

# Supplemental Materials

*Molecular Biology of the Cell*

Way *et al.*

# Predicting cell health phenotypes using image-based morphology profiling

Gregory P. Way<sup>1,+</sup>, Maria Kost-Alimova<sup>2,+</sup>, Tsukasa Shibue<sup>2</sup>, William F. Harrington<sup>2</sup>, Stanley Gill<sup>2,4</sup>,  
Federica Piccioni<sup>3,5</sup>, Tim Becker<sup>1</sup>, Hamdah Shafqat-Abbasi<sup>1</sup>, William C. Hahn<sup>2,4</sup>, Anne E. Carpenter<sup>1,^</sup>,  
Francisca Vazquez<sup>2,^</sup>, Shantanu Singh<sup>1,^</sup>

1. Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
2. Cancer Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
3. Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
4. Dana-Farber Cancer Institute, Department of Medical Oncology, Harvard Medical School, Boston, Massachusetts, USA
5. Present address: Merck & Co., Inc., Boston, Massachusetts, USA (F.P.)

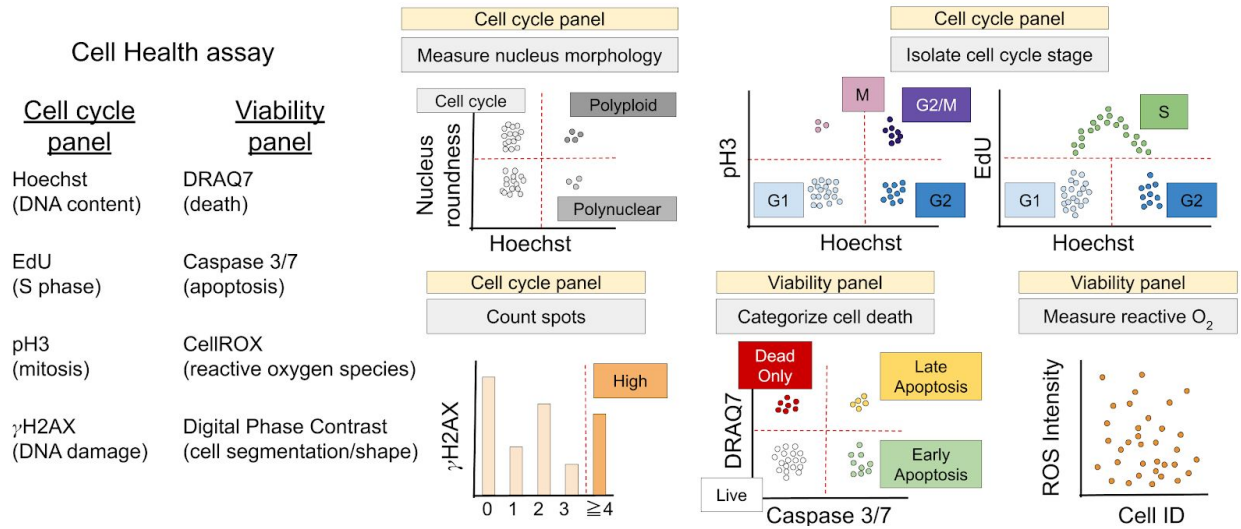
+Co-First Authors

^Co-Senior Authors

Corresponding Authors: G.P.W. ([gregory.way@gmail.com](mailto:gregory.way@gmail.com)), S.S. ([shsingh@broadinstitute.org](mailto:shsingh@broadinstitute.org))

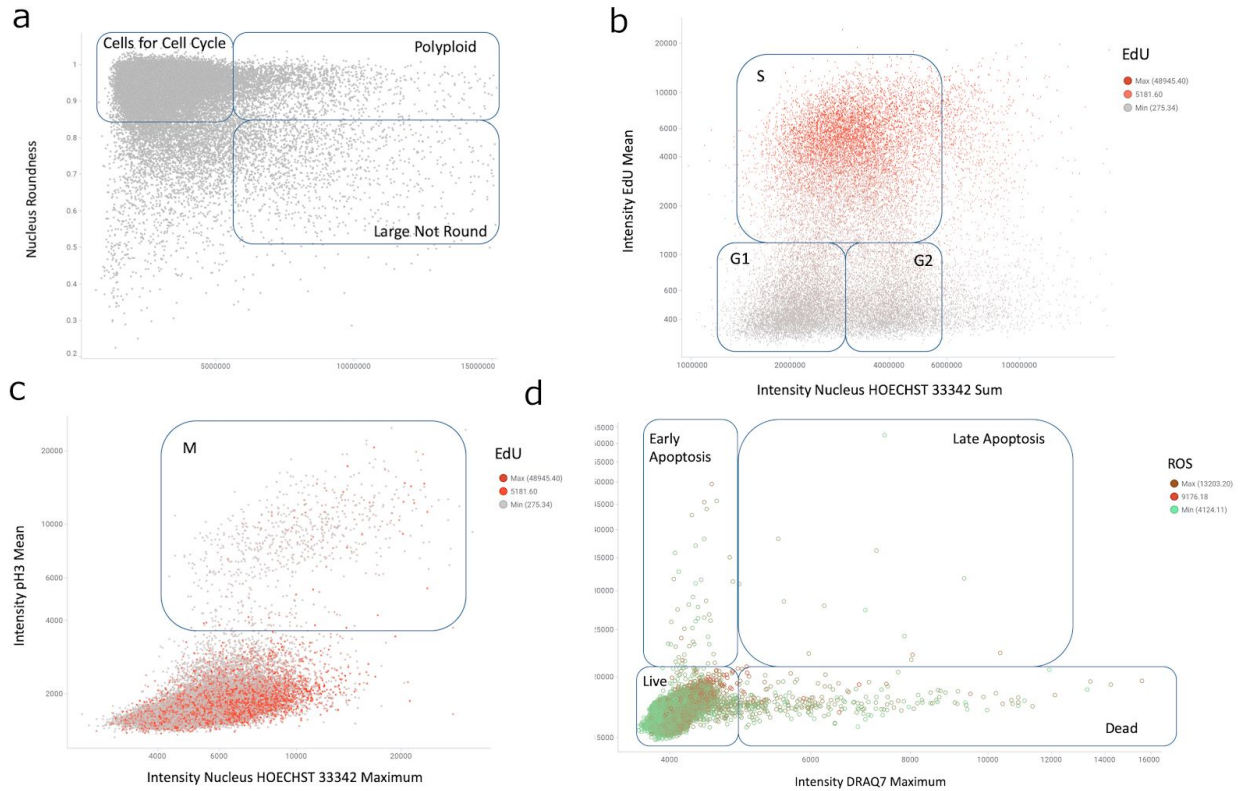
<b>Supplementary Figure S1. Illustration of the gating strategy in the Cell Health assays.</b>	<b>30</b>
<b>Supplementary Figure S2. Real data of manual gating in the Cell Health assays.</b>	<b>31</b>
<b>Supplementary Figure S3. Replicate correlation of CRISPR perturbations.</b>	<b>32</b>
<b>Supplementary Figure S4. Performance of predicting 70 cell health variables with independent regression models.</b>	<b>33</b>
<b>Supplementary Figure S5. Summarizing model performance compared to permuted data.</b>	<b>34</b>
<b>Supplementary Figure S6. Examples of six models in predicting cell health assay readouts.</b>	<b>35</b>
<b>Supplementary Figure S7. Model performance according to specific cell health reagents.</b>	<b>36</b>
<b>Supplementary Figure S8. Exploring the relationship between CRISPR infection efficiency and regression model performance.</b>	<b>37</b>
<b>Supplementary Figure S9. Regression model coefficients for predicting live cell area from digital phase contrast (DPC) measurements.</b>	<b>38</b>

