Novel multi-modal technology that generates integrated cell function and sequencing data at scale

Pier Federico Gherardini[1], Mehdi Mohseni[1], Roger Dettloff[1], Shreya Deshmukh[1], Shan Sabri[1], Niranjan Srinivas[1], Josh Lamstein[1], Tim Williams[1], Liz Wu[1], Filiz Yasar[1], Richard Yau[1], Ilya Kupershmidt[1], Mostafa Ronaghi[1], Tarun Khurana,[1] Gary Schroth[1]

1 Cellanome, Foster City, CA

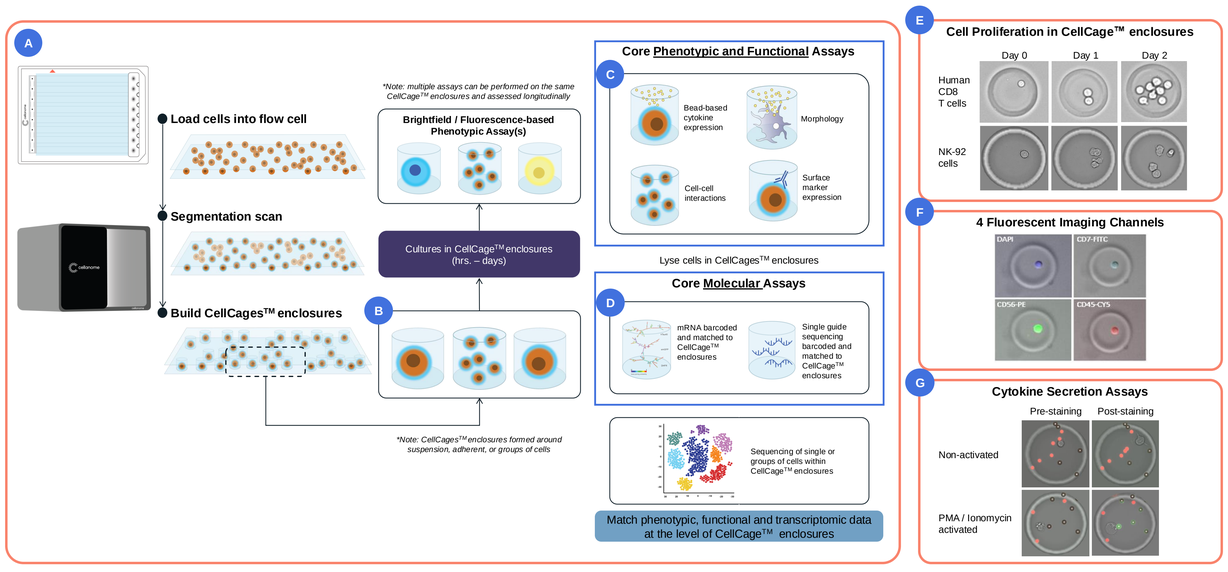
INTRODUCTION

Understanding how gene expression programs regulate cell functions is difficult due to the lack of technologies that capture both sequencing and live cell function data. Researchers tackle this by linking data across separate platforms, but this approach is limited by noise and variability and fails to align measurements to the same cell. To address these limitations, we have developed a novel technology for cell compartmentalization, analysis, and culture combining light-guided polymerization of hydrogels with advances in engineering and computer vision.

Here we describe the computational innovations that power the technology and the infrastructure we have developed to handle the scale and complexity of the resulting multi-modal data.

