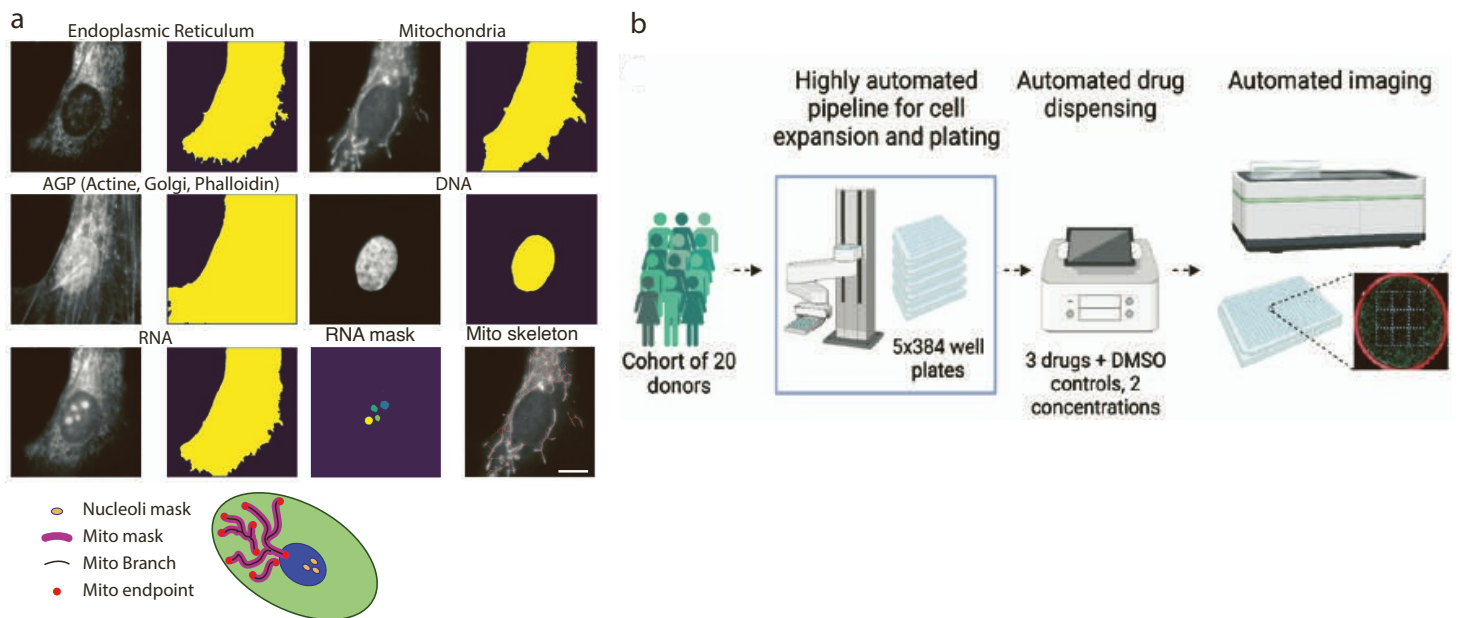


iScience, Volume 27

Supplemental information

**A highly efficient, scalable pipeline for fixed
feature extraction from large-scale
high-content imaging screens**

**Gabriel Comolet, Neeloy Bose, Jeff Winchell, Alyssa Duren-Lubanski, Tom
Rusielewicz, Jordan Goldberg, Grayson Horn, Daniel Paull, and Bianca Migliori**



