北京大学
PEKING UNIVERSITY

CVPR JUNE 3-7, 2026 DENVER COLORADO

TRIDENT: A Trimodal Cascade Generative Framework for Drug and RNA-Conditioned Cellular Morphology Synthesis

Rui Peng#, Ziru Liu#, Lingyuan Ye, Yuxing Lu, Boxin Shi*, Jinzhuo Wang*

# Equal Contribution, * Corresponding Authors

Author contact information:

The AI Virtual Cell vision

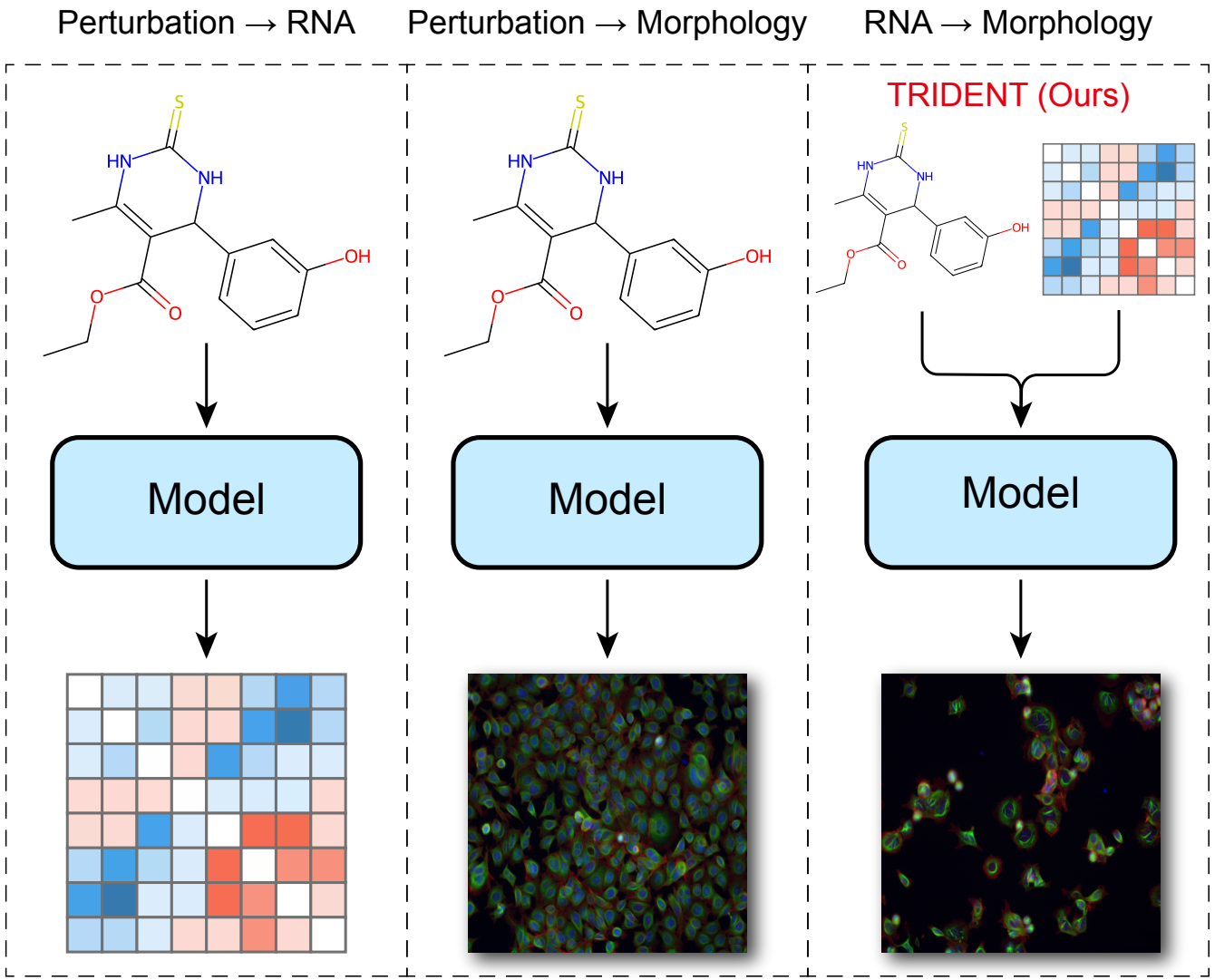


Fig. 1 — Comparison of cellular response modeling tasks.

