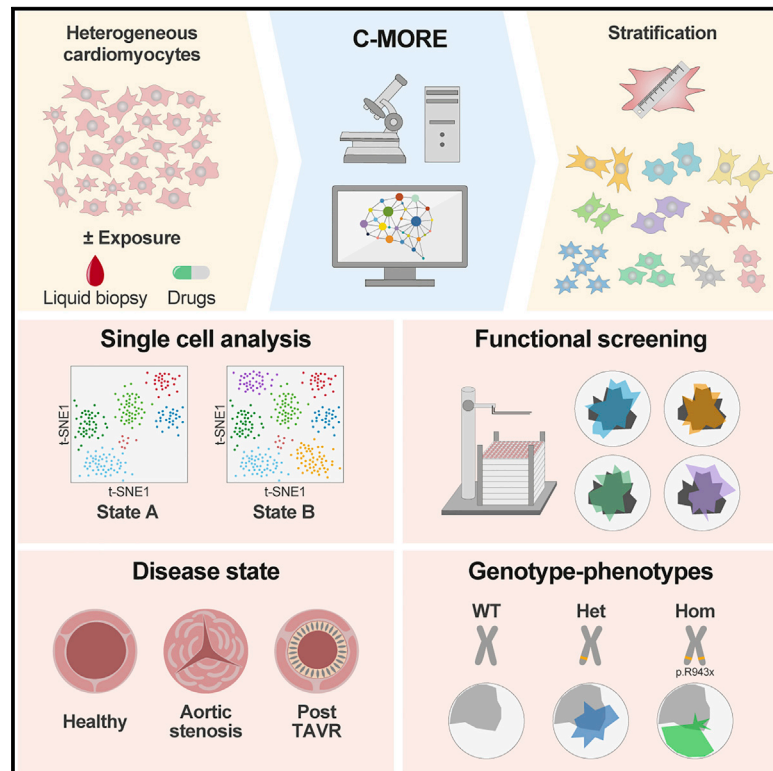


C-MORE: A high-content single-cell morphology recognition methodology for liquid biopsies toward personalized cardiovascular medicine

Graphical abstract



Authors

Jennifer Furkel, Maximilian Knoll, Shabana Din, ..., Amir Abdollahi, Hugo A. Katus, Mathias H. Konstandin

Correspondence

jennifer.furkel@med.uni-heidelberg.de (J.F.),
mathias.konstandin@med.uni-heidelberg.de (M.H.K.)

In brief

Furkel et al. present an integrative single-cell morphology-based strategy (C-MORE) to detect cardiomyocyte activation status and response to pharmacological treatment. Deconvolution of disease state in cardiomyocyte cultures exposed to liquid biopsies and in genetically engineered models by C-MORE may pave the way for development of next-generation personalized cardiovascular medicine.

Highlights

- C-MORE deconvolutes cardiomyocyte morphology on single-cell level
- C-MORE could be utilized for high-throughput pharmacological screens
- C-MORE detects pathological phenotypes caused by genetic alterations
- C-MORE liquid biopsy setup reflects disease state in individuals with aortic stenosis



Article

C-MORE: A high-content single-cell morphology recognition methodology for liquid biopsies toward personalized cardiovascular medicine

Jennifer Furkel,^{1,2,3,4,5,6,7,*} Maximilian Knoll,^{3,4,5,6} Shabana Din,^{1,2} Nicolai V. Bogert,^{1,2} Timon Seeger,^{1,2} Norbert Frey,^{1,2} Amir Abdollahi,^{3,4,5} Hugo A. Katus,^{1,2} and Mathias H. Konstandin^{1,2,*}

¹Department of Cardiology, Angiology and Pneumology, Heidelberg University Hospital, 69120 Heidelberg, Germany

²DZHK (German Center for Cardiovascular Research), Site Heidelberg/Mannheim, 69120 Heidelberg, Germany

³German Cancer Consortium (DKTK) Core Center Heidelberg, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

⁴Clinical Cooperation Unit Translational Radiation Oncology, National Center for Tumor Diseases (NCT), Heidelberg University Hospital (UKHD) and DKFZ, 69120 Heidelberg, Germany

⁵Division of Molecular and Translational Radiation Oncology, Department of Radiation Oncology, Heidelberg Faculty of Medicine (MFHD) and Heidelberg University Hospital (UKHD), Heidelberg Ion-Beam Therapy Center (HIT), Heidelberg, Germany

⁶These authors contributed equally

⁷Lead contact

*Correspondence: jennifer.furkel@med.uni-heidelberg.de (J.F.), mathias.konstandin@med.uni-heidelberg.de (M.H.K.)
<https://doi.org/10.1016/j.xcrm.2021.100436>

SUMMARY

Cellular morphology has the capacity to serve as a surrogate for cellular state and functionality. However, primary cardiomyocytes, the standard model in cardiovascular research, are highly heterogeneous cells and therefore impose methodological challenges to analysis. Hence, we aimed to devise a robust methodology to deconvolute cardiomyocyte morphology on a single-cell level: C-MORE (cellular morphology recognition) is a workflow from bench to data analysis tailored for heterogeneous primary cells using our R package *cmoRe*. We demonstrate its utility in proof-of-principle applications such as modulation of canonical hypertrophy pathways and linkage of genotype-phenotype in human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs). In our pilot study, exposure of cardiomyocytes to blood plasma prior to versus after aortic valve replacement allows identification of a disease fingerprint and reflects partial reversibility following therapeutic intervention. C-MORE is a valuable tool for cardiovascular research with possible fields of application in basic research and personalized medicine.

INTRODUCTION

Heart failure remains the leading cause of mortality in western countries and is the mutual final stage of a wide range of cardiac pathologies. To prevent progression, early diagnosis and adapted treatment with regard to the specific underlying cause is crucial.¹ However, diagnosis and estimation of prognosis remain challenging, and invasive diagnostics (e.g., biopsy) are often indispensable.² Novel methods with potential to elucidate the modes of action of cardiac diseases as well as clinical decision-making are needed.

Image-based high-content morphology assessment and its proficiency to serve as a surrogate for cell state and functionality has become an integral readout in cell line-based screenings in drug discovery:^{3,4} high-throughput fluorescence microscopes allow automated acquisition of images from 96- or 384-well plates in multiple fluorescence channels, and raw images can then be analyzed to identify cells and compute single-cell morphological multi-feature profiles (e.g., shape, intensity, and texture, including Haralick features).^{5,6}

However, in the cardiac context, heterogeneity and a higher variability of model cell types under investigation—namely, primary neonatal rat ventricular cardiomyocytes (NRVMs) or human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs)—have impeded implementation of robust morphology assessment.^{7,8} Morphological readout in cardiovascular research is still limited to cell size and a few selected features (e.g., perimeter, elongation, and form factor).^{9,10} Alternative methods focus on sarcomere organization, 3D volume/shape, proliferation, or reporter readouts.^{11–14} However, these established methods address only specific parts of morphology, and integrated approaches are not established for use in cardiovascular research.

In the clinical setting, manual morphological assessment of tissue and cellular biopsy material is a traditional method to classify disease state and estimate prognosis. In oncological research, first advances toward integration of automated morphological assays on tumor cells in clinical personalized medicine have been taken. For example, cellular morphology has enabled prediction of responsiveness to therapeutic