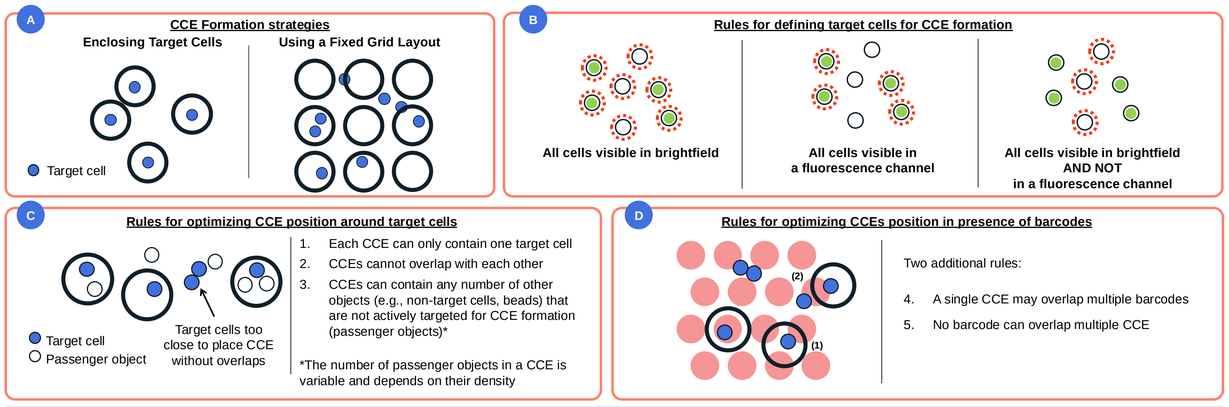
CELLANOME TECHNOLOGY AND WORKFLOW

Tens of thousands of suspension or adherent cells are mixed with hydrogel precursor and loaded on an 8-lane flow cell where a computer vision model is used to detect their position. An optimization algorithm then determines the optimal placement of compartments in real-time according to the position of the cells. Light is then projected on the fluidic channel in a corresponding pattern to induce polymerization of the hydrogel precursor, which results in the formation of hollow compartments around the cells called CellCage™ enclosures (CCEs) (A). CCEs are permeable to reagents (e.g., small molecules, immunofluorescent antibodies) enabling long-term culturing (B) and a variety of imaging-based, longitudinal phenotypic and functional assays to be performed on the same cells (C). Cells can be lysed within CCEs to generate positionally barcoded cDNA for downstream library prep and sequencing off instrument (D). Bio-compatible materials enable cell culturing and proliferation (E). The instrument supports 4-channel fluorescent imaging (F). Beads can be included in the experiment to perform various assays including cytokine detection (G).



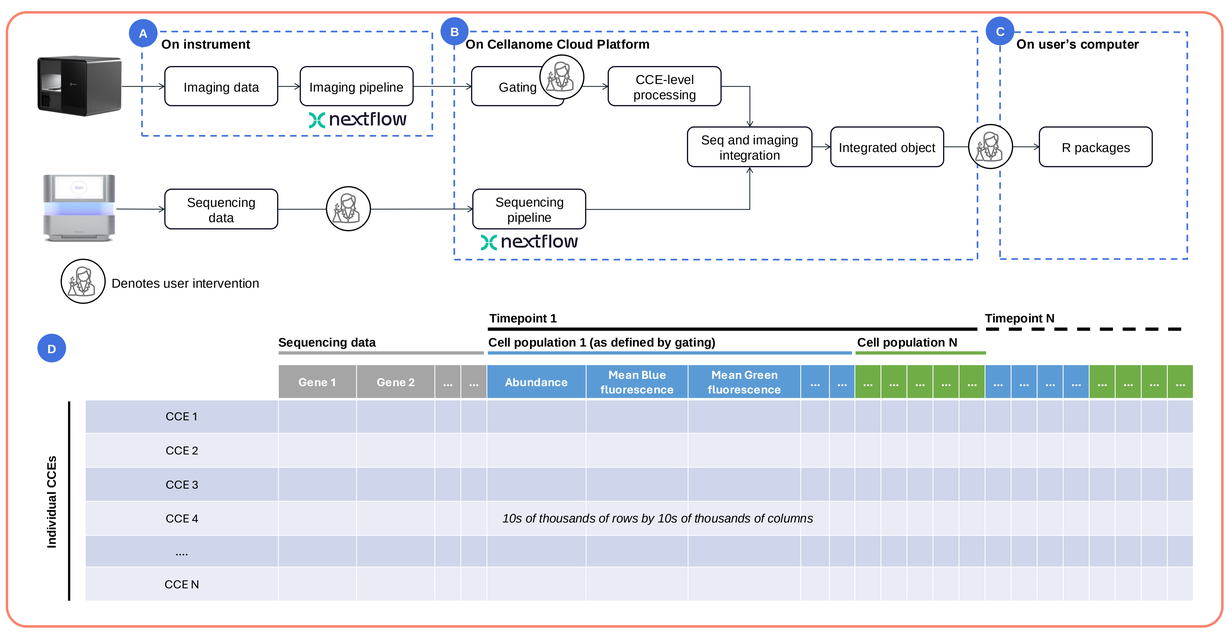
CELLCAGE™ ENCLOSURE (CCE) FORMATION RULES

CCEs can be formed around target cells, or in a fixed grid pattern (A). Target cells can be defined according to a variety of imaging-based criteria (B). The optimal placement of the CCEs is determined by an optimization algorithm according to a set of rules using as input the position of the cells, as detected by computer vision (C and D)