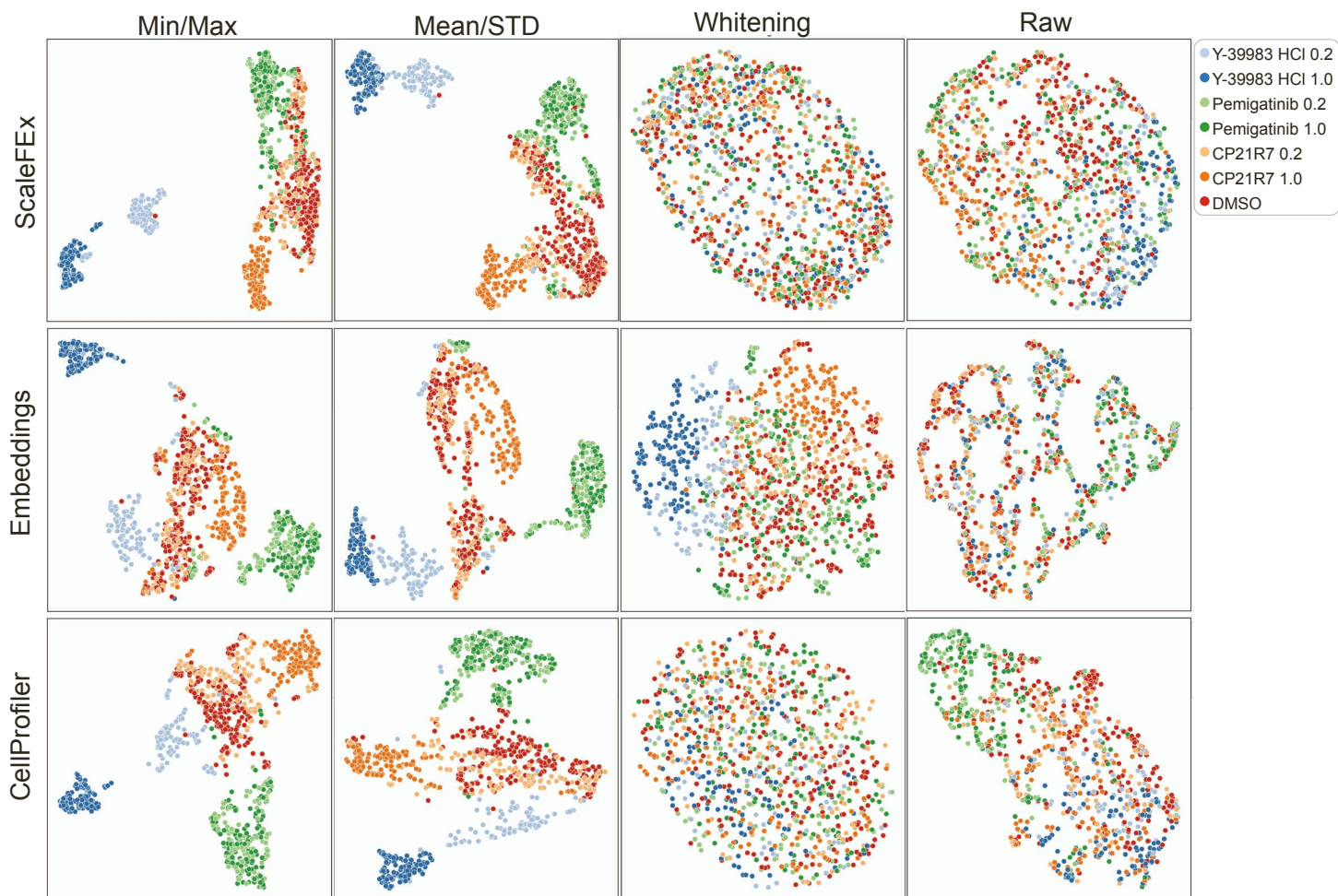
Supplementary Figure 1: Details of the computed cell masks and Mitochondria and RNA-specific features, and overview of the experimental pipeline

Related to Figure 1

a) Example of segmentation masks for each channel in a single cell and additional segmentation masks used for RNA and Mitochondria (skeleton only) and illustration to describe the additional features related to Mitochondria and RNA that are computed. An additional mask was first computed to segment the sub-structures of the channel. For RNA, the values related to the number of Nucleoli and volume were calculated, shown in the mask (bottom right, panel b - RNA Mask).