<b>Supplementary Figure S10. Summarizing Cell Painting feature importance scores for all 70 cell health regression models.</b>	<b>39</b>
<b>Supplementary Figure S11. Results from a cell line holdout analysis.</b>	<b>40</b>
<b>Supplementary Figure S12. Systematically removing classes of features has little impact on most models' performance.</b>	<b>41</b>
<b>Supplementary Figure S13. Dropping samples from training reduces test set model performance in high, mid, and low performing models.</b>	<b>42</b>
<b>Supplementary Figure S14. Applying a Uniform Manifold Approximation (UMAP) to Drug Repurposing Hub consensus profiles of 1,571 compounds across six doses.</b>	<b>43</b>
<b>Supplementary Figure S15. Applying a Uniform Manifold Approximation (UMAP) to the Cell Painting consensus profile data of CRISPR perturbations.</b>	<b>44</b>



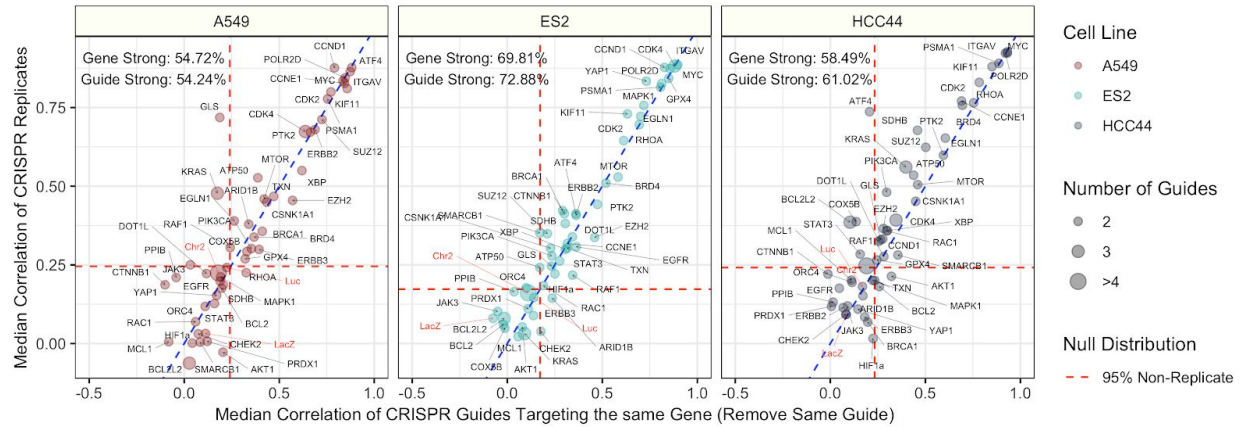
**Supplementary Figure S1.** *Illustration of the gating strategy in the Cell Health assays.*

We extract 70 different readouts from the Cell Health imaging assays. The assay consists of two customized reagent panels, which use measurements from seven different targeted reagents and one channel based on digital phase contrast (DPC) imaging; shown are five toy examples to demonstrate that individual cells are isolated into subpopulations by various gating strategies to define the Cell Health readouts.



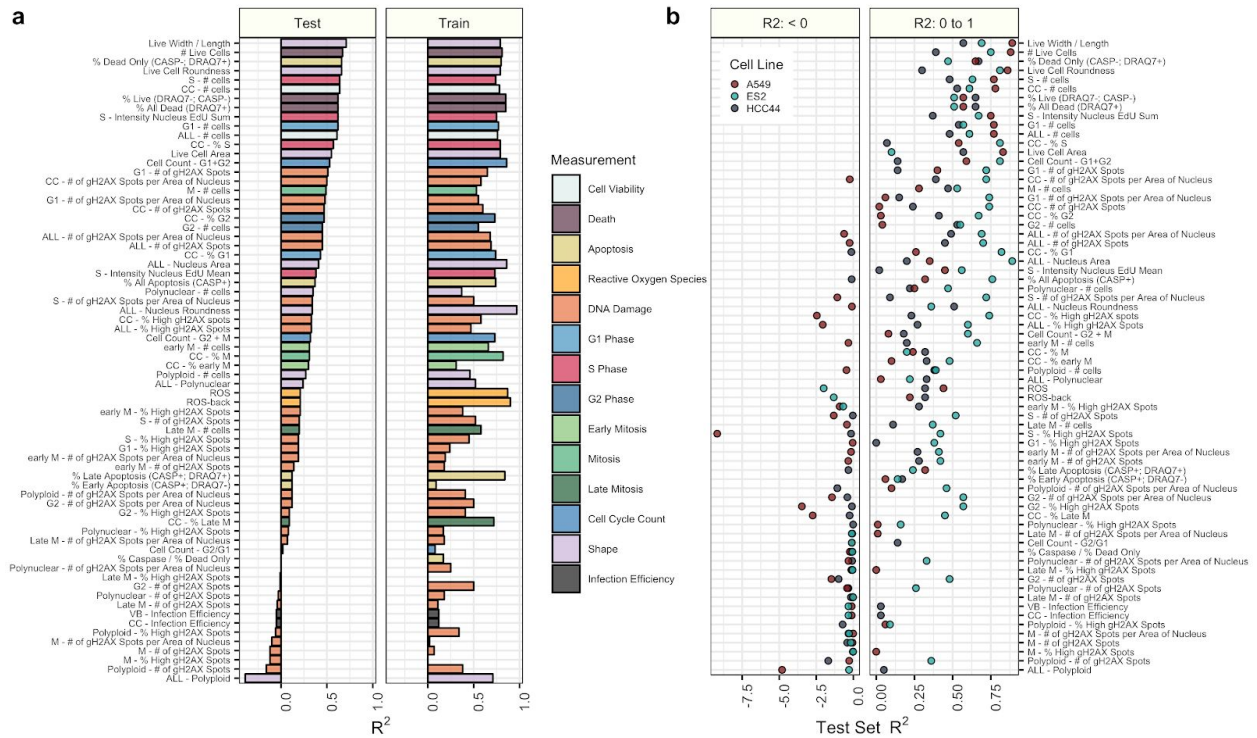
**Supplementary Figure S2.** Real data of manual gating in the Cell Health assays.

For each cell line, we apply a series of manual gating strategies defined by various stain measurements in single cells to define cell subpopulations. **(a)** In the cell cycle panel, we first select cells that are useful for cell cycle analysis based on nucleus roundness and Hoechst intensity measurements. We also identify polyploid and “large not round” (polynuclear) cells. **(b)** We then subdivide the cells used for cell cycle to G1, G2, and S cells based on total Hoechst intensity (DNA content) and EdU incorporation signal intensity. **(c)** We use Hoechst and PH3 nucleus intensity to define mitotic cells. The points are colored by EdU intensity in the nucleus in both (b) and (c). **(d)** Example gating in the viability panel. We use DRAQ7 and CellEvent (Caspase 3/7) to distinguish alive and dead cells, and categorize early or late apoptosis. See Methods for more details about how the Cell Health measurements are made.



