena under study to design effectively which profoundly limits translation of one such metric to other problems or experiments.

To evaluate how well our model characterizes spatial organization of cells in an unsupervised manner, we incorporate annotations from 7 expert biologists who graded 300 randomly selected MEMA spots as either *unorganized* (intermixed or single cell-type populations), *partially organized*, or *well organized* (centrally clustered luminal cells surrounded by basal cells). The inter-rater agreement is measured using the Fleiss kappa metric ($\kappa = 0.473$), which suggests moderate agreement between raters[17] while reflecting the inherently subjective challenge of characterizing multi-cellular organization. Downstream analyses assign the mode rating across all raters to each of the 300 scored spots as a simple majority vote decision. Associations between learned VAE features and annotated organization are illustrated in Figure 4, which suggests that certain features (particularly features 4, 7, 9, and 14) appear to exhibit shifts in their distribution with respect to organizational annotation.

To provide a fair comparison between the learned VAE space and the hand-crafted feature, we first reduce the sixteen learned VAE features into a single feature for comparison for both supervised and unsupervised settings. In this analysis, we used the first principal component (PC1) for unsupervised comparison and the first linear discriminant (LD1) for supervised comparison. The relationship between the first principal component (PCA) and first linear discriminant (LDA) of the VAE latent space, shown at left in Figure 5, illustrates that the fully unsupervised and supervised metrics are strongly associated (Pearson correlation $R = 0.9$) while neither the first linear discriminant nor principal component correlate particularly strongly with the hand-crafted organizational feature (Pearson correlation $R = 0.49$ and $R = 0.41$, respectively). However, clear class separability is evident for both the hand-crafted feature as well as the fully unsupervised characterization by the learned VAE space, shown at right in Figure 5.

To test the significance of these observations, ANOVA tests compute statistically-significant separation of the three expert annotation classes (unorganized, partially organized, well-organized) with respect to the first principal components, first linear discriminant, the hand-crafted organization feature, and the spot cell count. The resulting F values and associated p-values tabulated in Table 1 suggest that a fully unsupervised trained VAE model (F value = 717.9) improves class separability over a classically designed hand-crafted feature (F value = 254).

3.4 Characterizing Microenvironment Perturbation

This study was designed to evaluate the effect microenvironment perturbations (MEPs) have on cellular growth characteristics. If certain groups of MEPs (either ligands or extra-cellular matrix components) induce similar changes in growth morphology on the MEMA spots, and if the VAE feature space learns to capture those organizational characteristics, then similarly treated spots should be closely associated in the learned VAE feature space. This analysis first computes the mean principal latent space projection of spots treated with the same ligand-ECM combination and then performs hierarchical clustering on both ligand and ECM conditions which are shown as a heatmap in Figure 6. In addition to reflecting understanding that ligands have an overall more pronounced effect on cell spot organization than ECMs, this analysis highlights microenvironmental factors most strongly associated with multi-cellular organization characteristics. For example, this visualization associates certain ligands known to be highly associated with cellular growth and organization (TGFB, FGF2, FGF6, WNT3A, WNT10A, IL6, IL13, and BMP2). Interestingly, TGFB and BMP ligands– two closely related signaling molecules–tend to be associated with cellular organization in cancer.[18] They are also implicated in epithelial–mesenchymal transition, which is relevant to the shift in KRT markers. This observation is intriguing, as these molecules are also known to play a key role in cellular differentiation and morphogenesis. Additionally, this analysis also identifies independently observed technical artifacts in a few of the ECM conditions, which are clearly distinct as the furthest two right columns of the heatmap and shown as technical artifacts in Figure 2. Though preliminary, this type of analysis provides a rapid, unsupervised inference approach to evaluating sets of microenvironment perturbations that similarly affect cellular organization and allow prioritization of factors for more detailed experimental studies.

