Paired Single-Cell Imaging and Transcriptomics Provide a Functional Alignment Layer for Virtual Cell Models

Shan Sabri, Mehdi Mohseni, Annarita Scaramozza, Isaac Thomas, Richard Yau, Olaia Vila, Juan Pablo Romero, Pier Federico Gherardini, Gary Schroth, Mostafa Ronaghi

Cellanome, Foster City, CA, USA

INTRODUCTION Virtual cell foundation models such as scGPT and Geneformer learn rich gene–gene relationships from massive single-cell atlases, yet they train exclusively on static transcriptional snapshots of isolated cells – blind to the dynamic, context-dependent behaviors that define biological function. Cellanome's CellCage™ technology leverages light-guided hydrogel polymerization to enable longitudinal live-cell imaging coupled with endpoint scRNA-seq of the same cells. Same-cell linkage closes the training loop: every model prediction can be checked against per-cell ground truth. Using DINOv2 to extract morphological embeddings and pairing them with matched transcriptomes, we show across two systems – preadipocyte senescence and microglial phagocytosis – that the genes driving each functional phenotype are invisible to unsupervised transcriptomic clustering and are recovered only when a supervised model is anchored on the functional readout. We hypothesize that paired imaging and transcriptomics may provide an alignment layer that lets virtual cell models predict cellular function, not just gene expression.

CellCage™ Enclosures Unite Morphology and Transcriptomics

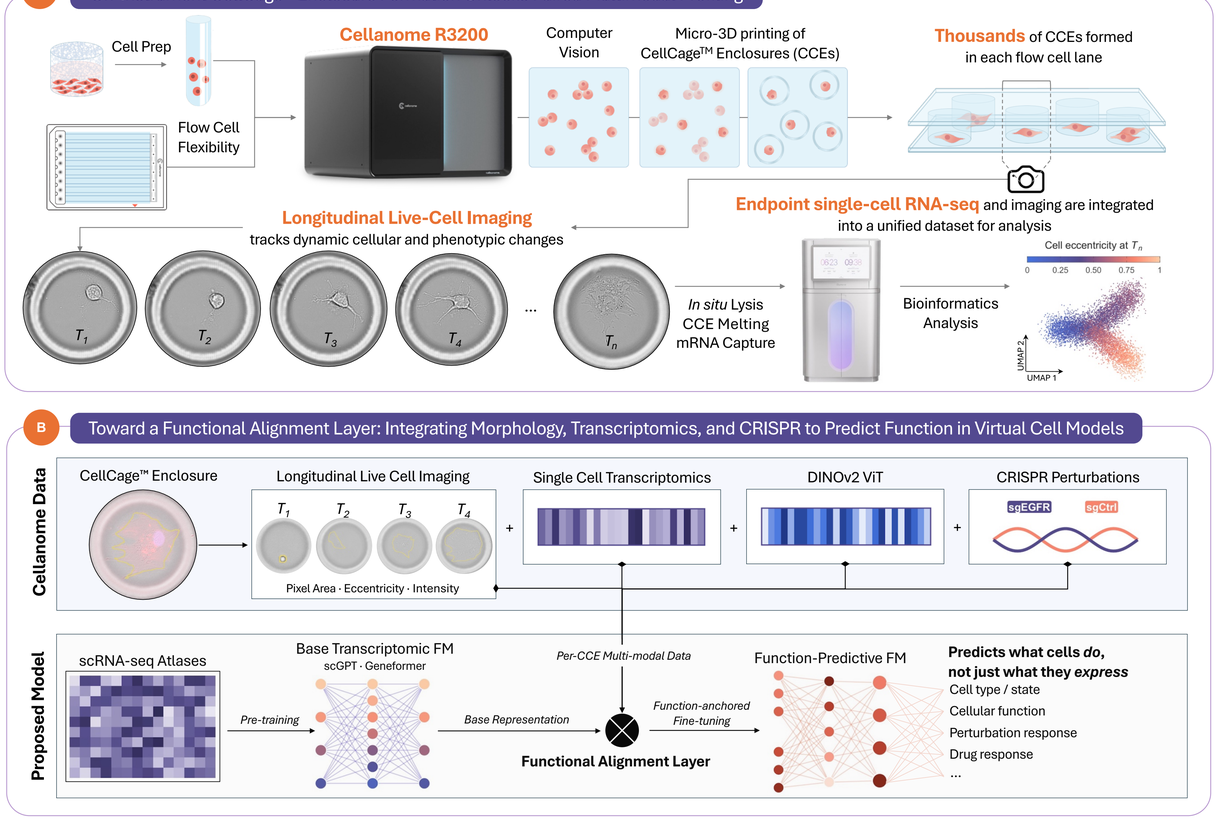
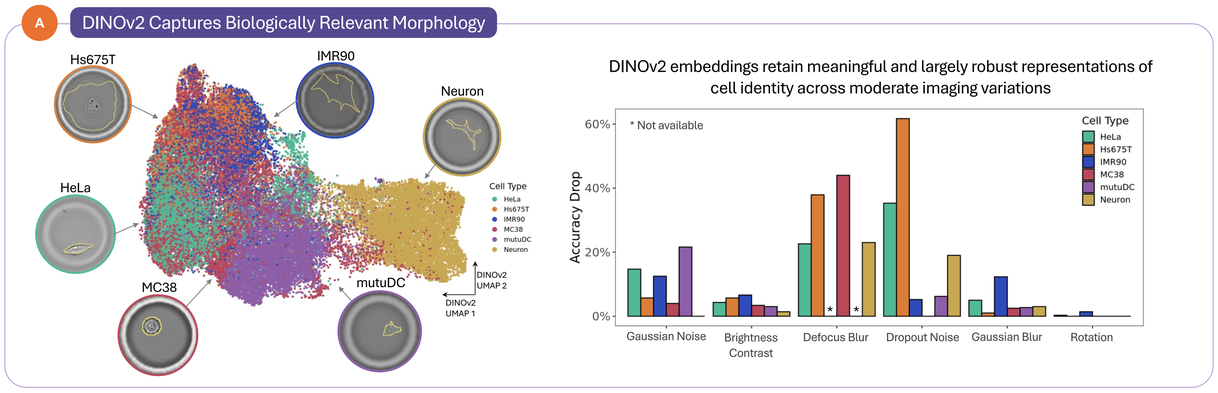