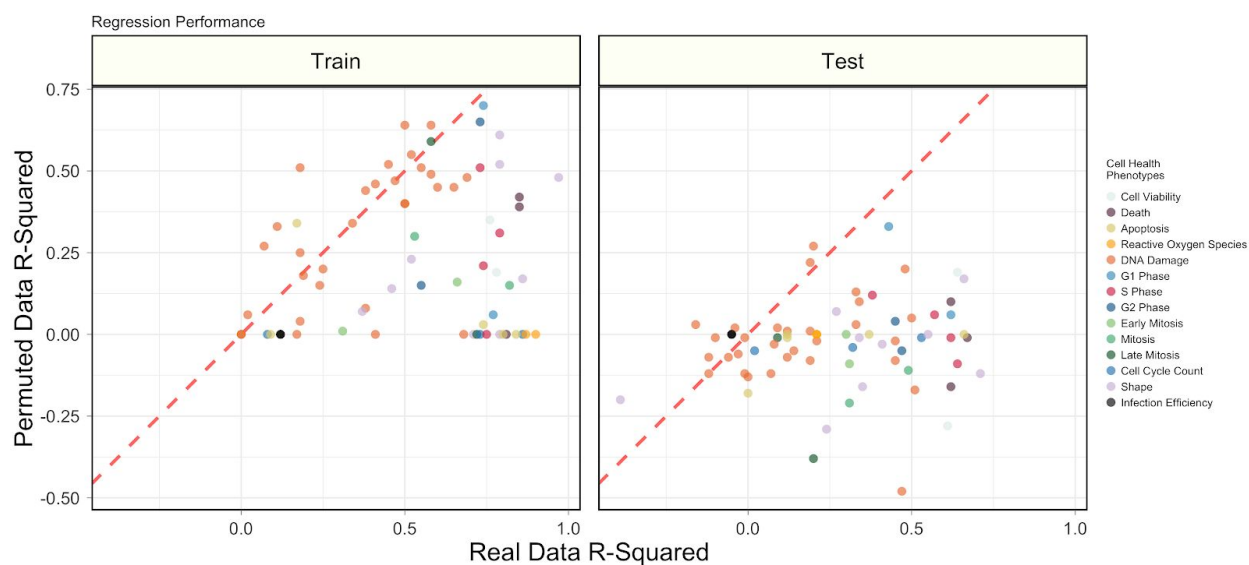
**Supplementary Figure S3. Replicate correlation of CRISPR perturbations.**

Median pairwise Pearson correlation of CRISPR guide replicate profiles (y axis) compared against Median pairwise Pearson correlation of CRISPR guides targeting the same gene or construct (x axis). We removed biological replicates when calculating the same-gene correlations. The three different cell lines (A549, ES2, and HCC44) are shown in different colors and in different facets of the figure. We generated the profiles by median aggregating CellProfiler measurements for all single cells within each well of a Cell Painting experiment (see Methods for more processing details). The text labels represent the proportion of gene and guide profiles with “strong phenotypes”. In other words, these profiles had replicate correlations greater than 95% of non-replicate pairwise Pearson correlations in the particular cell line. The dotted red line represents this 95% cutoff in the null distribution and the blue dotted line is  $y = x$ , which shows a strong consistency across CRISPR guide constructs.



**Supplementary Figure S4. Performance of predicting 70 cell health variables with independent regression models.**

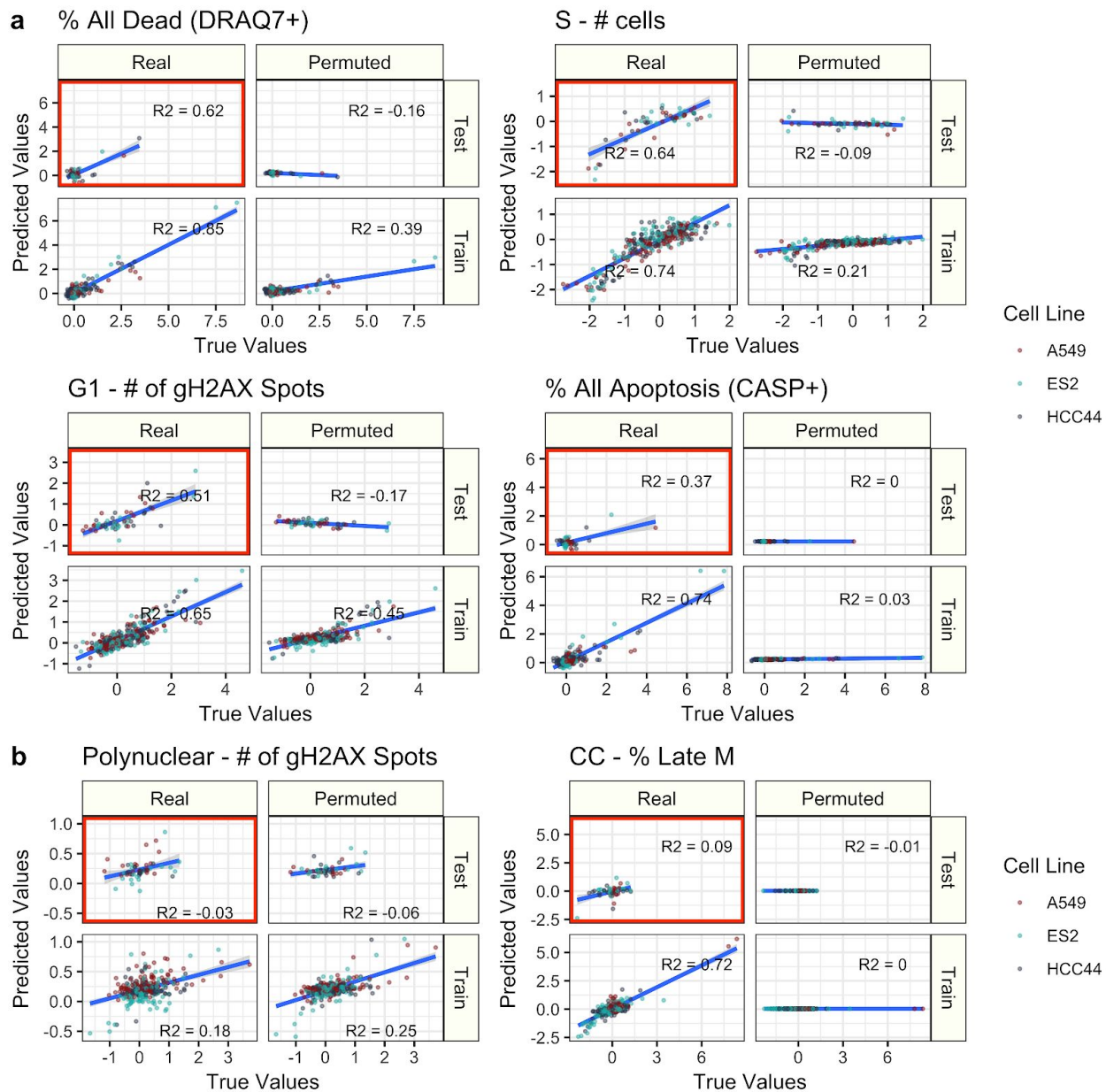
(a) Testing and training performance for each phenotype is shown, sorted by decreasing  $R^2$  performance on the test set, aggregated across the three cell lines and colored based on the primary measurement metadata (see Supplementary Table S3). (b) Test set  $R^2$  performance for each cell line independently, with the same ordering as in (a). Performance is highly variable across cell lines, with ES2 having the highest performance for most models. Note that the left and right facets of b have different x-axis scales and that the  $R^2$  values can lie below zero for the test set because the model is learned using the training set.