Building an **AI Virtual Cell** — a digital twin of the cell — requires modeling how cells respond to perturbations across three inseparable elements:

- **Perturbation** (e.g., a drug)
- **Gene expression (RNA)**
- **Cellular morphology**

Prior work only generates a single modality from perturbation. We combine perturbation with RNA to drive a cross-modal generation.

A missing causal link

Existing methods model only direct associations:

- **Perturbation → RNA** (e.g., scGen, GEARS)
- **Perturbation → Morphology** (e.g., MorphoDiff)

They **overlook** the crucial **RNA → Morphology** link. Without it, we cannot simulate the cell as an integrated system where molecular events mechanistically drive phenotype.

Our goal. Explicitly learn **(Perturbation + RNA) → Morphology**, so morphology generation is conditioned on both the drug and its molecular effect.

TRIDENT — a two-stage cascade

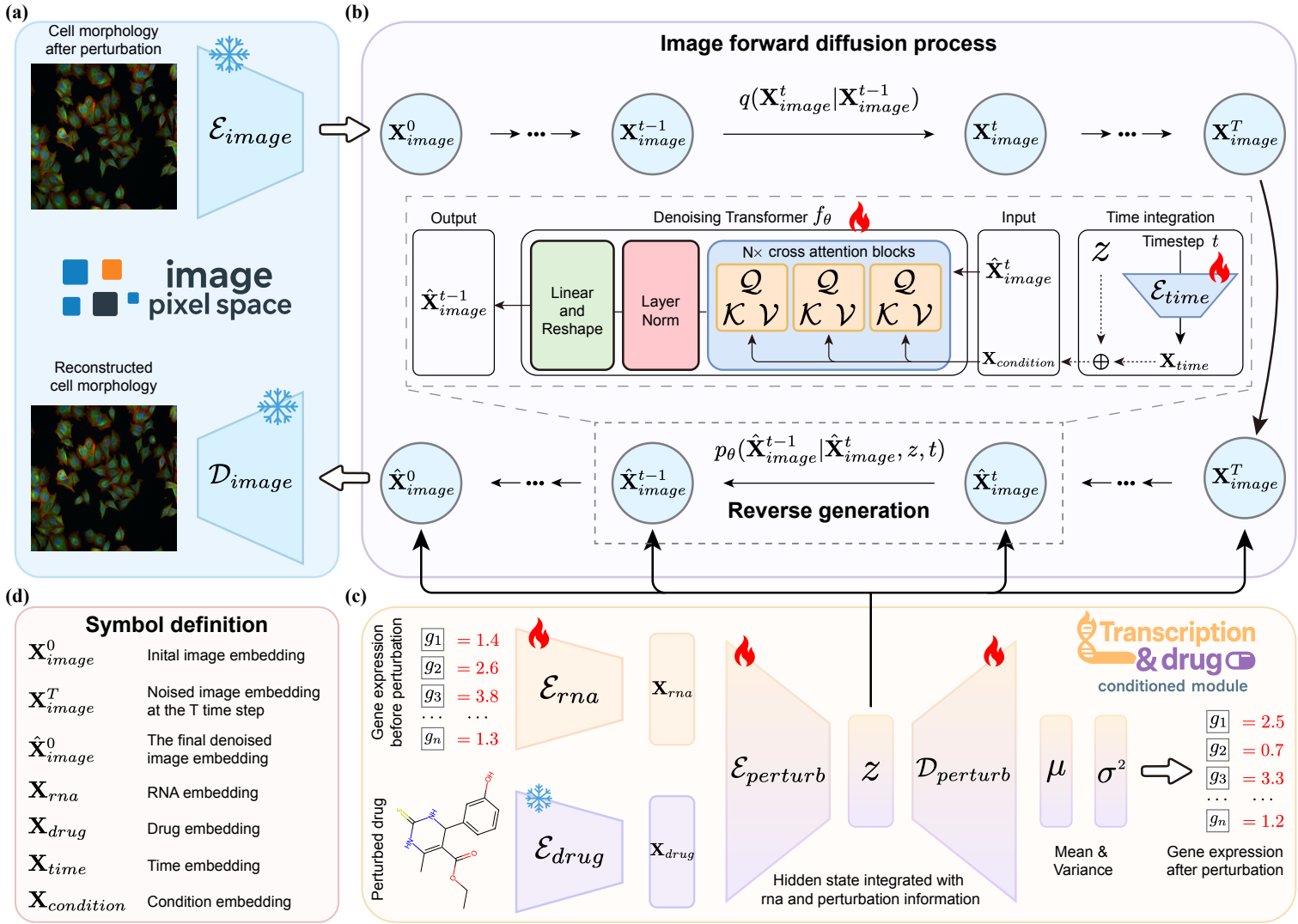


Fig. 2 — Overview of the TRIDENT framework.

