Bridging Self-Supervision and Mechanism of Action Discovery in Morphological Profiling

Anonymous CVPR submission

Paper ID *****

Abstract

In the quest to interpret complex cellular responses, self-supervised learning (SSL) methods have been developed, promising powerful generic representations. However, their performance on biologically critical tasks such as mechanism of action (MoA) classification remains limited. We argue that no single model—no matter how sophisticated or generalisable—can produce representations that are optimal for all downstream tasks, as different objectives impose conflicting requirements. To address this, we propose a novel framework called **T***ask-guided* **R***epresentation* **ex***aptation* **(TRex)**. *In TRex, a generic (possibly self-supervised) model first extracts broad and rich morphological embeddings, which are then refined by a lightweight adaptation network optimised for biological relevance linked to the specific downstream tasks. This modular design enables rapid and resource-efficient transformation of generic features into biologically meaningful, task-focused representations — without the need to retrain large-scale models. We evaluate TRex on a 20-plate Cell Painting dataset spanning two cell lines and show that MoA-based adaptation not only significantly improves MoA classification performance (doubling the mAP), but also enhances compound recognition. Our results highlight the limitations of static, generic representations and demonstrate the utility of task-aware adaptation for maximising the biological relevance of morphological profiling.*

1. Introduction

Cell Painting is a high-content imaging assay designed for morphological profiling, offering a rich, multiplexed readout of cellular responses to chemical and genetic perturbations [1, 4, 9]. By staining multiple cellular compartments with a combination of fluorescent dyes, the assay provides valuable insights into cellular structure and function. However, to fully exploit the information encoded in these images, robust and biologically meaningful methods for representation learning and analysis are required. Over the years, approaches to extracting and interpreting features have evolved rapidly from handcrafted descriptors [2] to deep learning-based techniques [8], and more recently, to self-supervised learning frameworks [6, 7], reducing the need for extensive annotations and promising improved representations.

The earliest approach to feature extraction from Cell Painting images relied on CellProfiler [2], an open-source image analysis platform that extracts hand-crafted morphological features. These features, which include measurements of shape, texture, intensity, and spatial organization, are computed at the single-cell level and aggregated into well-level feature vectors. Despite wide-scale adoption, this methodology presents several limitations. It typically requires manual parameter tuning for different datasets or tasks, and it is unlikely to capture the complex phenotypic variations present in cellular images. The reliance on predefined morphological descriptors may limit the discovery of novel patterns in large-scale profiling experiments.

To address these limitations, deep learning-based approaches have been introduced to learn potentially superior image features directly from data. One such approach is DeepProfiler [8], which uses convolutional neural networks (CNNs) to extract morphological features in a data-driven manner. DeepProfiler employs EfficientNet-B0, initialized with ImageNet weights and fine-tuned on a Cell Painting dataset comprising 232 plates and 488 perturbations. Feature extraction is performed using activations from the block6a layer, generating 672-dimensional feature vectors per cell. Training is conducted using a weakly supervised learning (WSL) approach, where classification loss enables the model to learn treatment-specific representations from single-cell images. To mitigate confounding technical variation and enhance the biological relevance of extracted features, DeepProfiler applies sphering transform-based batch correction.

To further improve the scalability and generalisability of morphological profiling, and to eliminate the need for extensive annotations required for training, researchers turned

to self-supervised learning (SSL). One such method is SubCell [6], which utilizes Vision Transformers (ViTs) to extract hierarchical and spatially-aware cellular features. SubCell learns from individual cells extracted from the Human Protein Atlas (HPA) [10], a near-proteome-wide dataset of high-resolution immunofluorescence images spanning 35 different cell lines. The model optimizes a three-component loss function as the basis of feature learning: a reconstruction loss to ensure general feature extraction, a cell-specific similarity loss to enforce consistency across augmented views of the same cell, and a protein-specific localization loss to minimize variation between images stained for the same protein across different cell types. Additionally, an attention pooling mechanism suppresses background artifacts and enhances the focus on cellular features, improving robustness across diverse microscopy tasks without requiring fine-tuning.

