ActiveSeq pairs same-well imaging and transcriptomics in a Z-Screen microchip to recover compound-resolved phenotypes

Abstract

High-throughput phenotypic discovery needs readouts that connect chemical identity to cell state without forcing a choice between scalable imaging and deeper molecular measurement. We evaluated ActiveSeq, the same-well image-plus-RNA arm of Z-Screen, in which cells in a one-bead-one-compound micro-well are imaged and then sequenced. Across two HEK293 pilots (ACT010 and ACT011), we analyzed 20,222 matched wells: 11,511 wells from a 35-compound control panel and 8,711 ZEL026 dummy-bead wells used only as negative-control QC. The image arm used an early 32-dimensional latent derived from generic image features and calibrated on this control-panel domain, so it tests a first assay representation rather than the full potential of raw multi-channel or target-aware imaging. Even with this early representation, both modalities recovered compound identity within and across runs: split-half top-1 retrieval reached 98.8% / 80.1% for RNA and 94.2% / 86.8% for image, and cross-run centroid transfer reached 71.4% for RNA and 85.7% for image. Image and RNA compound maps were aligned but non-identical, and merged image-plus-RNA features improved held-out classification to 55.5% balanced accuracy versus 41.4% for RNA and 29.2% for image. Same-compound shuffled fusion matched true same-well fusion on this task, while exploratory analyses detected modest above-random within-compound pairing structure. Low-pass RNA remained interpretable through compound-by-run pseudobulk signatures. These results establish ActiveSeq as a scalable same-micro-well multimodal readout linking morphology, RNA state, and compound provenance.

Significance

Most multimodal perturbation resources pair image and RNA profiles by treatment label across separate cell populations. ActiveSeq instead measures both readouts from the same small cell population in the same Z-Screen micro-well. In this pilot, a generic image-derived latent and low-pass RNA profile each recover compound-resolved information, and their merged representation improves held-out compound classification across a 35-compound control panel. The present image latent is an early representation, not the endpoint of the imaging strategy. The larger capability is that morphology, assay state, target-aware image channels, RNA state, and compound provenance can be linked by well in a scalable one-bead-one-compound screen.

Introduction

Phenotypic drug discovery is most useful when a screen can see both breadth and depth. Breadth is needed to explore many chemical structures; depth is needed to understand whether a response is reproducible, interpretable, and worth following. High-content imaging and Cell Painting make large phenotypic screens practical by measuring morphology and other visual features at high throughput [1]. Single-cell transcriptomic perturbation screens measure cell-state programs more directly, but usually at greater cost per perturbation and lower compound concurrency [3]. Recent public multimodal resources such as scGeneScope connect these readouts by profiling matched treatments across replicate populations, pairing Cell Painting and single-cell RNA sequencing profiles for the same treatment labels in U2-OS cells [4].

Z-Screen is organized around a different experimental unit: a high-density one-bead-one-compound microchip. ActiveSeq is the Z-Screen arm that reads image and RNA from the same micro-well. A small local cell population is imaged, the same well is then sequenced, and the resulting readouts can be linked to the well and its compound assignment. This design matters because it lets image-derived morphology and RNA state be interpreted as paired measurements from the same physical assay unit rather than as treatment-matched averages from separate wells.

The image arm should be understood positively but precisely. The present study does not ask whether a generic image embedding is the final imaging solution for Z-Screen. It asks whether an early image-derived representation, paired with low-pass RNA from the same micro-well, already carries reproducible compound information. Generic morphology can report assay state, colony state, quality-control structure, and low-cost triage signal. Future target-aware fluorescent or photoreactive image channels can be selected to report biology closer to the question being asked, while the same well still contributes RNA for molecular interpretation.

We tested ActiveSeq in two HEK293 pilot experiments, ACT010 and ACT011. The dataset contains 20,222 matched wells with both image and RNA measurements. The compound-response analyses focus on 11,511 matched wells assigned to a 35-compound control panel. A separate set of 8,711 matched wells labeled ZEL026 is used only as a dummy-bead negative-control QC condition; those wells support read-depth and detected-gene checks and are excluded from compound-response claims.

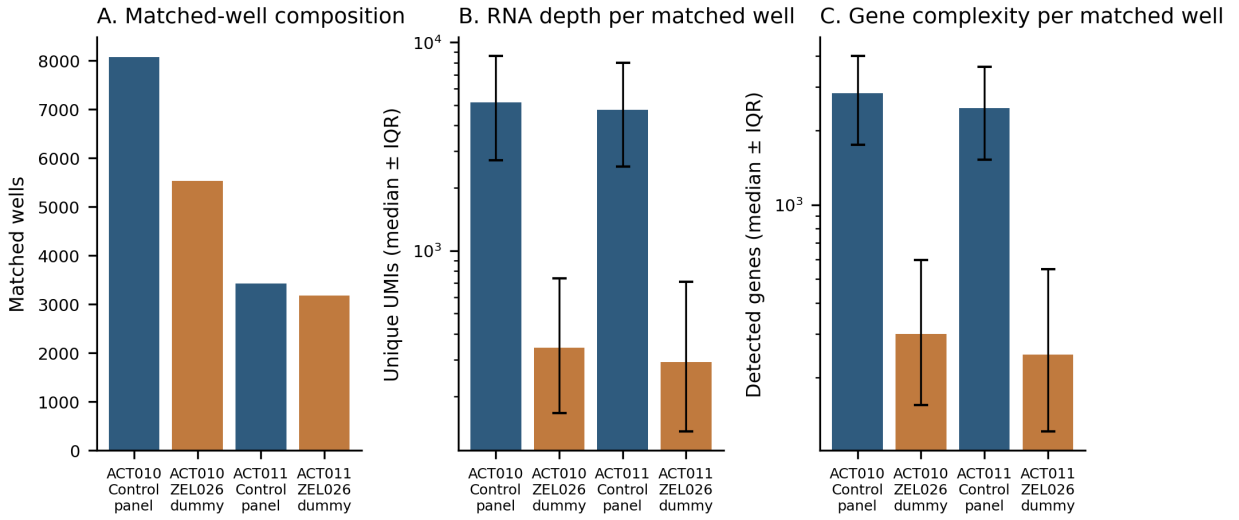
The analysis follows the decision path a scalable same-well workflow must pass. First, are the paired wells analyzable and is the image representation defined? Second, does each readout recover compound identity within and across experiments? Third, do image and RNA organize compounds in related but non-identical ways? Fourth, does merging the same-well readouts improve held-out compound classification? Fifth, how much exact well-to-well cross-modal structure is detectable after compound identity is held constant? Finally, does the low-pass RNA branch remain interpretable enough for triage?

Results

Are the same-well ActiveSeq readouts analyzable?

