A rapid phenomics workflow for the *in vitro* identification of antiviral drugs

Jonne Rietdijk[1], Marianna Tampere[2,3], Aleksandra Pettke[2], Polina Georgieva[1], Maris Lapins[1], Ulrika Warpman Berglund[2], Ola Spjuth[1], Marjo-Riitta Puumalainen[2] and Jordi Carreras-Puigvert[1*]

[1]Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Box 591, SE-75124, Uppsala, Sweden

[2]Department of Oncology and Pathology and Science for Life Laboratory, Karolinska Institutet, S-171 76, Stockholm, Sweden

[3]National Veterinary Institute, SE-756 51 Uppsala, Sweden

*Correspondence: Jordi Carreras-Puigvert(J.C-P)

Abstract

Morphological profiling of cells in the presence of perturbants, also known as phenomics, is gaining momentum given its successful implementation for drug discovery and compound profiling. The current COVID-19 pandemic has fueled the search for new and fast methods to identify novel or repurposed therapeutic drugs. A popular method to identify antiviral drugs is the use of antibody-based immunofluorescence to visualise infected cells. However, this method lacks depth towards the effect of such drugs on the host cells. Here we present a phenomics workflow for untargeted phenotypic drug screening of virus infected cells, combining Cell Painting with antibody-based detection of viral infection in a single and simple method and provide a semi-automated image analysis pipeline for classification and feature extraction of virus infected cells. Our phenomics workflow provides valuable information about the effect of both virus and drugs on the host cells. We validated our method using a panel of 9 antiviral compounds including known and novel compounds on MRC5 human lung fibroblasts infected with Human coronavirus 229E (CoV-229E). Two of the compounds showed strong antiviral efficacy concomitant with a recovery of the morphological profile towards non-infected.

Introduction

In the light of the current Covid-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), development of new therapies against emerging viruses is of high priority. Understanding how viruses affect the host cells is key to identify potential targets for treatment. RNA viruses are characterized by their fast mutation rate resulting in resilience against antiviral drugs, as well as heavy reliance on host pathways for replication (1,2). Coronaviruses are an exception to this norm given the existing RNA proofreading machinery within their genome, but nevertheless develop treatment escape mutants (3). Therefore, host targeting- as well as combination- therapies might be advantageous to battle current and future virus outbreaks. To this end, new methods to profile cellular responses during virus infection that can be leveraged for drug screening are greatly needed.

Image-based morphological cell profiling is a multiparametric method for studying biological and chemical perturbations at single-cell level. The Cell Painting approach is a high-content image-based assay for morphological profiling that uses six multiplexed fluorescent dyes. These dyes reveal eight relevant cellular components or organelles interrogating numerous biological pathways simultaneously (4). The resulting images can be analyzed using traditional image analysis techniques to provide information-rich profiles of individual cells, as well as serve as input for machine or deep learning approaches (5). Previous studies showed that analysis of the Cell Painting-derived features can be used to study drug toxicity, predict cell health indicators, detect morphological disease signatures, and give insights into the mechanism of action (MoA) of both existing and novel compounds (6–10). In contrast to target-based screens, where only a limited number of features are quantified to select for a known cellular phenotype, morphological profiling combined with statistical analysis can unravel subtle morphological patterns and provide insights into new pathways (11).

While morphological profiling is gaining ground in the field of pharmacology and toxicology, the employment of phenotypic profiling in the field of virology has been limited. One possible explanation is the methodological complexity when combined with high-throughput pipelines, as previously suggested (12). However, with the increased momentum in virology research caused by the spread of the SARS-CoV-2, phenotypic profiling is progressively being implemented as a strategy for the selection and optimization of drugs, and drug combinations, for the treatment of virus-infected cells (13,14).

Traditional immunofluorescence-based antiviral screening methods commonly rely on the use of specific antibodies for the visualisation and quantification of virus-infected cells. These methods have been proven successful for identifying drugs that directly affect cell survival and infectability of the virus, but lack the depth to provide detailed information about the effect on the host cell. In this paper we describe a novel phenomics workflow, combining morphological profiling with antibody-based detection of virus infection in a single and simple assay. We provide an automated image analysis pipeline for the identification of infected and non-infected cells, generating hundreds of data points, to study host-cell biology at a single-cell level.

To validate our method, we have applied our phenomics workflow on Human coronavirus 229E (CoV-229E) infected MRC5 primary lung fibroblasts, in the absence or presence of a number of known and novel antiviral drugs. The results of this study show that the extracted phenotypic features successfully capture effects of viral infection and that treatment with the compounds Remdesivir and E-64d could reverse the cellular phenotype from an “infected”- towards a “non-infected” morphological profile. Taken together, our phenomics workflow provides a novel method for the unbiased study of virus-induced effects on host cells that can be leveraged for drug screening of novel or repurposed antiviral drugs.

Results

A phenomics workflow to morphologically profile cells exposed to viruses

In order to investigate how viral exposure affects the host cells, the original morphological profiling protocol known as Cell Painting (4), was modified and combined with antibody-based staining. Incorporation of an antibody specific against coronavirus nucleoprotein, allowed for the identification of virus-infection, as well as the parallel measurement of hundreds of parameters from the organelles stained using the remaining Cell Painting dyes, which more specifically consisted of Hoechst (Nuclei), SYTO 14 (nucleoli and cytoplasmic RNA), Concanavalin A for visualisation of the Endoplasmic Reticulum (ER), Wheat Germ Agglutinin (plasma membrane and Golgi), and Phalloidin for F-actin (**Fig. 1A and Table 1**). Once the morphological profiling protocol was set up, an analysis workflow was designed to capture the in-depth morphological profiles induced by virus infection as well as treatment with selected antiviral compounds (**Fig. 1B**). To validate our method, we chose human lung fibroblasts (MRC5) as host cells and infected them with the Human coronavirus 229E (CoV-229E). The analysis part of our phenomics workflow included quality control of the images, feature extraction of the image data, data analysis on the extracted features, and finally data representation **(Suppl. Fig. 1)**.

Cell Painting features capture effects of viral infection

An image analysis pipeline was designed to distinguish between infected and non-infected cells using the coronavirus nucleoprotein antibody stain, implemented as a Cell Profiler pipeline (15). The pipeline is designed to be universal, requiring only adjustments for accurate segmentation when applied on different cell lines and viruses; in this manuscript we demonstrate it on MRC5 cells to evaluate the effect of CoV-229E infection.