Another SSL based approach, DINO [5], also uses ViTs to learn rich, task-agnostic representations of cellular morphology without manual annotations or supervision. Developed originally for natural images, DINO has proven highly effective for microscopy data, capturing biologically meaningful variation across subcellular, single-cell, and at population levels. The architecture uses a teacher-student framework, where both networks are trained to produce similar feature representations from different augmented views of the same image, encouraging the model to focus on the invariant, biologically relevant structures. Unlike DeepProfiler, which relies on weak supervision and convolutional architectures, DINO operates entirely self-supervised and benefits from the global contextual awareness of transformers. Compared to SubCell, which learns single-cell features using contrastive objectives over protein localization, DINO emphasizes semantic consistency across scales and has been shown to uncover hierarchical biological structure, including MoA, cell-cycle stages, and protein localization patterns. Its general-purpose embeddings make it a strong candidate for further task-specific adaptation, as explored in this work through the **T**ask-guided **R**epresentation **ex**aptation (TRex) framework.

More recently, SSL has also been explored for image-level feature extraction, eliminating the need for cell segmentation. In [7], self-supervised vision transformers were trained on a subset of the JUMP Cell Painting dataset [3] to extract morphological features directly from full-field images, bypassing segmentation-based workflows. While this segmentation-free approach significantly reduces computational costs (by skipping the cell segmentation stage), the biological relevance of the resulting features - particularly in the context of MoA prediction - remains an open question, as demonstrated by the reported MoA performance.

In [6], the authors presented a detailed evaluation of compound matching (Cmpd) and mechanism of action (MoA) prediction on the JUMP Cell Painting chemical perturbation dataset, comparing the performance of CellProfiler, DeepProfiler, DINO, and their own models. The results reveal minimal variation between methods, with notably low performance for MoA prediction across all approaches. The Cmpd mAP ranged from 0.27 (DeepProfiler) to 0.39 (SubCell), while the MoA mAP remained consistently poor, varying from 0.16 (DeepProfiler) to 0.18 (CellProfiler). Notably, hand-crafted CellProfiler features outperformed all machine learning-based methods, including SSL approaches, on the MoA task, suggesting that existing learned representations may not yet capture the biologically relevant features needed for this application. While self-supervised features promise to transfer to a wider range of tasks, their ability to effectively capture task-specific biological information remains unclear.

In this work, we address this limitation by introducing TRex, a novel framework for task-guided representation adaptation. TRex begins with a general-purpose feature extractor—preferably a pre-trained self-supervised model—and adds a lightweight adaptation network that refines these embeddings for a specific downstream task. Unlike traditional fine-tuning, our adaptation module operates on frozen features and requires minimal computation and data, enabling efficient adaptation of powerful SSL backbones.

We show that our approach yields substantial gains in both MoA classification and compound recognition, demonstrating that task-aware adaptation is key to unlocking the full potential of self-supervised representations for biological discovery. Our specific contributions are as follows:

1. We propose a novel, lightweight TRex architecture for adapting SSL-derived morphological features to biological prediction tasks, achieving substantial performance gains.
2. We design an efficient adaptation module composed of four fully connected layers with batch normalization, GELU activation, dropout regularization, and residual connections in selected layers to preserve signal flow and reduce overfitting.
3. We systematically evaluate the TRex framework and show that it can effectively double MoA classification performance over prior art, improving mAP from 0.15 to 0.32.
4. We evaluate three state-of-the art representations within TRex and show that SSL-derived features offer good performance, with DINO achieving the highest accuracy on MoA prediction tasks.
5. We demonstrate that MoA-based training leads to more biologically relevant representations, enhancing not only MoA classification but also compound recognition including generalisation to unseen compounds.

2. Introducing TRex

We propose a new two-stage architecture for efficient morphological profiling. In the first stage, self-supervised learning is used to generate a broad set of biologically meaningful features. In the second stage, a learnt task-specific feature transformation is applied to adapt this self-supervised feature space into a task-oriented representation optimised for specific downstream objective. We refer to our approach as **T**ask-Guided **R**epresentation **ex**aptation (TRex). We adopted the term *exaptation* from evolutionary biology because it aligns with how our method repurposes and adapts generic self-supervised features for specific tasks.

The motivation behind this approach stems from the fact that it is not possible for a single generic feature representation to be equally effective across all morphological profiling tasks. Take for instance, two compound perturbations, A and B, which coincide in mechanism of action (MoA). It therefore follows that their feature representations should aim to be distinct for compound identification (where the goal is to differentiate them), yet simultaneously identical for MoA classification (where their shared functional effect should be captured). Clearly, a static, task-agnostic representation cannot reconcile these conflicting requirements without additional adaptation. This limitation is evident in the poor MoA prediction performance reported in [6] across various methods, which plateau around 0.17, highlighting the need for a more tailored, task-aware approach.