- ① **Transcription-Drug Condition.** A VAE encodes (G_{pre}, Drug) into a latent z, regularized by predicting G_{post}.
- ② **Morphology Generation.** A Diffusion Transformer denoises image latents, guided by z via cross-attention.
- **Joint objective.** $L_{TRIDENT} = L_{VAE} + L_{LDM}$, trained end-to-end.

§ 04 · Dataset

MorphoGene — a new trimodal dataset

We pair Cell Painting images (BBBC021, MCF-7) with L1000 gene expression profiles, linked by 98 small-molecule drugs. Each sample contains **Drug · Image · G_{pre} · G_{post}**.

- **98 compounds**, 1,000 images per compound, **98,000 trimodal samples**
- **Train + ID test:** 44 shared compounds (8:2 split)
- **OOD test:** 54 fully held-out compounds — tests generalization to unseen drugs

§ 05 · Results · I

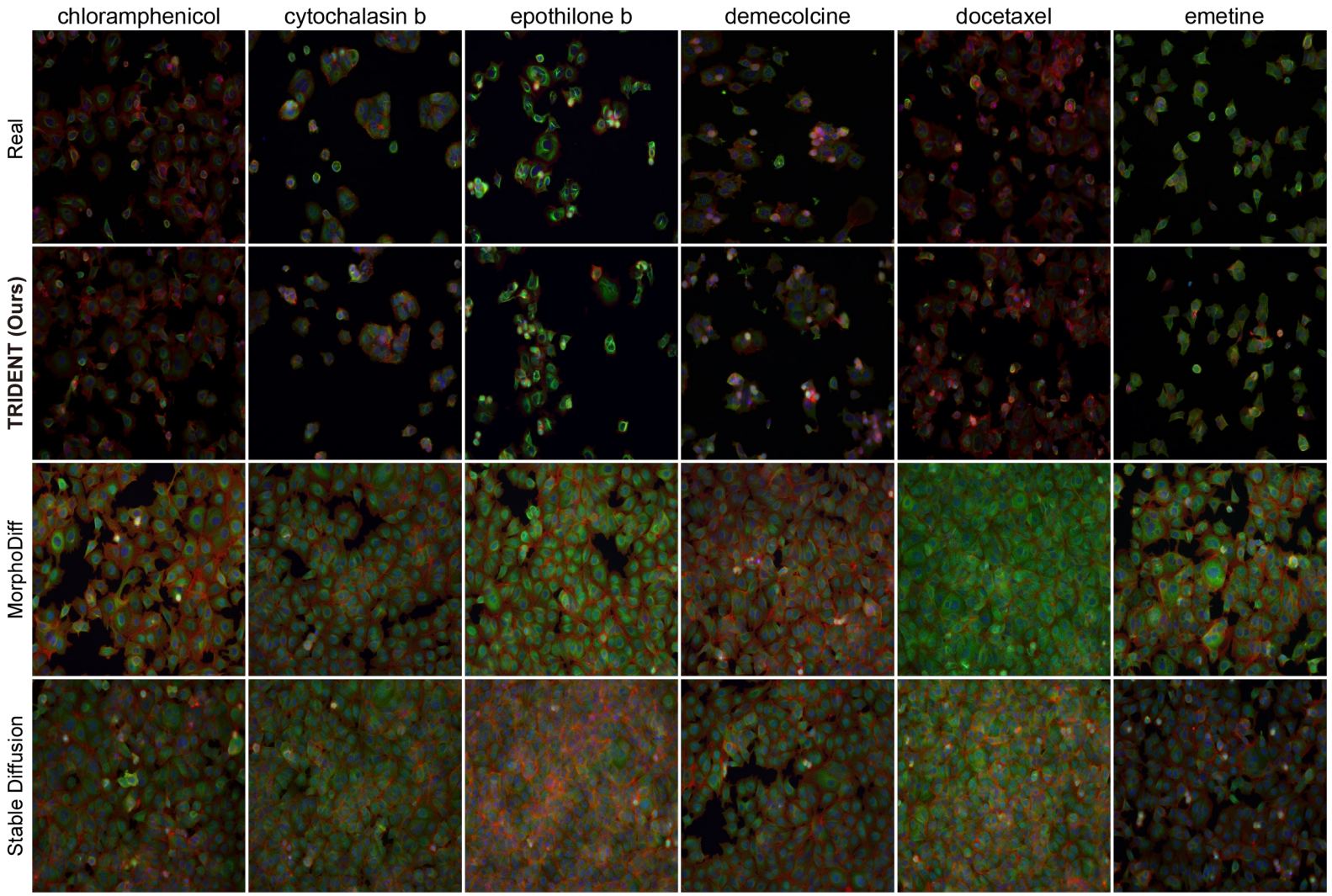
State-of-the-art fidelity

Comparison against MorphoDiff and Stable Diffusion (lower is better on all metrics).