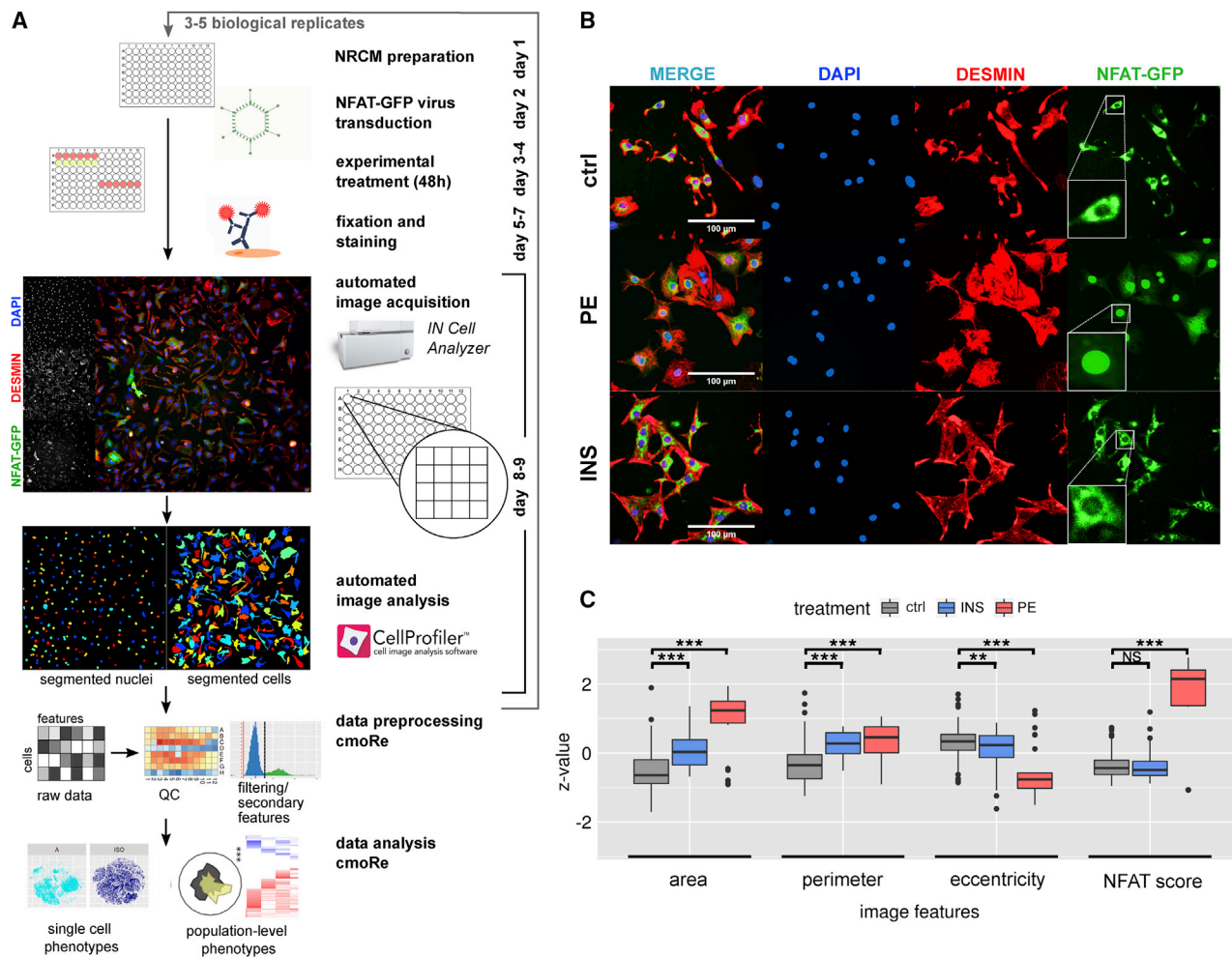


Figure 1. C-MORE experimental and data analysis workflow with NFAT-GFP reporter readout

(A) Overview of the C-MORE workflow. QC, quality control.

(B) Representative images of CMs without stimulation (ctrl) and with phenylephrine (PE) and insulin (INS) treatment. In the lower left corner of the GFP/FITC channel, we show a representative cell magnified to demonstrate NFAT-GFP enrichment in the nucleus after PE treatment. For ctrl and INS treatment, NFAT-GFP is localized predominantly in the cytoplasm, with low GFP-intense nuclei appearing as black holes within the GFP intense cytoplasm. Scale bar, 100 μm.

(C) The image features area, perimeter, and eccentricity and the NFAT score are shown for ctrl, INS, and PE. The boxplots show z-transformed feature values of 5 independent cell preparations, the box center shows the median value, and the box limits show the 25th and 75th percentiles. The p values were calculated with linear mixed effects models; **p < 0.01, ***p < 0.001. NS, not significant. The NFAT score was calculated via our cmoRe thresholding function on the feature Intensity_MedianIntensity_GFP of the nucleus.

agents.^{15–18} In cardiovascular research, however, no high-content morphology assay for translational applications has been established yet. Hence, we aimed to devise a robust *in vitro* methodology to deconvolute CM morphology for use in basic research and translational medicine.

We present C-MORE (cellular morphology recognition), a ready-to-use workflow for single-cell morphology analysis in CMs. The C-MORE workflow comprises a standardized experimental protocol including high-throughput image acquisition and morphological feature extraction with CellProfiler.⁵ With our R package cmoRe, we present data preprocessing and curation tailored for CM-specific biology, followed by data analysis for highly heterogeneous cell types such as NRCMs and hiPSC-CMs. We validate C-MORE against manual assessment and

benchmark C-MORE performance against alternative analysis approaches. Last, we evaluate C-MORE for use in basic and clinical cardiovascular research and demonstrate the applicability of C-MORE with four examples.

RESULTS

Experimental workflow, imaging, and feature extraction

We present an overview of our complete workflow in Figure 1A. NRCMs were isolated and seeded in 96-well imaging plates. On the following day, NRCMs can optionally be transduced with a reporter construct to add specific biological information to the assay; e.g., the subcellular location of a protein of interest. As a proof of principle, we chose a viral reporter for NFAT

(nuclear factor of activated T cells), a transcription factor known to translocate to the nuclear compartment upon activation of the pro-hypertrophic Calcineurin-NFAT pathway but otherwise located in the cytoplasm.¹⁹ Cells were then treated with test substances for 48 h; e.g., hypertrophic stimuli, inhibitors, or plasma. The cytoskeleton and the nuclei of cells were stained fluorescently, and NFAT was detected by native green fluorescence. After image acquisition with the IN Cell Analyzer 2200 microscopy system, segmentation and computing of single-cell morphological features were performed using the freely available software CellProfiler.⁵ We adapted a standard pipeline comprising nucleus segmentation with the Otsu algorithm and cell segmentation using the propagation algorithm (Data S1 and S2; Figures S1A–S1D). For initial development purposes, we worked with the two canonical hypertrophic stimuli phenylephrine (PE) and insulin (INS).^{20,21} In Figure 1B, we show representative immunofluorescence images for control, PE, and INS treatment together with representative feature values in Figure 1C. PE is known to activate the Calcineurin-NFAT pathway and, therefore, served as a positive control for the reporter readout (Figure 1B, GFP/fluorescein isothiocyanate [FITC] images).

Data preprocessing: *In silico* cell sorting and cell cycle features

CellProfiler output consists of a large data matrix with primary features for each measured cell (nucleus, cytoplasm, and custom readouts). These can be analyzed with our custom R package cmoRe. The main steps include (1) data loading and quality control, (2) cell cycle analysis and secondary feature calculation, (3) cell filtering, and (4) data analysis of aggregated or non-aggregated single cells, as outlined in Figure 2A (detailed outline in Figure S2).

cmoRe expects a specific folder structure for data loading (data organization). With these data, an initial quality control (QC) is performed: QC plots depict the number of cells per well on the plate layout (Figure 2Ab). To assure absence of potentially interfering edge effects or cytotoxicity, we checked whether at least 500 cells per well were present (Figure 2Aa). Further QC plots or criteria can be set generated on any feature or metric of interest.

In the next step, specific features of interest are assessed for filtering steps and deriving biologically relevant information; e.g., the cell cycle state. For cell cycle analysis, we utilized the well-established approach of quantifying DNA content via DAPI intensity (integrated DAPI intensity within the nucleus), as used commonly in flow cytometry.^{22,23} For filtering and cell cycle analysis, distributions of the respective feature measurements are evaluated per biological replicate, plate, and treatment to identify reasonable cutoffs under a set of constraints (prior knowledge; e.g., an expected range for a cutoff; Methods S1). Regarding the cell cycle, the most abundant fraction of cells are G1/G0 cells. Therefore, the highest peak of nuclear DAPI signal in a histogram/estimated density will correspond to G1/G0 cells. cmoRe identifies this peak as global maximum in the corresponding density and a (local) minimum right (mR) and minimum left (mL) from this value utilizing a resampling method. These identified minima are used as thresholds to

separate dead cells (left from mL) and G2/M cells (right from mR) (Figure 2Ac; Methods S1). Additional filters, using the same methodology, are implemented to remove improperly attached cells characterized by a high nuclear to cellular area ratio and median DNA intensity of the nucleus to discriminate CMs from non-CMs (see also Ali et al.²⁴). These filters were developed and validated on a manually rated dataset, as described in Figure S3 (filter steps). Cell cycle fractions are a prime example of secondary features and are calculated in cmoRe by default. Defined by the user, any feature can be processed analogously, yielding custom secondary features and, thus, extending the number of features originally provided by CellProfiler as, e.g., the NFAT score. The NFAT score was calculated via the cmoRe thresholding function on the feature Intensity_MedianIntensity_GFP of the nucleus to quantify NFAT translocation to the nucleus. For single-cell analysis, these cutoffs, as, e.g., the nuclear DNA intensity, can be used to assign single cells to specific groups (e.g., CMs or non-CMs) for cell-type-specific analysis/filtering (see below).

