SHOT-CCR: Biologically Guided Adversarial Training for Test-Time Adaptation in Cellular Morphology

William Dee [1,2], Aaron Wenteler[1], Srijit Seal[3,4], Otto Morris[2] and Gregory Slabaugh[1]

[1] Digital Environment Research Institute, Queen Mary University of London
[2] Recursion Pharmaceuticals
[3] Broad Institute of MIT and Harvard
[4] University of Cambridge
w.t.dee@qmul.ac.uk

Abstract. Pervasive batch effects are a common issue, especially in recent large-scale Cell Painting datasets, which have been produced to aid AI-enhanced drug discovery efforts. Technical differences arising from experiments carried out in different batches can cause models to fail to generalize to unseen batches, despite good predictive performance "within batch". We propose a biologically grounded test-time adaptation framework, SHOT-CCR, which uses cell-invariant gradient reversal to decouple morphological signal from experimental confounders. Our approach performs 4.5% better than the current RxRx1 benchmark, classifying 1,139 classes of siRNA genetic perturbations with 91.6% accuracy. We deliver consistent results over four distinct cell types and two prominent Cell Painting datasets – RxRx1 and a subset of JUMP-CP. Across 484 classes of CRISPR perturbations in JUMP-CP our method improves accuracy by 15.7%. All code will be made publicly available upon publication.

Keywords: Domain Adaptation, Microscopy, Batch Effect Correction.

1 Introduction

Over the past decade the availability of high-content screening (HCS) data has grown dramatically [1, 2, 3, 4]. High-throughput microscopy screening systems are now an important component within drug discovery, enabling biologists to image the responses of millions of cells to thousands of different chemical and genetic perturbations [5, 6]. This capability is relatively cheap when compared with mRNA or protein profiling and contributes distinct information regarding a cell's phenotypic response [7, 8]. Computer vision pipelines have been leveraged for a variety of cellular perturbation tasks, including discerning molecular mechanism of action [8, 9, 10, 11], compound toxicity [12, 13], gene function and disease mechanisms [7].

Multiple studies have documented how technical artifacts and other confounding batch effects can arise even under the most stringent experimental settings [8, 9, 10, 13]. To compound this, larger datasets necessitate a greater number of separate experimental batches. These batch effects typically obscure biological cellular processes, causing models to fail to generalize to out-of-batch samples.

Table 1. Summary of the datasets used in this work. Genetic perturbations include knockdown (KD) or knockout (KO).

Dataset	Cell Type	Pert. type	# Images	# Train / Test Batches	# Perturbations
RxRx1	U2OS, HepG2, HUVEC, RPE	Gene KD (siRNA)	125,510	33 / 18	1,139
JUMP-CP (Subset)	U2OS	Gene KO (CRISPR)	21,780	4 / 1	484

Sypetkowski et al. [1] showed that, in the large and diverse Cell Painting dataset RxRx1, adaptive batch normalization (AdaBN) can be an effective strategy to mitigate the impact of batch effects across experimental batches for certain cell types. Their model achieved a classification accuracy of 87.1% across 1,139 unique siRNA perturbations, with the model being able to generalize with 95.6% effectiveness to unseen batches. However, the approach underperformed for the U2OS cell type, which accounted for less than 10% of the training data, classifying those perturbations with 68.2% accuracy.

To improve these results, we extend domain adaptation techniques, originally developed for natural image analysis, to the problem of experimental batch variation in Cell Painting data. We incorporate recent advances such as entropy minimization, diversity regularization, and pseudo-labeling, and combine them with adversarial training designed to de-prioritize the cell count differences which commonly arise across experimental batches and cell types. Crucially our method, SHOT Cell Count Reversal (SHOT-CCR), does not aim to eliminate all batch information from the encoder embeddings. Instead, it selectively down-weights cell-count–driven batch differences while retaining residual batch-specific structure that remains predictive of the underlying perturbation. We focus on cell count in this work due to the fundamental biological nature of the feature and how integral it has proven to be in other Cell Painting research [2, 8, 11, 13]. Our main contributions are as follows:

1. Biologically informed test-time adaptation. We propose a framework that extends test-time adaptation techniques from computer vision to Cell Painting data, resulting in robust generalization under unseen batch and cell type shifts.
2. Cell count adversarial training. We introduce a novel biologically guided adversarial mechanism that reduces batch effects by discouraging the network from over-relying on cell count features - outperforming the general batch-effect gradient reversal of [1] when there are multiple cell types present.
3. Comprehensive evaluation across datasets and cell types. We demonstrate consistent gains across two large-scale Cell Painting datasets (RxRx1 and JUMP-CP) and four distinct cell types (Tab. 1), establishing a new benchmark for morphological batch correction.