Method	FID ↓ (ID)	KID ↓ (ID)	IS ↓ (ID)	FID ↓ (OOD)	KID ↓ (OOD)	IS ↓ (OOD)
MorphoDiff	250.29	0.248	2.614	387.14	0.436	2.747
Stable Diffusion	354.58	0.378	2.792	393.13	0.543	2.932
TRIDENT (ours)	**49.77**	**0.013**	**2.240**	**126.15**	**0.222**	**2.523**

- **5–7× FID improvement** on the ID test set.
- **3× improvement on OOD** — strong generalization to unseen compounds.

Qualitative comparison



TRIDENT reproduces **drug-specific** cell patterns — e.g., the low cell density induced by *cytochalasin B*.

In contrast, baselines **collapse to a generic high-density monolayer** and fail to respond to the conditioning signal.

Fig. 3 — Row 1: GT · Row 2: TRIDENT · Row 3: MorphoDiff · Row 4: Stable Diffusion.

MOA-specific phenotypic clusters

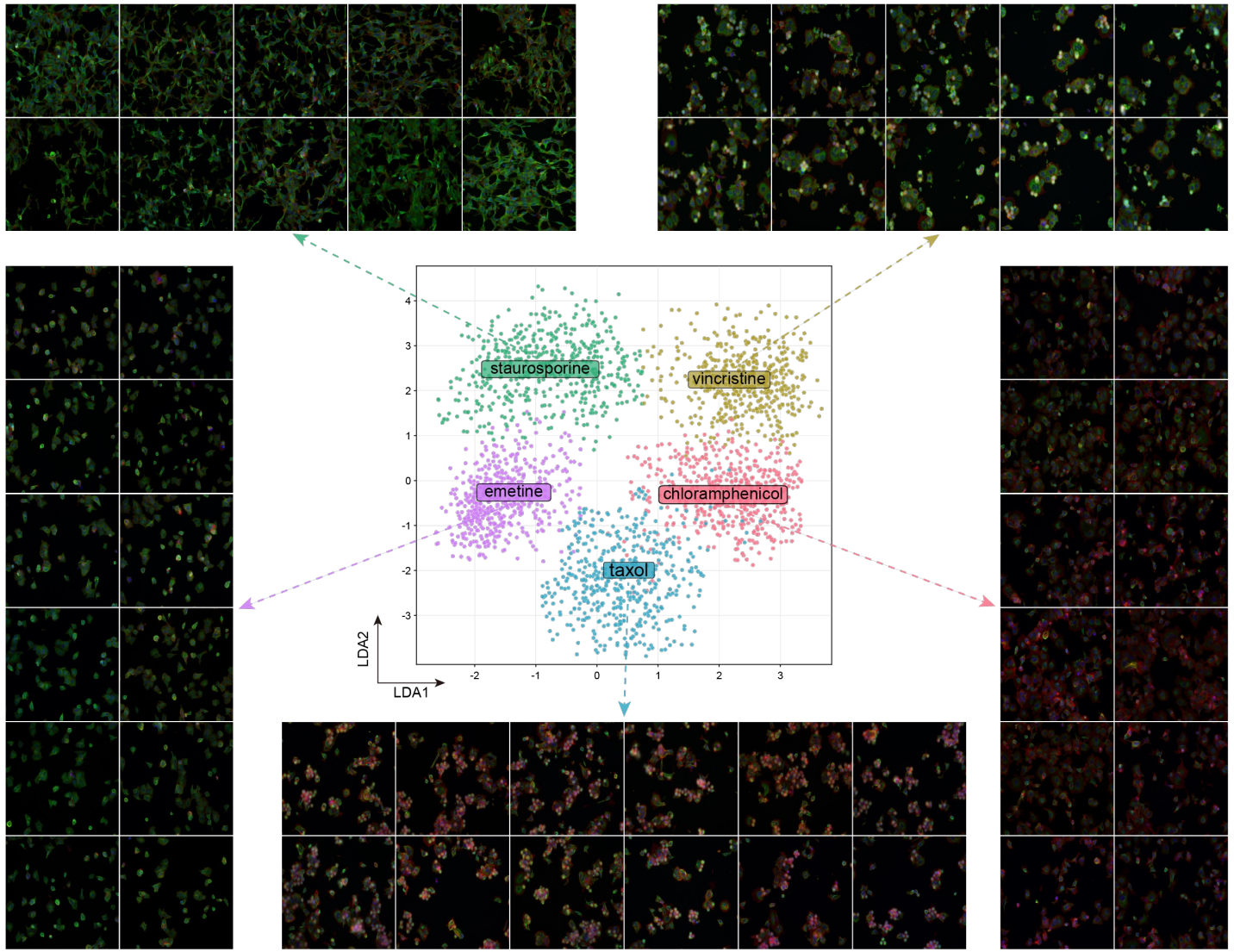
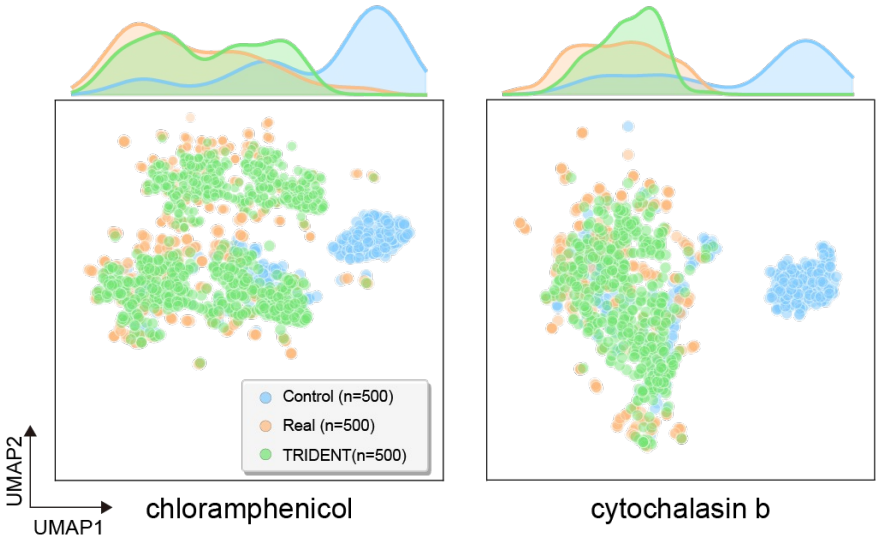