In our TRex framework, shown in Figure 1, the first stage (TRex Feature Extractor) derives a comprehensive pool of generic single-cell embeddings, most likely via a self-supervised approach or even by aggregating multiple representations from complementary methods. The second stage (TRex Adaptation Module) applies a lightweight transformation network to adapt these features into a task-specific final representation. Since the first stage is computationally intensive, it is trained only once. In contrast, the second-stage adaptation is fast, efficient, and can be performed even with limited resources. This makes our approach not only biologically meaningful - by aligning representations with task-specific biological distinctions such as compound mechanisms of action or phenotypic similarity— but also computationally practical, allowing for rapid adaptation to new tasks and datasets.

Let:

- $f_{\text{SSL}} : X \to Z$ be a **self-supervised feature extractor**, where X represents the input 5-channel microscopy images, and $Z \in \mathbb{R}^d$ is the high-dimensional feature representation learned via self-supervised pretraining.
- $t_{\text{TASK}} : Z \to Y$ be the **task-guided feature transformation network**, where Y is the final representation and $TASK$ is the task at hand, for example the perturbation or MoA classification.

For example, in MoA classification task, the goal of our second-stage network is to learn a transformation t_{MoA} such that the extracted features Z are mapped to an optimized, lower-dimensional space Y that retains biologically meaningful information linked to MoA, thus improving MoA classification accuracy. Unlike end-to-end learning, our framework operates on precomputed feature embeddings, significantly reducing computational complexity and data requirements.

2.1. TRex design details

The input features are extracted in the stage one model (TRex Feature Extractor), which, in our case, is from one of the three pre-trained models: DeepProfiler, DINO, or Subcell. The key properties of these models are summarised in Table 1. The DeepProfiler model [8], based on EfficientNet-B0, was trained in a weakly supervised manner on cellular images from the BBBC and LINCS datasets, generating a 672-dimensional embedding per cell. The DINO method [5], a vision transformer (ViT), was trained in a self-supervised manner on the same datasets and produces a 768-dimensional embedding per cell. Both models were trained on 224×224 images containing five fluorescence channels (DNA, ER, RNA, Golgi/Actin, and Mitochondria). The SubCell model, also based on ViT, was trained in a self-supervised manner on the Human Protein Atlas (HPA) dataset, which consists of three-channel images (ER, DNA, Protein). The model generates a 4608-dimensional embedding per cell. The HPA training set included 35 cell lines, including both U2OS and A549. All three models have seen U2OS and A549 cells during training, but SubCell was additionally trained on 33 other cell lines, which in theory may enable it to produce more generalisable features across diverse cellular contexts.

Method	Dataset	Training	Cell Types	Dim.
DeepProfiler	L+B	Weak-sup.	U2OS,A549	672
DINO	L+B	Self-sup.	U2OS,A549	768
SubCell	HPA	Self-sup.	Various	4608

Table 1. Summary of feature extraction methods tested in our framework. **L+B** denotes a combination of LINCS and BBBC datasets.

To make these diverse and generic embeddings task-relevant, we employ the TRex adaptation network, which transforms the output of each feature extractor into a biologically meaningful and task-optimised representation suitable for downstream prediction. This network is lightweight by design, enabling efficient adaptation regardless of the upstream embedding source. It consists of four fully connected layers, each followed by batch normalization, GELU activation, and dropout regularization to ensure stable gradient propagation, improved non-linearity, and reduced over-