Finally, data of pure, vital, and properly attached CMs can be analyzed on the single-cell level or after median aggregation of features per well. Typically, aggregated data are z-transformed per experiment.

Data analysis: Feature selection

To analyze well-aggregated data for differential treatment effects, we use a number of steps to ensure robustness of identified features and implement redundancy reduction by generation of meta-features. We utilize (1) cross-validation, (2) multiple test correction, and (3) dimensionality reduction in cmoRe. (1) and (2) are performed with linear mixed models with random effects of biological replicate and plate to adjust for unwanted variability. For meta-feature calculation, hierarchical cluster analysis with adaptive selection of numbers of clusters was utilized.

Figure 2A demonstrates analysis steps, and Figures 2B–2E show representative analysis results for identification of PE-induced morphological feature alterations. After preprocessing and secondary feature generation, the data matrix consisted of 1,338 morphological features for PE and approximately 110,000 cells per plate. These were each evaluated with a linear mixed model, assuming a linear equidistant dependency between the control (ctrl) and lowest concentration as well as between dose levels (Figure 2B; Methods S1). Cross-validation ensured only retaining features with high reproducibility (Figure 2E; Cross-validated coefficient of determination (R^2_{cv}) > 0.7), and multiplicity adjustment was performed (Benjamini-Hochberg). For PE, 112 features were retained (Figure 2E; adjusted p value (p-adj) < 0.05).

To summarize highly correlated features and to ease interpretability, dimensionality reduction was performed using hierarchical clustering (Figure 2C). Numbers of clusters were iteratively incremented within a fixed range, and the intra-cluster variability was calculated (Figures S4A–S4C). This knee plot gives a first orientation for a suitable cutoff, but we used the change in intra-cluster variability for cutoff selection (Figures S4A–S4C; Methods S1). For PE, this approach identified 38 clusters/meta-features (Figure 2E). The latter can be visualized with radar plots (Figure 2D).

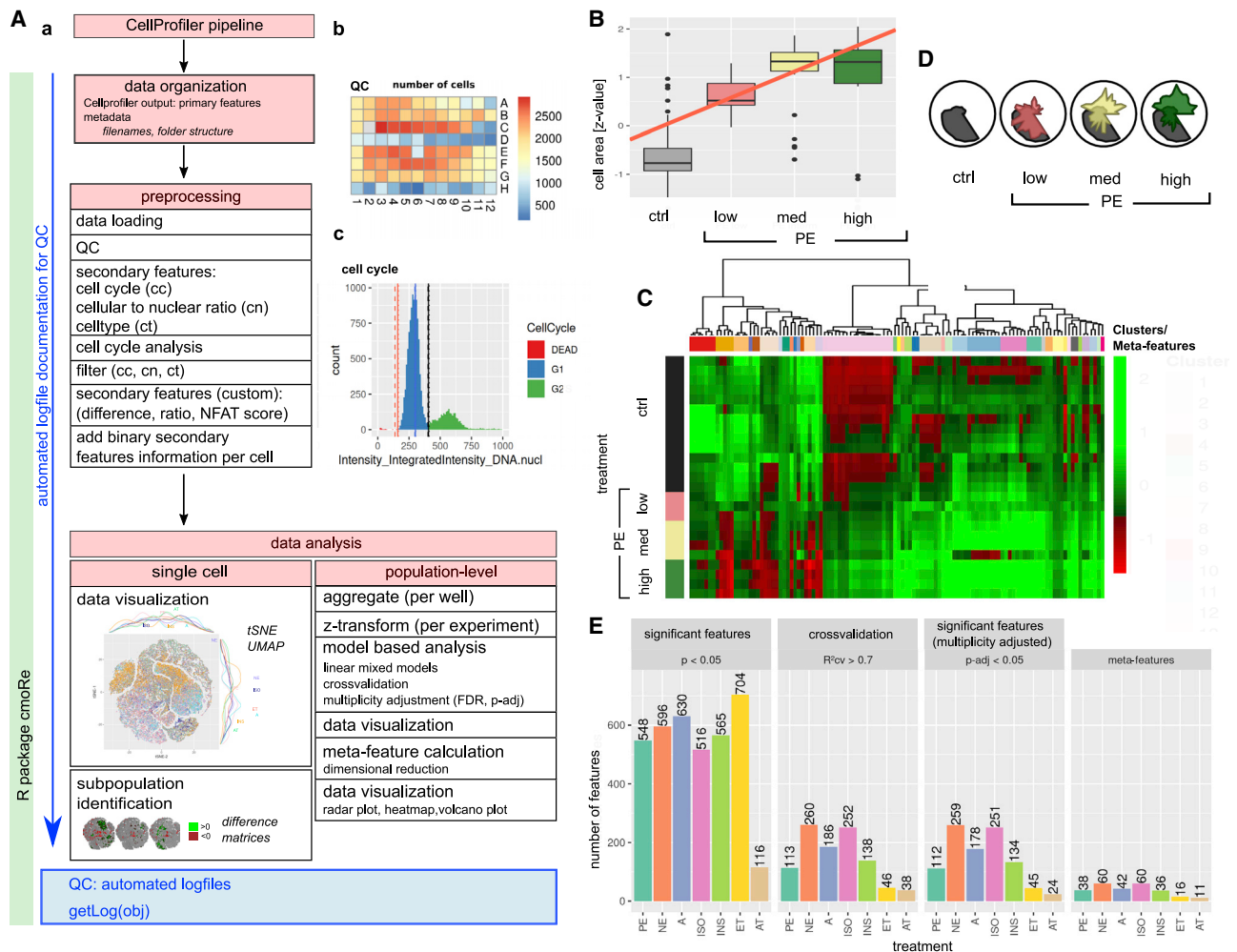


Figure 2. Preprocessing and data analysis with cmoRe

(A) Overview of cmoRe package functionality. (a) Preprocessing comprises data loading, QC, filtering steps, and addition of secondary features. Data analysis can be done at the single-cell level or bulk population level. (b) An exemplary QC plot for a 96-well plate regarding the number of cells per well. (c) A representative cell cycle plot as used for cell cycle-based filtering of apoptotic cells and to assign cell cycle status as a secondary feature to each cell. Vertical lines indicate determined cutoffs; dashed lines limit high-certainty regions of cutoffs.

(B) This boxplot shows the z-transformed cell area for NRCMs incubated with three concentrations of PE in comparison with the control (ctrl). The red line depicts a fitted linear mixed model assuming equidistance between ctrl/dose-levels, as applied in our analyses. The box center shows the median value, and the box limits show the 25th and 75th percentiles.

(C) Hierarchical cluster analysis of selected features (ward.D2) for meta-feature construction. Each cluster corresponds to a meta-feature (color coded).

(D) The PE phenotype. Radar plots show meta-feature values for PE (colored) against meta-feature values of an unstimulated control (gray) ordered in a circle. For PE, three concentrations are shown (red, low; yellow, intermediate; green, high; control, gray).

(E) Numbers of retained features after each selection step for the canonical hypertrophic stimuli: norepinephrine (NE), adrenaline (A), isoproterenol (ISO), endothelin (ET), and angiotensin (AT); Data of 5 independent cell preparations are shown.

An alternative to the aggregated bulk population analysis constitutes single cell analysis, which might help with identification of (functional) subpopulations of cells (see below).

Validation and benchmarking

We present an overview of established morphological tools with regard to cardiology or integrated analysis in general in Table S1. Because NRCMs are cells with high morphological heterogeneity, thus hindering established cell morphology analysis ap-

proaches, we aimed to assure that cmoRe-analyzed data are reliable and, therefore, manually evaluated key morphological features (1). Further, we compare cmoRe with naive analysis approaches on Cell Profiler output data (2).