ActiveSeq produced 20,222 image-plus-RNA matched wells across ACT010 and ACT011. Of these, 11,511 wells were assigned to the 35-compound control panel used for compound-response analyses. The remaining 8,711 ZEL026 dummy-bead wells were used only as negative-control QC stress-test wells.

The QC contrast behaved as expected. Relative to control-panel wells, ZEL026 dummy-bead wells showed sharply reduced RNA recovery: median unique molecular identifiers (UMIs) decreased from 5,167 to 342 in ACT010 and from 4,750 to 293 in ACT011. Median detected genes decreased from 2,826 to 301 in ACT010 and from 2,463 to 248 in ACT011. This separation confirms that the same-well RNA branch detects the intended low-recovery QC condition while keeping ZEL026 outside compound-response interpretation.



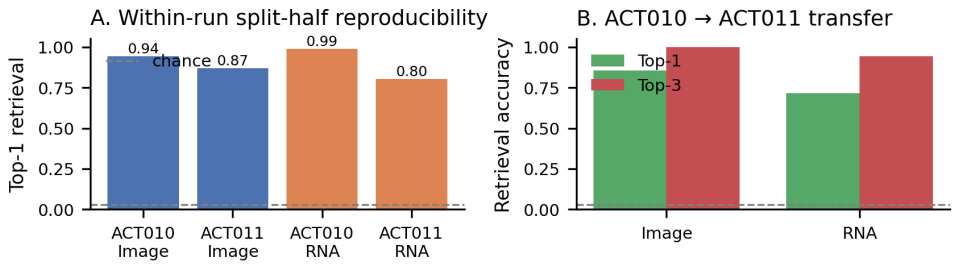
Supplementary Figure 1. QC overview for matched control-panel wells and ZEL026 dummy-bead negative-control QC stress-test wells in ACT010 and ACT011. ZEL026 wells are used only to show low RNA recovery in the QC stress-test condition and are excluded from compound-response analyses.

For the image branch, each well had an embedding derived from a frozen generic vision model. We analyzed a 32-dimensional calibrated image latent learned on the matched control-panel domain. The calibration objective retained compound-discriminating image information while reducing experiment-of-origin signal. This makes the image latent an early internal assay representation: it is useful for asking whether same-well imaging contributes compound-resolved information in this pilot, while leaving broader-library, cell-type, and imaging-channel generalization for larger studies.

Does each readout recover compound identity within and across experiments?

We first asked whether image and RNA independently encode reproducible compound information. Within each experiment, wells for each control compound were split into pseudo-replicates, and one half was used to retrieve the other among the 35 compounds. RNA latent space was highly stable, with split-half top-1 retrieval of 98.8% in ACT010 and 80.1% in ACT011. The calibrated image latent was also strong, with split-half top-1 retrieval of 94.2% in ACT010 and 86.8% in ACT011. Top-3 retrieval reached 100% for RNA in ACT010 and 99.6% for image.

Cross-run centroid transfer provided a stricter test. Compound centroids learned in ACT010 retrieved the same compounds in ACT011 at 71.4% top-1 and 94.3% top-3 for RNA, and at 85.7% top-1 and 100% top-3 for the calibrated image latent. Chance top-1 retrieval is 2.9%. Thus, both branches recover compound identity well above chance across independent ActiveSeq runs, and the image branch is not merely a QC accessory; even this early representation carries reproducible compound-level signal.


A. Within-run split-half reproducibility
Top-1 retrieval
1.00
0.75
0.50
0.25
0.00
0.94
chance 0.87
0.99
0.80
ACT010 Image
ACT011 Image
ACT010 RNA
ACT011 RNA
B. ACT010 → ACT011 transfer
Retrieval accuracy
Top-1
Top-3
Image
RNA