**Supplementary Figure S5.** *Summarizing model performance compared to permuted data.*

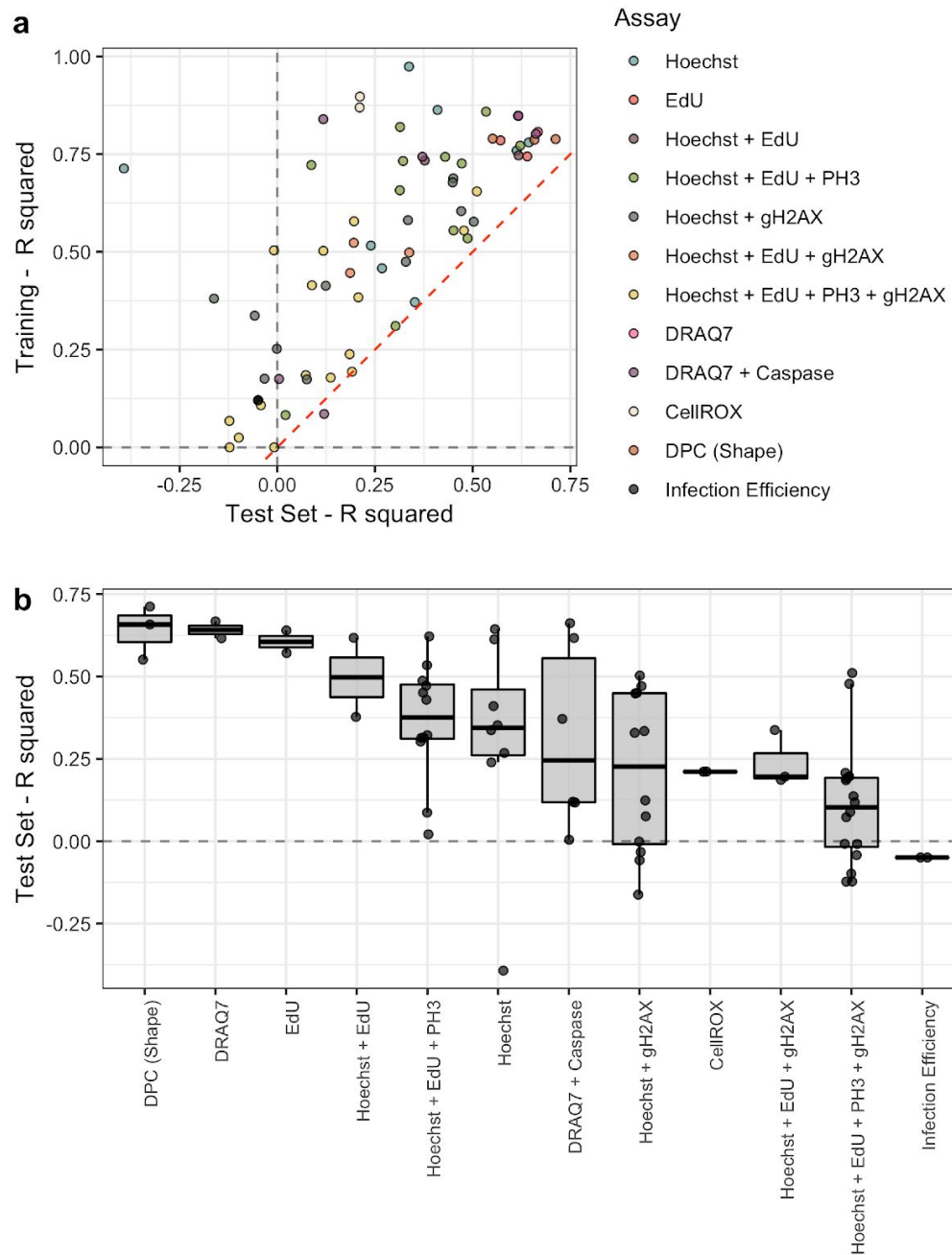
Performance of all 70 Cell Health assay models in train and test sets using real and permuted data. Data were permuted by randomly shuffling observed morphology features for each column independently (see Methods). Each point represents a Cell Health model, and the color represents the specific phenotype.





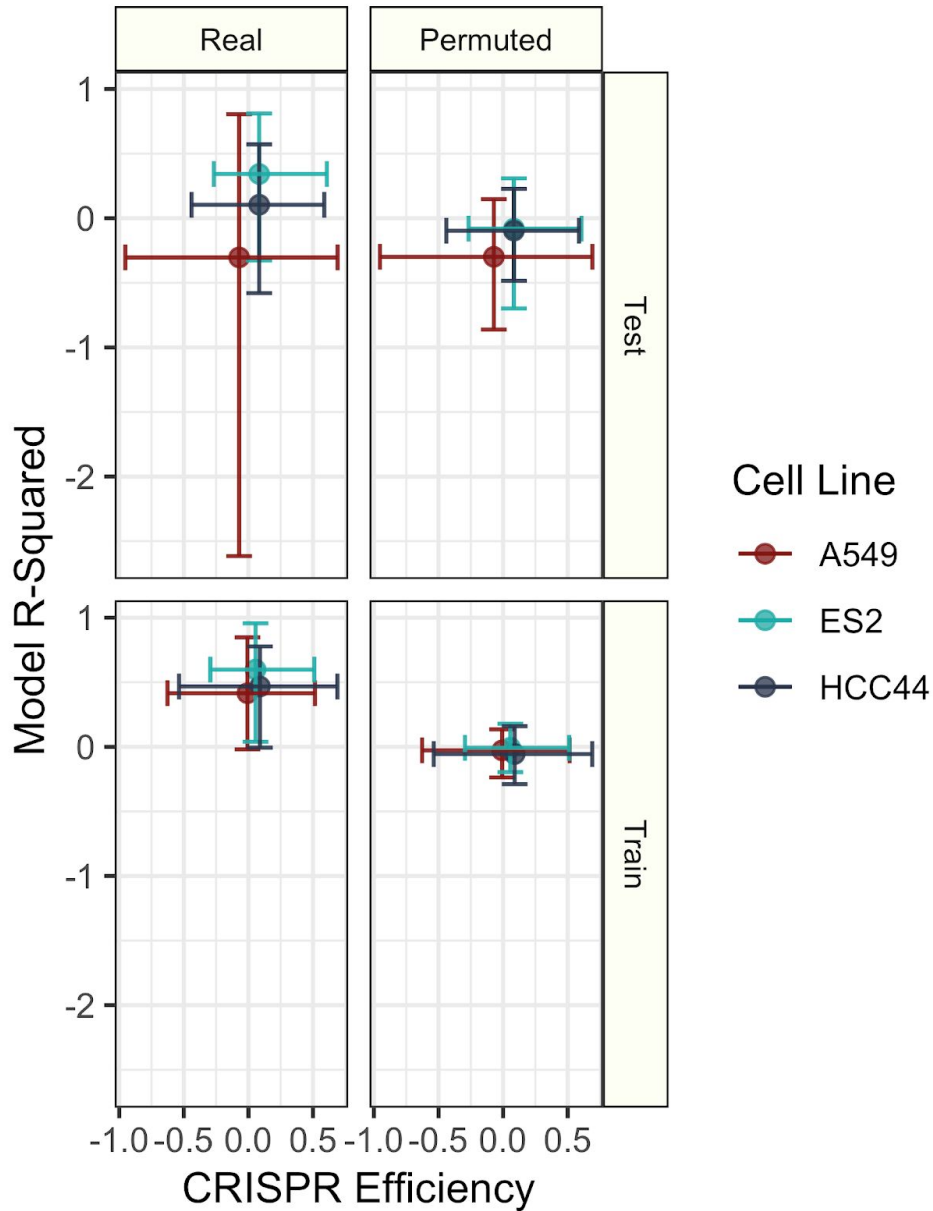
**Supplementary Figure S6. Examples of six models in predicting cell health assay readouts.**

**(a)** Four high performing models chosen to span different Cell Health assay readouts. **(b)** Predicting two example low performing models. R-squared ( $R^2$ ) was calculated using `sklearn.metrics.r2_score` and is bound by negative infinity to 1. The blue lines represent a smoothed linear model fit and the shading represents a 95% confidence interval.