2 Related Work

Cell Painting batch effect correction. The reduction of batch effect impact is a central theme in most Cell Painting microscopy research [1, 2, 3, 4, 5, 8, 27]. Image embeddings and other downstream representations, such as CellProfiler image-based profiles [28], have been shown to group tightly together according to these effects, making classification tasks difficult. There is a growing desire to amalgamate microscopy datasets from different sources (e.g. labs) or time periods, with the belief that more diverse data will result in robust models which generalize more effectively to unstudied cell types, drug compounds and genetic perturbations, facilitating more scalable and efficient drug discovery [2, 3, 4, 5].

Test-time adaptation (TTA) refers to methods that enable models to adapt to new, unseen data distributions during inference without requiring model re-training or access to the original training data [14]. There are several publicly available large-scale Cell Painting datasets [1, 2, 3, 4, 15] and released models pre-trained on those datasets [1, 5, 11], making it increasingly necessary to explore TTA methods that allow previously trained models to adapt to new experimental batch data with potentially different cell types. TTA approaches range from recalculating or updating batch normalization statistics at test time [20, 21], to adding pseudo-class labelling or prototypes based on predictive class confidence [23]. Adaptation can occur through a series of episodes, a single-pass epoch, or, as in our work, iterative multi-epoch adaptation using self-supervised objectives.

Gradient reversal is a technique for learning domain-invariant representations that retain relevant information for the goal task while minimizing the impact of domain-related variables, achieved by multiplying gradients from a domain-specific head by a negative factor [24, 25]. Our work utilizes gradient reversal to selectively reduce, but not eliminate, the impact of cell count differences across experimental batches. We demonstrate that adding gradient reversal learning parameters can improve model accuracy during training and improve subsequent TTA performance. While recent work has also applied domain-adversarial learning to Cell Painting batch correction using a generative model framework [27]; in contrast, SHOT-CCR requires no model retraining at test time and no access to source data during adaptation, making it directly applicable to scenarios where previously trained models must generalize to new experimental batches.

3 Datasets

3.1 RxRx1

RxRx1 was produced by Recursion using their high-throughput screening platform [1]. In each experimental well, cells are genetically perturbed with a fixed concentration of small interfering RNA (siRNA) to knockdown a single target gene [1, 16]. A proprietary implementation of the Cell Painting protocol [17] produces a six-channel 512×512-

pixel image, where each channel corresponds to a fluorescent dye representing specific cellular components, including DNA, RNA, mitochondria and Golgi apparatus (Fig. 1).

RxRx1 contains 51 experimental batches across four cell types — HUVEC (24 batches), RPE (11), HepG2 (11) and U2OS (5) - each conducted a minimum of one week apart. U2OS was considered the most biologically distinct and therefore most difficult to predict, particularly in a limited data scenario [1]. We use the same batch-separated splits as [1]: 33 training batches (16 HUVEC, 7 RPE, 7 HepG2, 3 U2OS) and 18 test batches (8 HUVEC, 4 RPE, 4 HepG2, 2 U2OS).