Fig. 4a — Phenotypic separability in LDA embedding space.

ViT embeddings of TRIDENT-generated images form **tight, MOA-specific clusters** in LDA space.

Different MOAs yield visibly distinct phenotypes — e.g., the filamentous morphology of the kinase inhibitor *staurosporine* vs. the rounded cells of the protein-synthesis inhibitor *emetine*.

Alignment with real data

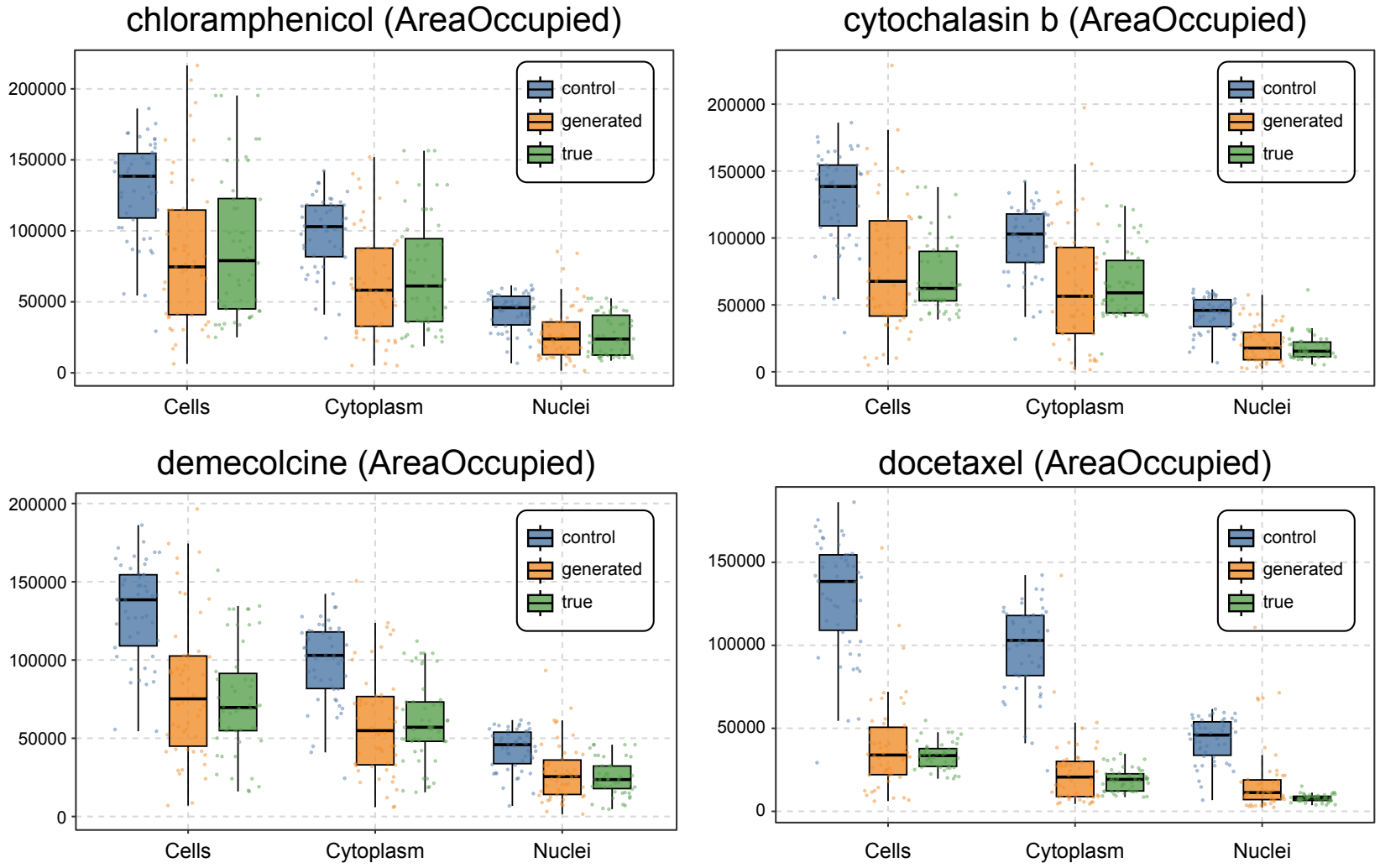


On UMAP, TRIDENT-generated (green) and real (orange) images are **tightly intermingled** on the same manifold.

Both are clearly separated from controls (blue) — generated morphologies are virtually indistinguishable from real data at the embedding level.

Fig. 4b — UMAP embedding of real, generated and control images.

Quantitative morphological features



CellProfiler analysis of the *AreaOccupied* feature shows **highly similar distributions** between generated and real images — across whole cell, cytoplasm and nucleus.

Both are consistently shifted away from controls, confirming TRIDENT captures the multi-scale morphological changes induced by drugs.

Fig. 4c — CellProfiler AreaOccupied across cellular compartments.