Cell classification based on the virus antibody intensity measures in the perinuclear space accurately discriminated infected from non-infected cells (**Fig. 2A**), which allowed us to further analyse the morphological features and differentiate both cell populations in an unbiased manner (**Fig. 2B**). While we collected as many as 1200 morphological features, there were clear changes in a subset of them, including Concanavalin A and SYTO 14 mean intensities, which were elevated in the infected cells compared to non-infected (**Fig. 2B and Suppl. Table 1**). Next, we performed a hierarchical clustering analysis on the extracted features, visualizing the results in a heat map and generated morphological profiles for each condition. Clearly, CoV-229E infection induced measurable morphological changes, as depicted by the distinct morphological profiles between DMSO conditions in the absence or presence of CoV-229E (**Fig. 2C**).

To visualize the distribution of phenotypic features of each condition, we performed Uniform Manifold Approximation and Projection (UMAP) and Principal Component (PCA) analyses. For this, we excluded the features corresponding to the anti-coronavirus nucleoprotein antibody, thus solely using the Cell Painting produced features to study host-cell morphology. Remarkably, PCA analysis revealed a separation of infected from noninfected cells in the first principal component (**Fig. 2D and E**). The associated PCA loadings revealed that phenotypic features relating to the ER and nucleoli/cytoplasmic RNA stains, as well as granularity, size and shape features were most important on the direction of the first principal component **(Suppl. Table 1)**. Overall, our analysis demonstrated the strength of phenomics profiling to identify morphological changes exclusive of virus-infected cells.

Assessment of antiviral activity of reference drugs by an antibody-based approach

To benchmark our phenomics workflow to the traditional methods to screen for antiviral drugs, we first explored the antiviral efficiency of a set of 9 compounds that included two direct-acting drugs targeting viral RNA polymerase (Remdesivir and Favipiravir), and seven host-targeting antiviral compounds (E-64d, Camostat, Cathepsin L inhibitor and Bafilomycin A1), including three novel in-house developed broad-spectrum antiviral compounds TH3289, TH6744 and TH5487 were tested (16,17) (**Fig. 3A**). MRC5 cells were either not infected or infected with CoV-229E and then exposed to the compounds at the indicated concentrations (**Fig. 3A**). We determined cell survival by nuclei count and infection rate by the number of anti-coronavirus nucleoprotein positive cells, both in respect to DMSO (infection 91%) (**Fig. 3B**). Antiviral activity with no cellular toxicity was observed upon Remdesivir and E-64d treatment. Treatment with clinically available Remdesivir reduced infection between 58% at 0.1μM and 91% at 1μM and 8μM concentrations. At the same time E-64d treatment resulted in a reduction in viral infection of 52%, 77% or 85% compared to DMSO at 1, 10 or 30 μM concentrations, respectively, providing an adequate therapeutic window (Fig. 3B, C). The recently developed broadly active host-targeting antiviral compounds TH3289, TH6744 and TH5487 showed 58%, 52% and 53% reduction in the levels of infected cells, respectively, upon 10 μM treatments. Neither FDA-approved Favipiravir nor pre-clinical compounds Camostat or Cathepsin L inhibitor reduced the number of infected cells, while Bafilomycin A1 was toxic at all tested concentrations (**Fig. 3B and C and Suppl. Fig 2A**).

Validation study: Phenomics approach identifies antiviral drugs against coronavirus that recover infected morphological profiles

To validate our method, we used the developed phenomics workflow to assess the antiviral activity of the drugs and compounds that were tested at the same concentration using a traditional antibody-based method (**Fig. 3**). We used unsupervised hierarchical clustering of the extracted morphological profiles to compare the phenotypic profiles among all the tested conditions (**Suppl. Fig. 3A**). The profiles from infected cells treated with Remdesivir or E-64d, which showed the most prominent antiviral effect, were similar to the non-infected cells in DMSO, suggesting a rescue or prevention of the virus-induced phenotypic profile (**Fig. 4A, Suppl. Fig 4A**). In the absence of the virus, the morphological profiles of the drugs with strong antiviral effect (Remdesivir and E-64d), were largely similar to the DMSO condition (**Suppl. Fig 2A**). The treatment with compound TH6744 induced morphological profiles that were closer to non-infected controls on the first component of the PCA, but deviated on the second component **(Fig 4D).** Finally, the profiles for Camostat, which exhibited no antiviral effect, and no toxicity, remained unchanged compared to DMSO, both in the presence or absence of the virus, confirming the lack of activity of this compound in the tested conditions. To facilitate the interpretation of the data, we next performed dimensionality reduction by PCA of the phenotypic features for each of the drugs and compounds tested. In the case of Remdesivir and E-64d, the PCA resulted in several clusters: one cluster corresponding to the DMSO condition in infected cells, followed by clusters corresponding to the treatments, which deviated towards the cluster belonging to the DMSO condition in non-infected cells (**Fig. 4B and C**). Analysis of the morphological profiles induced by TH6744 treatment also separated from the DMSO condition in the presence of the virus on the first principal component, but separated on the second principal component (**Fig. 4D**). Camostat treatment clusters were indistinguishable from the DMSO-treated infected cells (**Fig. 4D**). In summary, these results validate our phenomics method to identify antiviral drugs.

Discussion

Traditional targeted screens for antiviral drugs, such as antibody-, reporter-, or biochemical-based assays, can be costly and time-consuming, and are often biased for discovery constrained by what we know. Cell profiling enables an untargeted approach for characterization of cellular responses with a high spatiotemporal resolution, high throughput, and on a single-cell level. Here, we present a method utilizing Cell Painting combined with antibody staining to profile host cell responses upon virus infection and validate its potential to serve as a platform for finding effective antiviral compounds.

Utilizing our phenomics workflow, we successfully captured distinct morphological profiles upon CoV-229E infection. Drug treatment with Remdesivir and E-64d, which have been previously shown to ameliorate coronavirus infections (18,19), successfully reversed the morphological profile from a disease (infected) to a control (non-infected) one. In fact, hierarchical clustering of the morphological profiles by similarity among the tested compounds, resulted in Remdesivir and E-64d clustering with the non-infected conditions (**Suppl. Fig 4A**).

On the other hand, under our tested conditions we could not reproduce the previously reported antiviral activities from Favipiravir, Camostat, Cathepsin L inhibitor or Bafilomycin A1 (19–22). These results may be explained by cell-dependent effects, the timing of treatment or by viral infection, which could be addressed by optimisation of the experimental conditions. However this is out of the scope of this work. While this could be considered a caveat of this method, assay and conditions optimisation are a requisite for any type of screening method.

The antiviral activity of TH6744, TH3289 and TH5487 has previously been characterized on viral progeny release (16), which differs from the primary virus infection used in this work. This methodological difference might explain the mild antiviral activity found by both the traditional antibody-based approach and our phenomics workflow. However, the mild antiviral activity of these newly developed compounds, in particular of TH6744, was reflected by its morphological profile and PCA, which showed that TH6744 partially rescued the virus-induced phenotype.

