Self-supervised representation learning enables genomic prediction at single-organoid resolution

Geoffrey F Schau[1], Stanislaw Sydlo[1], Brandon Mapes[1], Elle Moore[1], Timothy Don Lopez[1], Mary Flaherty[1], Michael Streit[1], Justin Guinney[1], Sonal Khare[1], Madhavi Kannan[1], Chi-Sing Ho[1]

[1] Tempus AI, Inc., Chicago, IL

Poster/Abstract #1241

INTRODUCTION

Patient-derived organoids (PDOs) have emerged as compelling cancer models due in part to their conserved properties of tumor heterogeneity and 3D structure of the tumor-immune microenvironment. A principal challenge in high-throughput screening experiments is tracking and labeling the somatic states of TOs during clonal and subclonal selection and expansion in the presence of immune effector cells. In this work, we have developed a self-supervised learning model of tumor organoid morphology in the presence of immune cells and have employed learned features to predict gene mutation state from individual organoid instances.

METHODS

We developed a computational pipeline to segment 1.5 million individual organoids from 14880 brightfield images taken from 62 distinct PDOs as part of the Tempus Pan Cancer Solid Tumor Screen cohort. Each organoid was cropped and processed through a pair of featurization pipelines: one designed to capture human-interpretable features and another to capture learned features from a SimCLR self-supervised representation learning model. Feature sets from both pipelines were employed to predict gene mutation status by a linear logistic regression model.

Figure 1. Laboratory and computational workflow

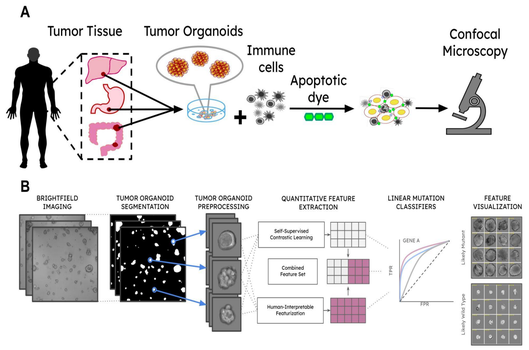

Figure 1. (A) PDOs are plated, treated, and imaged in the presence of immune effector cells. (B) Our computational pipeline segments and measures each organoid's morphology using both human- and machine-interpretable methods.

SUMMARY

- We trained a self-supervised SimCLR computer vision model for learning meaningful representations of organoid morphology in the absence of human labels while simultaneously employing human-interpretable features to capture structural content of the organoids
- Learned morphological features were validated through a clinically-relevant task by predicting gene mutation state in organoids stratified by cancer type
- We observed different utilities for different feature sets depending on the gene being predicted, and identified the best-performing predictor as one which predicts *KRAS* mutations from a combined feature set (AUROC = 0.742)

RESULTS

Figure 2. SimCLR for representation learning

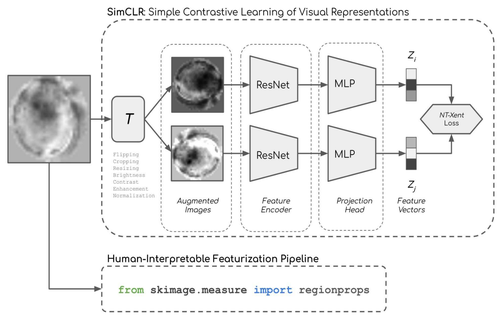

Figure 2. SimCLR learning architecture and regionprops invocation for robust organoid featurization

Figure 4. UMAP projection of learned organoid features

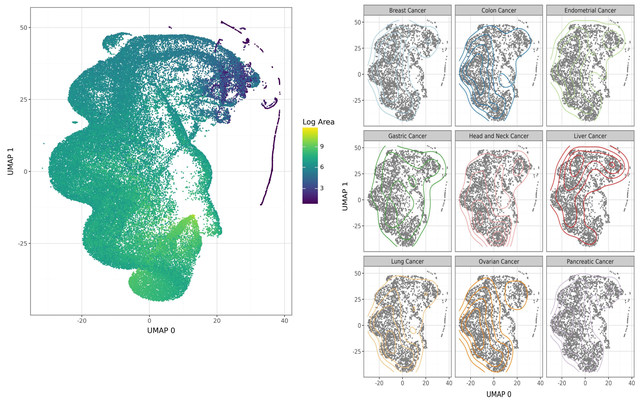

Figure 4. Universal manifold projection (UMAP) of a combined feature set composed of both human-interpretable features and self-supervised learning morphology features colored by organoid area (Left). Same UMAP projection colored by cancer type (Right).