4. CONCLUSION

This work evaluates the role of variational autoencoding models to learn latent space representations of high-throughput imaging screens of human mammary epithelial cells in response to microenvironment perturbations. We illustrate that convolutional VAE architecture provides a powerful approach for capturing high-level features that associate with expert human annotation and hand-crafted features designed to measure cellular organization. Additionally, we introduce the Principal Feature Manifold technique designed to visualize interactions between learned VAE features beyond typical latent space walking. These analyses represent a preliminary exploration into the utility of deep learning systems to capture experimentally meaningful features of spatial organization with which to characterize tissue growth patterns in response to microenvironment perturbation. Future investigations are extending this approach towards the study of breast cancers to begin quantifying changes in growth response characteristics in competing biological contexts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

REFERENCES

[1]. Lin CH, Jokela T, Gray J, and LaBarge MA, "Combinatorial Microenvironments Impose a Continuum of Cellular Responses to a Single Pathway-Targeted Anti-cancer Compound," Cell Reports 21(2), 533–545 (2017). [PubMed: 29020637]

[2]. Junttila MR and de Sauvage FJ, "Influence of tumour micro-environment heterogeneity on therapeutic response," Nature 501, 346 (9 2013). [PubMed: 24048067]

[3]. Pegoraro G and Misteli T, "High-Throughput Imaging for the Discovery of Cellular Mechanisms of Disease," Trends in Genetics 33(9), 604–615 (2017). [PubMed: 28732598]

[4]. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, Wawer M, Paavolainen L, Herrmann MD, Rohban M, Hung J, Hennig H, Concannon J, Smith I, Clemons PA, Singh S, Rees P, Horvath P, Linington RG, and Carpenter AE, "Data-analysis strategies for image-based cell profiling," Nature Methods 14(9), 849–863 (2017). [PubMed: 28858338]

[5]. Lin C-H, Lee JK, and LaBarge MA, "Fabrication and Use of MicroEnvironment microArrays (MEArrays)," Journal of Visualized Experiments (68), 1–7 (2012).

[6]. Watson SS, Dane M, Chin K, Tatarova Z, Liu M, Liby T, Thompson W, Smith R, Nederlof M, Bucher E, Kilburn D, Whitman M, Sudar D, Mills GB, Heiser LM, Jonas O, Gray JW, and Korkola JE, "Microenvironment-Mediated Mechanisms of Resistance to HER2 Inhibitors Differ between HER2+ Breast Cancer Subtypes," Cell Systems 6(3), 329–342.e6 (2018). [PubMed: 29550255]

[7]. Goldsborough P, Pawlowski N, Caicedo JC, Singh S, and Carpenter A, "CytoGAN: Generative Modeling of Cell Images," bioRxiv (Nips), 227645 (2017).

[8]. Eulenberg P, Köhler N, Blasi T, Filby A, Carpenter AE, Rees P, Theis FJ, and Wolf FA, "Reconstructing cell cycle and disease progression using deep learning," Nature Communications 8(1), 1–6 (2017).

[9]. Christiansen EM, Yang SJ, Ando DM, Javaherian A, Skibinski G, Lipnick S, Mount E, O'Neil A, Shah K, Lee AK, Goyal P, Fedus W, Poplin R, Esteva A, Berndl M, Rubin LL, Nelson P, and Finkbeiner S, "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images," Cell 173(3), 792–795.e19 (2018). [PubMed: 29656897]

[10]. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, and Bengio Y, "Generative Adversarial Networks," arXiv preprint, 1–9 (2014).

[11]. Kingma DP and Welling M, "Auto-Encoding Variational Bayes," arXiv preprint (Ml), 1–14 (2013).

[12]. Burlingame EA, Margolin A, Gray JW, and Chang YH, "SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks," in [SPIE], 10581, 1058105–1058107 (2018).

[13]. Hu Z, Yang Z, Salakhutdinov R, and Xing EP, "On unifying deep generative models," arXiv preprint arXiv:1706.00550 (2017).

[14]. Chollet F et al., "Keras." https://github.com/fchollet/keras (2015).

[15]. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X, and Brain G, "TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning," 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), 265–284 (2016).