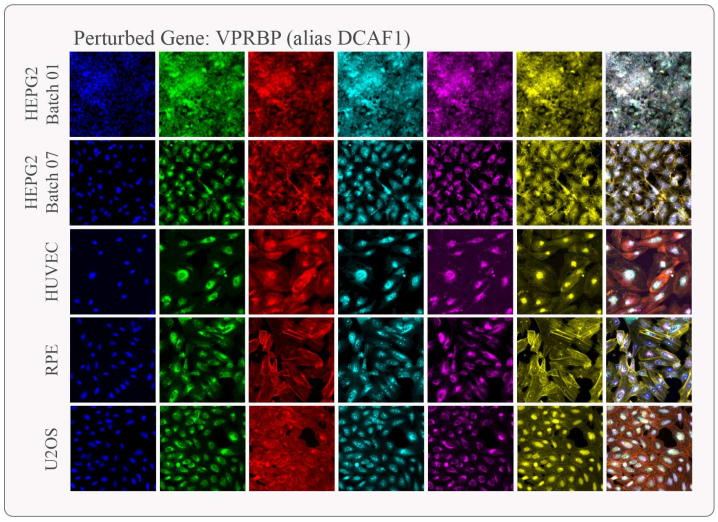

Fig. 1. Six-channel faux-colored images from the RxRx1 dataset with a composite image shown on the right. Different stains reveal a range of cellular components across each channel (columns). Each set of images displays the *same* perturbation class (the VPRBP gene knockdown) but imaged in different batches and/or different cell types. The variety of cellular responses to the *same* perturbation, which can be seen by eye, show how difficult it can be for a model to adapt to both biological and technical batch effects.

3.2 JUMP cpg0016 (JUMP-CP)

We filtered the JUMP-CP dataset [3, 4] to include only CRISPR knockout perturbations present in more than one batch with at least five replicates across five separate plates, ensuring sufficient data per perturbation. For comparability with [1] we retained only genes also present in RxRx1, resulting in 484 unique perturbations across five experimental batches (Tab. 1). Images were centrally cropped to 512×512 pixels. Only the U2OS cell type is present in JUMP-CP, and the original Cell Painting protocol produces five-channel images rather than six [17], though the stains target functionally similar cellular components to RxRx1. Due to the smaller diversity of data compared with RxRx1 (Tab. 1), and the known difficulty of predicting CRISPR knockouts versus siRNA knockdowns [18], we expect models to perform more poorly on this data.

4 Methodology

4.1 Model architecture

An overview of our model architecture is shown in Fig. 2. The architecture is based on the DenseNet-161 backbone proposed in the original RxRx1 paper [1]. The backbone is adapted for Cell Painting images by altering the first convolutional layer to accept either five (JUMP-CP) or six-channel (RxRx1) images as input, before a Kaiming normal initialization [29] is used to initialize the additional channel weights. Global adaptive average pooling is applied to the backbone feature maps before passing through two fully connected layers, each utilizing batch normalization and the ReLU activation function.

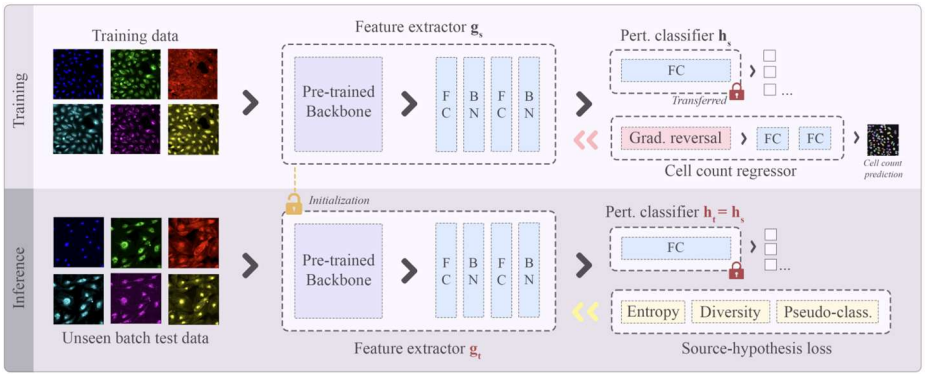

Figure 2: Our SHOT-CCR architecture. During training, the feature extractor (g_s) feeds both a perturbation classifier (h_s) and a cell count regression head, which applies gradient reversal to adversarially reduce cell count alignment in the feature maps. At test time, the classifier is frozen while the feature extractor adapts using unsupervised SHOT loss.

4.2 Cell-count gradient reversal (CCR)

In contrast to [1], we use a gradient reversal layer to adversarially train the model to de-prioritize cell count. A separate regression head, consisting of two linear layers, is used for this task and the gradient reversal layer feeds back the inverse of the loss to the feature extractor (Fig. 2). Since cell count is trivial for the model to learn to predict, we introduce separate learning rate and gradient reversal alpha (α) parameters to avoid both over fitting as well as training the model to be completely agnostic to cell count.