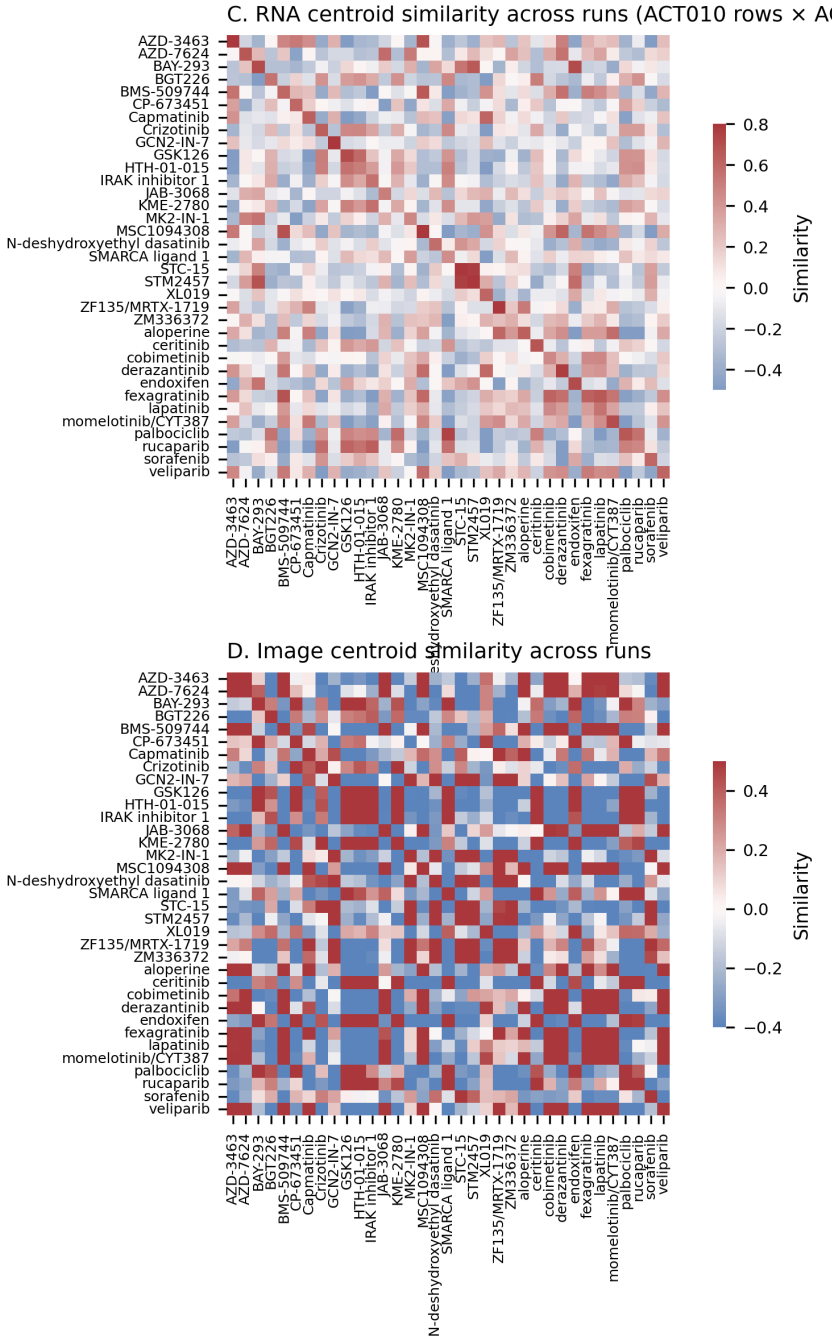

C. RNA centroid similarity across runs (ACT010 rows × ACT011 cols)
AZD-3463
AZD-7624
BAY-293
BGT226
BMS-509744
CP-673451
Capmatinib
Crizotinib
GCN2-IN-7
GSK126
HTH-01-015
IRAK inhibitor 1
JAB-3068
KME-2780
MK2-IN-1
MSC1094308
N-deshydroxyethyl dasatinib
SMARCA ligand 1
STC-15
STM2457
XL019
ZF135/MRTX-1719
ZM336372
aloperine
ceritinib
cobimetinib
derazantinib
endoxifen
fexagratinib
lapatinib
momelotinib/CYT387
palbociclib
rucaparib
sorafenib
veliparib
Similarity
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
D. Image centroid similarity across runs
Similarity
0.4
0.2
0.0
−0.2
−0.4

Figure 1. Compound-level reproducibility and cross-run centroid transfer in the 35-compound control panel. Both modalities are reproducible within run, and both transfer well above chance from ACT010 to ACT011. These metrics establish compound-level signal, not the incremental value of exact same-well pairing.

Do image and RNA encode related but non-identical compound maps?

A useful paired readout should not simply duplicate the same information twice. We therefore compared pairwise compound-distance matrices in image and RNA space within each run. The maps were aligned but distinct: Spearman correlations between image and RNA compound-distance matrices were 0.718 in ACT010 and 0.607 in ACT011. The strongest perturbations also recurred across experiments, with cross-run effect-magnitude rankings correlated at Spearman 0.799 for image and 0.662 for RNA.

Several compounds, including GCN2-IN-7, KME-2780, BGT226, ZF135 / MRTX-1719, and CP-673451, appeared among the stronger responses in both modalities in at least one run. The important result is the geometry, not any single example: same-well image and RNA measurements organize the control panel in related but non-identical ways. This is the pattern expected from complementary readouts of the same perturbation state.

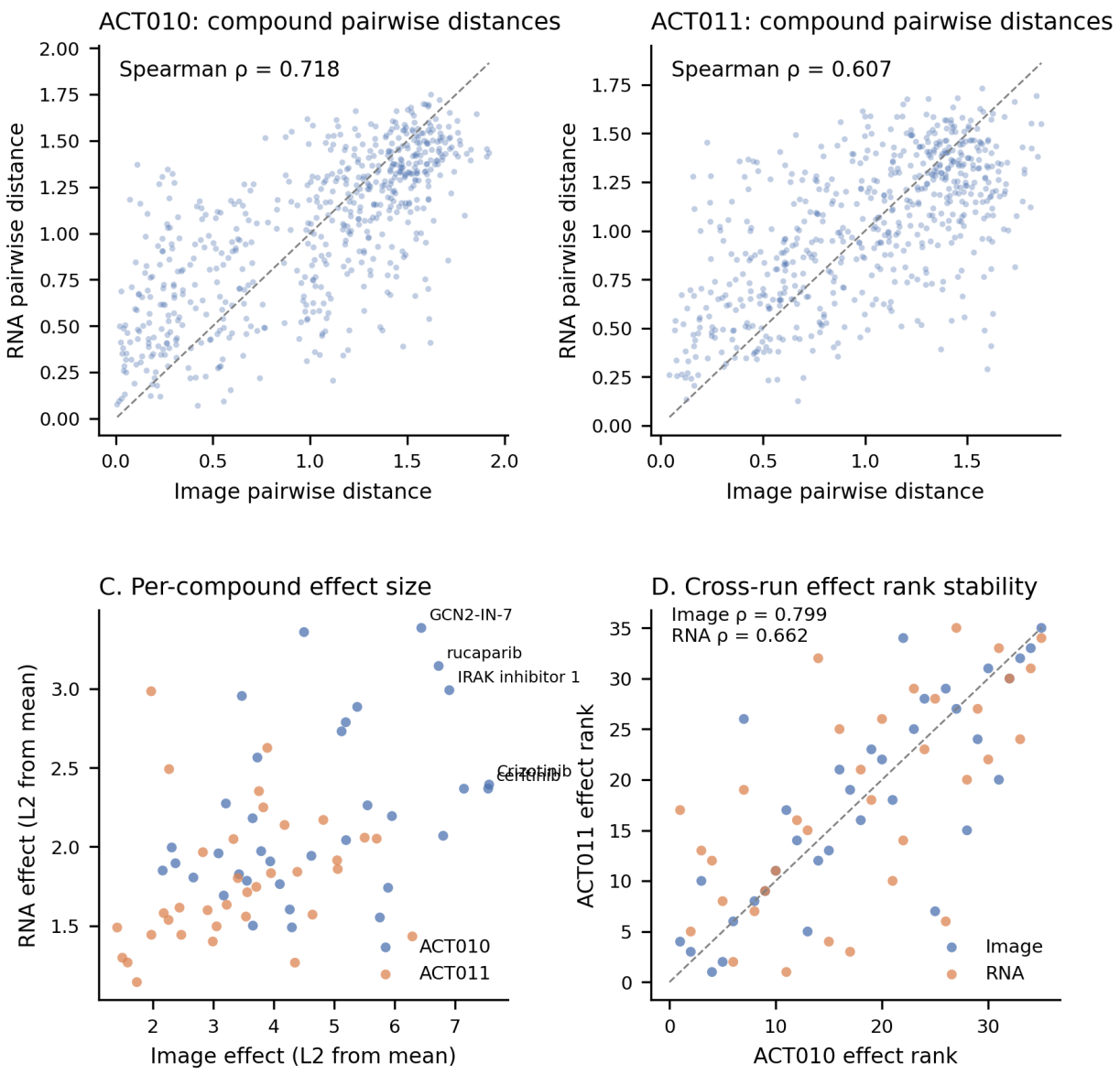


Figure 2. Image and RNA preserve a shared but non-identical compound geometry within run, and effect-size rankings are stable across runs. Pairwise compound-distance correlations are 0.718 in ACT010 and 0.607 in ACT011; cross-run per-compound effect magnitudes correlate at Spearman 0.799 for image and 0.662 for RNA.