Figure 6. Organoids by probable mutation state

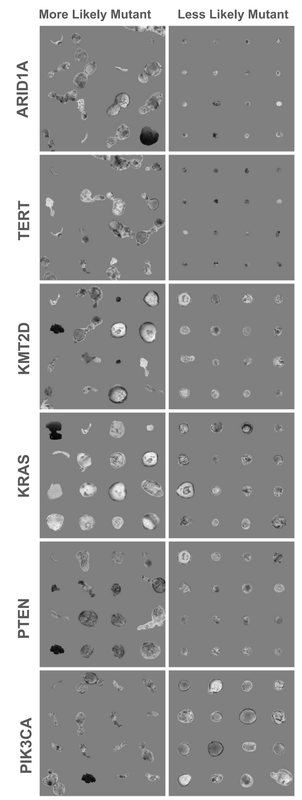

Figure 6. Each 4x4 mosaic depicts those organoids deemed most- and least-likely as having a specific mutation based on the predicted logic from the SSL feature models for each gene. Generally, organoids predicted to be mutants are larger, more heterogeneous, and more textured than those more likely to be wild-type, which are in general smaller and rounder than their likely mutant counterparts.

Figure 3. Distribution of segmented organoids from negative control conditions (Left) Mutation calls available for classification (Right)

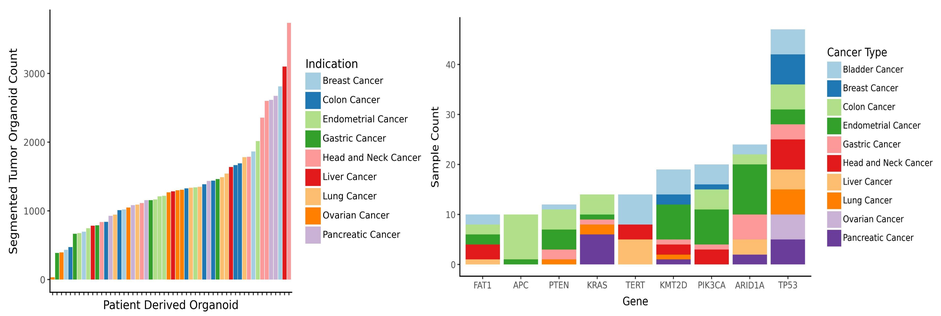

Figure 3. Distribution of detected gene mutations present in this cohort colored by cancer type (Left) . Each predictive model was stratified its training, and test split by cancer type and mutation state. Total counts of segmented tumor organoids across each of the sixty organoid lines used in this study (Right).

Figure 5. AUROC for gene targets by feature set

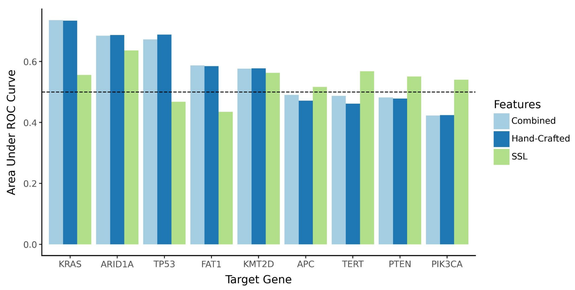

Figure 5. The height of each bar represents the area under the receiver operator characteristics curve (AUROC) for a logistic model trained on each of three feature sets stratified by cancer type. The horizontal dotted line at 0.5 represents a random model.

ACKNOWLEDGMENTS

We thank Amrita A. Iyer from the Tempus Science Communications team for poster development

Correspondence: geoffrey.schau@tempus.com