For mitochondria, measurements related to volume on the mask (purple portion in the schematics) are first computed, before computing the mask's skeleton and subsequently branches, endpoints, total length and number. Scale bar: 10 μ m

b) Experimental pipeline: Using a collection of skin fibroblasts from a cohort of 20 donors, 5, reproducibly seeded, 384 well plates were generated for compound profiling. Three drugs at 2 different concentrations were applied, alongside DMSO controls via automated dispensing. After 72 hours of growth in the presence of the drugs, the cells were fixed, Cell-Painted and imaged using an automated microscope.



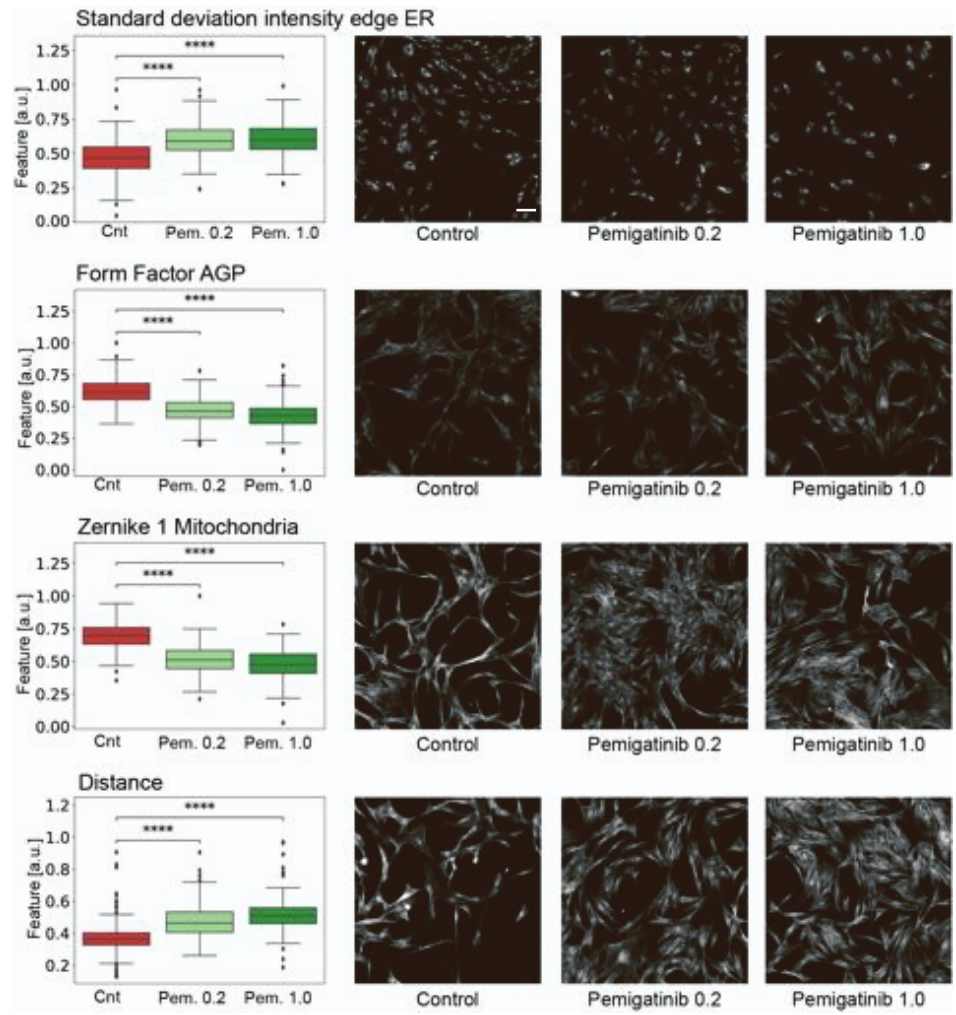
Supplementary Figure 2: UMAP of the feature vectors color-coded by drug and condition. Related to STAR methods section 6.