Accurate transcriptome prediction

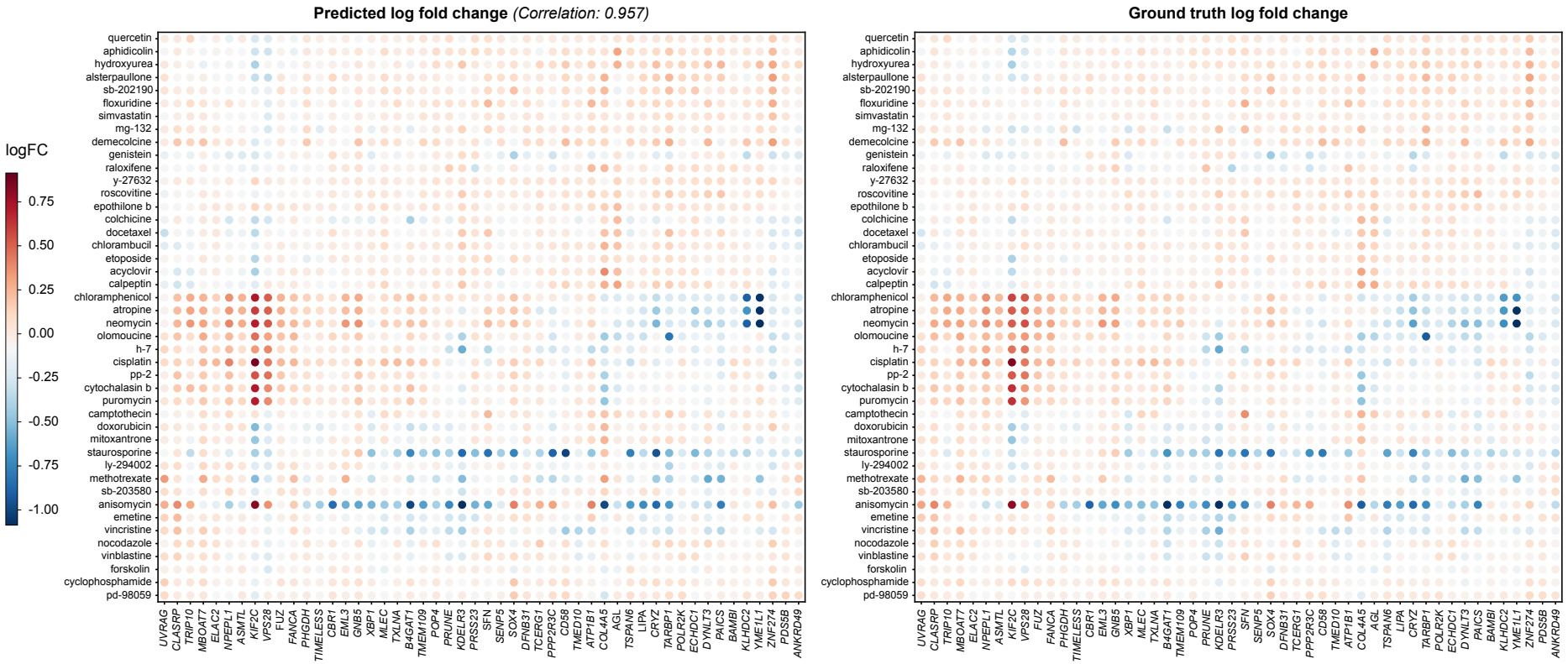


Fig. 5a — Predicted vs. ground-truth gene LFC heatmap.

Predicted log fold change patterns closely match ground truth across **44 compounds**.

Pearson correlation **r = 0.957** — the Transcription-Drug Condition Module generates biologically faithful transcriptional profiles.