[16]. Van Der Maaten LJP and Hinton GE, "Visualizing high-dimensional data using t-sne," Journal of Machine Learning Research 9, 2579–2605 (2008).

[17]. Fleiss JL, "The Equivalence of Weighted Kappa and the Interclass Correlation Coefficient as Measures of Reliability," Education and Psychological Measurement 33, 613–619 (1973).

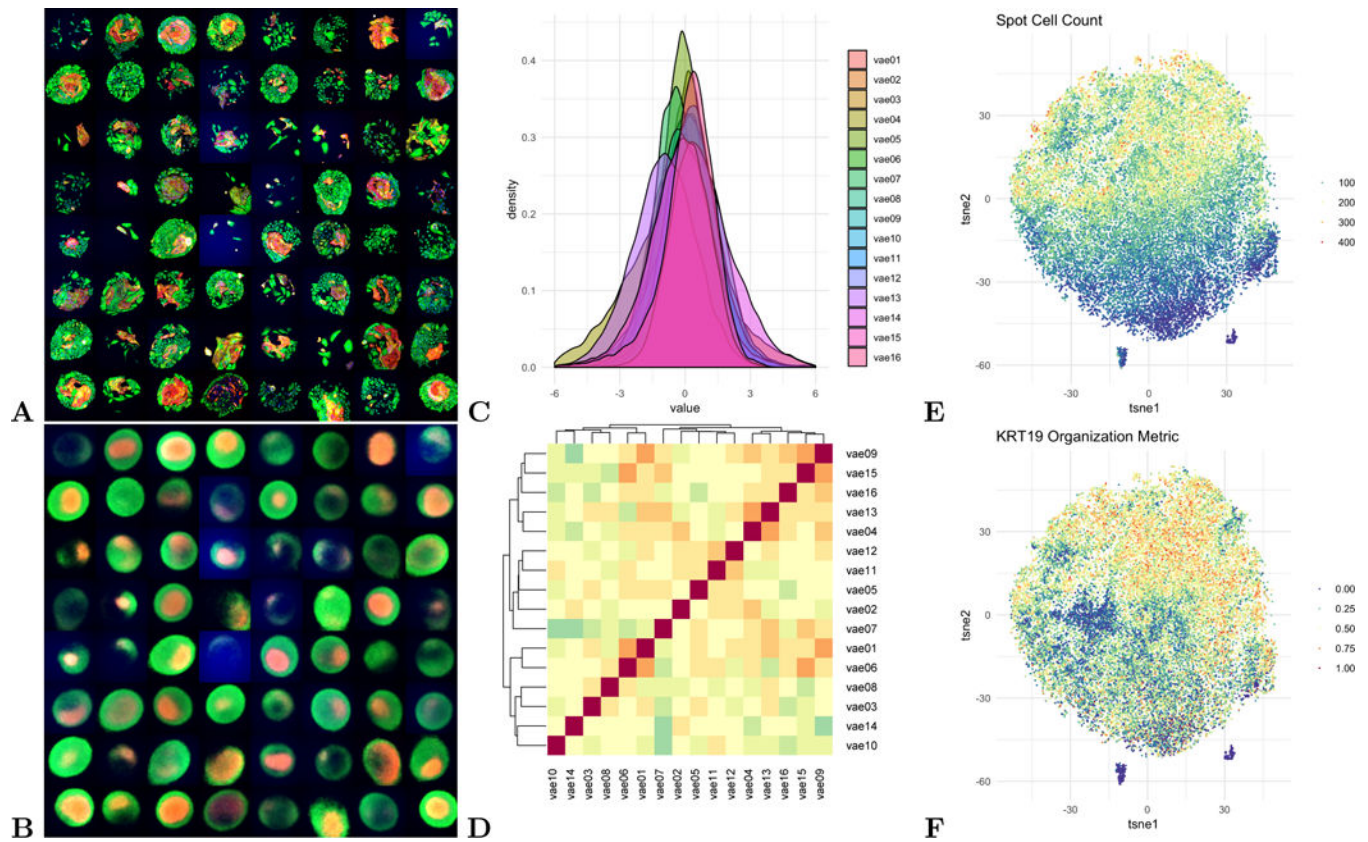[18]. Massagué J, "Tgfβ in cancer," Cell 134(2), 215–230 (2008). [PubMed: 18662538]