A clear advantage of using morphological profiling to screen for antiviral drugs is the in-depth analysis of the host cell status upon infection as well as compound treatment. Thus, despite not being able to recapitulate the antiviral activities of the above mentioned compounds, using our phenomics workflow, we have obtained important information that can be used to further understand the biological effect of both a given virus or a given compound, and importantly it validates our method to identify potential novel antiviral compounds.

Overall, we demonstrated that our phenomics workflow, which includes an adapted Cell Painting protocol, complemented with a custom image analysis pipeline, can be utilized the unbiased study of virus-induced effects on host cells that can be leveraged for drug repurposing, as well as for the discovery of novel antiviral compounds. We anticipate that this untargeted approach, with only minor adjustments of the image analysis pipeline, will enable other applications using diverse human-derived cell lines, as well as different viruses and antibodies.

Materials and Methods

Biosafety

All experiments in presence of infectious virus were performed under respective biosafety laboratory (BSL) conditions according to Swedish Work and Health Authorities. Experiments using CoV-229E were performed at BSL2 at Karolinska Institutet.

Cell culture

MRC5 cells (from ATCC, Manassas, VA, USA) were cultured in Minimum Essential Media supplemented with 10% (v/v) fetal bovine serum (FBS; Thermo Fisher Scientific, Waltham, MA, USA), 50 U/ml penicillin and 50 µg/ml streptomycin (Thermo Fisher Scientific, Waltham, MA, USA). The cells were maintained at 37 °C under 5% CO2 and the cell culture was routinely tested for Mycoplasma using a luminescence-based MycoAlert kit (Lonza).

Virus production

CoV-229E (VR-740; from ATCC, Manassas, VA, USA) stocks were amplified in Huh7 cells. Virus titers were determined in Huh7 cells by end point dilution assay combined with high-throughput immunofluorescence imaging of viral protein staining as described previously (16).

Compounds

Remdesivir and Favipiravir were purchased from Carbosynth. TH6744, TH3289 and TH5487 were in-house produced and recently described (16,17). Compounds Camostat, E-64d, Cathepsin L inhibitor, Bafilomycin A (Sigma-Aldrich). Cell painting positive and negative controls included the Etoposide (E1383), Fenbendazole (F5396), Metoclopramide (M0763), and DMSO (Sigma Aldrich).

Virus infection and compound treatments

Antiviral compounds and cell painting positive controls were dispersed in 96-well source plates at 500× of the final concentration using the D300e Digital Dispenser (Tecan, Männedorf, Switzerland) and stored at - 20 °C until use. Compound and virus infection conditions were randomized within the plates and multiple plate layouts were used to compensate for systematic effects related to well position. Replicates were placed both on the same plate and in independent plates. To further avoid plate and edge-effects, the outer wells were excluded. Before performing the assay, 198 µL culture medium was added on the source plate reaching

200 µL of 5× dilution of the final concentration. Three doses were chosen for each antiviral compound (final concentration; max DMSO 0.3% *v*/*v*). A total of 2500 MRC5 cells per well were seeded in black 96 multiwell plates 24 hours before infection (Costar 3603, Corning Life Science, Corning, New York, USA) in 100 µL culture medium, kept at room temperature for 20 minutes to aid homogeneous spreading in the wells and incubated overnight at 37 ℃ at 5% CO_2 atmosphere. Cells were infected with a MOI 10 or 15 in 50 µL culture medium for 1 h and at least four wells per plate were left uninfected as a mock control. Virus-containing medium was removed, replaced with 80 µL culture medium and 20 µL compound-containing medium in at least technical duplicates, reaching 1× of the final compound concentration. Cells were incubated for 24 or 48 hours. Cells were fixed in 4% paraformaldehyde (Thermo Fisher Scientific, Waltham, MA, USA) in PBS (Gibco, USA) for 20 min, washed twice with PBS and stored at 4℃ prior processing for the Cell Painting and antibody detection.

Cell Painting and antibody detection

To profile morphological features of coronavirus infected cells an adapted version of the cell painting assay was used (11). Specific fluorescent probes or fluorophore-conjugated small molecules were used for the detection of the nucleus, nucleoli, endoplasmic reticulum (ER), actin cytoskeleton, Golgi apparatus and plasma membrane (Table 1). To accommodate virus-specific antibody detection in parallel to Cell Painting, one dye of the Cell Painting assay was replaced with coronavirus-nucleoprotein specific antibody (4). Compared to the original cell painting assay consisting of 6 dyes targeting eight components/ cell organelles, here 5 dyes and a virus-specific antibody were used to target virus nucleoprotein in parallel to 7 cell organelles and cell components in a single multiplexed staining assay (table 1). Example images of each of the stains can be found in **Fig 1A**. Coronavirus pan monoclonal antibody was combined with a secondary antibody using Alexa Fluor 647 fluorophore, chosen for its narrow emission and excitation spectra to limit interference with the other channels. In this assay, the mitochondrial dye was removed. To avoid fading of the chemical stains, in our experience, the secondary antibody is best added together or after staining with chemical dyes. Concentrations of secondary antibodies might need to be adjusted for a good signal-to-noise ratio in the multiplexed assay. In **Table 1**, working concentrations for the different cell lines and antibodies are presented. A detailed pipeline can be found in **Suppl. Fig 1**.

Fixed cells were washed twice (BioTek, 405 LS washer) and permeabilized by incubation with 80 µl of 0.1% Triton X-100 for 20 minutes at room temperature, followed by two washes with 1x PBS. Then, 80 µl of 0.2% BSA in PBS (Multiflo FX, multimode dispenser, BioTek) was added and incubated for 30 minutes. Next, the plates were washed twice with 1X PBS and

50µl primary antibody was added. Primary antibody was incubated for 4 hours at room temperature after which the plates were washed 3 times for 5 minutes. A staining solution of Hoechst 33342 (Invitrogen, cat.no H3570), SYTO 14 green (Invitrogen, cat.no S7576), Concanavalin A/Alexa Fluor 488 (Invitrogen, cat.no C11252), Wheat Germ Agglutinin/Alexa Fluor 555 (Invitrogen, cat.no W32464) and Phalloidin/Alexa Fluor 568 (Invitrogen, cat.no A12380) was prepared in PBS in addition to Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 (Invitrogen Catalog # A-21235). 50 µl of staining mixture was added to each well and incubated for 30 minutes, after which the plates were washed a final 3 times for 5 minutes. Plates were either imaged immediately or sealed and kept at 4°C to be imaged later.

All washing steps were performed using a Biotek 405 LS washer, dispensing was done using the Biotek Multiflo FX, multimode dispenser. Plates were protected from light as much as possible.

Table 1. Overview of ImageXpress Micro XLS channels and stains imaged in the assay.