Does combining same-well image and RNA improve held-out classification?

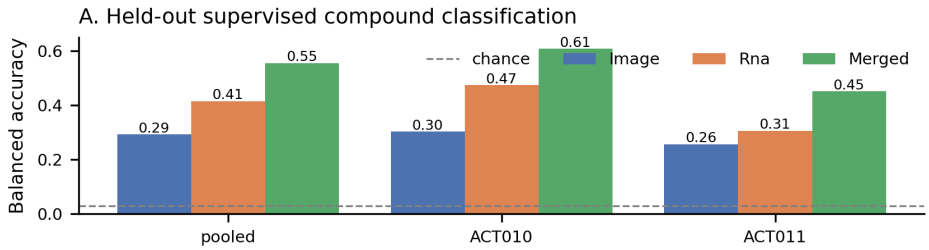
We next asked whether paired readouts help a decision model classify held-out wells. Each matched well was represented by one calibrated image vector, one RNA vector, or their concatenation. A linear discriminant model was trained on image alone, RNA alone, or merged image-plus-RNA features and evaluated on held-out wells.

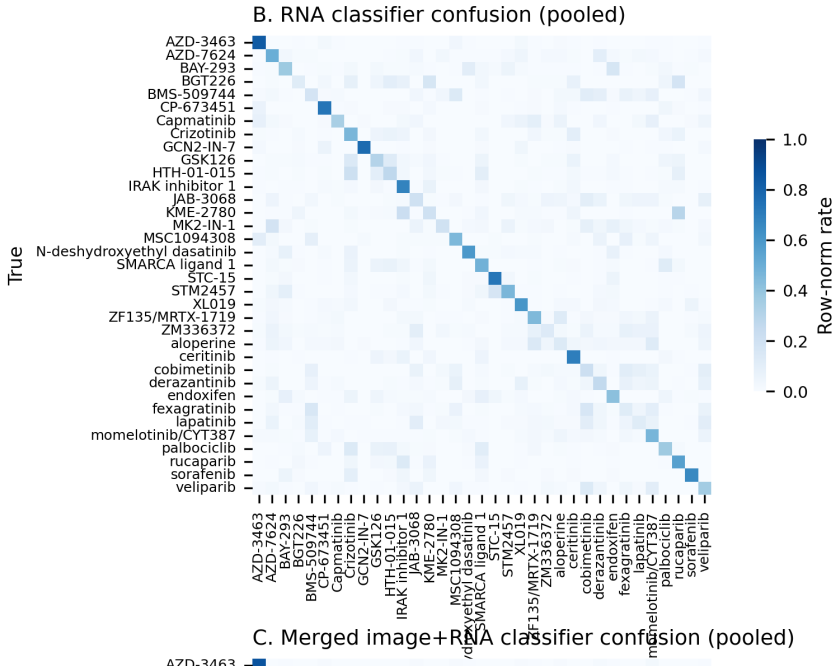
Merged features outperformed either single modality. In the pooled analysis, balanced accuracy

increased to 55.5% for merged features, compared with 41.4% for RNA alone and 29.2% for image alone. Within ACT010, merged features reached 60.8% balanced accuracy versus 47.3% for RNA and 30.3% for image. Within ACT011, merged features reached 45.1% versus 30.5% for RNA and 25.6% for image. Overall accuracy followed the same pattern, rising to 54.9% pooled, 60.2% in ACT010, and 45.5% in ACT011.

The gain was broad across the panel. Merged features improved per-compound recall for 35 of 35 compounds in the pooled analysis, 33 of 35 in ACT010, and 30 of 35 in ACT011. Confusion matrices show that the merged representation reduces many near-confusions rather than relying on a few easy compounds.

We then tested whether the classification gain required exact image/RNA pair identity. In a same-compound unpaired-fusion control, image and RNA vectors were shuffled among wells sharing the same compound and experiment before fitting the same classifier. This shuffled control matched the true same-well merged result: across 100 shuffles, pooled balanced accuracy averaged 55.4% (95% interval 54.7% to 56.1%), compared with 55.5% for true same-well fusion. ACT010 and ACT011 showed the same pattern. The classification result therefore supports a strong compound-level conclusion: same-well ActiveSeq produces complementary image and RNA features that improve held-out compound classification. On this label task, the added value comes from combining modality-level compound information, not from exact well-to-well pairing alone.


A. Held-out supervised compound classification
Balanced accuracy
0.0
0.2
0.4
0.6
chance
Image
Rna
Merged
0.29
0.41
0.55
0.30
0.47
0.61
0.26
0.31
0.45
pooled
ACT010
ACT011


B. RNA classifier confusion (pooled)
True
AZD-3463
AZD-7624
BAY-293
BGT226
BMS-509744
CP-673451
Capmatinib
Crizotinib
GCN2-IN-7
GSK126
HTH-01-015
IRAK inhibitor 1
JAB-3068
KME-2780
MK2-IN-1
MSC1094308
N-deshydroxyethyl dasatinib
SMARCA ligand 1
STC-15
STM2457
XL019
ZF135/MRTX-1719
ZM336372
aloperine
ceritinib
cobimetinib
derazantinib
endoxifen
fexagratinib
lapatinib
momelotinib/CYT387
palbociclib
rucaparib
sorafenib
veliparib
Row-norm rate
1.0
0.8
0.6
0.4
0.2
0.0