- (1) Cell sizes used in cmoRe analyses were compared with manually contoured cell size measurements, revealing a high overlap (Figures S1E and S4E). The number of NFAT-GFP-positive nuclei showed high concordance between manual assignment and cmoRe-based labels (Figure S1F).

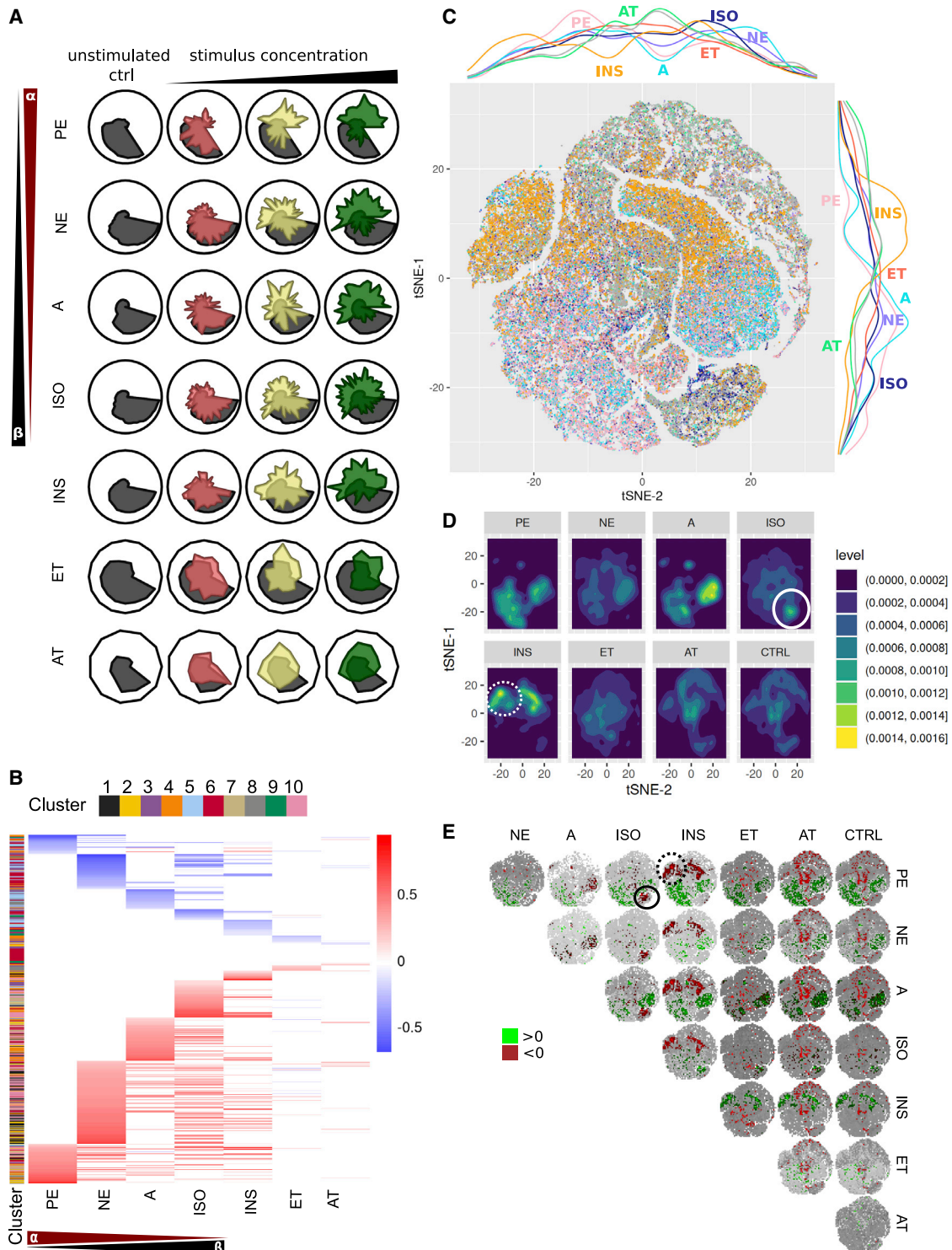


Figure 3. Population-level and single-cell phenotyping of canonical hypertrophic stimuli

(A) Radar plots show meta-feature values for the canonical hypertrophic stimuli aggregated on a population level (three concentrations: red, low; yellow, intermediate; green, high; untreated control, gray).

(B) Comparative analysis of the features selected for the canonical hypertrophic stimuli. For each stimulus, the dose-dependent dynamic is color coded (red indicates an increase and blue a decrease; clusters were calculated on an unstimulated control). See also Figure S4D.

(legend continued on next page)

CellProfiler is a versatile tool for cell segmentation and extraction of morphological features from images and an integral element in our C-MORE workflow. However, we aimed to assess the added value of combined use of CellProfiler and cmoRe.

- (2) We evaluated pairwise comparison results of ctrl and INS/PE with three dose levels against the full cmoRe workflow (preprocessing and filtering of cells, linear mixed model analyses to naive linear model analyses using filtered and non-filtered Cell Profiler data with per-well aggregated data) (Figure S5).

Cell size alterations induced by a strong stimulus such as PE were detected with all analysis approaches (Figures S5A and S5B). For weaker stimuli such as INS, only cmoRe was able to identify significant differences between control and treated cells for all tested dose levels. Assessment of all cellular and nuclear morphological features in PE/INS-treated cells revealed a small benefit of using filtered versus non-filtered data with naive analysis approaches in detecting differences, but a major increase in detected differential features was observed for cmoRe (Figure S5C).

As a proof of principle, after establishing C-MORE, we aimed to test its usability for basic research questions.

Morphological patterns of canonical hypertrophic stimuli

A potentially reversible preliminary step of heart failure is hypertrophic growth of the heart, the primary response to an increased workload.²⁵ Key signaling pathways involved in cardiac hypertrophy have been identified.²⁶ We chose these well-described canonical hypertrophy stimuli to model differing modes of actions and to generate a broad range of phenotypes to train C-MORE.^{9,26} Following the C-MORE workflow, we incubated NRCMs with PE, INS, norepinephrine (NE), epinephrine/adrenaline (A), isoproterenol (ISO), endothelin ET 1, and angiotensin (AT) II in three respective concentrations (Table S2; Methods S1). Figure 2E shows differences in the numbers of selected features for each selection step with cmoRe. Initially, hundreds of altered features are identified. However, fewer of these features meet the cmoRe criteria for robustness ($R^2_{cv} > 0.7$). Further dimensional reduction is achieved by multiplicity adjustment and meta-feature aggregation. In Figure 3A, we show meta-features for all stimuli under three concentrations (analogous to Figure 2D). We ordered the stimuli by similarity in previously described adrenergic receptor activation (alpha, beta, and none). Direct comparison of the feature pattern between the tested stimuli revealed that each stimulus pattern comprises a set of stimulus-specific features and a set of shared features with other stimuli (Figure 3B; Fig-

ure S4D). Biologically related stimuli, (A and ISO, for example) share a great amount of features (Figure 3B). Additionally, we could validate that the NFAT reporter was functional because it showed translocation to the nucleus for stimuli known to activate the respective pathway (Figure S4F).

Single-cell phenotyping

High inherent variability of the primary cell types and the observed non-uniform morphology following stimulation highlight the importance of single-cell analysis for detection and characterization of subpopulations or subtypes of cells. Especially less abundant cell populations might easily be missed with per-well aggregated population analyses. Therefore, we assessed low-dimensional t-distributed stochastic neighbor embedding (t-SNE) (Figures 3C–3E) and uniform manifold approximation and projection (UMAP) (Figures S6A and S6B) representations of our high-dimensional morphology measurements of NRCMs treated with canonical stimuli. Different areas in these representation plots represent specific morphological phenotypes, and pairwise comparisons allow fast identification of enriched or reduced regions (corresponding to specific subpopulations) following treatment (e.g., circles and dotted circles in Figures 3D and 3E).