Label	Excitation spectra	Emission spectra	Provider	Cat-No	Stock concentration	Concentration used	Target organelle
Hoechst 33342	377/50	447/60	Invitrogen	H3570	10 mg/ml	10 µg/ml	nucleus
Coronavirus pan Monoclonal Antibody (FIPV3-70)	-	-	Invitrogen	11500123	1 mg/mL	1:1000	Coronavirus viral nucleoprotein
Goat Anti-Mouse IgG H&L secondary antibody	628/40	692/40	Invitrogen	10739374	2 mg/mL	1:500	
Wheat Germ Agglutinin	562/40	624/40	Invitrogen	W32464	5 mg/m in PBS	15 µg/ml	Golgi, plasma membrane,
Phalloidin			invitrogen	10135092	40x (200U)	10 µl/ml	F-actin
SYTO 14 green	531/40	593/40	Invitrogen	S7576	5mM in DMSO	4 µM	nucleoli, cytoplasmic RNA
Concanavalin A/Alexa Fluor 488	482/35	536/35	Invitrogen	C11252	5 mg/ml in 0.1M sodium bicarbonate	80 µg/ml	Endoplasmic Reticulum

Imaging

Fluorescent images were acquired using an Image Xpress Micro XLS (Molecular Devices) microscope with a 20X objective using laser-based autofocus. In **Table 1**, the specifications of each of the stains are listed. In total, 9 sites per well were captured for 5 fluorescence channels. To avoid crosstalk between channels, images were taken with decreasing wavelengths. Examples of images from the different channels are shown in **Figure 1A.**

Image Analysis pipeline

Image Quality control and preprocessing

Images were processed and analysed with the open-source image analysis software CellProfiler (available at https://cellprofiler.org/), version 4.0.6 (29). Prior to analysis, a quality control pipeline was run on the raw images to detect images with artefacts, which may corrupt the data with false values (5). We computed various measures to represent a variety of artefacts and used statistical analysis to detect outliers. Images deviating more than 5 standard deviations from the median of FocusScore, MaxIntensity, MeanIntensity, PercentMaximal, PowerLogLogSlope and StdIntensity were flagged, inspected and removed if necessary. To remove blurred images, the PercentMaximal score and PowerLogLogSlope was computer and images with percentMaximal values higher than 0.25, or PowerLogLogSlope values lower than -2.3 were removed from the dataset. The QC detected 95 images as outliers, as indicated by one or multiple of the control measures. To assess the reproducibility of the assay and detect possible plate and drift effects across multiple plates, we visualized the quality measures and performed visual inspection (**Supp. Fig. 5)**.

In addition, we used cell-level quality control based on the Hoechst staining to detect bright artefacts and clumped cells. Bright artefacts in the DAPI channel were identified using the Indentify Primary Object module in CellProfiler, artefacts and surrounding pixels were masked to avoid interference with segmentation.

Segmentation

Nu segmentation based on the DAPI channel was performed by applying a gaussian blur followed by Otsu thresholding to segment the outlines of each nucleus. The cell objects were segmented using Watershed segmentation using minimum Cross-entropy of the cytoskeleton stain, using the nuclei as a seed. The cytoplasm cell compartment was defined as the cell object subtracted from the nuclei. Segmentation of the perinuclear cell compartment was accomplished by the expansion of 20 pixels in the concanavalin A channel and subtraction of the nuclei. The pipeline can be found as a supplement to this paper. Adjustments on the parameters might be needed to be used between different experimental setups and on different cell lines.

Feature extraction

After identification of the four cell compartments, phenotypic characteristics were measured using shape (AreaShape), intensity-based (Intensity, Radial-distribution), texture (granularity),

as well as microenvironment and context features (location, Neighbors) [4]. A total of 1211 features were extracted from each object which were exported into csv format for downstream analysis.

Classification

To classify the cells, the mean intensity of the antibody channel in the perinuclear cell compartment was used as measurement in the Classification module in CellProfiler. Visual inspection and validation using CellProfiler Analyst, confirmed that this measure was highly indicative for the infection state of the cells. Background levels of antibody signals were determined based on analysis of the histograms of noninfected cells, and used as a set-off for the infection status.

Antiviral dose-response and cytotoxicity

To capture relevant cellular phenotypes, cytotoxic compounds were removed from the analysis. Cell count was used to check the half maximal inhibitory concentration (IC50), defined as a 50% reduction in cell count compared to solvent control conditions. Antiviral efficacy was calculated by dividing the number of infected cells by the infection rate in solvent controls.

Statistics

Statistical analysis was done using GraphPad Prism version 9.0 (GraphPad Software Inc). Two-way-ANOVA was performed for all the datasets that required comparison among multiple data points within a given experimental condition. Dunnet was applied for the correction of multiple comparisons, and the family-wise alpha threshold and confidence level were 0.05 (95%confidence interval).

Feature analysis

In order to facilitate the broad use of our phenomics approach, an image analysis pipeline is provided with the paper, which is tailored to the application here presented, and once executed will provide the necessary feature data for subsequent analysis. A step-by step narrative of the methods used to analyse and visualize the feature data is presented below. Downstream analysis of the feature data was performed using Python 3. Prior analysis, feature data was cleaned from images with artefacts as identified by our QC pipeline described above. Next, the feature values were centered and normalized. More specifically, the feature values were fit to the data and then transformed using the fit.transform method of StandardScaler of scikit-

learn module (version 0.22.1). Normalization was done across all plates. Invariant features were removed.

All analyses were performed on the total of all extracted features, excluded from the metadata columns and features representing counts (e.g. number of children objects), and percentile counts. The analysis studying virus-induced morphological profiles were excluding features relating to the virus antibody.

To compare morphological profiles the features were presented as cluster maps. Features were scaled using the sklearn StandardScaler method, by subtracting the mean and scaling to unit variance. This ensured that each feature (clustermap column) has a mean of 0 and variance of 1. Unsupervised hierarchical clustering on the morphological profiles was computed using the Euclidean metric, and the clustering algorithm using Ward's method and visualized using the seaborn data visualization library (version 0.11.0).

Visualization of the high dimensional data was done using dimensional reduction methods including Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) using sklearn (version 0.22.1). UMAP visualization was done using 30 neighbors (30).

Data availability

Image Data will soon become publicly available at Figshare. Image analysis pipelines can be found as supplementary information.

Author contributions

J.R., M.T., A.P., M.-R.P. and J.C.-P. designed experiments. M.T. and A.P. performed experiments with BSL2 viruses. J.R and P.G performed morphological profiling experiments, J.R and P.G, M.L performed image analysis, M.L, J.R and J.C-P performed data analysis. O.S. and J.C-P. initiated the project. J.R., M.T., M.-R.P. and J.C-P. wrote the manuscript which was revised by all authors. All authors have read and agreed to the published version of the manuscript.