Figure 1.
(**A**) randomly sampled input images from the full dataset (**B**) lossy reconstructions of the sampled input images after training (**C**) distributions of each of the 16 features across the entire dataset (**D**) correlation heatmap of the learned VAE features (**E**) t-SNE projection of VAE space colored by the number of cells on each spot (**F**) t-SNE projection of VAE space colored by a hand-crafted feature designed to evaluate cell spot organization.
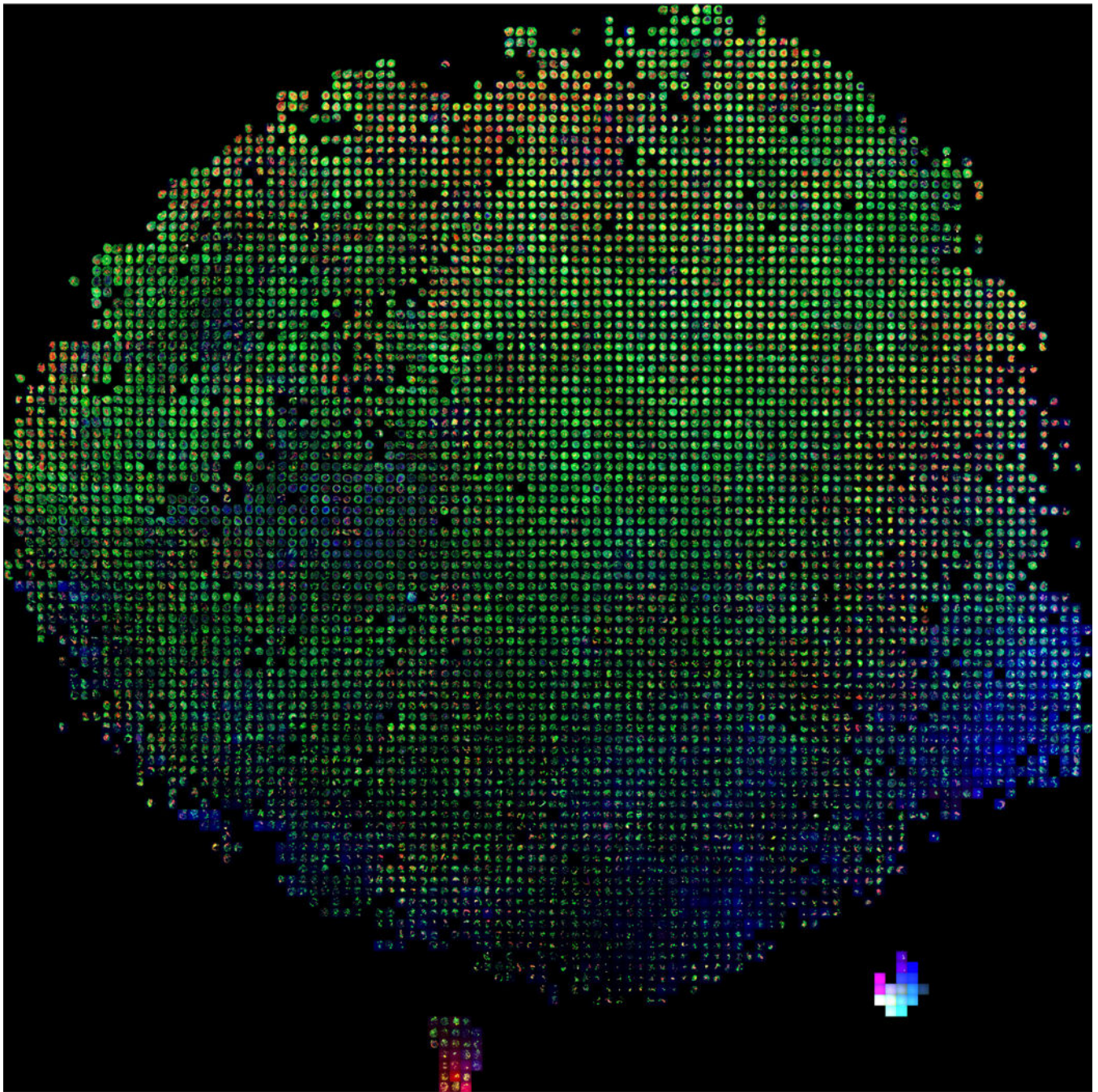
Figure 2.