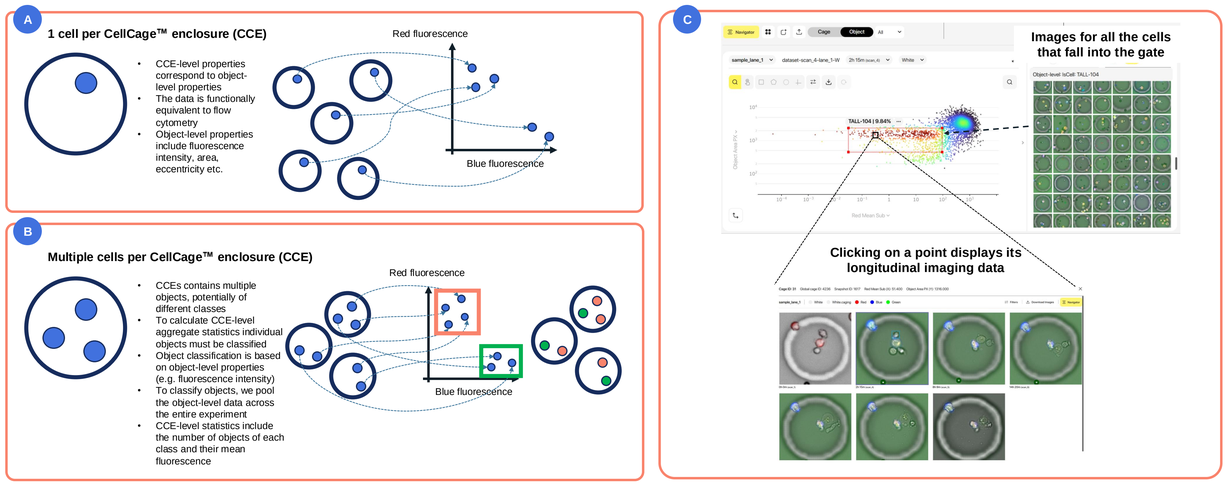
DATA ANALYSIS INFRASTRUCTURE

Imaging data is processed locally on the instrument by an analysis pipeline written in Nextflow. The pipeline segments all the objects in an image and calculates various features including the intensity of each object across all the fluorescent channels, and morphological properties such as size, eccentricity, etc. (A). The results are then uploaded to the Cellanome Cloud Platform where they are further processed in an interactive workflow to define populations of interest (see below) and integrated with sequencing data results (B). The integrated data can be downloaded from the Cellanome Cloud Platform and further processed using R packages internally developed (C). The integrated data contains gene expression measurements and longitudinal imaging-based features matched at the level of individual CCEs (D).



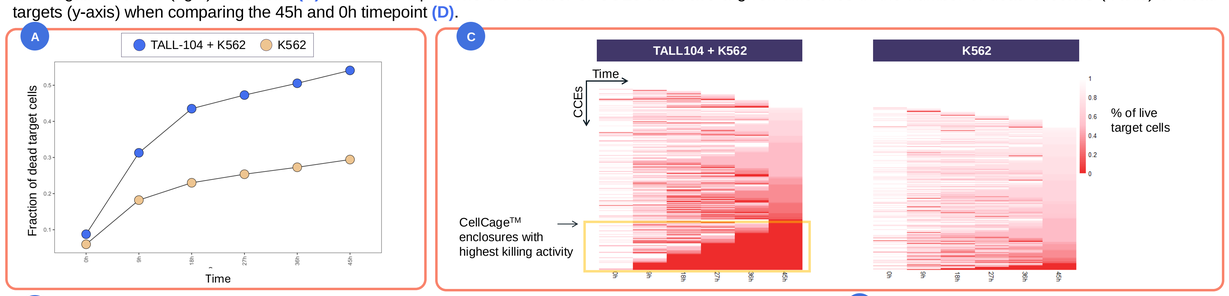
QUANTITATIVE ANALYSIS OF IMAGING DATA

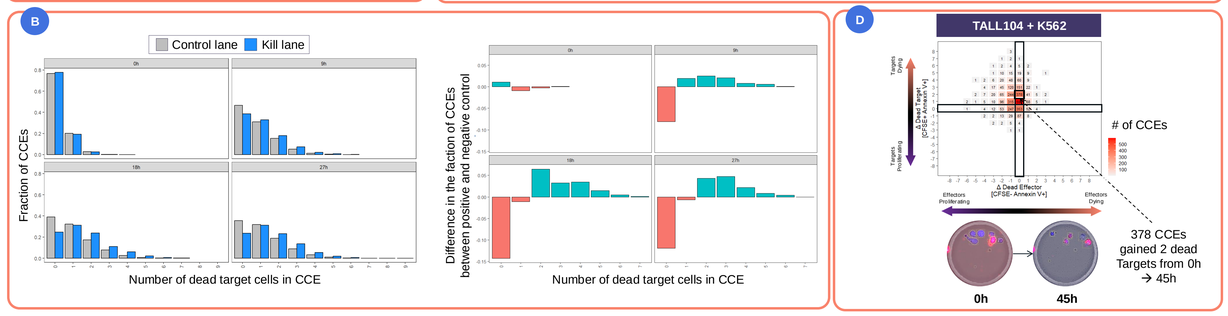
The image analysis pipeline calculates quantitative features for all objects (beads, cells etc.), following segmentation. In experiments where each CCE contains a single cell, object-level statistics correspond to CCE-level statistics (A). In experiments where CCEs contain multiple objects, the object-level data is pooled across the entire experiment to identify populations of interest. The classification is then used to calculate CCE-level statistics (e.g., the number of objects of each class and their average fluorescent intensity) (B). The data can be interactively analyzed in the cloud platform to define populations of interest and inspect the longitudinal images associated with each data point (C).