Funding

This project received funding within the SciLifeLab National COVID-19 Research Program and Knut och Alice Wallenbergs stiftelse (KAW 2020.0182). This project also received funding from the Swedish Research Council (2017-05631 for M-R.P; 2020-03731 for OS and JCP), FORMAS (2018-00924 for OS and JR) and the European Union’s Horizon 2020 research and

innovation programme under the Marie Skłodowska-Curie grant agreement no 722729. M.T. was supported by SSF (FID15-0010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Professor Tomas Bergström for providing the CoV-229E virus.

Competing interests

The authors declare no competing interests.

References

1. Bloom JD, Gong LI, Baltimore D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. Science. 2010 Jun 4;328(5983):1272–5.
2. Hai R, Schmolke M, Leyva-Grado VH, Thangavel RR, Margine I, Jaffe EL, et al. Influenza A(H7N9) virus gains neuraminidase inhibitor resistance without loss of in vivo virulence or transmissibility. Nat Commun. 2013 Dec 10;4(1):2854.
3. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. bioRxiv. 2020 Dec 1;2020.11.30.405472.
4. Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016 Sep;11(9):1757–74.
5. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, et al. Data-analysis strategies for image-based cell profiling. Nat Methods. 2017 Sep;14(9):849–63.
6. Way GP, Kost-Alimova M, Shibue T, Harrington WF, Gill S, Piccioni F, et al. Predicting cell health phenotypes using image-based morphology profiling [Internet]. Bioinformatics; 2020 Jul [cited 2020 Sep 11]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.07.08.193938
7. Bray M-A, Gustafsdottir SM, Rohban MH, Singh S, Ljosa V, Sokolnicki KL, et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. GigaScience [Internet]. 2017 Dec 1 [cited 2020 Sep 2];6(12). Available from: https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giw014/2865213
8. Schiff L, Migliori B, Chen Y, Carter D, Bonilla C, Hall J, et al. Deep learning and automated Cell Painting reveal Parkinson's disease-specific signatures in primary patient fibroblasts [Internet]. Cell Biology; 2020 Nov [cited 2021 Jan 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.11.13.380576
9. Christoforow A, Wilke J, Binici A, Pahl A, Ostermann C, Sievers S, et al. Design, Synthesis, and Phenotypic Profiling of Pyrano-Furo-Pyridone Pseudo Natural Products. Angew Chem Int Ed. 2019 Oct 7;58(41):14715–23.
10. Spjuth, Alexander Kensert. Transfer Learning with Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes.
11. Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016 Sep;11(9):1757–74.
12. Aulner N, Danckaert A, Ihm J, Shum D, Shorte SL. Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases. Trends Parasitol. 2019 Jul 1;35(7):559–70.
13. Heiser K, McLean PF, Davis CT, Fogelson B, Gordon HB, Jacobson P, et al. Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of

human cells infected with SARS-CoV-2 [Internet]. Systems Biology; 2020 Apr [cited 2020 Sep 2]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.04.21.054387
14. Mirabelli C, Wotring JW, Zhang CJ, McCarty SM, Fursmidt R, Frum T, et al. Morphological Cell Profiling of SARS-CoV-2 Infection Identifies Drug Repurposing Candidates for COVID-19 [Internet]. Cell Biology; 2020 May [cited 2020 Sep 2]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.05.27.117184
15. Analysis of a complex of statistical variables into principal components. - PsycNET [Internet]. [cited 2021 Jan 11]. Available from: https://psycnet.apa.org/doiLanding?doi=10.1037%2Fh0071325
16. Tampere M, Pettke A, Salata C, Wallner O, Koolmeister T, Cazares-Körner A, et al. Novel Broad-Spectrum Antiviral Inhibitors Targeting Host Factors Essential for Replication of Pathogenic RNA Viruses. Viruses. 2020 Dec;12(12):1423.
17. Visnes T, Cázares-Körner A, Hao W, Wallner O, Masuyer G, Loseva O, et al. Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. Science. 2018 Nov 16;362(6416):834–9.
18. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr 16;181(2):271-280.e8.
19. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020 Mar;30(3):269–71.
20. Zhou Y, Vedantham P, Lu K, Agudelo J, Carrion R, Nunneley JW, et al. Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res. 2015 Apr 1;116:76–84.
21. Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. Proc Natl Acad Sci. 2005 Aug 16;102(33):11876–81.
22. Kawase M, Shirato K, Matsuyama S, Taguchi F. Protease-Mediated Entry via the Endosome of Human Coronavirus 229E. J Virol. 2009 Jan 15;83(2):712–21.
23. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006 Sep 29;313(5795):1929–35.
24. Caicedo JC, Singh S, Carpenter AE. Applications in image-based profiling of perturbations. Curr Opin Biotechnol. 2016 Jun;39:134–42.
25. Hughes RE, Elliott RJR, Munro AF, Makda A, O'Neill JR, Hupp T, et al. High-Content Phenotypic Profiling in Esophageal Adenocarcinoma Identifies Selectively Active Pharmacological Classes of Drugs for Repurposing and Chemical Starting Points for Novel Drug Discovery. SLAS Discov Adv Sci Drug Discov. 2020 Aug;25(7):770–82.
26. Parang K, El-Sayed NS, Kazeminy AJ, Tiwari RK. Comparative Antiviral Activity of Remdesivir and Anti-HIV Nucleoside Analogs Against Human Coronavirus 229E (HCoV-229E). Mol Basel Switz. 2020 May 17;25(10).
27. Kawase M, Shirato K, Matsuyama S, Taguchi F. Protease-Mediated Entry via the Endosome of Human Coronavirus 229E. J Virol. 2009 Jan 15;83(2):712–21.
28. Willis C, Nyffeler J, Harrill J. Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. SLAS Discov Adv Sci Drug Discov. 2020 Aug;25(7):755–69.
29. McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. PLoS Biol. 2018 Jul;16(7):e2005970.
30. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv180203426 Cs Stat [Internet]. 2020 Sep 17 [cited 2021 Jan 11]; Available from: http://arxiv.org/abs/1802.03426

Figure Legends

Figure 1. A phenomics workflow for identification of antiviral drugs. A. MRC5 lung fibroblast cells infected with Human coronavirus 229E, stained using Hoechst (Nuclei), SYTO 14 (nucleoli and cytoplasmic RNA), Concanavalin A (Endoplasmic Reticulum (ER)), Wheat Germ Agglutinin (plasma membrane and Golgi), and Phalloidin (F-actin), in combination with an anti-coronavirus nucleoprotein antibody. Note the presence of non-infected (asterisk) and infected cells. **B.** Overview of the phenomics workflow. Cells are seeded in multiwell plates, incubated for 24 hours followed by virus infection. Subsequently, virus-containing media is removed and compounds are added and incubated for 24 or 48h followed by fixation fixed. Next, antibody staining and Cell Painting is performed on the cells, followed by high content imaging, image analysis and data analysis of the extracted phenotypic features (a more detailed description of the workflow can be found in **Suppl. Fig. 1**).