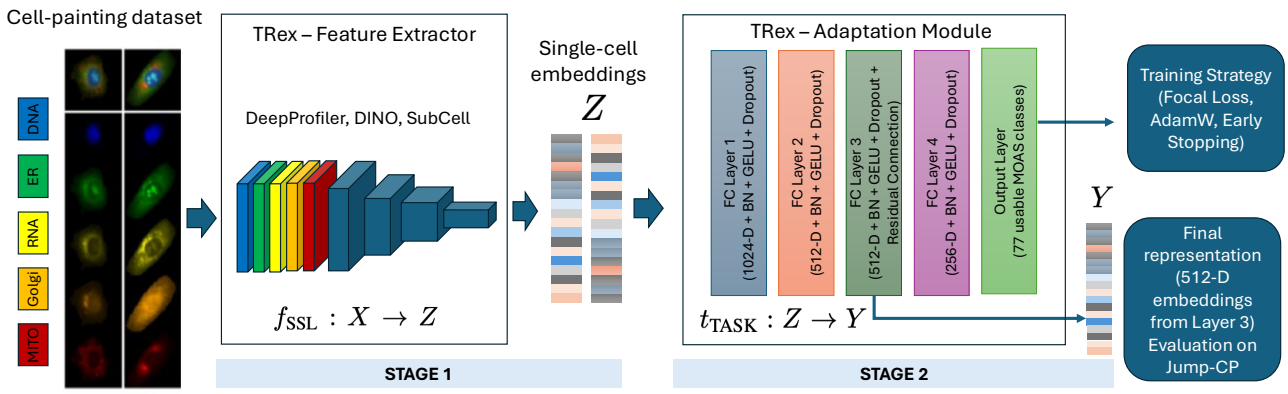

Figure 1. Overview of the TRex framework. Single-cell embeddings are extracted from five-channel Cell Painting images using pre-trained DeepProfiler, DINO, or SubCell models (STAGE 1). These embeddings are passed to the TRex adaptation module, which is trained using focal loss for the mechanism of action (MoA) classification task (STAGE 2). During evaluation/inference, features are extracted from the adaptation module's third layer and used for two tasks: (1) replicate retrieval — identifying wells treated with the same compound across experimental batches, and (2) MoA identification — detecting compounds that share the same mechanism of action.

fitting.

The first layer of the proposed model processes the embeddings from one of these pre-trained models through a fully connected transformation (1024-dimensional output), stabilizing activations with batch normalization, introducing non-linearity with GELU, and applying dropout for regularization. The second layer projects the transformed features into a 512-dimensional latent space, retaining the same activation and normalization mechanisms. To enhance gradient flow and preserve learned representations, a residual connection is incorporated in the third layer, allowing the input to be directly added to the output before further transformation. The fourth layer (256-dimensional output) refines the feature representation before passing it to the final classification layer, which is optimised for a specific downstream task. In our experiments, this final layer was trained either for compound classification (a single-label task) or for Mechanism of Action (MoA) prediction, formulated as a multi-label classification problem to account for cases where a single cell may be associated with multiple MoAs. More broadly, the TRex framework is compatible with other downstream objectives, depending on the structure of the final classification head and the chosen optimisation criterion.

2.2. Key Design Principles of TRex

The TRex framework is built around several core principles that distinguish it from prior approaches to morphological profiling:

1. Feature Refinement Rather than Direct End-to-End Training – Rather than training directly from raw images, TRex operates on high-quality embeddings produced by large self-supervised models. This approach preserves expressive morphological information while reducing redundancy and computational cost..
2. Task-Specific Optimisation – The adaptation module is explicitly trained using a task-specific loss (e.g., MoA classification), ensuring that the learned representation is aligned with biologically meaningful distinctions rather than generic morphological similarity.
3. Multi-Label Adaptability – TRex naturally supports multi-label classification, enabling the representation of complex phenotypes such as compounds with multiple mechanisms of action — a common scenario in biological data.
4. Computational Efficiency and Modularity – By freezing the upstream feature extractor and training only the lightweight adaptation module, TRex achieves efficient task adaptation with minimal labelled data and compute resources. In contrast to full SSL training, which requires multi-GPU server infrastructure and days of compute time, the TRex adaptation module can be trained in under an hour on a single consumer-grade GPU (e.g., NVIDIA RTX 3090). This makes it practical for scaling across new tasks, cell lines, or experimental conditions.

3. Experimental Evaluation

3.1. Datasets

Our training and evaluation were performed on a selection of 20 plates from the JUMP-CP Pilot dataset, representing two distinct human cell lines: U2OS, which consists of human bone osteosarcoma epithelial cells, and A549, which consists of human alveolar basal epithelial cells. To ensure a representative distribution of morphological variations, separate subsets of plates were designated

for training and testing for each cell line. For the A549 cell line, the training set included plates BR00116991, BR00116992, BR00117050, BR00117052, BR00117054, and BR00117055, while the testing set comprised plates BR00116993, BR00116994, BR00117008, BR00117009, BR00117051, and BR00117053. For the U2OS cell line, the training set consisted of plates BR00116995, BR00117013, and BR00117025, while the test set included plates BR00117011, BR00117012, BR00117024, and BR00117026.