CELL-CELL INTERACTION ASSAYS IN CELLCAGE™ ENCLOSURES

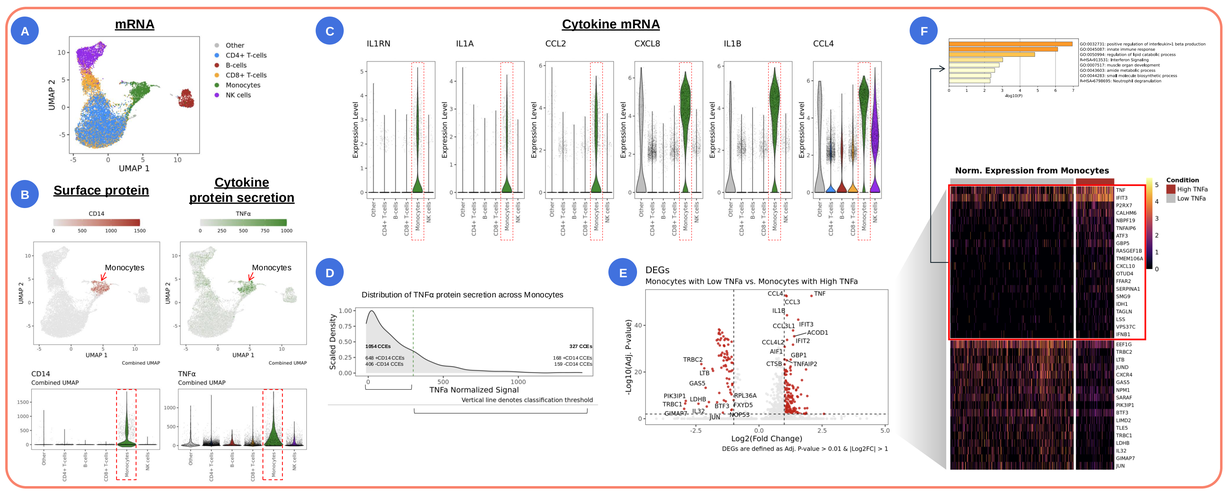
Cell killing assay performed by encapsulating a single effector cell (TALL-104) with one or more target cells (K562). CCEs were observed over a period of 45 hours. Overall, the experimental condition (TALL-104 + K562) shows a higher proportion of dead K562 cells than the negative control (K562 only) (A). Moreover, the number of CCEs with a given amount of dead target cells increases over time, specifically in the experimental condition when compared to the control (B). The data can also be analyzed at the level of individual CCEs: the heatmaps show the proportion of live target cells over time (x-axis) in individual CCEs (y-axis) for the experimental (left) and negative control (right) conditions (C). The heatmap shows the number of CCEs that have a given difference in the number of dead effectors (x-axis) or dead targets (y-axis) when comparing the 45h and 0h timepoint (D).





MULTIMODAL ANALYSIS OF PBMCs FOLLOWING STIMULATION WITH LPS

PBMCs from a healthy donor were rested overnight and stimulated with LPS for 75 minutes. Cell type annotation based on gene expression data was performed using SingleR and reveals an expected distribution of cell types (A). Monocytes (as identified by gene expression) express CD14 (surface receptor assay) and secrete TNFα (cytokine secretion assay) following LPS stimulation, as expected (B). Expression of multiple pro-inflammatory cytokines is enriched specifically in the monocyte compartment (C). Monocytes were divided in two groups according to whether they secrete TNFα or not in response to simulation (D). Differential Expression (DE) analysis between the two groups identifies TNFα as the most over-expressed gene, as expected (E). Gene Set Enrichment Analysis on the full set of DE genes identifies enrichment for pathways associated with IL1-B and the innate immune response (F).



CONCLUSION

Cellanome has developed a platform powered by computer vision, micro-3D printing, and advances in hydrogel chemistry that enables the generation of rich multi-modal longitudinal datasets in live cells, at a scale of tens of thousands of individual cells or groups of interacting cells. Integration of multiple analysis modalities can be used to resolve molecular programs associated with functional readouts of interest in heterogeneous cellular populations.