Figure 2. Morphological profiling of virus-infected MRC5 cells. A. A representative composite image of infected cells with F-actin in green, nuclei in blue and anti-coronavirus nucleoprotein antibody in red. Segmentation and classification of individual cells visualized by an outline with infected cells in purple and non-infected cells in yellow. **B.** Boxplots of a few of the morphological features that were altered in infected cells compared to non-infected. **C.** Morphological profiles of infected (+CoV-229E) and non-infected cells (-CoV-229E). Columns in the matrix represent extracted features after clustering, colour intensity represents the magnitude of increase or decrease in a measured feature. **D and E.** Dimensionality reduction using UMAP and PCA was applied to the extracted Cell Profiler features to reduce redundancy and correlation of the features (15). The results were then plotted as single cells and coloured according to their infected or non-infected classification based on antibody staining.

Figure 3. Antiviral efficacy assessment. A. A set of two reference drugs **(**Remdesivir and Favipiravir), seven antiviral compounds (E-64d, Camostat, Cathepsin L inhibitor and Bafilomycin A1), including three in-house developed broad-spectrum antiviral compounds (TH3289, TH6744 and TH5487) were used to assess antiviral efficacy by anti-coronavirus nucleoprotein staining. **B.** Percentage of survival (red) and percentage of infected cells (blue) normalised to DMSO (solvent control). Coloured intervals denote standard deviation. **C.** Representative images of MRC5 cells non-infected, or infected with CoV-229E. Hoechst staining for nuclei are visualized in blue, and anti-coronavirus nucleoprotein antibody against CoV-229E in yellow. Two-way ANOVA was performed to assess the statistical significance of each condition (*p<0.02, **p<0.002, ***p<0.0002).

Figure 4. Identification of phenomics profiles of antiviral drugs efficient against coronavirus. A. Similarity matrix of phenomic profiles of the indicated samples based on their proximity in feature space. Two main groups are shown, non-infected (-229E) and infected (+229E), in the presence or absence of Remdesivir, E-64d, TH6477, Camostat or DMSO (vehicle) at the indicated concentrations. **B, C and D.** PCA analysis of morphological features upon treatment with Remdesivir, E-64d, TH6477 and Camostat in non-infected (-229E) and infected (+229E) conditions compared to DMSO control. Each dot in the PCA represents one image by taking the mean of all objects in the image.

Supplementary Figure 1. Schematic overview of the phenomics workflow here described.

Supplementary Figure 2. Survival of cells exposed to antiviral compounds. A. Nuclei count was used to assess the survival of MRC5 cells exposed to the indicated compounds for 48h. Two-way ANOVA was performed to assess the statistical significance of each condition (*p<0.02, **p<0.002, ***p<0.0002).

Supplementary Figure 3. Morphological profiles presented as heatmaps of each tested compound at non-toxic concentrations.

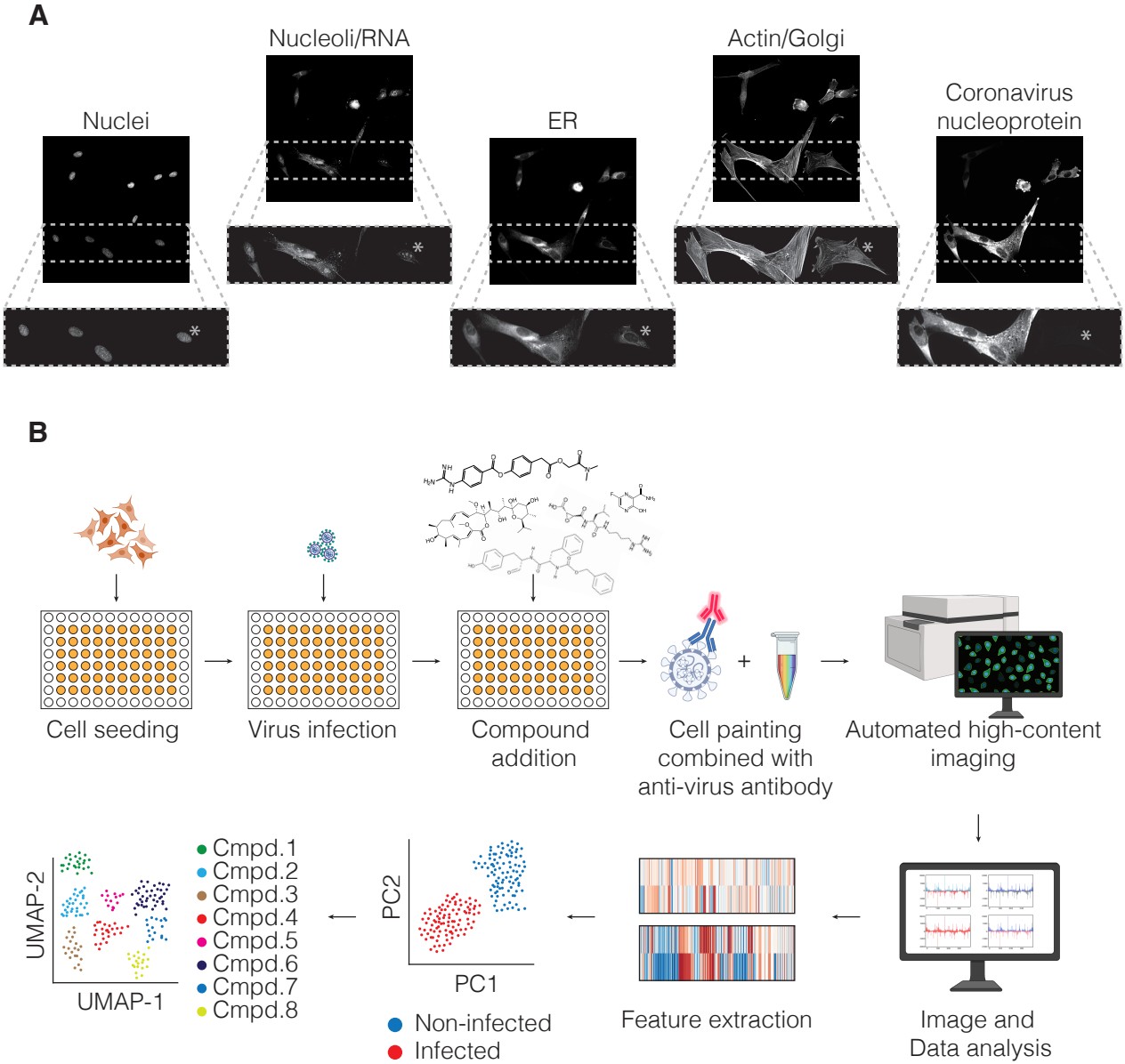
Supplementary Figure 4. Hierarchical clustering and PCA of the morphological features of each indicated compound. A. heatmap representing the morphological profiles for selected compounds and controls. Profiles of compounds E64 and Remdesivir cluster together with non-infected conditions, indicating a recovery from infected towards a non-infected profile. **B, C and D.** PCA of morphological features upon treatment with TH3289, TH5487 and

Cathepsin L inhibitor in infected (+CoV-229E) conditions compared to DMSO control. Each dot in the PCA represents one image by taking the mean of all objects in the image.