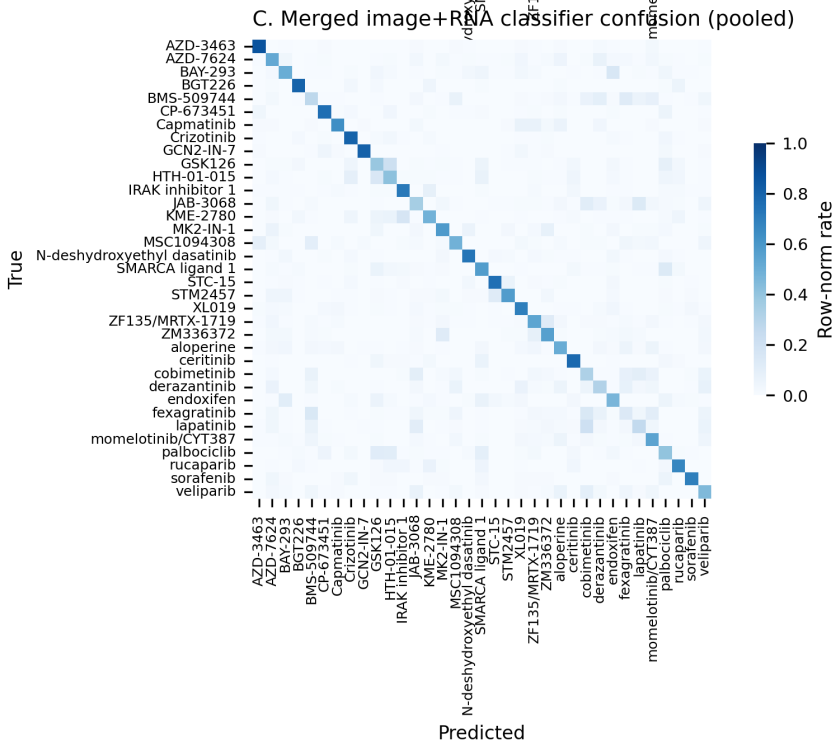

C. Merged image+RNA classifier confusion (pooled)
True
Predicted
Row-norm rate
1.0
0.8
0.6
0.4
0.2
0.0

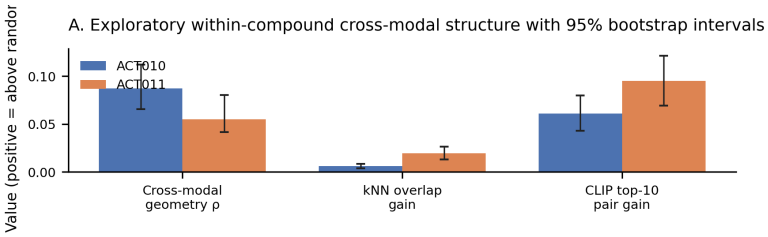
Figure 3. Same-well multimodal classification on held-out wells. Merged image-plus-RNA features outperform RNA alone in pooled and per-run analyses, and per-compound confusion matrices show the gain is from reduced class confusion across most compounds rather than from a few easy classes.

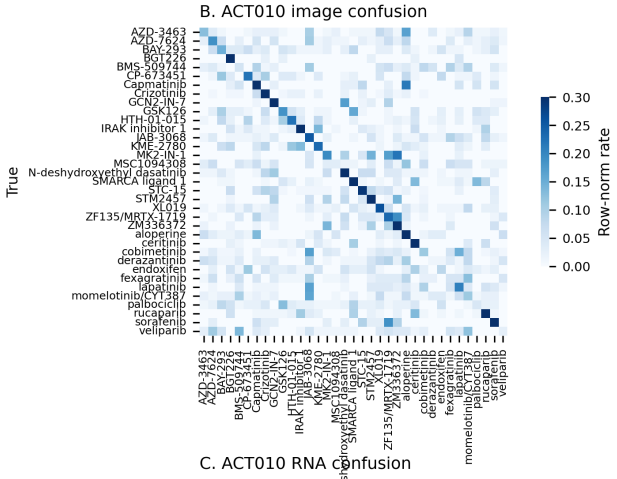
Is exact same-well structure detectable within compounds?

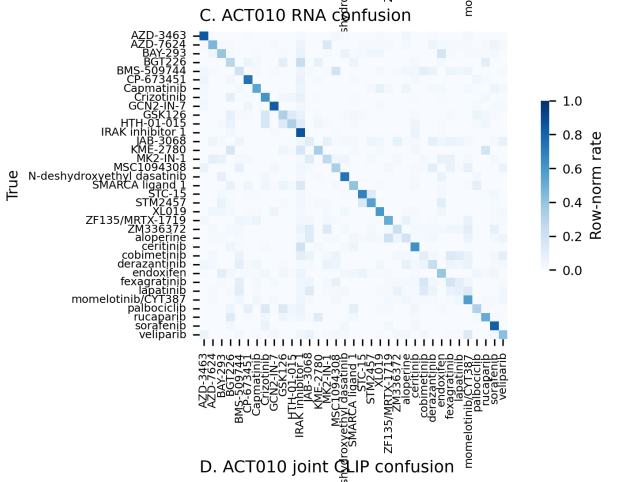
The previous analyses establish compound-level multimodal value. We next asked the stricter same-well question: after compound identity is held fixed, do image and RNA preserve matching well-to-well structure? This is a high bar because each well contains only a small local cell population, and many sources of biological and technical variation are not expected to align perfectly across modalities.

In a held-out centroid-retrieval task, image alone reached 25.1% / 21.7% top-1 across ACT010 / ACT011, RNA alone reached 44.8% / 24.3%, and naive image-plus-RNA concatenation reached 49.1% / 33.7%, exceeding RNA alone in both runs. A lightweight contrastive image-RNA pairing objective on top of the calibrated latent did not improve retrieval further, suggesting that the current representation already captures most of the readily available compound-identity information.

Exact pair retrieval and within-compound geometry showed small but consistent signal. Top-10 exact image-to-RNA retrieval after contrastive training was 2.1% in ACT010 and 2.5% in ACT011, compared with random expectations of 0.4% and 1.0%. Within each compound, image and RNA distance structure was positive in 34 of 35 compounds in both runs, with median cross-modal distance correlations of 0.087 in ACT010 and 0.055 in ACT011. Mean k-nearest-neighbor overlap gains were 0.006 and 0.020. These effects are modest in absolute size, but they indicate that same-well structure is measurable in the pilot. Larger panels, richer image channels, and perturbations designed to produce within-compound response heterogeneity should provide a clearer test of when exact pairing becomes operationally important.


A. Exploratory within-compound cross-modal structure with 95% bootstrap intervals
Value (positive = above randor
ACT010
ACT011
0.10
0.05
0.00
Cross-modal geometry ρ
kNN overlap gain
CLIP top-10 pair gain


B. ACT010 image confusion
True
Row-norm rate
0.30
0.25
0.20
0.15
0.10
0.05
0.00


C. ACT010 RNA confusion
True
Row-norm rate
1.0
0.8
0.6
0.4
0.2
0.0

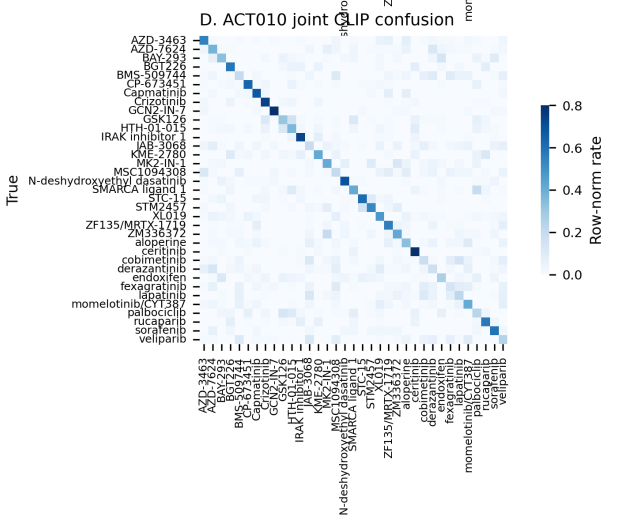

D. ACT010 joint CLIP confusion
True
Predicted
Row-norm rate
0.8
0.6
0.4
0.2
0.0

Figure 4. Exploratory same-well structure beyond supervised classification. Within-compound cross-modal agreement is usually above random but modest in magnitude. These data support the possibility of same-well response-state information, while leaving the incremental value of exact pairing to a dedicated control analysis.