**Supplementary Figure S7. Model performance according to specific cell health reagents.**

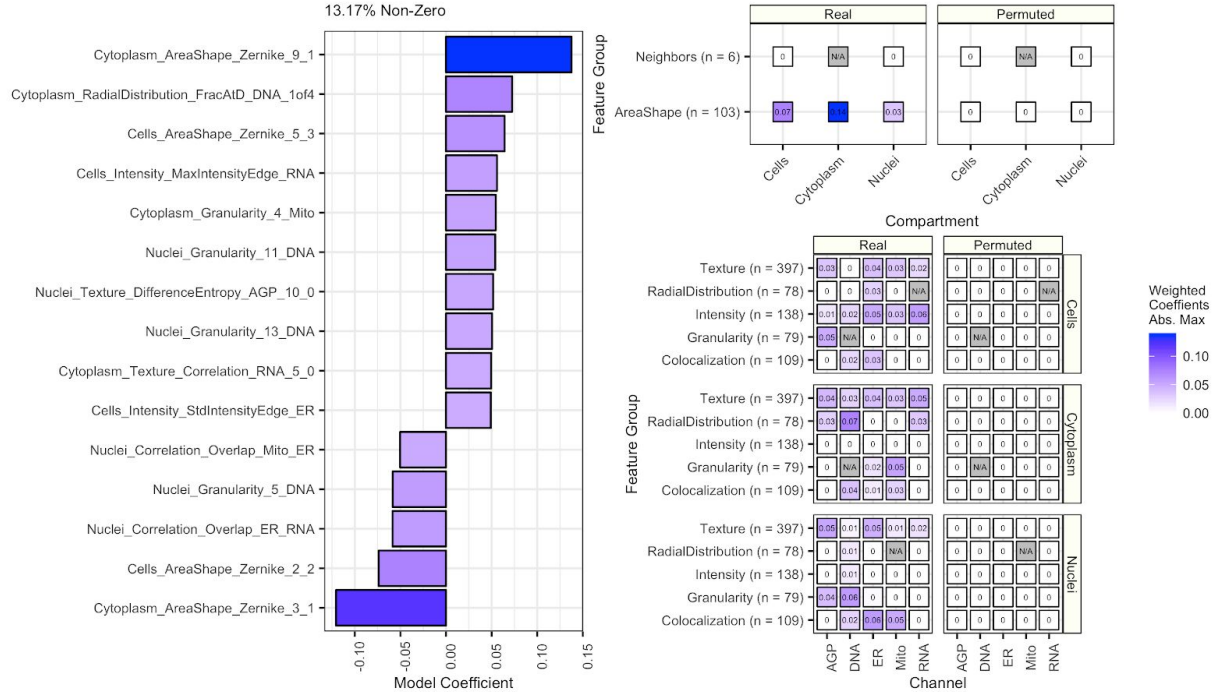
**(a)** Training and testing model performance based on  $R^2$ . The dotted red line is the line  $y = x$  and depicts a small amount of model overfitting. **(b)** Test set  $R^2$  grouped by cell health reagents used to form the cell health indicators. The cell health variables are sorted by median test set  $R^2$  performance.



**Supplementary Figure S8.** Exploring the relationship between CRISPR infection efficiency and regression model performance.

Cell line specific training and testing model  $R^2$  performance in real and permuted data compared against CRISPR infection efficiency readouts. Infection efficiency is measured by comparing cell count in wells treated with and without puromycin (see Methods). We generated infection efficiency measurements for all individual wells and then aggregated by MODZ to form consensus measurements (see Methods). The points represent mean values and the extended bars represent 5% and 95% of the observed cell-line specific distributions. ES2 has the highest test set model  $R^2$  and the highest CRISPR efficiency.

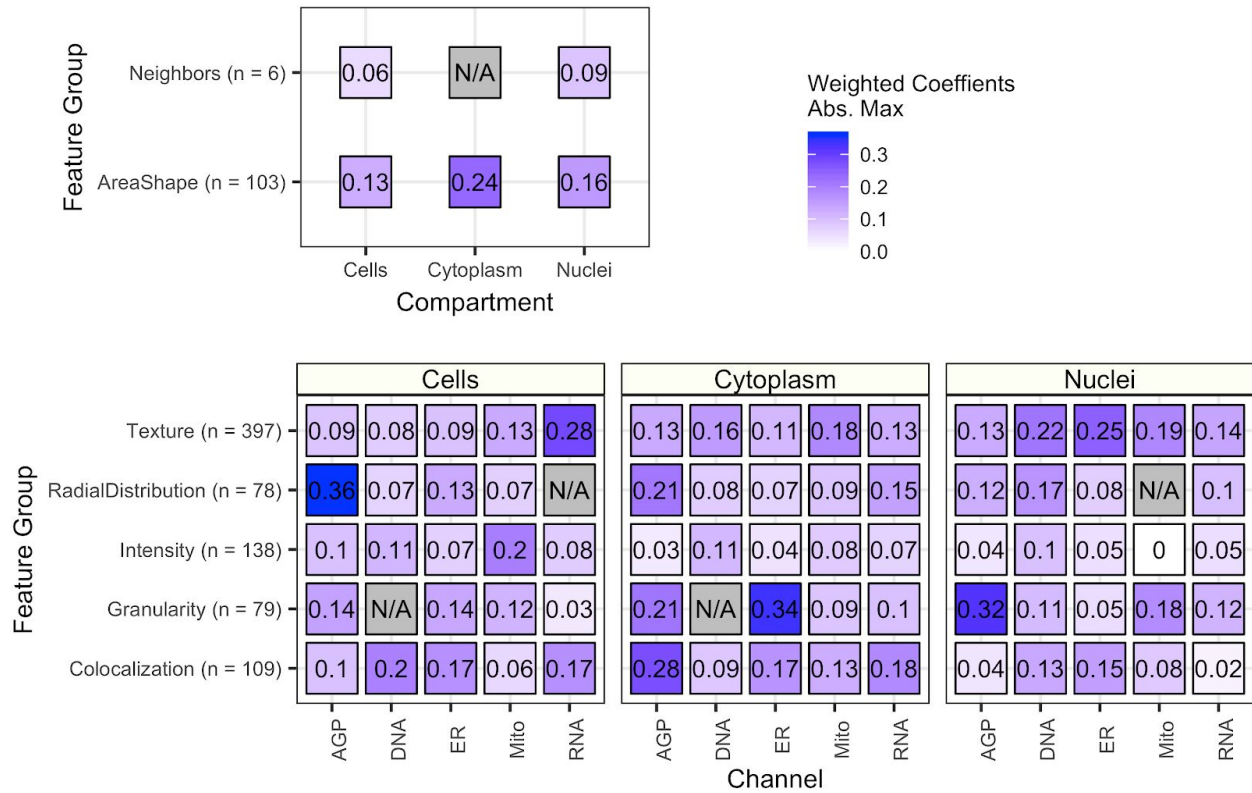
## Live Cell Area



**Supplementary Figure S9.** Regression model coefficients for predicting live cell area from digital phase contrast (DPC) measurements.