There were 303 different perturbations present, representing 77 unique usable modes of action. We followed the MoA definitions and evaluation methodology defined in [6].

The images of selected plates were pre-processed and segmented into individual cells using DeepProfiler. Default parameter values were used for segmentation to maintain generability and consistency with prior studies. Following segmentation, the data set consisted of a total of 1,079,388 cells for training and 1,144,682 cells for testing in the U2OS cell line, and 3,179,429 cells for training and 2,626,695 cells for testing in the A549 cell line. These single-cell images were subsequently used for feature extraction and downstream performance analyses in our study.

3.2. Learning Framework

To address class imbalance in the dataset, the model is trained using focal loss, which dynamically adjusts the loss contribution to emphasize hard-to-classify samples. The training process runs for 100 epochs, utilizing the AdamW optimizer with an initial learning rate of 1e-3, which is reduced by a factor of 0.5 if the validation mAP does not improve for five consecutive epochs. Early stopping is applied if no improvement is observed for 10 consecutive epochs, ensuring efficient convergence while preventing overfitting.

For evaluation of the JUMP-CP test dataset, we extract the output from layer 3 of our model, generating a 512-dimensional embedding that serves as the learned feature representation. This representation supports two key downstream tasks: replicate retrieval, which identifies wells treated with the same compound across experimental batches, and MoA identification, which detects compounds that share the same mechanism of action (MoA). The architecture design of the proposed method is presented in Figure 1.

To generate well-level profiles, single-cell embeddings from our model are first aggregated into Field of View (FOV) profiles using mean aggregation, followed by a second mean aggregation step to obtain well-level representations. Post-processing includes Principal Component Analysis (PCA) for dimensionality reduction, followed by Median Absolute Deviation (MAD)-robust standardization to enhance data robustness.

Well-level profiles were used to evaluate replicate retrieval. For mechanism of action (MoA) prediction, well-level profiles from the same compound treatment were averaged across plates to generate a consensus profile for each compound. Mean average precision (mAP) was calculated using cosine similarity between well-level and consensus profiles, for replicate retrieval and MoA identification, respectively.

3.3. Experimental Evaluation

We performed a series of experiments to evaluate the effectiveness of the proposed TRex framework. First, we assessed the impact of the training objective by comparing the models trained for compound classification versus the mechanism of action classification. Next, we applied TRex to three different base representations: one obtained via weak supervision (DeepProfiler) and two derived from distinct self-supervised learning strategies (DINO, SubCell). This analysis provides insight into whether self-supervised representations may be more effective for downstream biological tasks. Finally, we investigated the generalizability of the method by training on datasets restricted to a single cell type and evaluating its performance on data from an unseen cell type.

Training on compound recognition. In the first set of experiments, we trained TRex on the task of compound recognition (Table 2). Performance was evaluated on both compound recognition and MoA classification tasks, using datasets for individual cell lines as well as the combined dataset. We note that TRex offers significant improvements in compound recognition across all representations, providing gains of $+0.09$ for A549 and $+0.05$ for U2OS. The gain for the combined dataset is 0.07. We observe that self-supervised representations offer some limited performance gains over DeepProfiler, and these gains are maintained when using TRex. The best performing combination, TRex + SubCell, achieved compound recognition of 0.34, compared to 0.29 for the SubCell representation alone. We also note that training on compound recognition results in only marginal improvements for MoA classification, with gains between 0.01 and 0.02. The best MoA result is 0.17, which may not be sufficient for many applications. This suggests that compound-based training does not promote biologically relevant representations and is suboptimal for MoA tasks, and the mAP for MoA remains disappointingly low.

Training with the MoA objective. In contrast, training with the MoA objective leads to a substantial improvement in MoA classification performance, effectively more than doubling the mAP score from 0.15 to 0.32 for TRex + DINO (Table 3). We note that DINO offered the best base representation for distillation by TRex, outperforming TRex + SubCell by 0.05 on the combined cell dataset. Interestingly, this approach also results in a notable improvement

Table 2. Compound recognition (Cmpd) and MoA classification performance (mAP) for models trained with compound supervision on both A549 and U2OS cell lines.