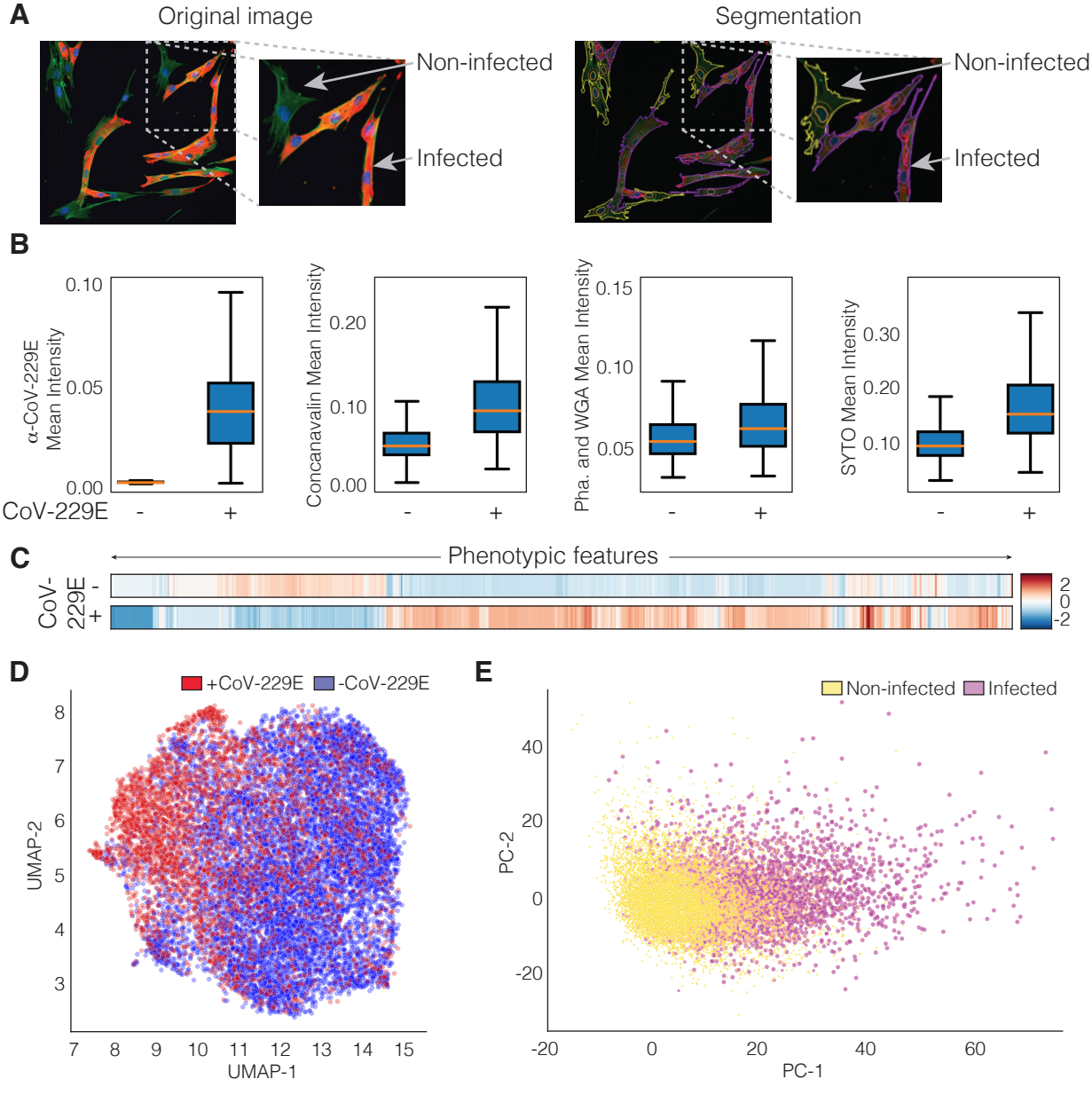
Supplementary Figure 5. Quality control measures plotted on images from all plates in the experiment. Images deviating more than 5 standard deviations from the median of FocusScore, MaxIntensity, MeanIntensity, PercentMaximal, PowerLogLogSlope and StdIntensity were flagged as outliers and removed from the analysis.

Supplementary Table 1. Loadings of Principal component analysis of infected and non-infected cells. The top 20 positively correlated, as well as 20 most negatively correlated features including their loadings on the first principal component.