Cell counts were obtained by feeding the nuclei-stained channel through CellposeSAM (cellpose v4.0.1) [19] and normalized using training set statistics. The regression head minimizes MSE between predicted and actual normalized cell counts during training; cell counts are not used at inference or test-time adaptation to avoid data leakage.

4.3 Test-time adaptation

Adaptive batch normalization (AdaBN). AdaBN [20] utilizes batch normalization layers to standardize samples only using other samples produced in the same domain. In practice, this means that during training a sampler is used to ensure that each mini batch only contains data from a single domain (an experimental batch in our case) – referred to as batch effects normalization (BEN [30]). At inference time, the running statistics accumulated on the training set are discarded, and instead normalization is performed according to test mini batch statistics (which are also sampled from a single domain).

Test entropy minimization (Tent). Tent [21] adapts a pretrained model solely at inference time by minimizing the Shannon entropy [22] of model predictions, gradually increasing prediction confidence throughout adaptation. Unlike AdaBN, Tent updates both the scale (γ) and shift (β) parameters of the batch normalization layers through entropy backpropagation, remaining fully unsupervised since it relies only on model predictions rather than labels.

Source hypothesis transfer with information maximization (SHOT). SHOT [23] extends Tent by freezing the pretrained classifier (the hypothesis) and adapting only the feature extractor during test time. Adaptation is driven by an information maximization loss combining entropy minimization and a diversity term to ensure predictions remain globally distributed across classes. Additionally, samples exceeding a confidence threshold τ receive a pseudo-label, and the resulting classification loss is incorporated into the objective. In our work, we freeze the perturbation classifier after initial training (Fig. 2) and update the feature extractor using this unsupervised SHOT loss during test-time adaptation.

For test-time adaptation we use a learning rate of $1e^{-5}$, the Adam optimizer and a confidence threshold τ of 0.95. We train over multiple epochs, using an early stopping criterion of 15 epochs and selecting the model where SHOT loss reaches a minimum.

4.4 Evaluation

We apply average perturbation class classification accuracy, predicting the genetic perturbation present in each image. We evaluate results using batch-separated splits for both RxRx1 and JUMP-CP where each test batch has not been seen during training. This therefore mimics the scenario whereby new experimental data has been generated for an adjacent task, so a technique like test-time adaptation would allow our previously trained model to adapt to the new batch(es).

For RxRx1, we use the same training and test splits as the original paper (see Section 3.1), also repeating each run five times with different seeds to show model performance variance. For JUMP-CP, since there are only five batches, we repeated model training five times, each time using a different batch as the held-out test set. This approach is necessitated by the low number of individual batches in the JUMP-CP subset compared with RxRx1.

The same baseline models as [1] are used to ensure fair comparability. Results significance was assessed via one-sample t-test against the reported AdaBN baseline mean

for RxRx1, as individual run results were unavailable from [1]. However, pairwise comparisons between methods *within* this paper were conducted using Welch's t-test.

5 Results and experiments

5.1 Reducing batch effects

The results in Tab. 2 show a consistent improvement in perturbation classification accuracy when employing the architectural and test-time adaptation additions discussed in Sec. 4.1-4.3. Using Tent to minimize entropy loss at inference time while only changing the scale and shift parameters of the model increases test set accuracy for RxRx1 by 2% compared with the prior state-of-the-art baseline set by [1]. When diversity and pseudo-label classification losses of SHOT are included, the model improves by 2.9% over the baseline, before our cell count gradient reversal adds an additional 1.6%. Overall, this significantly improved accuracy by 4.5% over the original RxRx1 AdaBN benchmark ($t=63.64$, $p<0.0001$). The 1.6% improvement gain from the addition of CCR to SHOT was also assessed as statistically significant ($t=9.74$, $p<0.0001$).

As expected, the smaller and less diverse JUMP CRISPR dataset was classified with much lower accuracy overall. However, in this scenario, the test-time adaptation methods had a much larger impact on model performance. AdaBN improved classification accuracy by 17.6% over the baseline, while our approach utilizing SHOT and cell count gradient reversal improved classification by a further 15.7%. For this dataset, both gradient reversal methods (batch and CCR) only provide marginal improvement over vanilla SHOT. Fig. 3 contextualizes these results. Unlike RxRx1, the JUMP-CP subset contains only a single cell type across batches with broadly consistent cell count distributions (Fig. 3b), meaning cell count provides little discriminative batch signal for CCR to suppress. This is consistent with our hypothesis that CCR's benefit scales with cell count heterogeneity across batches. Crucially, the datasets where this heterogeneity is greatest - multi-cell-type datasets like RxRx1 - are also the most clinically relevant, as generalizing across cell types is a prerequisite for scalable drug discovery pipelines [1, 2, 3, 11].