Video 1: t-SNE embedding of MEMA spots used in this study based on learned latent features illustrates distinct sub-regions of feature space populated by spots of similar morphology, including a set of technical errors in the bottom right (best viewed at full 10k digital resolution). http://dx.doi.org/10.1117/12.2512660
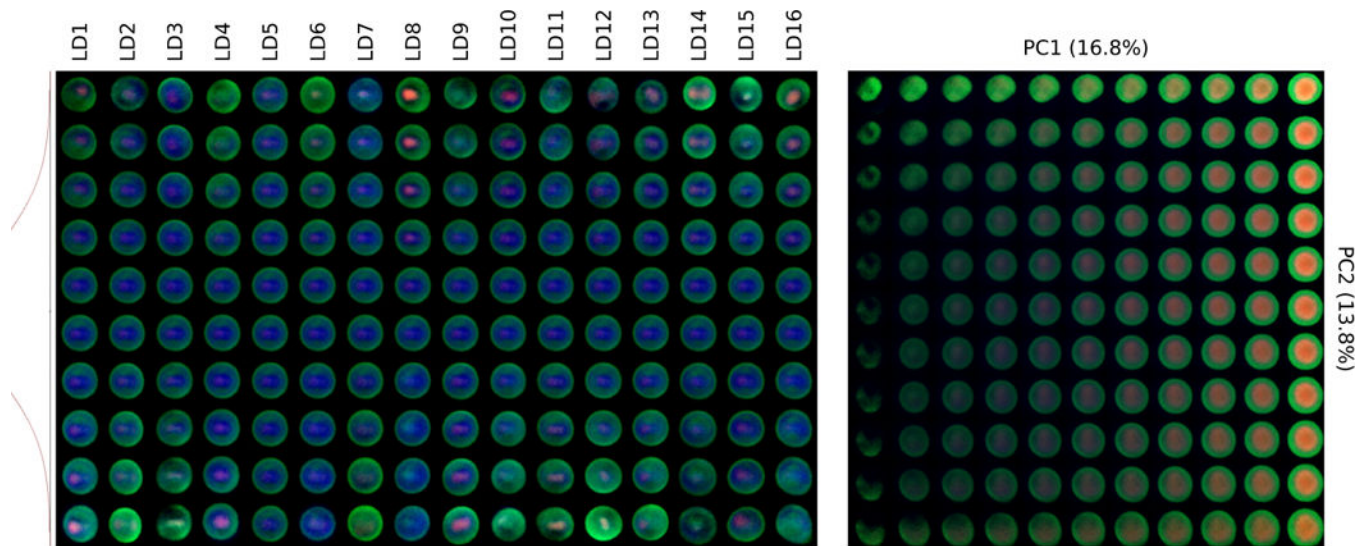
Figure 3.
(left) Latent space walking where each column represents one of 16 latent variables in the VAE model and each row represents uniformly spaced samples along the CDF of the latent variable distribution. (right) The principal feature manifold