We use the Cell Painting measurements to predict live cell area readouts from DPC measurements as a positive control. As expected, cytoplasm and cell shape contribute to live cell area predictions. **(left)** The top 15 most influential Cell Painting features by absolute value model coefficient. It is important to note that because the machine learning procedure automatically removes many features, not all explanatory features are selected. **(right)** The maximum absolute value model coefficient (weight) for compartments, channels, and feature groups. Coefficients for a model trained with real data is contrasted with a model trained with permuted data. For a complete description of all features, see the handbook:

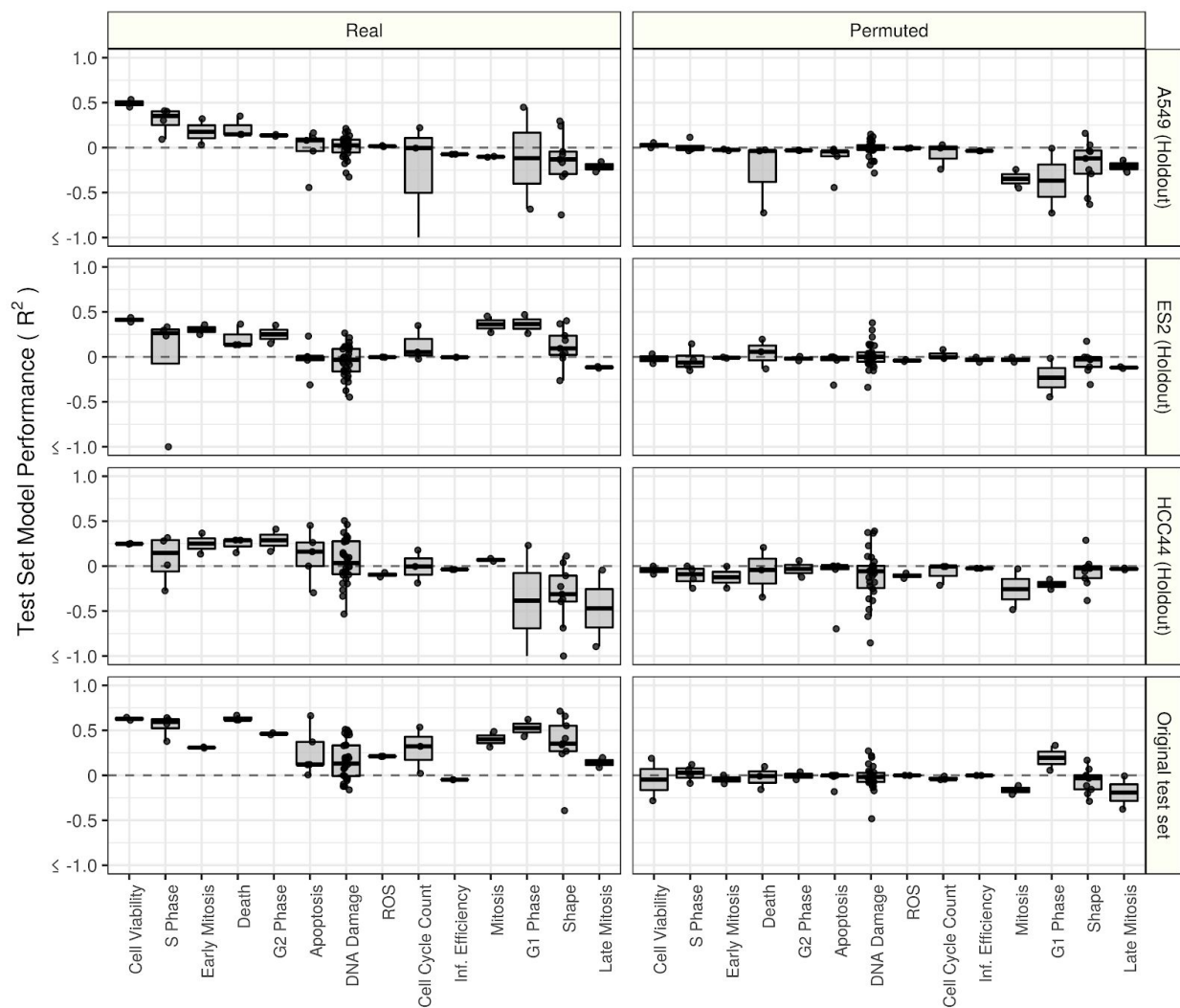
<http://cellprofiler-manual.s3.amazonaws.com/CellProfiler-3.0.0/index.html>



**Supplementary Figure S10.** Summarizing Cell Painting feature importance scores for all 70 cell health regression models.

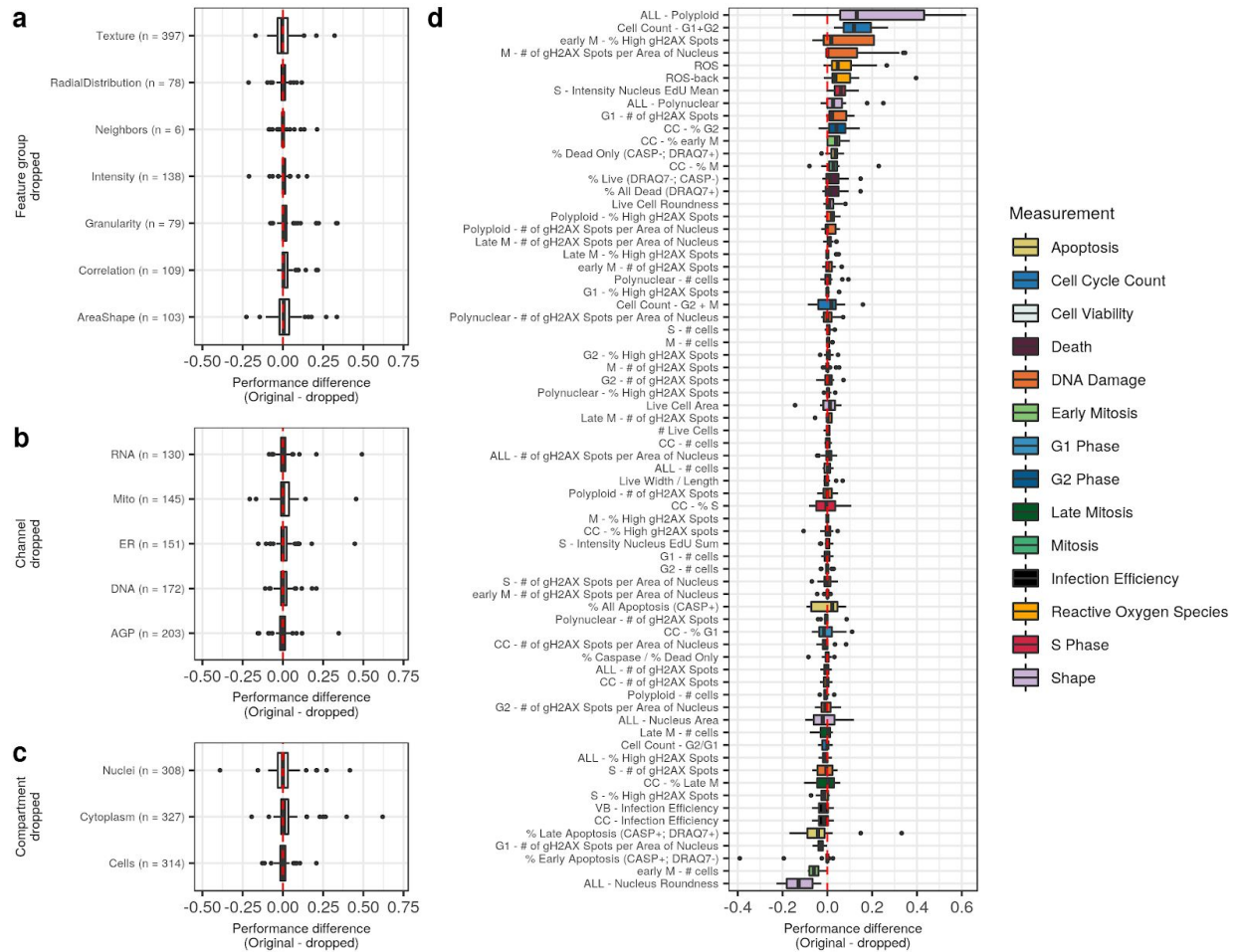
Each point represents the maximum absolute value of the model coefficients weighted by test set  $R^2$  across all Cell Health models. The features are broken down by compartment (Cells, Cytoplasm, and Nuclei), channel (AGP, Nucleus, ER, Mito, Nucleolus/Cyto RNA), and feature group (AreaShape, Neighbors, Channel Colocalization, Texture, Radial Distribution, Intensity, and Granularity). For a complete description of all features, see the handbook:

<http://cellprofiler-manual.s3.amazonaws.com/CellProfiler-3.0.0/index.html>. Dark gray squares indicate “not applicable”, meaning either that there are no features in the class or the features did not survive an initial preprocessing step.



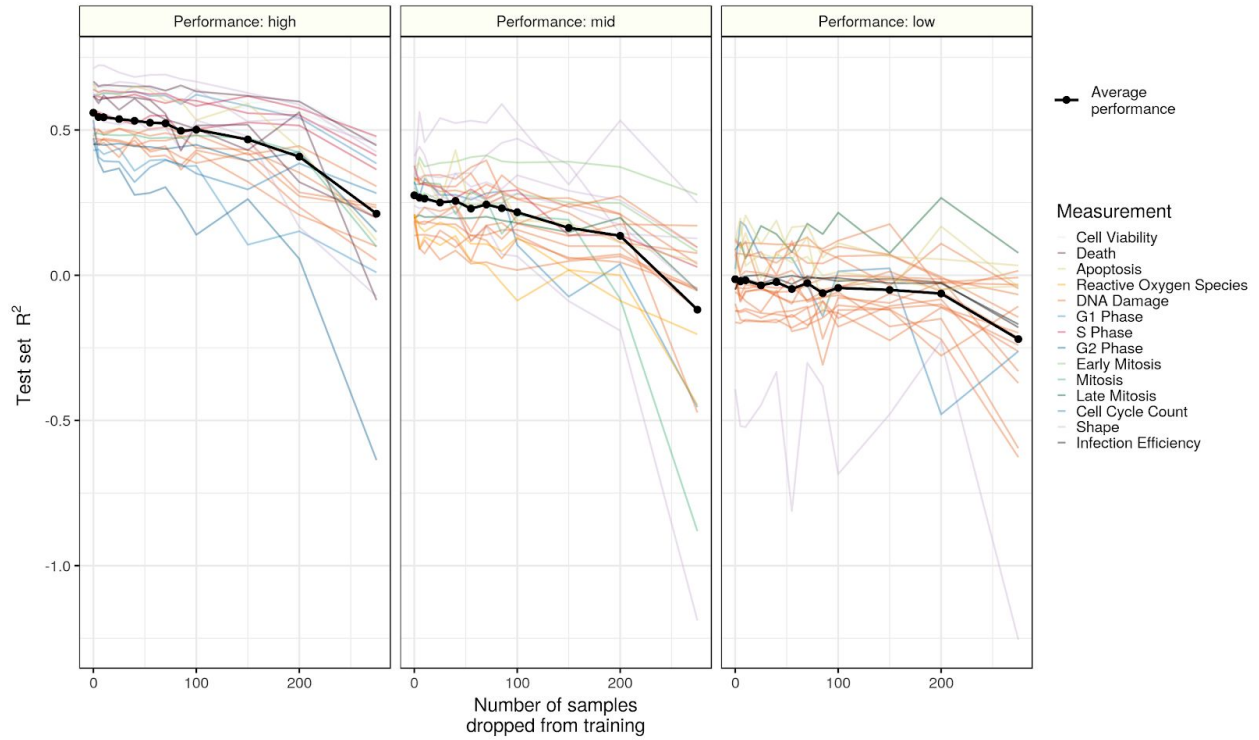
**Supplementary Figure S11.** Results from a cell line holdout analysis.

We trained and evaluated all 70 cell health models in three different scenarios using each combination of two cell lines to train, and the remaining cell line to evaluate. For example, we trained all 70 models using data from A549 and ES2 and evaluated performance in HCC44. We bin all cell health models into 14 different categories (see Supplementary Table S3 and <https://github.com/broadinstitute/cell-health/6.ml-robustness> for details about the categories and scores). We also provide the original test set (15% of the data, distributed evenly across all cell types) performance in the last row, as well as results after training with randomly permuted data. This cross-cell-type analysis yields worse performance overall. Nevertheless, despite the models never encountering certain cell lines, and having fewer training data points, many models still have predictive power across cell line contexts. Note that we truncated the y axis to remove extreme outliers far below -1. The raw scores are available on <https://github.com/broadinstitute/cell-health>.



**Supplementary Figure S12.** *Systematically removing classes of features has little impact on most models' performance.*