Is the low-pass RNA branch interpretable for triage?

Classification alone is not enough for discovery; a useful RNA branch should also yield interpretable signatures. We aggregated raw counts from matched control wells into per-compound, per-run pseudobulk signatures, converted them to log-counts per million (log-CPM), and centered each run by subtracting its experiment-wide mean. This uses the compound-by-run pseudobulk as the robust unit at the present depth.

Centered signatures retrieved the correct compound across runs at 48.6% top-1 and 68.6% top-3. Representative compounds carried reproducible ranked gene signatures, including palbociclib (ARC, RASD1, CDKN1C, EGR1), BAY-293 (MT1X, MT1F, MT2A, SLC30A1), rucaparib (ITPRIP, SLC37A4, NACC2, COL4A1), crizotinib (SORCS2, ZNF34, AC253572.2), and GSK126 (DUSP10, SPATA6L, CCN2). These signatures are best interpreted as triage features rather than validated mechanisms in HEK293. They show that low-pass ActiveSeq RNA can cluster hits, flag recurring expression programs, and nominate follow-up assays while remaining paired to same-well imaging.

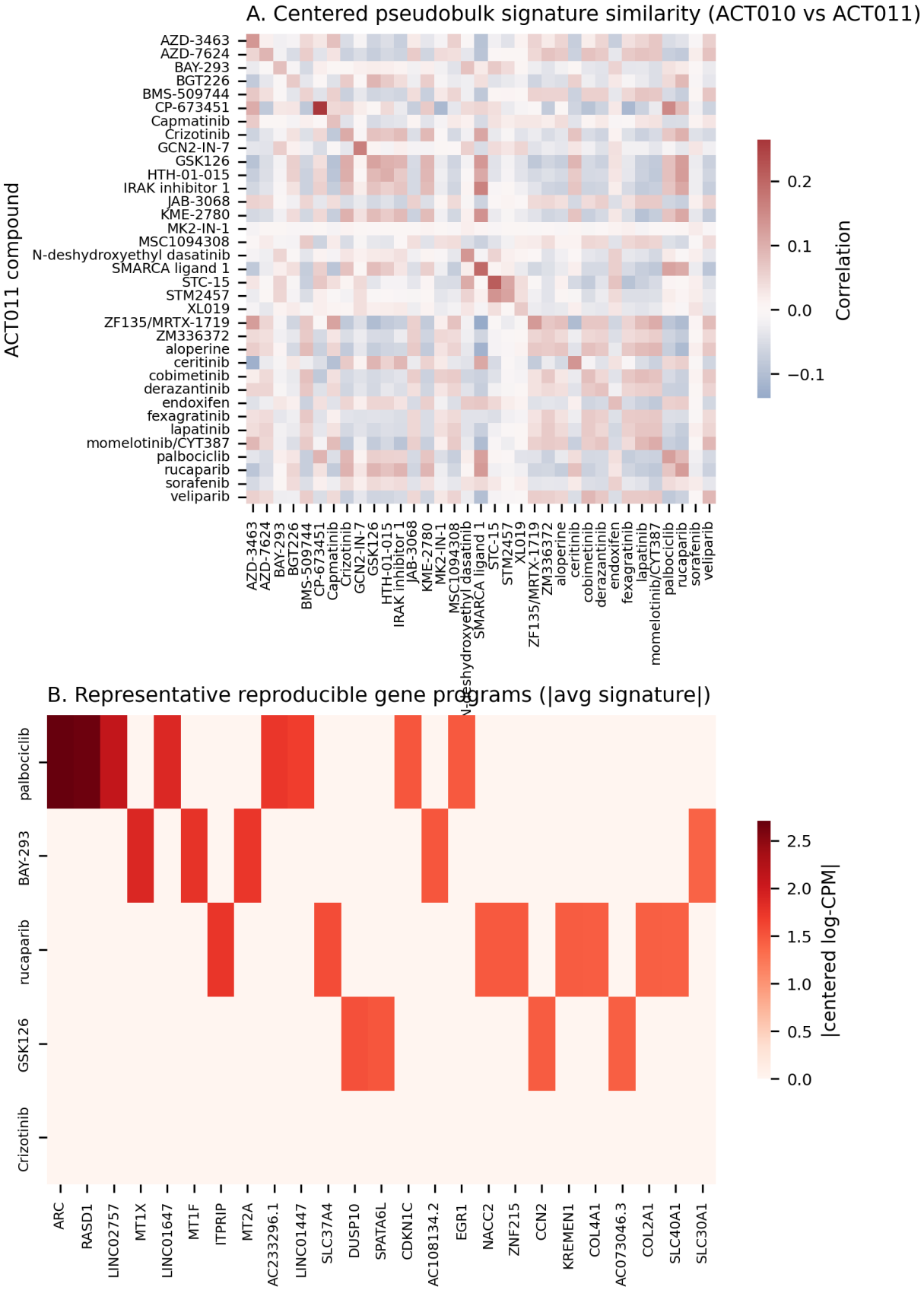

A. Centered pseudobulk signature similarity (ACT010 vs ACT011)
ACT011 compound
AZD-3463
AZD-7624
BAY-293
BGT226
BMS-509744
CP-673451
Capmatinib
Crizotinib
GCN2-IN-7
GSK126
HTH-01-015
IRAK inhibitor 1
JAB-3068
KME-2780
MK2-IN-1
MSC1094308
N-deshydroxyethyl dasatinib
SMARCA ligand 1
STC-15
STM2457
XL019
ZF135/MRTX-1719
ZM336372
aloperine
ceritinib
cobimetinib
derazantinib
endoxifen
fexagratinib
lapatinib
momelotinib/CYT387
palbociclib
rucaparib
sorafenib
veliparib
Correlation
0.2
0.1
0.0
−0.1
B. Representative reproducible gene programs (|avg signature|)
palbociclib
BAY-293
rucaparib
GSK126
Crizotinib
ARC
RASD1
LINC02757
MT1X
LINC01647
MT1F
ITPRIP
MT2A
AC233296.1
LINC01447
SLC37A4
DUSP10
SPATA6L
CDKN1C
AC108134.2
EGR1
NACC2
ZNF215
CCN2
KREMEN1
COL4A1
AC073046.3
COL2A1
SLC40A1
SLC30A1
|centered log-CPM|
2.5
2.0
1.5
1.0
0.5
0.0

Figure 5. Centered pseudobulk gene signatures preserve compound-specific ranked expression patterns across runs and recover the correct compound at 48.6% top-1 and 68.6% top-3 retrieval. The gene lists support RNA interpretability and triage rather than standalone mechanism calls.