sampled from the first two principal components of the learned VAE feature space embedding visualizes sources of significant variation in the VAE encoding dataset in two dimensions. In this analysis, the first two components explain 16.8% and 13.8% variance, respectively.
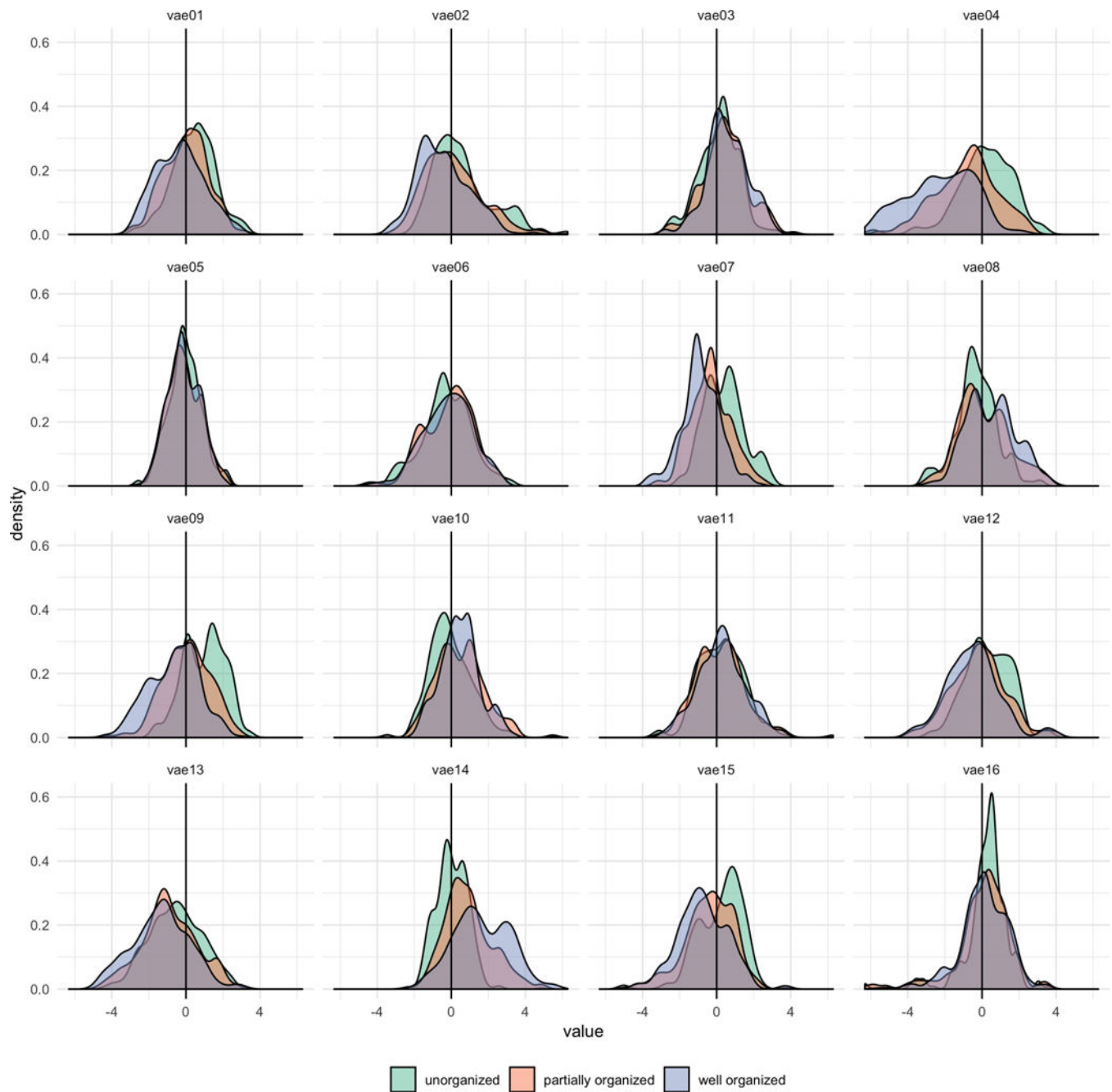
Figure 4.
Density separation of human annotation for 300 images across 16 learned latent features.
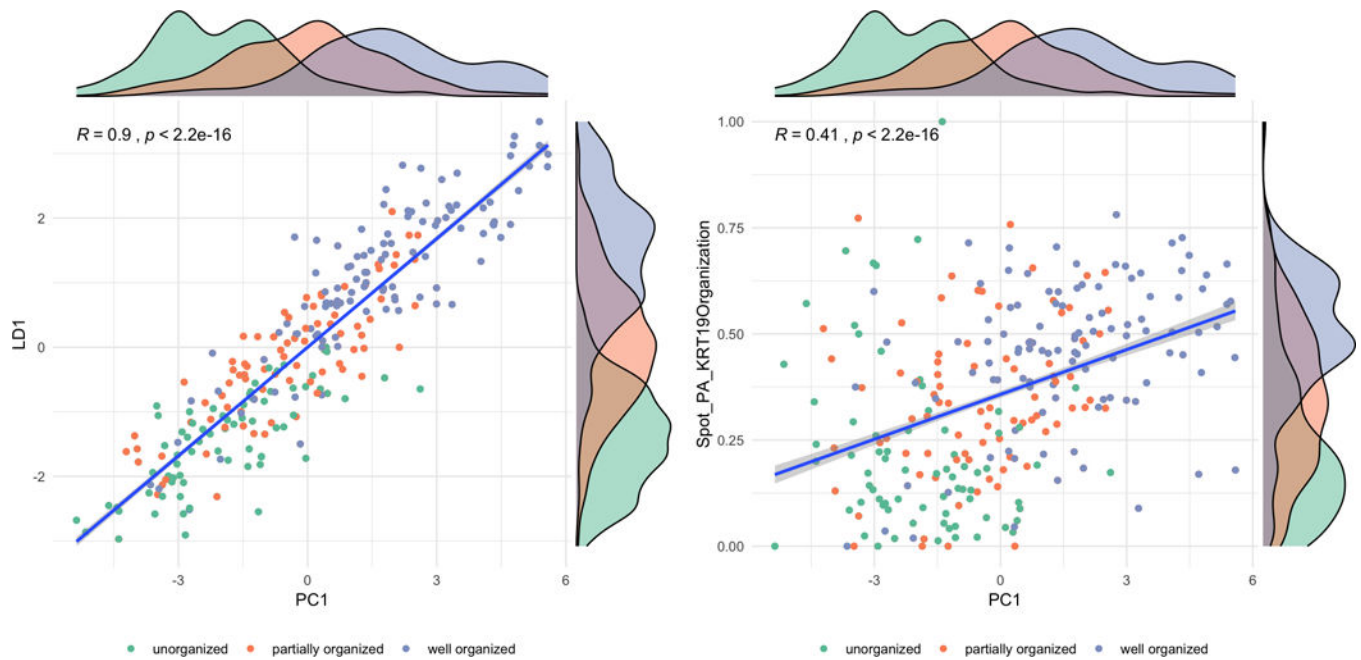
Figure 5.
(left) The first principal component (PC1) and first linear discriminant (LD1) of the latent space are tightly