Each column shows a different normalization method, and each row is for each different model.

Pemigatinib 0.2 μ M							
ER	1	0	0	0	2	1	0
Mito	0	0	0	0	0	0	0
AGP	1	0.5	0	0.5	0	0	0
DNA	1	0	0	0	0	1	0
RNA	2	0	1	0.5	1.5	1	0
None	0	0	0	0	0	0	2

Pemigatinib 1.0 μ M							
ER	1	0	0	0	3	4	0.5
Mito	0	0	0	0	0	0	0
AGP	1	0	0	0	1	1	0.5
DNA	0	0	1	0	0	3	0
RNA	2	0	0	0	1	2	0
None	0	0	0	0	0	0	2

	Shape	Granularity	Texture	Concentric	Intensity	Zernike	Channel_correlation	Global
--	-------	-------------	---------	------------	-----------	---------	---------------------	--------



Supplementary Figure 3: Value distribution and image visualization of the most important features to classify Pemigatinib at 0.2 μ M and 1.0 μ M.

Related to Figure 3.

The pictures depict the wells with the closest value to the mean of the feature. In interest of space, we only visualize 9 out of the 49 total tiles that make a well.

In the heatmaps, the darker the color, the higher the number of features. The number states the actual number of features per class and channel. The half numbers are for those features with double class or double channel. The feature distribution values are all normalized between 0 and 1.

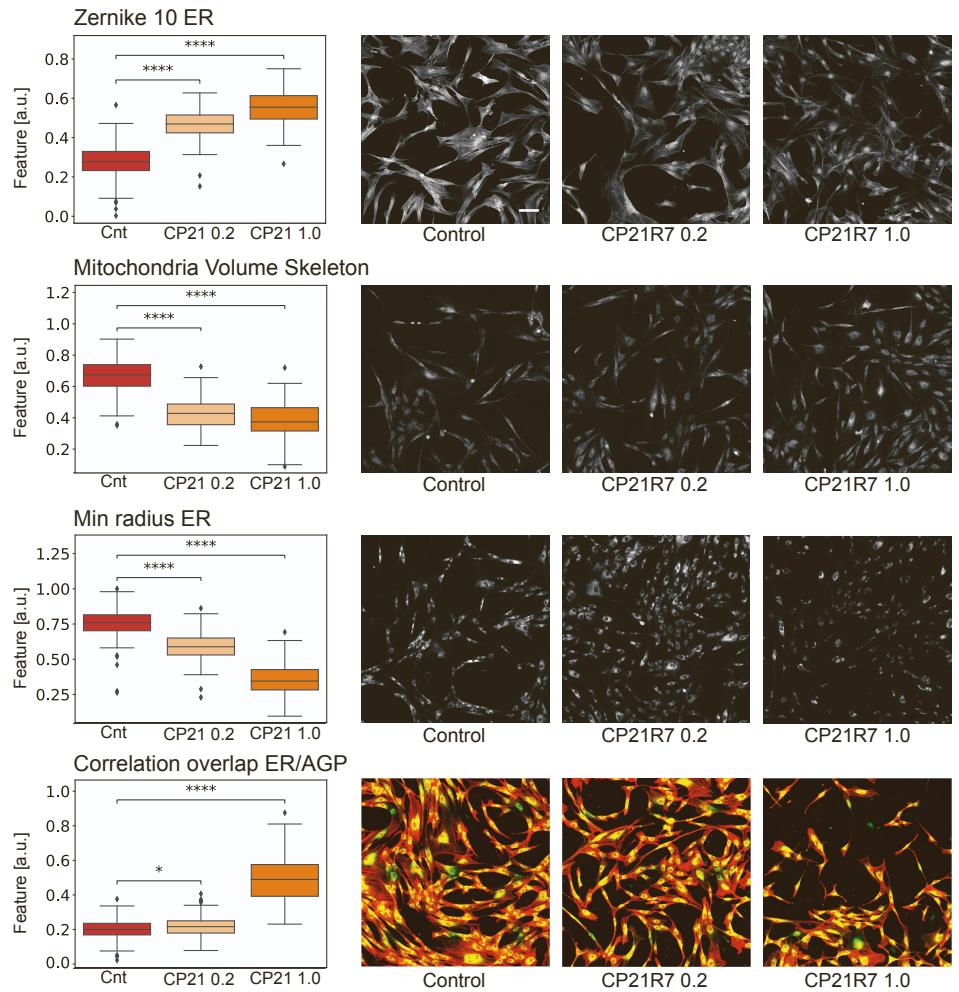
Scale bar: 100 μ m.