Screening in primary CMs

Automated image analysis opens the possibility of high-throughput screening. As a second exemplary application, we set up a screening setting for which we treated PE-stimulated NRCMs with a selection of inhibitors of canonical hypertrophy pathways²⁷ (Methods S1; Table S2). The resulting phenotypes are shown in Figure 4A, and with increasing inhibitor concentration, a phenotypic shift from the PE phenotype back to the control phenotype was observed. For screening purposes, we evaluated differences between meta-features in positive control (PE) versus positive control and inhibitor-treated cells (Figure 4B). Plotting a volcano plot with effect size and p value further simplifies evaluation, and the user is able to set the p value cutoff depending on the screening purpose (Figure 4B). C-MORE provides tools for hit detection in large-scaled screening approaches as commonly used for drug development.

Liquid biopsy: Aortic valve stenosis phenotype

In a third example, we aim to demonstrate the potential clinical applicability of C-MORE using blood-treated NRCMs. Aortic valve stenosis (AS) is a cardiac disease with high prevalence in elderly patients and commonly treated with transcatheter aortic valve replacement (TAVR)²⁸. We incubated NRCMs with blood plasma from individuals with AS before and within 1 week

(C) Single-cell phenotyping of the canonical hypertrophic stimuli: t-distributed stochastic neighbor embedding (t-SNE) of all single cells measured in one experiment (highest substance concentrations, substances are color coded). Marginal densities for each stimulus over the t-SNE plot are shown by colored lines to the right and at the top of the t-SNE plot.

(D) Densities per stimulus on the t-SNE plots from (C). Visually, areas of notably higher density represent subpopulations of potential interest and are circled for ISO and for INS are indicated with a dotted line.

(E) Quantification of stimulus-induced differences, pairwise differences of binned and normalized data (50×50 grid, rate differences were calculated: $x - y$ axis). The most extreme differences are colored in red (upper 5%) and green (lowest 5%). Quantitatively identified substance-specific subpopulations are circled for ISO and for INS are circled with a dotted line.

In (A) and (B), data are shown for 5 independent experiments. In (C)–(E), single-cell analysis of 1 experiment is shown.

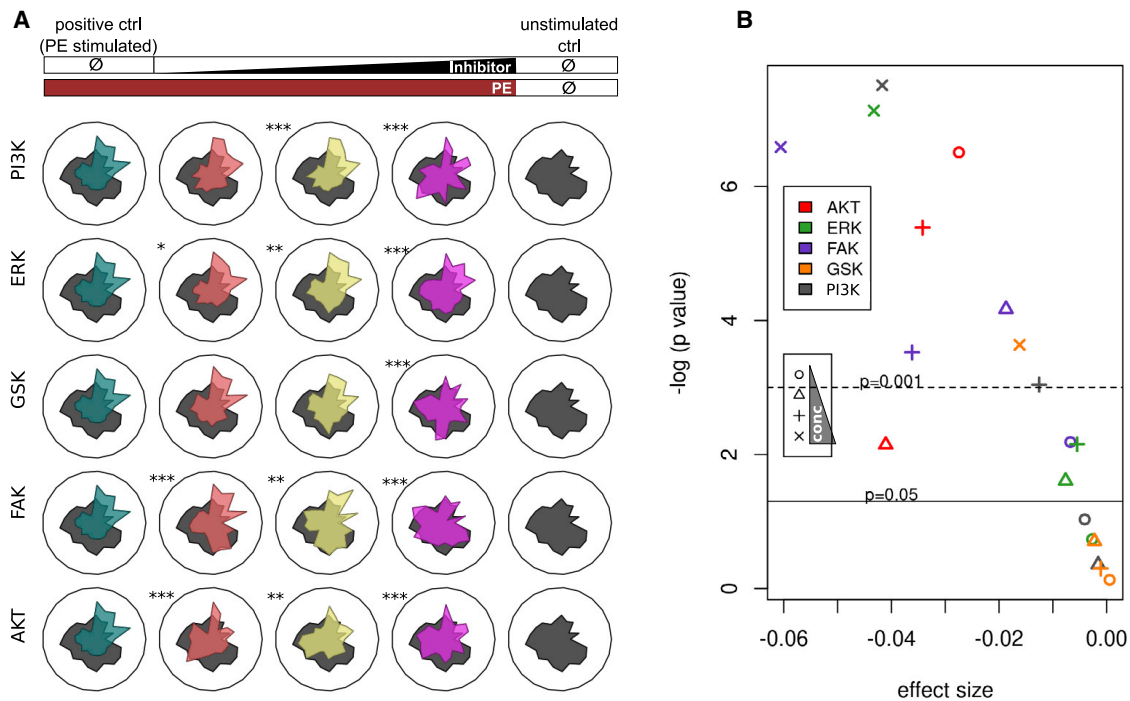


Figure 4. Inhibition of PE-induced hypertrophic growth by canonical pathway inhibitors

(A) Radar plots depict phenotypical changes upon concomitant application of PE and inhibitors of main cardiac hypertrophy pathways in three concentrations (low, red; intermediate, yellow; high, pink). A positive PE-stimulated control (PE stimulation without any inhibitor) is shown in turquoise, and an untreated control condition is shown in gray. Stars indicate whether significant differences between PE-stimulated and inhibitor-treated cells with PE stimulation can be observed (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, linear mixed effects model; evaluation of absolute differences in meta-features between PE stimulated \pm inhibitor-treated cells). (B) Volcano plot of effect size and p value (linear mixed effects model) of concomitantly inhibitor- and PE-treated phenotype versus PE-treated phenotype. Data are shown for 4 independent experiments.

following TAVR; blood plasma from healthy individuals served as controls (ctrl). Applied selection steps with C-MORE and numbers of selected features are presented in Figure 5A. The prominent aortic valve stenosis phenotype and the attenuated post-TAVR phenotype are shown in Figures 5C and 5D. For individuals with aortic valve stenosis in comparison with controls, 224 features were altered significantly; after TAVR treatment, 88 of these 224 features (39%) did not differ significantly from the control. In the next step, as a tool for systematic investigation of disease mechanisms, cmoRe allows cross-experimental comparison. As shown in Figure 5B, we quantified shared features of the canonical hypertrophic stimulus experiment and the aortic valve stenosis phenotype as well as TAVR-reversible features. To evaluate soundness of the selected features, we conducted reclassification using a random forest classifier that showed excellent performance in separating and predicting the conditions AS, post-TAVR, and ctrl (Figures S6C and S6D). In the setting of a liquid biopsy, C-MORE allowed us to detect morphological correlates of a cardiac disease and therapeutic effects with predictive properties.

Phenotypic characterization of hiPSC-CMs

Generation of organ-specific tissue *in vitro* from specific induced progenitor stem cells from affected individuals opens up a multitude of applications.²⁹ However, hiPSC-CMs are highly hetero-

geneous in their appearance and, like NRCMs, challenge morphological analysis. hiPSC-CM differentiation is a delicate process and not a specialty of our group; therefore, we re-analyzed plates from an already published research article for this fourth example. Seeger et al.³⁰ published an article about the mutation p.R943x in myosin binding protein C3 (MYBPC3) in individuals with hypertrophic cardiomyopathy regarding calcium handling and contractility. In this study, hiPSC-CMs derived from individuals with the heterozygous (het) mutation genotype were compared with healthy wild-type (WT) genotype controls. Additionally, in this study, hiPSC-CMs were genetically modified to generate an induced het genotype (het-ind) and an artificial homozygous genotype (hom) from WT hiPSC-CMs. Comparison of cell size showed no difference in this study; however, no high-content morphology analyses were conducted. We were kindly provided with the very exact same plates and cells used for this publication and re-evaluated cell size with cmoRe (Figures 6A and 6C). We could confirm no significant change in cell size; however, applying full cmoRe analysis, we observed a prominent phenotype for hom and an attenuated form for het in comparison with the WT (Figures 6B, 6D, and 6E). cmoRe was applied successfully in a setting, where traditional morphological readouts such as cell size show no change; however, we were able to identify a prominent genotype-related phenotype using cmoRe.