Rietdijk J *et al.* **Fig.1**



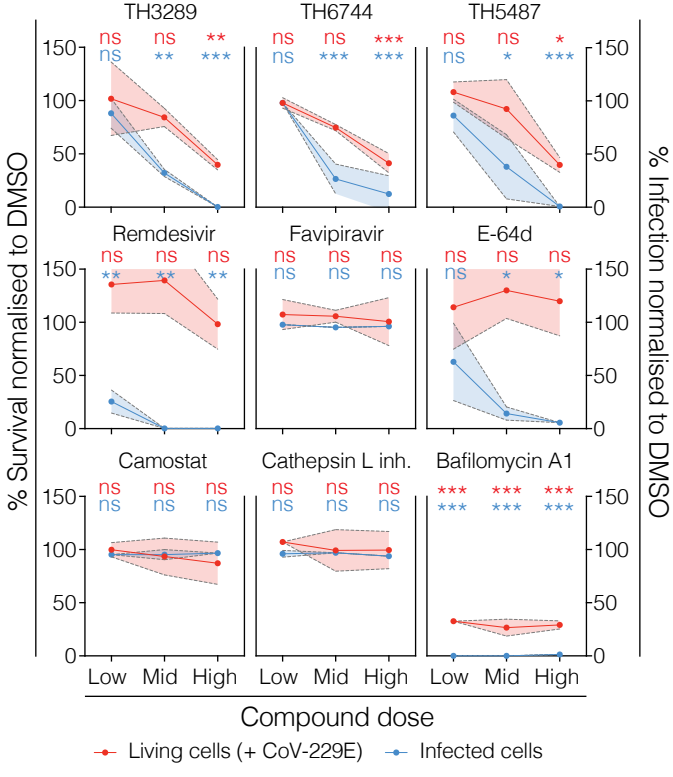
Rietdijk J *et al.* **Fig.2**



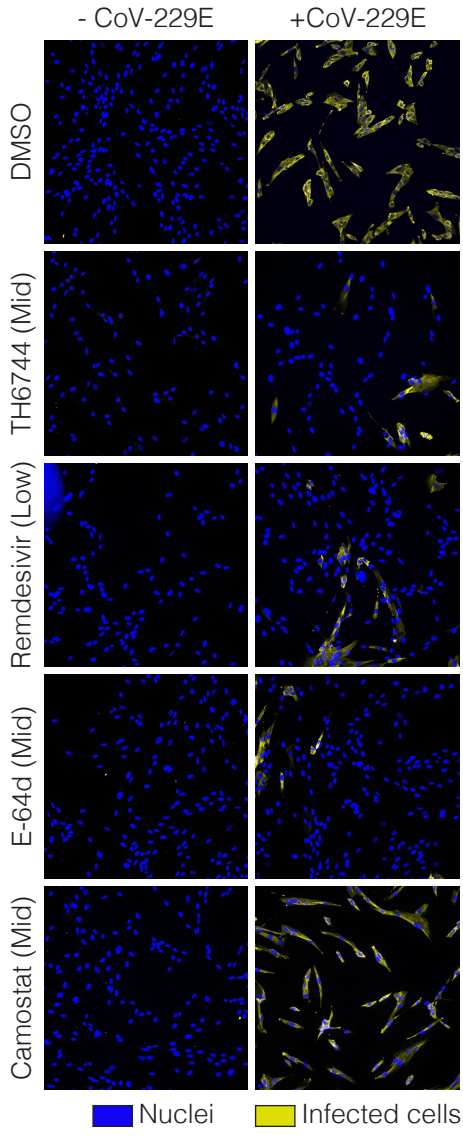
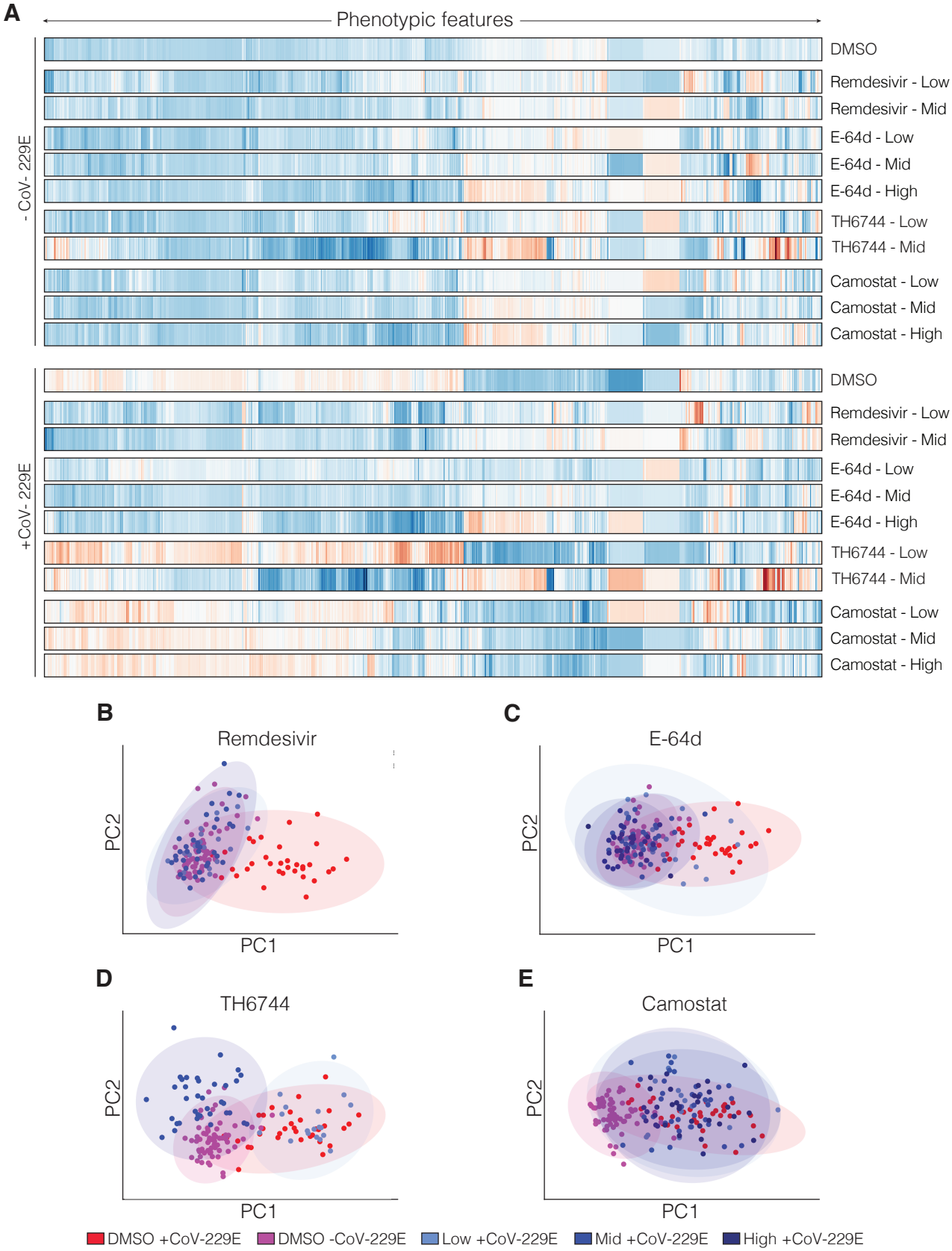
A

Compound /drug name	MoA	Low	Mid	High
TH3289	Host-targeting antiviral	1	10	30
TH6744	Host-targeting antiviral	1	10	30
TH5487	Host-targeting antiviral	1	10	30
Remdesivir	Nucleotide analogue	0.1	1	8
Favipiravir	Nucleoside guanine analogue	10	15	30
E-64d	Protease inhibitor	1	10	30
Camostat	TMPRSS family protease inhibitor	0.2	10	30
Cathepsin L inhibitor	Lysosomal cysteine proteinase	0.2	10	30
Bafilomycin A1	Vacuolar V-ATPase type inhibitor	0.1	1	3

B



C



Rietdijk J *et al.* **Fig.4**