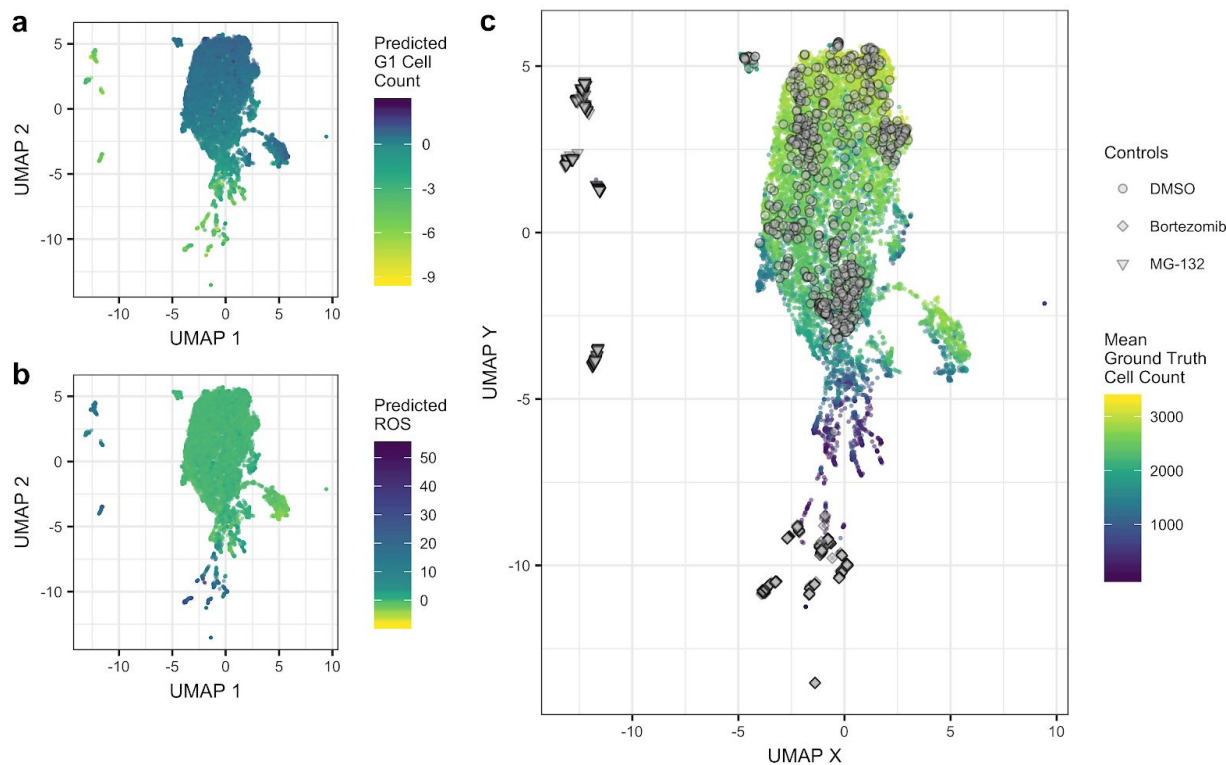
We retrained all 70 cell health models after dropping features associated with specific **(a)** feature groups, **(b)** channels, and **(c)** compartments. Each dot is one model (predictor), and the performance difference between the original model and the retrained model after dropping features is shown on the x axis. Any positive change indicates that the models got worse after dropping the feature group. **(d)** Individual model differences in performance after dropping features. Each dot is one class of features removed (as in a-c).



**Supplementary Figure S13.** *Dropping samples from training reduces test set model performance in high, mid, and low performing models.*

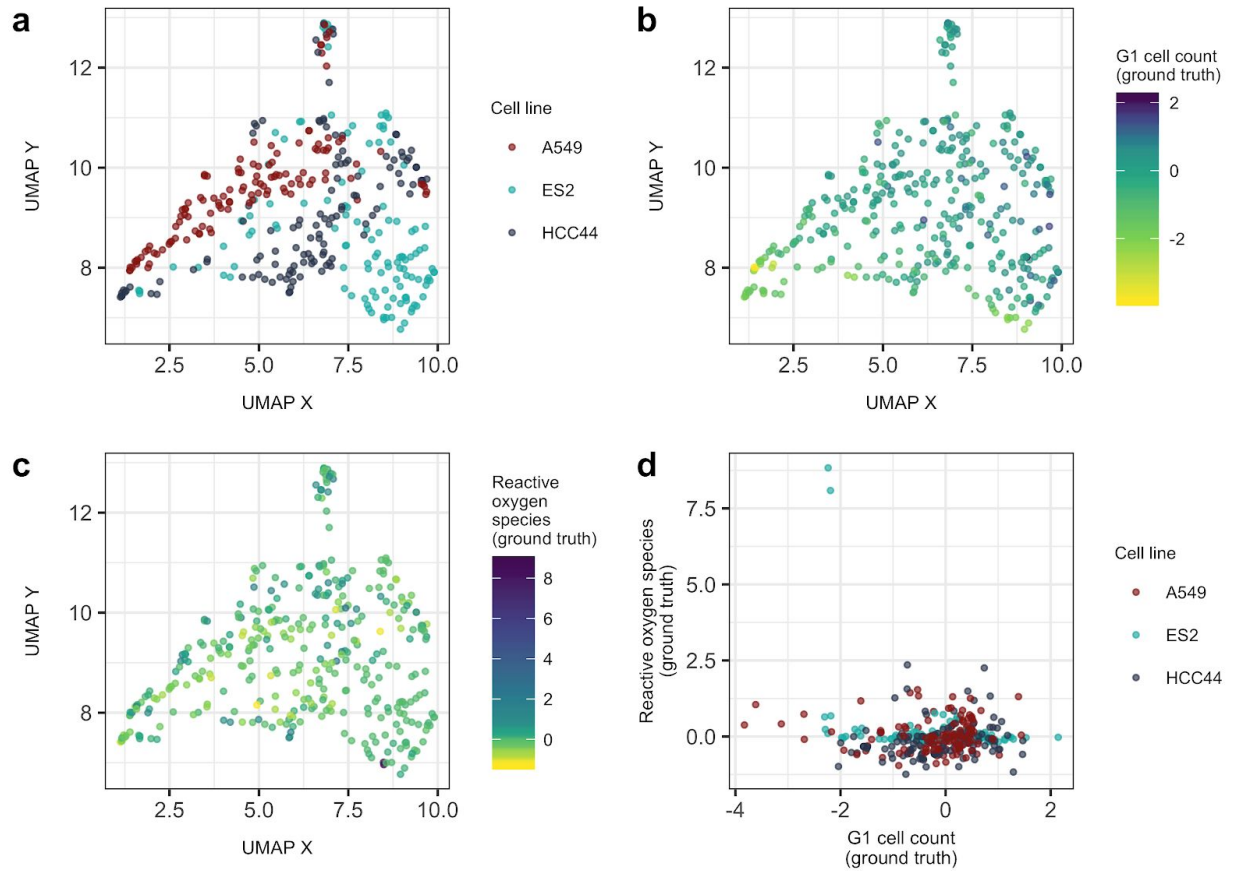
We determined model performance stratification by taking the top third, mid third, and bottom third of test set performance when using all data. We performed the sample titration analysis with 10 different random seeds and visualized the median test set performance for each model.





**Supplementary Figure S14.** *Applying a Uniform Manifold Approximation (UMAP) to Drug Repurposing Hub consensus profiles of 1,571 compounds across six doses.*

The models were not trained using the Drug Repurposing Hub data. **(a)** The point color represents the output of the Cell Health model trained to predict the number of cells in G1 phase (*G1 cell count*). **(b)** The same UMAP dimensions, but colored by the output of the Cell Health model trained to predict reactive oxygen species (*ROS*). **(c)** In the UMAP space, we highlight DMSO as a negative control, and Bortezomib and MG-132 as two positive controls (proteasome inhibitors) in the Drug Repurposing Hub set. We observe moderate batch effects in the negative control DMSO profiles, based on their spread in this visualization. The color represents the predicted number of live cells. The positive controls were acquired with a very high dose and are expected to result in a very low number of predicted live cells.



**Supplementary Figure S15.** *Applying a Uniform Manifold Approximation (UMAP) to the Cell Painting consensus profile data of CRISPR perturbations.*

UMAP coordinates visualized by **(a)** cell line, **(b)** ground truth G1 cell counts, and **(c)** ground truth ROS counts. **(d)** Visualizing the distribution of ground truth ROS compared against G1 cell count. The two outlier ES2 profiles are CRISPR knockdowns of GPX4, which is known to cause high ROS.