The box plots components are: horizontal line, median; box, interquartile range; whiskers, 1.5 \times interquartile range. arb. units: arbitrary units. Two-sided Mann–Whitney U test: ns: $P > 0.05$; $0.01 < P \leq 0.05$; $****P \leq 0.0001$.

		CP21R7 0.2 μ M							
ER	0	0	0	0	1	2	0.5	0	
Mito	3	0	1	0	0	0	0	0	
AGP	0	0	0	0	1	0	1.5	0	
DNA	2	0.5	3	1.5	1	1	1.5	0	
RNA	0	0	0	0.5	0.5	0	0.5	0	
None	0	0	0	0	0	0	0	1	

		CP21R7 1.0 μ M							
ER	1	0	0	0	1	1	1	0	
Mito	0	0	0	0	0	0	0	0	
AGP	1	0	0	0.5	0.5	0	1.5	0	
DNA	0	0	0	0	0	2	1	0	
RNA	0	0	0	0.5	0.5	0	0.5	0	
None	0	0	0	0	0	0	0	0	

	Shape	Granularity	Texture	Concentric	Intensity	Zernike	Channel_correlation	Global
ER	1	0	0	0	1	1	1	0
Mito	0	0	0	0	0	0	0	0
AGP	1	0	0	0.5	0.5	0	1.5	0
DNA	0	0	0	0	0	2	1	0
RNA	0	0	0	0.5	0.5	0	0.5	0
None	0	0	0	0	0	0	0	0



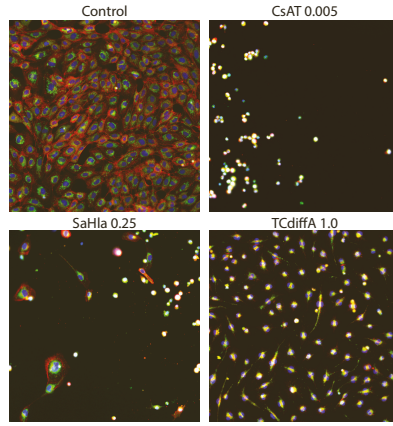
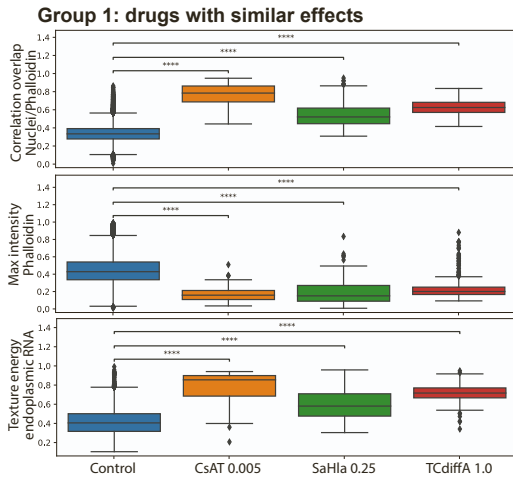
Supplementary Figure 4: Value distribution and image visualization of the most important features to correctly classify CP21R7 at 0.2 μ M and 1.0 μ M.

Related to Figure 3.

The features of interest are shown on top of the images. The pictures depict the wells with the closest value to the mean of the feature. In interest of space, we only visualize 9 out of the 49 total tiles that make a well.