Table 2. Perturbation classification accuracy (%) comparing different approaches to batch correction and test-time adaptation across two datasets – RxRx1 and JUMP-CP.

Method	**RxRx1**	**JUMP-CP (Subset)**
	Perturbation classification accuracy	
Baseline [1]	75.1% ± 0.2%	10.4% ± 8.3%
AdaBN [1]	87.1% ± 0.2%	28.0% ± 7.5%
AdaBN + Batch gradient reversal [1]	86.2% ± 0.3%	32.4% ± 7.2%
Ours (Tent [21])	89.1% ± 0.2%	35.0% ± 4.9%
Ours (SHOT [23])	90.0% ± 0.3%	43.4% ± 6.4%
Ours (SHOT + Batch gradient reversal)	90.1% ± 0.2%	43.2% ± 5.9%
Ours (SHOT-CCR)	**91.6%** ± 0.2%	**43.7%** ± 5.6%

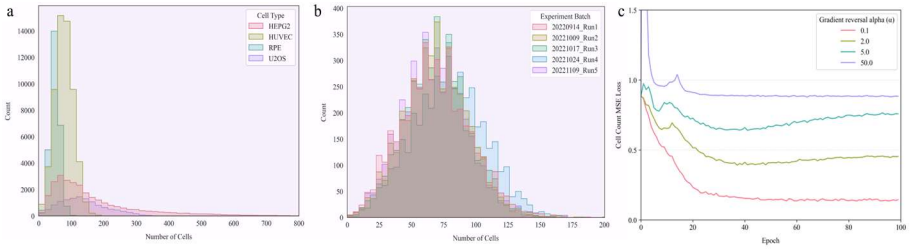

Figure 3. a) cell count distribution variation across cell types in RxRx1. b) relative uniformity of cell count distribution across the batches in the JUMP-CP subset. c) cell count regression head MSE loss throughout training under different gradient reversal α parameters.

5.2 Consistency across cell types

Our results show consistent improvement across all cell types in RxRx1 (Tab. 3). Encouragingly, performance gain was the strongest in the U2OS cell type (+8.0% accuracy), the cell type with the fewest training batches (3). This was the cell type established as most difficult to predict by [1]. This result suggests that TTA methods, guided by cell-count gradient reversal, are relatively agnostic to cell type, and can be most impactful where there is limited cell-type-specific training data.

Table 3. Perturbation classification accuracy (%) per cell type. Numbers in parentheses indicate training/test batch counts.

Method	HUVEC (16/8)	RPE (7/4)	HepG2 (7/4)	U2OS (3/2)
Baseline [1]	84.2% ± 0.2%	79.0% ± 0.4%	76.2% ± 0.1%	26.1% ± 1.5%
AdaBN [1]	92.1% ± 0.2%	87.2% ± 0.0%	86.2% ± 0.2%	68.2% ± 0.1%
Ours (SHOT-CCR)	**95.1%** ± 0.1%	**91.9%** ± 0.1%	**92.0%** ± 0.2%	**76.2%** ± 0.3%

5.3 Gradient reversal

While prior research found gradient reversal to be ineffective when trying to adversarially reduce the batch effect signal by predicting what experimental batch each image was produced in [1], here we show that gradient reversal can be effective if it's focused on a biological prior – cell count. Consistent with this, CCR showed no significant improvement over vanilla SHOT in JUMP-CP, where cell count distributions are homogeneous across batches (Fig. 3b), but delivered meaningful gains in RxRx1 where cell count varies substantially across cell types (Fig. 3a). We found that separate gradient reversal parameters ($\alpha = 5.0$ and regression head learning rate = $1e^{-5}$) were critical for model performance, whereby classification accuracy was dependent on reversal strength. Fig.3 shows how the cell count head regressor MSE loss is impacted by α. The notion of a trade-off between performance and invariance is supported by prior work, showing partial invariance is often the optimal solution [24, 25]. By controlling

how much cell-count related information is retained by the embeddings we strike a balance between removing batch-specific noise while preserving task-relevant information.