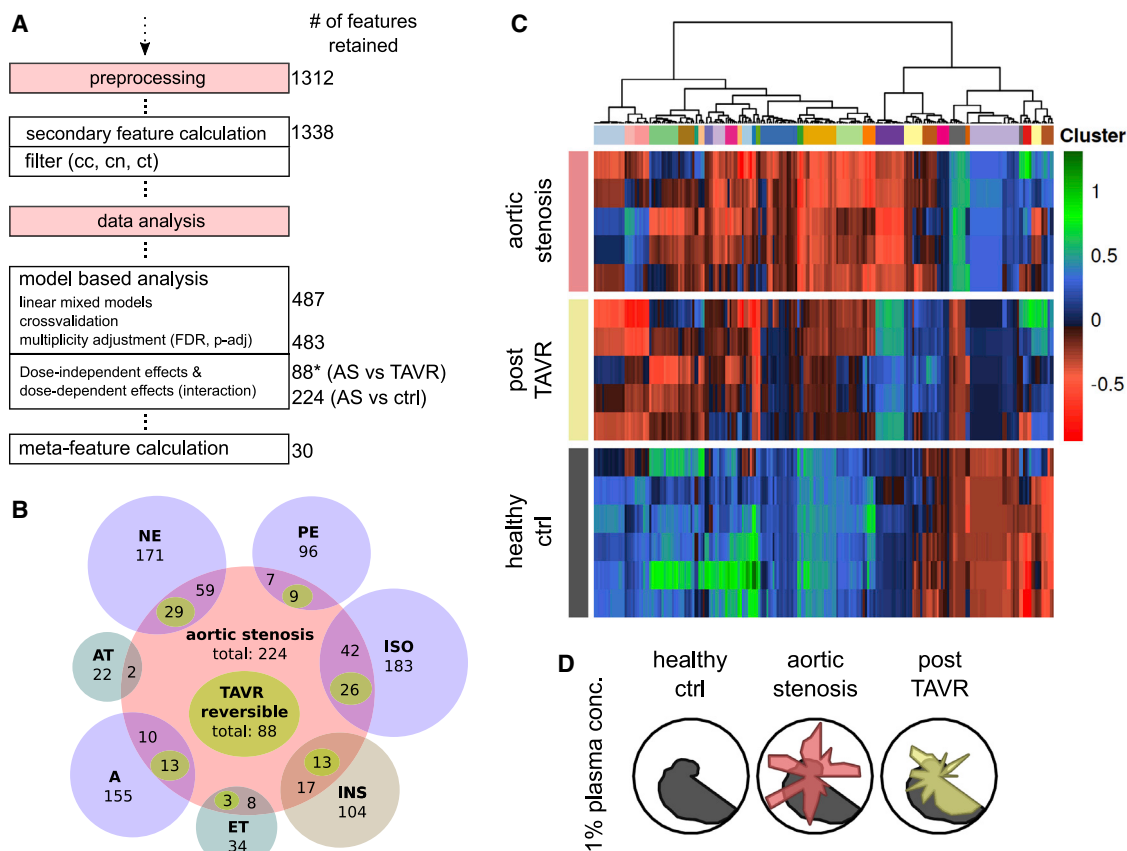


Figure 5. Morphological pattern of aortic valve stenosis (aortic stenosis/AS) and reversibility upon TAVR

(A) Preprocessing and data analysis steps and retained number of features in each step (*, reversible after TAVR).

(B) Venn-like diagram of substance-specific and AS-specific features. AS-specific TAVR-reversible features are counted separately and highlighted in lime. Overlaps within substance-specific features are not shown.

(C) Heatmap of model-predicted batch-corrected differential morphological features (hierarchical clustering, ward.D2).

(D) Radar plots show differentially regulated meta-features for AS and TAVR. Data are shown for 3 independent experiments.

DISCUSSION

Cellular morphology and its proficiency to serve as a surrogate for cell state and functionality is a powerful tool typically used for drug screening and increasingly for personalized medicine in oncology.^{4,15} In cardiology research, primary NRCMs are the most frequently used cell model for preclinical studies. NRCMs and hiPSC-CMs, however, exhibit a high degree of intercellular morphological variability, making identification of specific morphological phenotypes challenging.⁷ We therefore developed C-MORE, an experimental pipeline and analysis platform, to enable researchers to make use of the full potential of CM morphology assessment. Main advantages comprise the high number of morphological features being made available (including customizable secondary features) and the high power of the approach to detect small but nevertheless robust effects in highly variable data.

For the experimental part, we aimed to establish a lean workflow and therefore chose cytoskeletal and nucleic staining suitable for shape and texture analysis. This marker panel can be adapted to the specific purpose of C-MORE application; e.g.,

staining of specific intracellular proteins of interest. NRCM isolation optimally yields a highly pure population of more than 90% CMs. However, the presence of a minor fraction of non-CMs needs to be considered for accurate analysis. In CM-specific morphology readouts, this has been taken into account, as reported, e.g., by Jentzsch et al.³¹ However, these approaches typically require additional fibroblast-specific staining and were limited to a few morphological features.³² In contrast, cmoRe identifies CMs and non-CMs based on DAPI intensity, as reported previously by Ali et al.,²⁴ and further curates data from dead and non-attached cells; the automated and adaptive thresholding function is able to accommodate changing cell densities and highly heterogeneous cell sizes following specific treatments. By applying the cmoRe threshold function on the integrated intensity of DAPI as a surrogate for DNA content, we integrate a classical cell cycle analysis approach into our automated analysis, circumventing the need for manual gating as in flow cytometry.^{22,23}

Recently, the importance of cross-talk between different cell types in the heart has been investigated.⁷ With the ability to robustly differentiate CMs from non-CMs, cmoRe opens possibilities to analyze cross-talk between the different cell types in

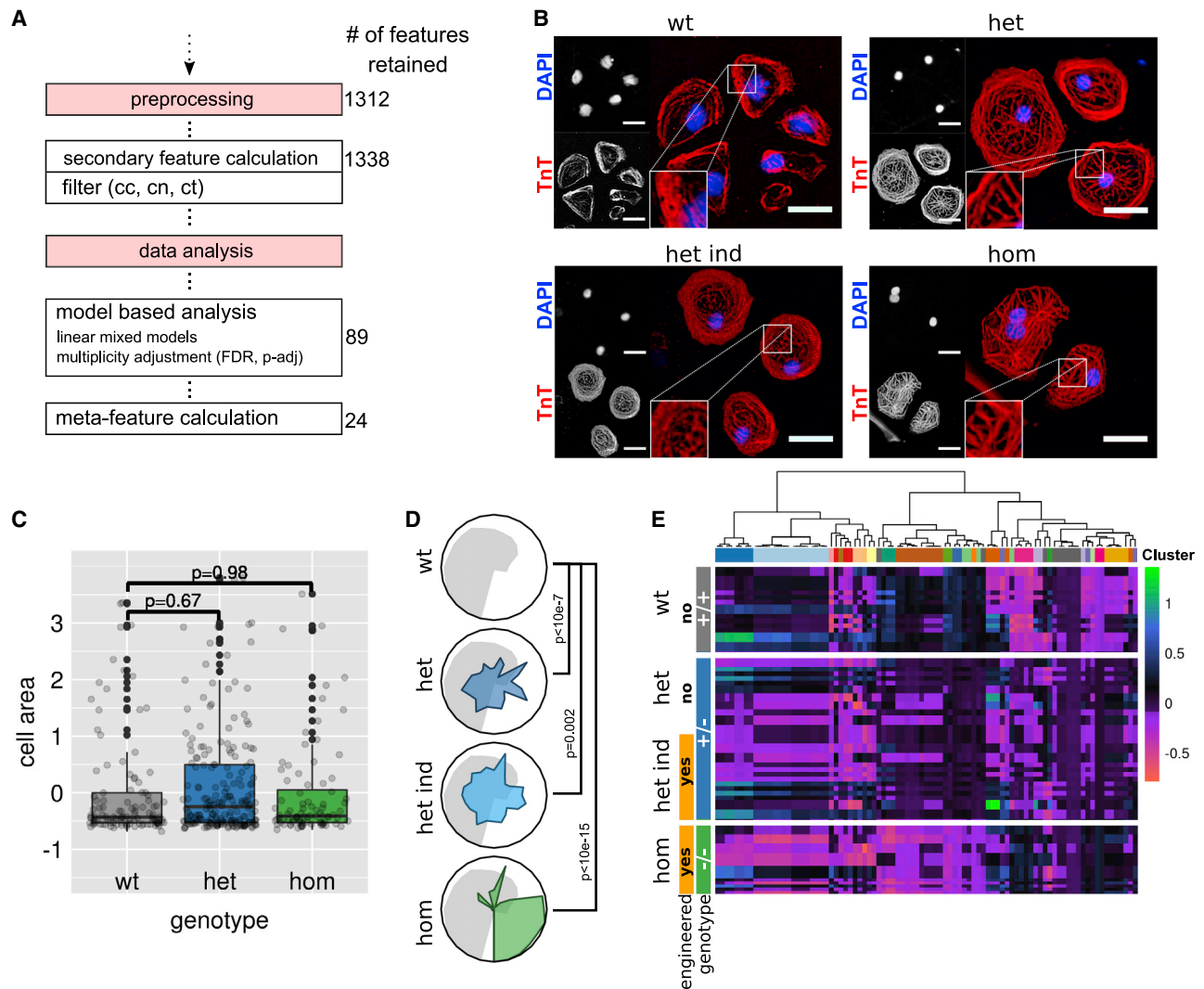


Figure 6. Morphological pattern of wild-type (WT), heterozygous (het), and homozygous (hom) MYBPC3 pR943x genotype hiPSC-CMs

(A) Preprocessing and data analysis steps with retained number of features in each step.