Method	Cmpd A549	MoA A549	Cmpd U2OS	MoA U2OS	Cmpd Both	MoA Both
DeepProfiler	0.35	0.15	0.24	0.15	0.22	0.14
DINO	0.38	0.16	0.27	0.16	0.26	0.15
SubCell	0.36	0.15	0.29	0.16	0.27	0.14
TRex + DeepProfiler	0.44	0.15	0.30	0.16	0.29	0.16
TRex + DINO	**0.47**	**0.18**	0.32	0.16	0.33	**0.17**
TRex + SubCell	0.45	0.17	**0.34**	**0.17**	**0.34**	**0.17**

Table 3. Compound recognition (Cmpd) and MoA classification performance (mAP) for models trained with MoA supervision on both A549 and U2OS cell lines.

Method	Cmpd A549	MoA A549	Cmpd U2OS	MoA U2OS	Cmpd Both	MoA Both
DeepProfiler	0.35	0.15	0.24	0.15	0.22	0.14
DINO	0.38	0.16	0.27	0.16	0.26	0.15
SubCell	0.36	0.15	0.29	0.16	0.27	0.14
TRex + DeepProfiler	0.42	0.27	0.28	0.25	0.29	0.26
TRex + DINO	**0.44**	**0.34**	0.30	**0.28**	0.30	**0.32**
TRex + SubCell	0.42	0.28	**0.32**	0.25	**0.31**	0.27

in compound classification performance for TRex across all representations. The improvement was most significant for DeepProfiler (0.07), followed by $+0.04$ for the self-supervised representations. These findings suggest that training with an MoA objective using the proposed multi-label formulation provides a more generalizable solution, as it enhances both MoA classification and compound recognition. This indicates that when tackling tasks related to MoA or compound classification, MoA-driven training should be preferred, as it leads to greater performance gains in both objectives.

Generalisation across compounds and cell lines. To evaluate the generalisation properties of the TRex framework, we conducted experiments in two settings: generalisation to unseen chemical compounds and generalisation to an unseen cell line. Training with the MoA objective provides a natural starting point, as 111 compounds are not used in TRex MoA training due to the lack of usable MoA annotations. Table 4 shows the mAP for the compounds that were unseen during the adaptation stage. As shown, TRex improves performance even on these previously unseen compounds, increasing mAP from 0.24 to 0.28.

Table 4. Compound recognition performance (mAP) for compounds unseen during TRex adaptation, using MoA-based training

Method	Cmpd (unseen)
DINO	0.24
TRex + DINO	0.28

For the second setting, we trained the adaptation module with MoA objective using data from only one of the two available cell lines (A549 or U2OS), and evaluated performance on both (Table 5). While TRex provides a clear performance boost on the cell line it was trained on, we observe no significant improvement on the unseen cell line. This outcome is not surprising: since training was performed on a single cell line, the adaptation module had no exposure to examples from other cellular contexts and therefore no basis for learning cell-line-invariant MoA features. In other words, we would not expect generalisation to emerge from a single-cell-line training setup. TRex is intended to specialise general-purpose features for a specific downstream task and cellular context, optimising performance where it is needed most.

Table 5. Compound recognition (Cmpd) and MoA performance (mAP) for models trained on a single cell line and evaluated on both seen and unseen cell lines. TRex was trained with MoA supervision.

Method	Train Cell	Eval Cell	Cmpd	MoA
DINO	–	A549	0.38	0.16
TRex + DINO	A549	A549	0.44	0.36
TRex + DINO	U2OS	A549	0.36	0.16
DINO	–	U2OS	0.27	0.16
TRex + DINO	U2OS	U2OS	0.30	0.31
TRex + DINO	A549	U2OS	0.25	0.17

In future work, we plan to investigate whether training TRex on multiple cell lines can support better generalisation to unseen cellular contexts by encouraging the adaptation module to learn cell-line-invariant biological features.

4. Conclusions

In this work, we addressed the limitations of self-supervised learning (SSL) for morphological profiling by introducing a two-stage learning framework that refines generic SSL features for task-specific objectives. While SSL provides scalable feature extraction, our results confirm that its representations are suboptimal for important tasks such as mechanism of action (MoA) classification. This is because a single, task-agnostic model cannot effectively satisfy conflicting requirements across different end tasks.