6 Conclusion

We have demonstrated that adversarial training targeting a biologically motivated confounder, cell count, combined with test-time adaptation methods, can effectively dampen batch effects in Cell Painting data and improve perturbation classification accuracy across multiple cell types and datasets, outperforming the previous benchmark of [1]. We set out the framework by which this approach can be applied to other Cell Painting data in future. Where cell count distributions vary substantially across experimental batches, particularly in multi-cell-type datasets, we encourage researchers to incorporate this as an explicit confounder in their modeling approach and to account for it when interpreting results. Finally, we introduce a curated JUMP-CP subset of 484 CRISPR perturbations as an additional public benchmark for the field and demonstrate that our approach transfers effectively to this distinct dataset and perturbation type.

Acknowledgments. This work was supported by the UKRI/BBSRC Collaborative Training Partnership in AI for Drug Discovery, led by Recursion Pharmaceuticals in partnership with Queen Mary University of London. We are extremely grateful to Maciej Sypetkowski for his guidance and help throughout this project, especially when replicating the original results from RxRx1. The authors would like to thank the Broad Institute and the JUMP Consortium for making the JUMP-CP Cell Painting data freely available and accessible. We would also like to thank Safiye Celik and Inês Sequeira for their valuable feedback on the draft manuscript. This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/Y009800/1], through funding from Responsible Ai UK (KP0016). This work acknowledges the support of the National Institute for Health Research Barts Biomedical Research Centre (NIHR203330). This research utilized Queen Mary's Apocrita and Andrena HPC facilities, supported by QMUL Research-IT. http://doi.org/10.5281/zenodo.438045. The Collaborative Training Partnership was funded by the Biotechnology and Biological Sciences Research Council, grant reference BB/X511791/1.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Sypetkowski, M., Rezanejad, M., Saberian, S., Kraus, O., Urbanik, J., Taylor, J., Mabey, B., Victors, M., Yosinski, J., Rezazadeh Sereshkeh, A., et al.: RxRx1: A dataset for evaluating experimental batch correction methods. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4284–4293 (2023)
2. Chandrasekaran, S.N., Ceulemans, H., Boyd, J.D., Carpenter, A.E.: Image-based profiling for drug discovery: due for a machine-learning upgrade? Nature Reviews Drug Discovery 20(2), 145–159 (2021)
3. Chandrasekaran, S.N., Cimini, B.A., Goodale, A., Miller, L., Kost-Alimova, M., Jamali, N., Doench, J.G., Fritchman, B., Skepner, A., Melanson, M., et al.: Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. bioRxiv (2022)
4. Chandrasekaran, S.N., Ackerman, J., Alix, E., Ando, D.M., Arevalo, J., Bennion, M., Boisseau, N., Borowa, A., Boyd, J.D., Brino, L., et al.: JUMP cell painting dataset: morphological impact of 136,000 chemical and genetic perturbations. bioRxiv (2023)
5. Kraus, O., Kenyon-Dean, K., Saberian, S., Fallah, M., McLean, P., Leung, J., Sharma, V., Khan, A., Balakrishnan, J., Celik, S., Beaini, D., Sypetkowski, M., Cheng, C.V., Morse, K., Makes, M., Mabey, B., Earnshaw, B.: Masked autoencoders for microscopy are scalable learners of cellular biology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11757–11768 (2024)
6. Fraietta, I., Gasparri, F.: The development of high-content screening (HCS) technology and its importance to drug discovery. Expert Opinion on Drug Discovery 11(5), 501–514 (2016)
7. Arevalo, J., Su, E., Ewald, J.D., van Dijk, R., Carpenter, A.E., Singh, S.: Evaluating batch correction methods for image-based cell profiling. Nature Communications 15(1), 6516 (2024)
8. Dee, W., Sequeira, I., Lobley, A., Slabaugh, G.: Cell-vision fusion: A Swin transformer-based approach for predicting kinase inhibitor mechanism of action from Cell Painting data. iScience 27(8), 110511 (2024)
9. Ljosa, V., Caie, P.D., ter Horst, R., Sokolnicki, K.L., Jenkins, E.L., Daya, S., Roberts, M.E., Jones, T.R., Singh, S., Genovesio, A., Clemons, P.A., Carragher, N.O., Carpenter, A.E.: Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. Journal of Biomolecular Screening 18(10), 1321–1329 (2013)
10. Palma, A., Theis, F.J., Lotfollahi, M.: Predicting cell morphological responses to perturbations using generative modeling. Nature Communications 16(1), 505 (2025)
11. Seal, S., Trapotsi, M.A., Spjuth, O., Singh, S., Carreras-Puigvert, J., Greene, N., Bender, A., Carpenter, A.E.: Cell Painting: A decade of discovery and innovation in cellular imaging. Nature Methods 22(2), 254–268 (2025)
12. Seal, S., Mahale, M., García-Ortegón, M., Joshi, C.K., Hosseini-Gerami, L., Beatson, A., Greenig, M., Shekhar, M., Patra, A., Weis, C., et al.: Machine learning for toxicity prediction using chemical structures: Pillars for success in the real world. Chemical Research in Toxicology 38(5), 759–807 (2025)
13. Seal, S., Dee, W., Shah, A., Cerisier, N., Zhang, A., Miglietta, E., Titterton, K., Cabrera, Á.A., Boiko, D., Beatson, A., Slabaugh, G., Taboureau, O., Carreras Puigvert, J., Singh, S., Spjuth, O., Bender, A., Carpenter, A.E.: Counting cells can accurately predict small-molecule bioactivity benchmarks. Nature Communications (2026).
14. Liang, J., He, R., Tan, T.: A comprehensive survey on test-time adaptation under distribution shifts. International Journal of Computer Vision 133(1), 31–64 (2025)