(B) Representative images of WT, het, induced/engineered het (het-ind), and hom genotype hiPSC-CMs regarding the MYBPC3 pR943x mutation. In the lower left corner of the Troponin T (TnT) images, the indicated image section is shown magnified. Scale bar, 30 μ m.

(C) Cell area in MYBPC3 pR943x WT, het, and hom cells ($n = 4$ patients, linear mixed effects model, z-transformed data). The boxplots show z-transformed values of cell area, the box center shows the median value, and the box limits show the 25th and 75th percentiles.

(D) Radar plots showing meta-feature values for the pR943x phenotype (het, het-ind, and hom genotypes, colored) against the WT (gray). p values: linear mixed effects model.

(E) Heatmap of model-predicted differential morphological features for the genotypes WT, het, and hom (hierarchical clustering, ward.D2). $n = 4$ hiPSC-CM cell lines derived from affected individuals, $n = 2$ from pR943x het mutated individuals, $n = 2$ from pR943x WT and phenotypically healthy persons.

the heart; e.g., using co-culture experiments or even whole-heart cell suspensions.

Our data preprocessing tools ensure identification of cells of interest and appropriate handling of features with skewed or variable distributions and can acquire additional biologically valuable information.

For data analysis, we used linear mixed models to account for variability between biological replicates. Canonical hypertrophy stimulus and inhibitor experiments were performed with a number of increasing dose levels for which we assumed a linear dose

dependency over ascending treatment dosages. Although this serves as a robust base analysis, alternative dependencies (foremost sigmoidal) should be tested in further work. We implemented a cross-validation step, multiplicity adjustment and dimensional aggregation to meta-features, to provide a highly sensitive and robust analysis tool. Benchmarking showed superior performance of our proposed C-MORE workflow compared with naive analysis approaches using CellProfiler output data but lacking essential characteristics of C-MORE. Application of C-MORE was especially beneficial in settings where morphological

changes were very subtle (a weaker stimulus such as INS or lower concentrations of stimuli), cmoRe allows automated data curation (filtered cells) and is characterized by powerful detection of subtle changes in settings of low signal to noise. We might be able to apply C-MORE to other types of heterogeneous primary cells in the future.

Cardiac hypertrophic growth is a heterogeneous process and can involve mechanistically different pathways. However, more work is needed to better understand hypertrophic growth as a potential target for therapy.³³ Because our understanding of cardiac remodeling deepens on the molecular level, measuring a few morphological parameters without application of appropriate preprocessing steps risks missing relevant morphological effects.³⁴ We applied C-MORE to four exemplary applications for basic research and translational medicine to evaluate its performance but without focusing on biological interpretation.

To evaluate C-MORE as a diagnostic tool, NRCMs were incubated with blood plasma from individuals with aortic valve stenosis before and after TAVR and from healthy persons. In this pilot study, we observed excellent classification accuracy in a representative random forest analysis. This implicates that the morphology of cells incubated with material from affected individuals might reflect a complex disease state better than conventional blood tests and could be utilized to support non-invasive diagnostics and clinical decision-making in the future. Linking results cross-experimentally, as demonstrated for aortic valve stenosis and the canonical hypertrophy stimuli, can be a valuable tool for investigating mechanisms of actions in cardiological disease.

hiPSCs from affected individuals are an important tool for disease modeling and can facilitate implementation of precision medicine. However, differentiation of hiPSCs into CMs to study cardiovascular disease is a lengthy and highly variable process.^{34,35} Seeger et al.³⁰ demonstrated that hiPSC-CMs derived from MYBPC3 mutant individuals showed no significant difference in cell size compared with healthy controls. However, the ability of cmoRe to detect a prominent phenotype for the hom genotype and an attenuated phenotype (het) emphasizes the benefits of our approach. Applying C-MORE to patient-derived hiPSC-CMs may help deepen our understanding of individual pathology and enable individualized drug screening in the frame of personalized medicine. Furthermore, culture and maturation of hiPSC-CMs is a very delicate process under intense investigation. In general, fine-tuning of the attachment as lamellipodium formation, spreading or establishing focal adhesion are biological meaningful process of cell biology, reflecting the functional status of a cell in a detectable morphological readout. C-MORE might help to further our understanding of the underlying biology and standardize these processes.

We present C-MORE as an open-source tool to identify morphological patterns in basic research and cardiovascular disease in translational approaches. C-MORE is compatible with both primary cell types used most frequently in cardiac research and comprises preprocessing steps and data analysis suited for extracting relevant information out of data with a low signal-to-noise ratio. The demonstrated versatility of C-MORE makes it ideal for quick and easy application in drug screening

and in-depth characterization of single cells for investigating CM biology. Future applications of C-MORE may include investigation of disease mechanisms and drug screening in CMs, and might also be applicable to other primary cell types beyond the niche of CMs. In translational medicine, C-MORE shows applicability for material from affected individuals with potential to support clinical decision-making in the frame of personalized medicine.

Limitations of the study

With our experimental setup, we present a lean C-MORE experimental workflow, but it is possible to integrate additional custom settings, such as staining for other target proteins of interest or other reporter readouts. As a proof of principle, we integrated the NFAT reporter only. Furthermore, staining of the cell membrane might improve exact quantification of cell area or cell volume compared with staining of structural proteins such as Desmin or Troponin as used here. The applied microscope setup only allows acquisition of three fluorescence channels. Using more markers or even applying a multiplexing approach would be possible and would only be limited by the specifics of the microscope in use and computational power available.

Our density-based algorithm used for the preprocessing steps offers adaptive thresholding and can accommodate variation and skewed populations. However, in its current implementation, it is limited to detection of a maximum of three subpopulations (Methods S1). Also, in our study, the mixed linear effects models assume linear dose dependency over different treatment dosages. Although this shows superior performance over analyses lacking cmoRe characteristics, other dose-effect dependencies (e.g., sigmoidal) are currently not modeled explicitly.

The four representative use cases of C-MORE were designed as pilot studies to demonstrate C-MORE's versatile fields of application and require further intense experimentation.

For the liquid biopsy approach, we currently use samples of healthy control individuals as a reference, but we expect to enable computing a virtual control sample in the future based on the growing database.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - NRCM isolation and treatment
 - hiPSC-CMs
- **METHOD DETAILS**
 - AdNFATc3-GFP adenovirus amplification and purification
 - Immunocytochemistry
 - Image acquisition (IN Cell Analyzer 2200)
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

- Image processing, cell segmentation and feature extraction
- Preprocessing
- Single cell phenotyping
- Population-level phenotyping

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2021.100436>.

ACKNOWLEDGMENTS

We thank Dr. Katrin Rein for excellent support in creating the graphical abstract. This work was supported by the Deutsche Forschungsgemeinschaft (DFG; Ko3900/2-1 to M.H.K., Unite: SFB-1389, Project 404521405), the German Centre for Cardiovascular Research (DZHK), the Klaus-Tschira-Stiftung, and intramural funds of the National Center for Tumor Disease (NCT) and the German Cancer Consortium (DKTK). M.K. and J.F. received stipends in the frame of the MD/PhD program of the Heidelberg Faculty of Medicine (MFHD), Heidelberg University. Studies of human samples were performed under a protocol approved by the institutional review board in accordance with the declaration of Helsinki. Experimentation with primary rat cardiomyocytes was done under an approved animal protocol.