Fig. 1. Paired multi-modal data from CellCage™ Enclosures align virtual cell models to cellular function.
(A) Same-cell longitudinal imaging plus endpoint scRNA-seq on the Cellanome R3200 platform. **(B)** We propose that paired multimodal data could feed a Functional Alignment Layer to yield a function-predictive FM with task heads for functional behaviors.

Extracting Deep Morphological Features from CellCage™ Enclosure Images Using DINOv2



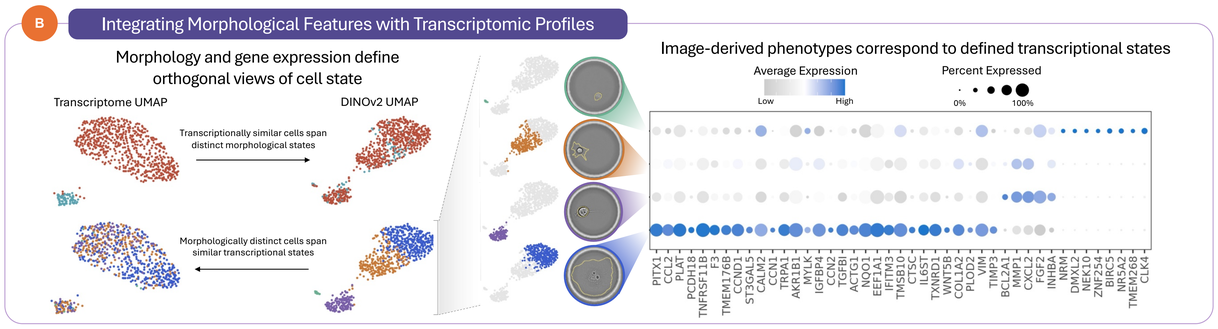


Fig. 2. Validation of DINOv2 morphological embeddings.
(A) DINOv2 features from six cell lines clustered by cell identity, confirming capture of meaningful morphological differences. Embedding similarity predicted the generalization performance of a Mask R-CNN (Region-based Convolutional Neural Network) segmentation model. The embeddings remained robust under common image augmentations, supporting their use for integration with transcriptomic data. **(B)** Clustering of DINOv2 features from fibroblast CCE images identified four distinct morphological states – ranging from fully adhered, ECM-enriched cells (Blue cluster) to weakly adhered, inflammatory cells (Orange cluster), with intermediate adhesion–stress profiles. These morphology-defined phenotypes were undetectable by transcriptomic clustering alone, demonstrating that image-derived embeddings capture additional, orthogonal, functionally relevant cellular heterogeneity.