15. Fay, M.M., Kraus, O., Victors, M., Arumugam, L., Vuggumudi, K., Urbanik, J., Hansen, K., Celik, S., Cernek, N., Jagannathan, G., et al.: RxRx3: Phenomics map of biology. bioRxiv (2023)
16. Tuschl, T.: RNA interference and small interfering RNAs. ChemBioChem 2(4), 239–245 (2001)
17. Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., Carpenter, A.E.: Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nature Protocols 11(9), 1757–1774 (2016)
18. Lazar, N.H., Celik, S., Chen, L., Fay, M.M., Irish, J.C., Jensen, J., Tillinghast, C.A., Urbanik, J., Bone, W.P., Gibson, C.C., Haque, I.S.: High-resolution genome-wide mapping of chromosome-arm-scale truncations induced by CRISPR–Cas9 editing. Nature Genetics 56(7), 1482–1493 (2024)
19. Pachitariu, M., Rariden, M., Stringer, C.: Cellpose-SAM: superhuman generalization for cellular segmentation. bioRxiv (2025)
20. Li, Y., Wang, N., Shi, J., Hou, X., Liu, J.: Adaptive batch normalization for practical domain adaptation. Pattern Recognition 80, 109–117 (2018)
21. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: Proceedings of the International Conference on Learning Representations (2021)
22. Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal 27(3), 379–423 (1948)
23. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In: Proceedings of the International Conference on Machine Learning, pp. 6028–6039 (2020)
24. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F.: A theory of learning from different domains. Machine Learning 79(1–2), 151–175 (2010)
25. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. In: Domain Adaptation in Computer Vision Applications, pp. 189–209 (2017)
26. Szaflarski, W., Leśniczak-Staszak, M., Sowiński, M., Ojha, S., Aulas, A., Dave, D., Malla, S., Anderson, P., Ivanov, P., Lyons, S.M.: Early rRNA processing is a stress-dependent regulatory event whose inhibition maintains nucleolar integrity. Nucleic Acids Research 50(2), 1033–1051 (2022)
27. Yan, C., Zhang, Y., Feng, J., Hua, H., Ruan, Z., Li, Z., Li, S., Yan, C., Li, P., Liu, J., Chen, S.: Triple-effect correction for Cell Painting data with contrastive and domain-adversarial learning. Nature Communications 16(1), 6886 (2025)
28. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., Sabatini, D.M.: CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biology 7(10), R100 (2006)
29. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
30. Lin, A., Lu, A.: Incorporating knowledge of plates in batch normalization improves generalization of deep learning for microscopy images. In: Proceedings of the 17th Machine Learning in Computational Biology Meeting, pp. 74-93. PMLR (2022)