By introducing a lightweight transformation network that adapts SSL features to specific predictive tasks, TRex enables efficient, task-aware optimization without requiring the retraining of large-scale SSL models. Our experiments on a dataset of 20 Cell Painting plates from two cell lines demonstrate that training for MoA prediction in a multi-label setting significantly improves both MoA classification and compound recognition, highlighting the benefits of learning task-specific refinements over purely static SSL representations.

Because TRex is computationally efficient and can be trained with fewer labeled examples, it provides a practical and scalable solution for large-scale drug discovery applications. More broadly, our findings suggest that self-supervised feature representations should not be treated as fixed but rather as a foundation for adaptive transformations that better align with specific biological objectives. Future work could explore extending TRex to other high-content imaging assays and integrating additional domain-specific constraints to further improve biological interpretability.

References

[1] Mark-Anthony Bray, Shantanu Singh, Han Han, Chadwick T Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M Gustafsdottir, Christopher C Gibson, and Anne E Carpenter. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature Protocols*, 11(9):1757–1774, 2016. 1

[2] Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Jason H. Chang, Ruth A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10):R100, 2006. 1

[3] Srinivas Niranj Chandrasekaran, Beth A Cimini, Amy Goodale, Lisa Miller, Maria Kost-Alimova, Nasim Jamali, John G Doench, Briana Fritchman, Adam Skepner, Michelle Melanson, Alexandr A Kalinin, John Arevalo, Marzieh Haghighi, Juan C Caicedo, Daniel Kuhn, Desiree Hernandez, James Berstler, Hamdah Shafqat-Abbasi, David E Root, Susanne E Swalley, Sakshi Garg, Shantanu Singh, and Anne E Carpenter. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nature Methods*, 21(6):1114–1121, 2024. 2

[4] Beth A Cimini, Srinivas Niranj Chandrasekaran, Maria Kost-Alimova, Lisa Miller, Amy Goodale, Briana Fritchman, Patrick Byrne, Sakshi Garg, Nasim Jamali, David J Logan, John B Concannon, Charles-Hugues Lardeau, Elizabeth Mouchet, Shantanu Singh, Hamdah Shafqat Abbasi, Peter Aspesi, Justin D Boyd, Tamara Gilbert, David Gnutt, Santosh Hariharan, Desiree Hernandez, Gisela Hormel, Karolina Juhani, Michelle Melanson, Lewis H Mervin, Tiziana Monteverde, James E Pilling, Adam Skepner, Susanne E Swalley, Anita Vrcic, Erin Weisbart, Guy Williams, Shan Yu, Bolek Zapiec, and Anne E Carpenter. Optimizing the cell painting assay for image-based profiling. *Nature Protocols*, 18(7): 1981–2013, 2023. 1

[5] Michal Doron, Thibault Moutakanni, Zihan S. Chen, et al. Unbiased single-cell morphology with self-supervised vision transformers. *bioRxiv*, 2023. Preprint. 2, 3

[6] Ankit Gupta, Zoe Wefers, Konstantin Kahnert, Jan N. Hansen, Will Leineweber, Anthony Cesnik, Dan Lu, Ulrika Axelsson, Frederic Ballllosera Navarro, Theofanis Karaletsos, and Emma Lundberg. Subcell: Vision foundation models for microscopy capture single-cell biology. *bioRxiv*, 2024. 1, 2, 3, 5

[7] Vladislav Kim, Nikolaos Adaloglou, Marc Osterland, Flavio M. Morelli, Marah Halawa, Tim König, David Gnutt, and Paula A. Marin Zapata. Self-supervision advances morphological profiling by unlocking powerful image representations. *Scientific Reports*, 15(1):4876, 2025. 1, 2

[8] Nikita Moshkov, Malte Bornholdt, Sylvain Benoit, et al. Learning representations for image-based profiling of perturbations. *Nature Communications*, 15(1):1594, 2024. 1, 3

[9] Srijit Seal, Maria-Anna Trapotsi, Ola Spjuth, Shantanu Singh, Jordi Carreras-Puigvert, Nigel Greene, Andreas Bender, and Anne E. Carpenter. Cell painting: a decade of discovery and innovation in cellular imaging. *Nature Methods*, 22(2):254, 2025. 1

[10] Mathias Uhlén, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M. Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, and Fredrik Pontén. Tissue-based map of the human proteome. *Science*, 347(6220):1260419, 2015. 2