Same-Cell Imaging Anchors a Transcriptomic Definition of Senescence

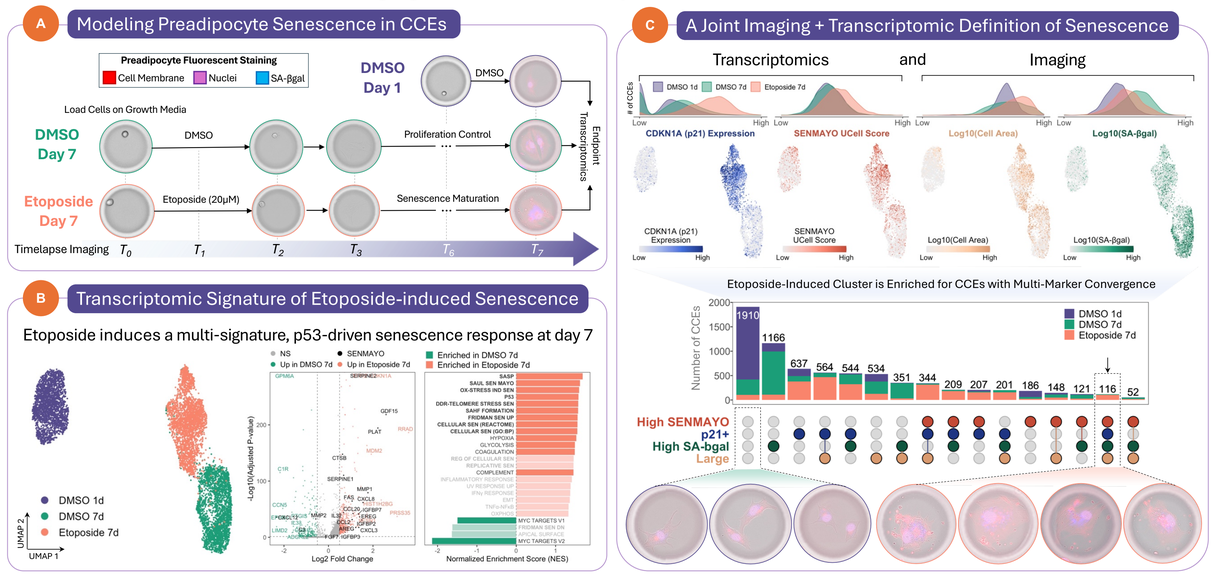
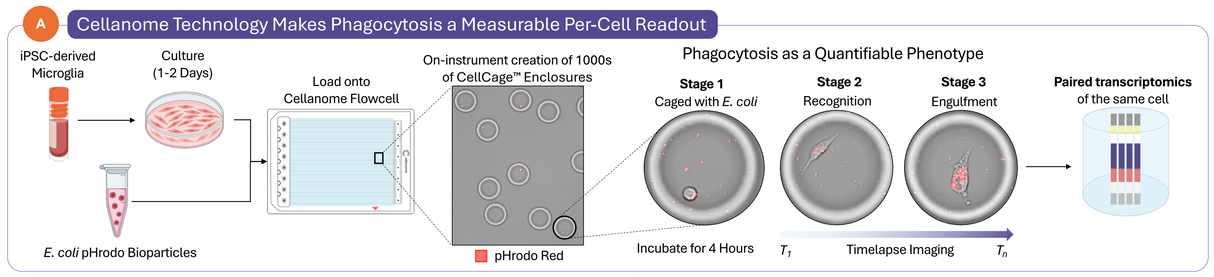
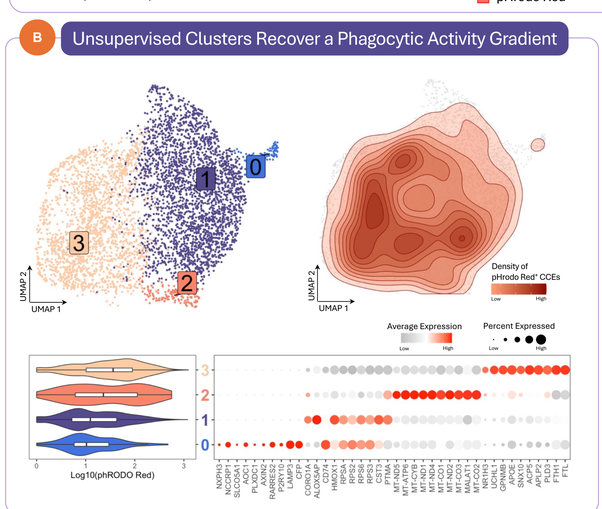
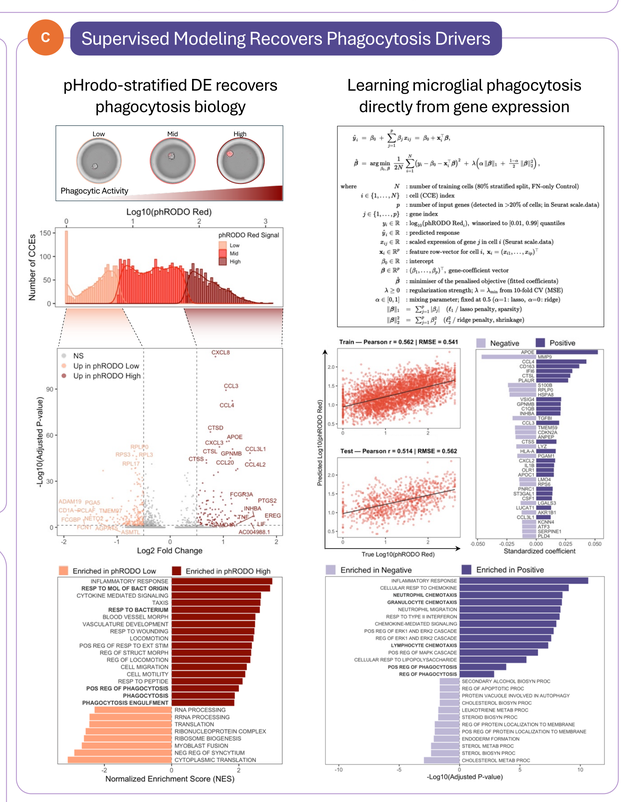