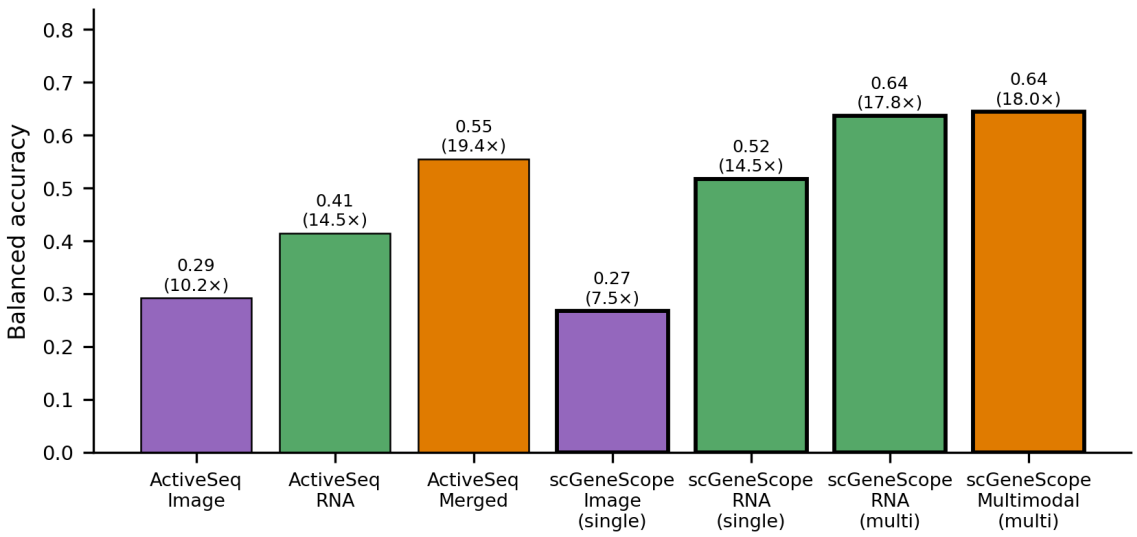
What context does scGeneScope provide?

scGeneScope is the closest public reference point for treatment-matched image and RNA perturbation modeling, but it is not a direct head-to-head benchmark. scGeneScope profiles approximately 716,767 single-cell images and 627,704 single-cell RNA profiles across 28 treatments in U2-OS cells, matching treatments across replicate populations [4]. ActiveSeq profiles 11,511 control-panel matched wells across two HEK293 experiments, each well measured as a 5- to 10-cell pseudobulk, with image and RNA read from the same micro-well and with a simpler early image representation.

The comparison is still useful for magnitude context. ActiveSeq trails the strongest scGeneScope multimodal reference in raw within-experiment balanced accuracy, consistent with the differences in scale, cell line, and imaging richness. After normalizing by random-chance label performance, pooled ActiveSeq RNA reaches 14.5 times chance, similar to the scGeneScope single-profile RNA value of 14.5 times chance. Pooled ActiveSeq image reaches 10.2 times chance, compared with 7.5 times for the scGeneScope single-profile image value. Pooled ActiveSeq merged features reach 19.4 times chance, compared with 18.0 times for the scGeneScope multimodal multiprofile reference.

The correct conclusion is that the same-well ActiveSeq pilot sits in the same broad performance range after chance normalization while measuring a different experimental unit. This supports scale-up and richer imaging rather than a claim of benchmark superiority.

A. Within-experiment compound classification (ActiveSeq filled, scGeneScope outlined)



B. Context for interpreting the gap

Aspect	**ActiveSeq (Z-Screen)**	**scGeneScope**
Single-cell scale	~60-110k cells (5-10 × 11.5k wells)	~717k images, ~628k RNA
Label space	35 compounds	28 treatments
Pairing	Same matched micro-well	Treatment-matched replicate
Imaging	Brightfield + calibrated latent	5-ch Cell Painting + ViT
Pooled image	29.2% (10.2× chance)	26.7% single (7.5×)
Pooled RNA	41.4% (14.5× chance)	51.7% single (14.5×)
Pooled multimodal	55.5% (19.4×)	64.4% multi (18.0×)
Best ActiveSeq run	ACT010 merged 60.8% (21.3×)	n/a
Takeaway	Smaller pilot; same range after chance no	Larger public benchmark on raw accuracy

Figure 6. Within-experiment compound classification for ActiveSeq versus the published scGeneScope reference, annotated with fold over random-chance performance, plus a context table summarizing the design differences that make the two benchmarks complementary rather than directly interchangeable.

Discussion

ActiveSeq establishes that the Z-Screen micro-well can support paired image and RNA readouts from the same small cell population. Across two HEK293 pilot experiments, both branches recovered compound identity within and across runs, image and RNA compound maps were aligned but non-identical, and merged features improved held-out compound classification across nearly the entire 35-compound panel. The same-compound shuffled-fusion control clarifies the mechanism of that gain: the current supervised classification improvement is driven by complementary compound-level image and RNA information rather than by exact well-to-well pair identity.

This is a positive result for scalable multimodal screening. A same-well workflow does not have to prove large exact-pair effects in a small control-panel classification task to be valuable. It first needs to show that both measurements can be collected from the same micro-well, that each readout is reproducible, that their compound maps are related but not redundant, and that their combination improves decisions. ActiveSeq passes those tests in this pilot.

The imaging result is especially important for the next version of the platform. The 32-dimensional latent used here is an early representation derived from generic image features and calibrated on the same control-panel domain. It already contributes reproducible compound information and improves classification when combined with RNA. At the same time, it should not define the ceiling for imaging in Z-Screen. Raw multi-channel images, assay-state features, colony morphology, and target-aware fluorescent or photoreactive channels can all be paired with the same RNA branch in future ActiveSeq screens. Target-aware channels are likely to be more informative for specific biological questions than a generic latent alone because they can report selected pathway, localization, abundance, binding, or activity states directly.

The RNA branch gives the paired screen interpretability. Low-pass per-well RNA is noisy, but compound-by-run pseudobulk signatures retrieved compounds across experiments and produced ranked gene signatures suitable for triage. These signatures do not establish mechanisms by themselves. They provide a way to cluster hits, recognize recurring cell-state programs, and choose follow-up assays while retaining the same-well connection to imaging.