correlated and illustrate clear separation between annotation class. (right) Associations between the first principal component of the VAE feature space and the hand-crafted organizational feature are weak. However, ANOVA analysis suggests that the learned VAE space improves discriminatory power between the three annotated classes.
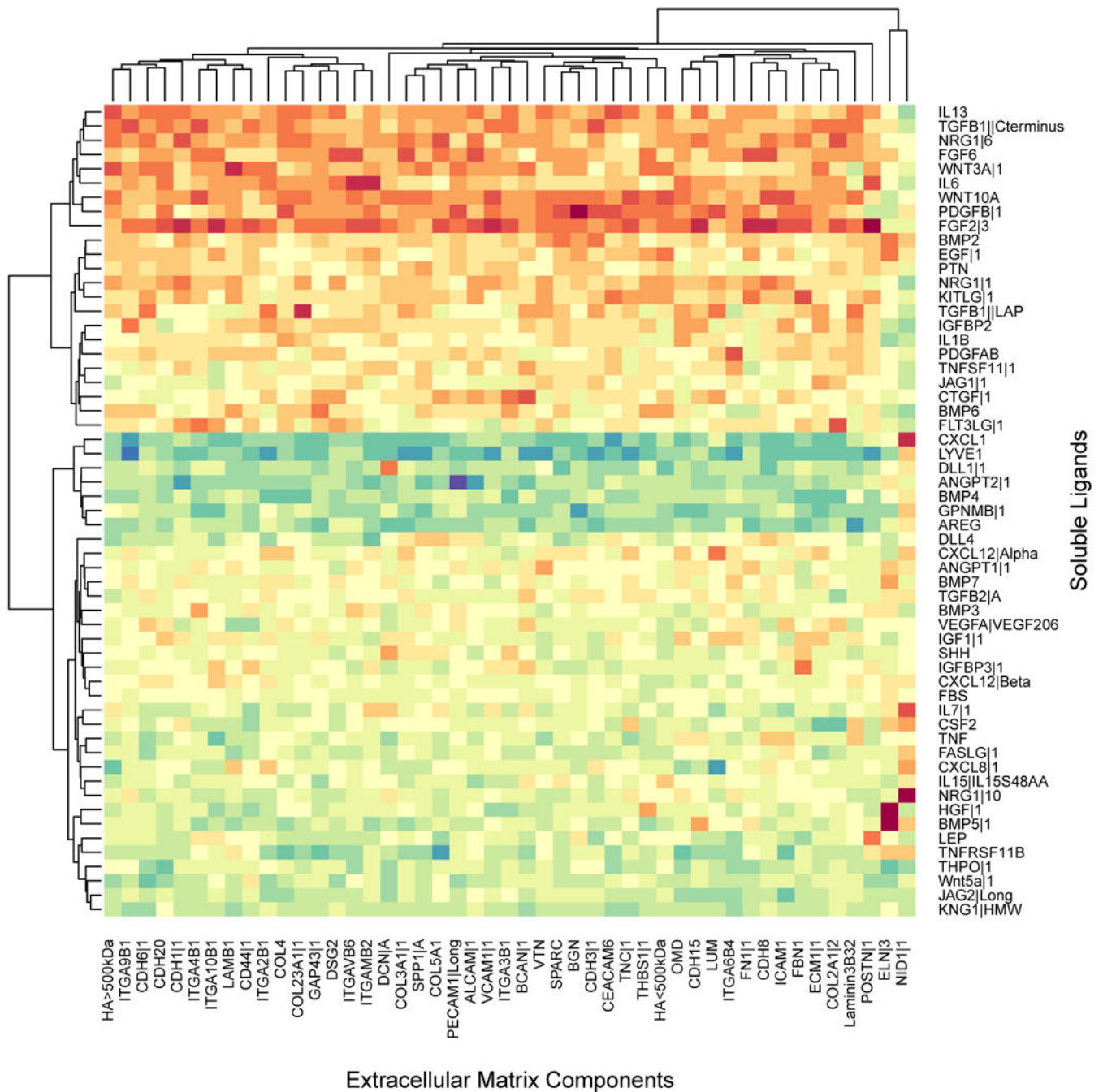
Figure 6.
Hierarchically clustered MEPs by the mean encoding of their treated MEMA spot images by extracellular matrix (x-axis) and soluble ligand (y-axis) conditions. Each square represents the mean projection of encoded images for given ligand-ECM conditions onto the first principal component of the VAE feature space. In this illustration, red colors are more highly associated with cell spot organization and blue colors are more highly associated with cell spot disorganization (see Figure 3).

Table 1.

ANOVA results of class separation

	VAE				
	PC1	PC2	LD1	HCF	Cell Count
F value	717.9	8.2	1073	254	431.5
Pr (>F)	<2e-16[*]	2.83e-4[*]	<2e-16[*]	<2e-16[*]	<2e-16[*]

[*] (VAE vs. hand-crafted feature (HCF) and Cell Count significant)

[*] significant)