Fig. 3. A multimodal definition of cellular senescence.
(A) Preadipocytes were cultured in CCEs in three conditions, stained for SA-βgal, cell membrane and nuclei, then profiled by endpoint scRNA-seq paired with longitudinal imaging of the same CCEs. **(B)** Etoposide 7d cells are transcriptionally distinct from DMSO and enriched for SASP/p53 targets. **(C)** Joint p21, SENMAYO, SA-βgal, and cell area criteria identify 116 multi-marker senescent CCEs.

Phagocytosis Drivers Hide in Plain Sight: Same-cell Imaging Surfaces What Unsupervised Clustering Misses







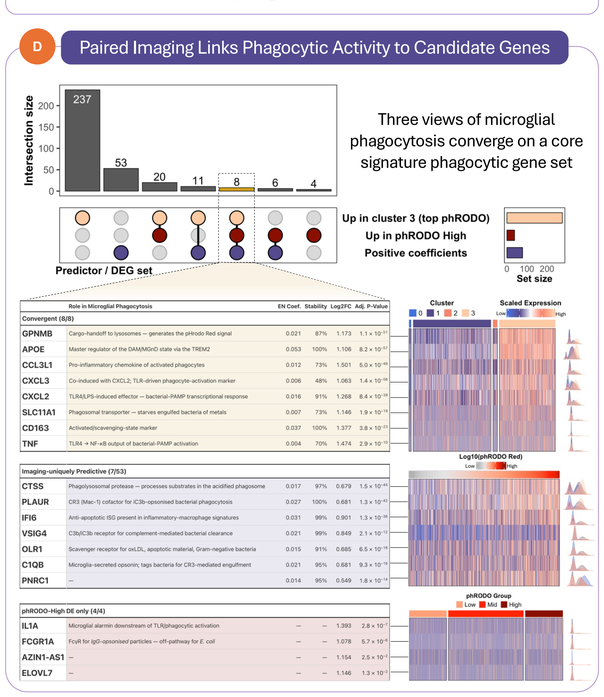


Fig 4. Imaging-supervised modeling recovers phagocytosis drivers that unsupervised clustering misses.
(A) Human iPSC-derived microglia were compartmentalized with pHrodo Red bioparticles on Cellanome's R3200 platform, imaged for 4 hours from compartmentalization through engulfment and paired with endpoint single-cell RNA-seq on the same cells. **(B)** Unsupervised clustering recovers a phagocytic-activity gradient, but its markers describe a generic activated state, not phagocytic machinery. **(C)** Anchoring the analysis on the imaging readout – stratified differential expression and an elastic-net regression onto pHrodo intensity – predicts phagocytic activity on held-out cells and enriches for phagocytosis pathways. **(D)** Intersecting markers, stratified DE, and supervised predictors yields a convergent phagocytic core plus a set of genes that only the imaging-supervised model surfaces.