The pilot also defines the next experiments. Broader control panels can test whether multimodal gains hold across more chemical and phenotypic diversity. New libraries can test how image calibration generalizes beyond the panel used here. Additional cell types can determine which compound-response programs are shared or context-specific. Image channels selected for target-relevant biology can ask whether exact same-well pairing becomes stronger when the image readout is designed to match the biological question. These are scale-up questions, not reasons to discount the present result.

In summary, ActiveSeq is the same-well image-plus-RNA arm of Z-Screen. In this first HEK293 pilot, it links morphology-derived signal, low-pass RNA state, and compound assignment in the same micro-well; both modalities recover compound-resolved phenotypes; and their combination improves held-out classification. The strategic value is the experimental unit: as Z-Screen scales, the same well can carry chemistry provenance, image-derived phenotypes, target-aware optical

measurements, and transcriptomic state in one linked assay record.

Methods

Dataset

The analysis uses the canonical Z-Screen ActiveSeq workspace under `data/ZScreen_Canonical_Dataset/`. Each ActiveSeq well is annotated with an experiment id (ACT010 or ACT011), a global well id, a compound assignment, an image embedding, and an RNA latent. Gene-level pseudobulk counts come from the matched H5AD object that backs the same-well RNA pipeline, repaired as documented in `DATA_REPAIR_REPORT.md`. The scGeneScope comparison uses published benchmark values transcribed from OpenReview Table 1 into `paper1/tables/scgenescope_reference.csv`.

Per-well image embedding calibration

The original per-well image embedding was produced by a frozen pretrained vision model on raw imaging tiles. That generic representation captures visual structure but does not guarantee that compound response is aligned with the principal directions of variation. Experiment-of-origin, position, microscopy, and colony-state effects can all contribute to the raw geometry.

A 32-dimensional control-panel-calibrated latent was learned from the original embedding by a lightweight domain-adversarial network trained on the matched control panel. The objective rewards compound discrimination and penalizes experiment-of-origin discrimination, encouraging compound-discriminating variation without retraining a vision model from scratch. Calibration was learned on the same control-panel domain studied here, so the calibrated-latent results should be read as an internal representation for this pilot. Future ActiveSeq studies can pair RNA with richer raw image representations and target-aware image channels.

Matched-well preprocessing

Imaging detections were aggregated by experiment_id and gwid so each matched well contributed one image vector. Wells with an assigned control name were treated as the 35-compound panel. ZEL026 wells were treated as dummy-bead negative-control QC stress-test wells and were excluded from compound-response analyses.

Reproducibility and transfer analyses

Within-run reproducibility was measured by repeated split-half pseudo-replication within each compound. Cosine similarity between pseudo-replicate centroids was computed after standardization within each matrix. Cross-run transfer used compound centroids learned in ACT010 and queried in ACT011.

Cross-modal geometry

For each run, pairwise distances among compound centroids were computed separately in image and RNA space. Agreement between distance matrices was summarized with Spearman correlation over

upper-triangular entries, which does not assume the two scales are matched. Cross-run effect-size stability used Spearman correlation over per-compound effect magnitudes.

Supervised multimodal classification

Held-out classification used a 50:50 compound-stratified split, either pooled across runs or fit within each run. Linear discriminant analysis with the eigen solver and automatic shrinkage was trained on calibrated image features alone, RNA features alone, or their raw concatenation. Performance was summarized with overall accuracy, balanced accuracy, macro F1, and per-compound recall.

To test whether the supervised gain depended on exact image/RNA pair identity, a same-compound unpaired-fusion control shuffled RNA vectors among wells sharing the same compound and experiment before fitting the same merged classifier. The control used 100 shuffle iterations with the same train/test split.

Same-well representation learning

A lightweight dual encoder with a symmetric contrastive image-RNA pairing objective was trained on same-well pairs from the training split, evaluated by nearest-centroid compound retrieval and by exact held-out pair retrieval. Within-compound analyses compared image and RNA distance structure among wells assigned to the same compound and summarized distance-correlation, k-nearest-neighbor overlap, and same-compound pair-retrieval gains over random baselines. These analyses were treated as exploratory because the absolute effects were small in the pilot.

Gene-level pseudobulk analysis

Raw counts were summed per experiment and control compound, converted to log-counts per million (log-CPM), and centered by subtracting the experiment-wide mean signature. Cross-run retrieval used correlation similarity between centered signatures. Per-cell differential expression was not performed because each well averages a 5- to 10-cell population; pseudobulk plus centering is the unit at which the signal is robust at the present depth.

External benchmark context

Within-run supervised metrics from `supervised_multimodal_metrics.csv` were compared to scGeneScope within-experiment values transcribed from OpenReview Table 1 into `scgenescope_reference.csv`. Because the label spaces differ (35 versus 28), the comparison reports both raw balanced accuracy and fold over random-chance performance.

Limitations

This pilot uses one cell line (HEK293), 35 control compounds, and two matched experiments. The calibrated image latent was learned on the same control panel studied here, so broader libraries, additional cell lines, and richer image channels are needed to evaluate generalization. The supervised fusion result shows additive value for combined image and RNA readouts, while the same-compound

shuffled-fusion control indicates that exact well-to-well pair identity does not add measurable classification value on this compound-label task. Exact same-well cross-modal retrieval and within-compound structure are above random but modest in absolute terms. Gene-level signatures are summarized through pseudobulk aggregation rather than full differential-expression modeling with biological replicates. These limits define the next scale-up studies while preserving the central result: ActiveSeq supports reproducible same-micro-well multimodal profiling with interpretable RNA and useful image-plus-RNA compound classification.

Data availability

All data tables, derived latent representations, manuscript figures, and analysis inputs needed to reproduce this manuscript are organized in this repository under `paper1/` and `data/ZScreen_Canonical_Dataset/`. A persistent-archive deposition with an assigned DOI will accompany the corresponding preprint posting.

Code availability

Analysis and figure-generation scripts are in `paper1/scripts/`. Package-level dependencies are in the repository root `requirements.txt`. Tested with Python 3.14.3 on Windows.

References

1. Bray MA, Singh S, Han H, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016;11:1757-1774. doi:10.1038/nprot.2016.105
2. Lopez R, Regier J, Cole MB, et al. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15:1053-1058. doi:10.1038/s41592-018-0229-2
3. Dixit A, Parnas O, Li B, et al. Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell. 2016;167(7):1853-1866.e17. doi:10.1016/j.cell.2016.11.038
4. Dapello J, Nassar M, Eksi R, et al. scGeneScope: A treatment-matched single-cell imaging and transcriptomics dataset and benchmark for treatment response modeling. NeurIPS 2025 Datasets and Benchmarks. OpenReview. https://openreview.net/pdf/f7d541dae38bcf88a79789a4c6440aadfec123c7