In the heatmaps, the darker the color, the higher the number of features. The number states the actual number of features per class and channel. The half numbers are for those features with double class or double channel.

The box plots components are: horizontal line, median; box, interquartile range; whiskers, 1.5 \times interquartile range. arb. units: arbitrary units. Two-sided Mann–Whitney U test: ns: $P > 0.05$; $0.01 < P \leq 0.05$; **** $P \leq 0.0001$.



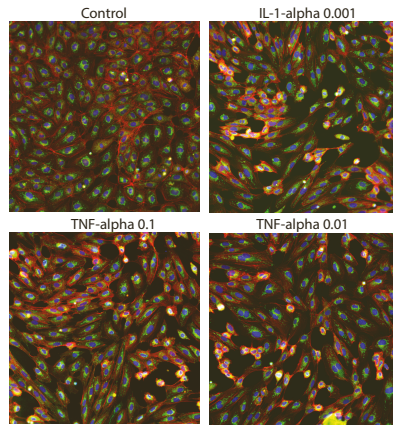
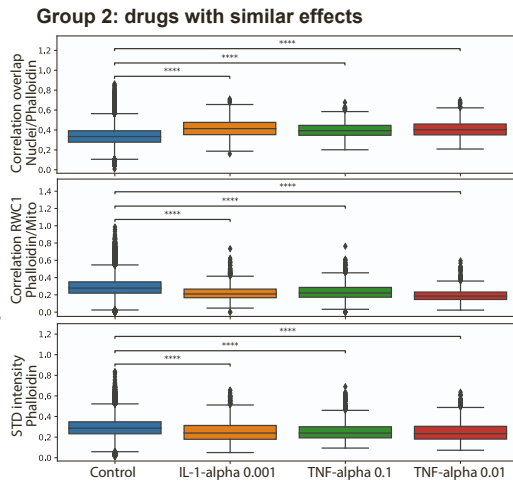
Supplementary Figure 5: Value distribution (box plots) and image visualization of the most important features that distinguish similar drug groups from the control (Group 1,2 and 3), and groups with unique effects (Group 4).

Related to Figure 4.

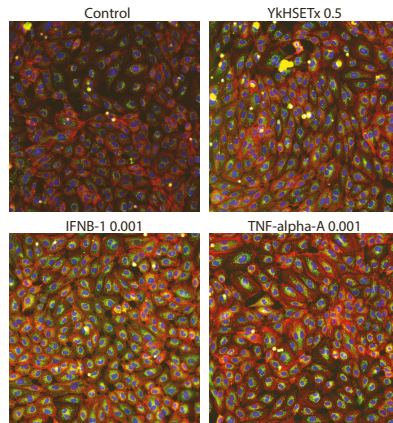
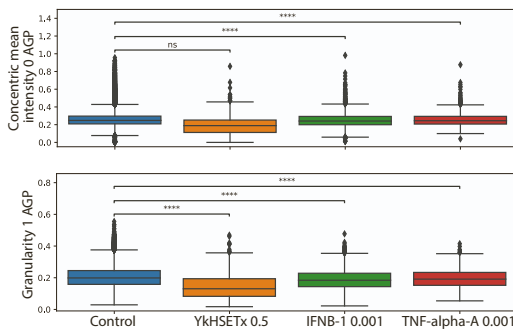
The features of interest are shown on the x axis of the box-plot. The pictures depict the wells with the closest value to the mean of the feature. In interest of space, we only visualize 9 out of the 49 total tiles that make a well.

The feature distribution values are all normalized between 0 and 1. The images were acquired at 20X.

The box plots components are: horizontal line, median; box, inter-quartile range; whiskers, 1.5× inter-quartile range. arb. units: arbitrary units. Two-sided Mann–Whitney U test: ns: $P > 0.05$; $0.01 < P \leq 0.05$; $****P \leq 0.0001$.



Group 3: drugs with similar effects



Group 4: example of drugs with unique effects