Case study: docetaxel enrichment

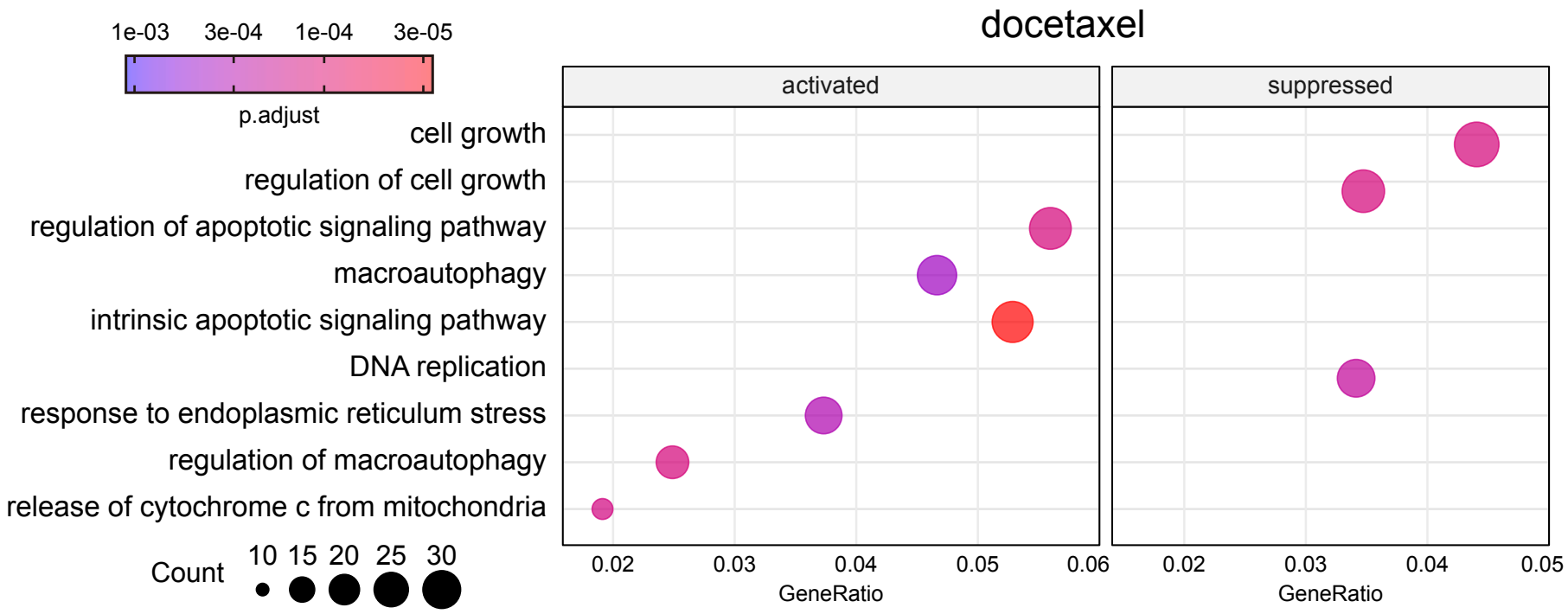


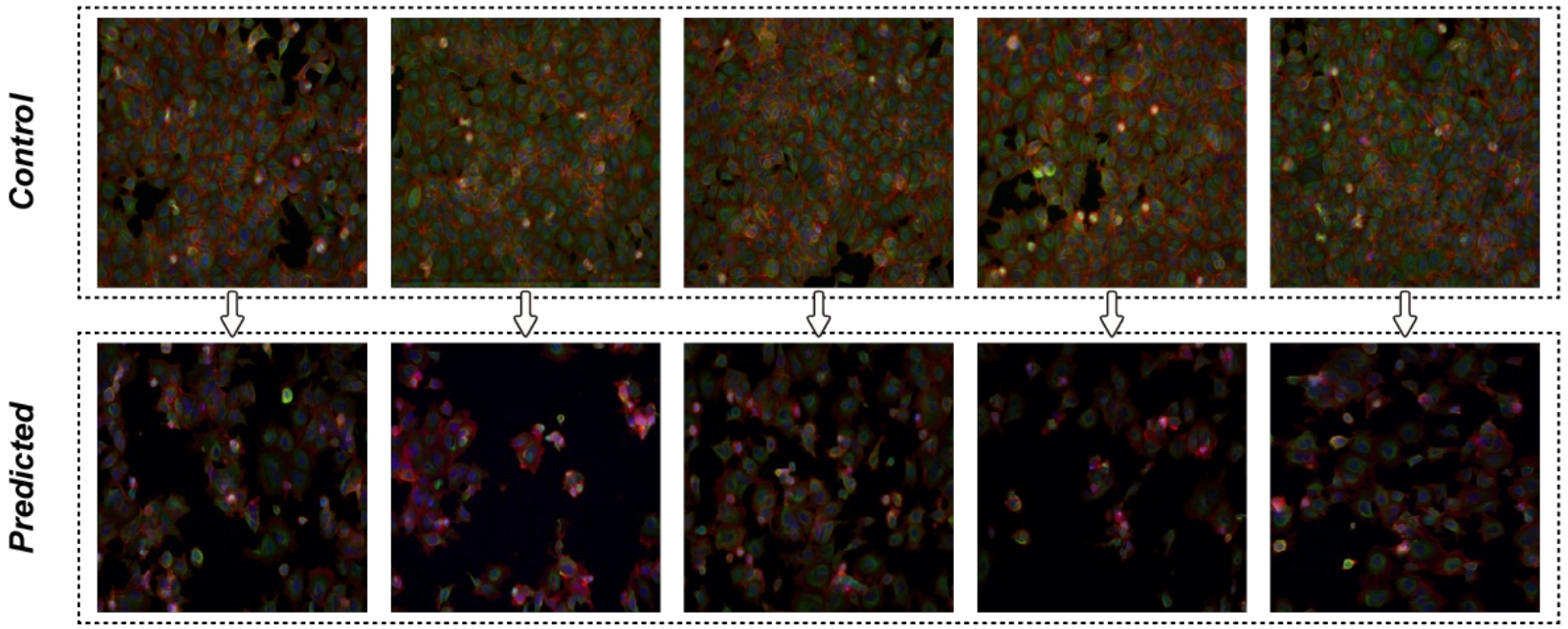
Fig. 5b — Functional enrichment analysis for docetaxel.

Docetaxel is a well-known **mitotic inhibitor**. Genes predicted by TRIDENT show enrichment consistent with its MOA:

- ↓ **suppressed** — *cell growth*, *DNA replication*
- ↑ **activated** — *apoptotic signaling pathway*

The predicted transcriptome is functionally and mechanistically correct.

RNA guides the morphology



The predicted biological program (suppressed growth, induced apoptosis) is **reflected in morphology** — docetaxel images show sparse, dying cells.

Fig. 5c — Control (top) vs. TRIDENT-predicted docetaxel (bottom).

§ 06 · Conclusion

Toward a predictive virtual cell

TRIDENT is the first framework to explicitly model **Perturbation → RNA → Morphology**. Together with the MorphoGene dataset, it achieves up to **7× FID improvement**, strong OOD generalization, and biologically faithful morphologies.

Thanks for your listening