AUTHOR CONTRIBUTIONS

Conceptualization, J.F., M.K., and M.H.K.; methodology, J.F., M.K., and M.H.K.; investigation, J.F., M.K., S.D., N.V.B., T.S., and M.H.K.; writing – original draft, J.F., M.K., and M.H.K.; writing – review & editing, J.F., M.K., S.D., N.F., A.A., H.A.K., and M.H.K.; funding acquisition, M.H.K. and H.A.K.; resources, M.H.K. and H.A.K.

DECLARATION OF INTERESTS

A.A. reports research grants (Merck and EMD, Fibrogen, and Bayer) and SAB (Merck and EMD, Fibrogen, BMS, Roche, and Bayer) and consulting roles (Bayer and BioMedX).

INCLUSION AND DIVERSITY

We worked to ensure sex balance in the selection of non-human subjects. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: March 10, 2021

Revised: August 4, 2021

Accepted: October 11, 2021

Published: November 3, 2021

REFERENCES

1. Cleland, J.G.F., van Veldhuisen, D.J., and Ponikowski, P. (2019). The year in cardiology 2018: heart failure. *Eur. Heart J.* *40*, 651–661.
2. Chimenti, C., and Frustaci, A. (2013). Contribution and risks of left ventricular endomyocardial biopsy in patients with cardiomyopathies: a retrospective study over a 28-year period. *Circulation* *128*, 1531–1541.
3. Bougen-Zhukov, N., Loh, S.Y., Lee, H.K., and Loo, L.H. (2017). Large-scale image-based screening and profiling of cellular phenotypes. *Cytometry A* *91*, 115–125.
4. Boutros, M., Heigwer, F., and Laufer, C. (2015). Microscopy-Based High-Content Screening. *Cell* *163*, 1314–1325.
5. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* *7*, R100.
6. Haralick, R.M., Shanmugam, K., and Dinstein, I.H. (1973). Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* *SMC-3*, 610–621.
7. Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., et al. (2020). Cells of the adult human heart. *Nature* *588*, 466–472.
8. Biendarra-Tiegs, S.M., Secreto, F.J., and Nelson, T.J. (2020). Addressing Variability and Heterogeneity of Induced Pluripotent Stem Cell-Derived Cardiomyocytes. *Adv. Exp. Med. Biol.* *1212*, 1–29.
9. Ryall, K.A., Bezzerides, V.J., Rosenzweig, A., and Saucerman, J.J. (2014). Phenotypic screen quantifying differential regulation of cardiac myocyte hypertrophy identifies CITED4 regulation of myocyte elongation. *J. Mol. Cell. Cardiol.* *72*, 74–84.
10. Hein, S., Furkel, J., Knoll, M., Aus dem Siepen, F., Schönland, S., Hegenbart, U., Katus, H.A., Kristen, A.V., and Konstantin, M.H. (2021). Impaired in vitro growth response of plasma-treated cardiomyocytes predicts poor outcome in patients with transthyretin amyloidosis. *Clin. Res. Cardiol.* *110*, 579–590.
11. Toepfer, C.N., Sharma, A., Cicconet, M., Garfinkel, A.C., Mücke, M., Neyazi, M., Willcox, J.A.L., Agarwal, R., Schmid, M., Rao, J., et al. (2019). SarcTrack. *Circ. Res.* *124*, 1172–1183.
12. Rupert, C.E., Chang, H.H., and Coulombe, K.L. (2017). Hypertrophy changes 3D shape of hiPSC-cardiomyocytes: Implications for cellular maturation in regenerative medicine. *Cell. Mol. Bioeng.* *10*, 54–62.
13. Woo, L.A., Tkachenko, S., Ding, M., Plowright, A.T., Engkvist, O., Andersson, H., Drowley, L., Barrett, I., Firth, M., Akerblad, P., et al. (2019). High-content phenotypic assay for proliferation of human iPSC-derived cardiomyocytes identifies L-type calcium channels as targets. *J. Mol. Cell. Cardiol.* *127*, 204–214.
14. van Rooij, E., Doevendans, P.A., de Theije, C.C., Babiker, F.A., Molkentin, J.D., and de Windt, L.J. (2002). Requirement of nuclear factor of activated T-cells in calcineurin-mediated cardiomyocyte hypertrophy. *J. Biol. Chem.* *277*, 48617–48626.
15. Snijder, B., Vladimer, G.I., Krall, N., Miura, K., Schmolke, A.S., Kornauth, C., Lopez de la Fuente, O., Choi, H.S., van der Kouwe, E., Gültekin, S., et al. (2017). Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label, pilot study. *Lancet Haematol.* *4*, e595–e606.
16. Gorshkov, K., Chen, C.Z., Marshall, R.E., Mihatov, N., Choi, Y., Nguyen, D.-T., Southall, N., Chen, K.G., Park, J.K., and Zheng, W. (2019). Advancing precision medicine with personalized drug screening. *Drug Discov. Today* *24*, 272–278.
17. Breinig, M., Klein, F.A., Huber, W., and Boutros, M. (2015). A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol. Syst. Biol.* *11*, 846.
18. Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J.K., Gustin, E., Chupakhin, V., Chong, Y.T., Vialard, J., Buijnsters, P., et al. (2018). Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chem. Biol.* *25*, 611–618.e3.
19. Wilkins, B.J., and Molkentin, J.D. (2004). Calcium-calcineurin signaling in the regulation of cardiac hypertrophy. *Biochem. Biophys. Res. Commun.* *322*, 1178–1191.
20. Shimizu, I., and Minamino, T. (2016). Physiological and pathological cardiac hypertrophy. *J. Mol. Cell. Cardiol.* *97*, 245–262.
21. Nakamura, M., and Sadoshima, J. (2018). Mechanisms of physiological and pathological cardiac hypertrophy. *Nat. Rev. Cardiol.* *15*, 387–407.
22. Darzynkiewicz, Z., and Juan, G. (2001). DNA content measurement for DNA ploidy and cell cycle analysis. *Curr Protoc Cytom. Chapter 7*, Unit 7.5.
23. Ferro, A., Mestre, T., Carneiro, P., Sahumbaiev, I., Seruca, R., and Sanches, J.M. (2017). Blue intensity matters for cell cycle profiling in fluorescence DAPI-stained images. *Lab. Invest.* *97*, 615–625.

24. Ali, S.R., Nguyen, D., Wang, B., Jiang, S., and Sadek, H.A. (2020). Deep Learning Identifies Cardiomyocyte Nuclei With High Precision. *Circ. Res.* *127*, 696–698.
25. Hein, S., Arnon, E., Kostin, S., Schönburg, M., Elsässer, A., Polyakova, V., Bauer, E.P., Klövekorn, W.P., and Schaper, J. (2003). Progression from compensated hypertrophy to failure in the pressure-overloaded human heart: structural deterioration and compensatory mechanisms. *Circulation* *107*, 984–991.
26. Mishra, S., and Kass, D.A. (2021). Cellular and molecular pathobiology of heart failure with preserved ejection fraction. *Nat. Rev. Cardiol.* *18*, 400–423.
27. Heineke, J., and Molkentin, J.D. (2006). Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat. Rev. Mol. Cell Biol.* *7*, 589–600.
28. Coffey, S., Cairns, B.J., and Lung, B. (2016). The modern epidemiology of heart valve disease. *Heart* *102*, 75–85.
29. Mummery, C.L. (2018). Perspectives on the Use of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes in Biomedical Research. *Stem Cell Reports* *11*, 1306–1311.
30. Seeger, T., Shrestha, R., Lam, C.K., Chen, C., McKeithan, W.L., Lau, E., Wnorowski, A., McMullen, G., Greenhaw, M., Lee, J., et al. (2019). A Premature Termination Codon Mutation in MYBPC3 Causes Hypertrophic Cardiomyopathy via Chronic Activation of Nonsense-Mediated Decay. *Circulation* *139*, 799–811.
31. Jentzsch, C., Leierseder, S., Loyer, X., Flohrschütz, I., Sassi, Y., Hartmann, D., Thum, T., Lagerbauer, B., and Engelhardt, S. (2012). A phenotypic screen to identify hypertrophy-modulating microRNAs in primary cardiomyocytes. *J. Mol. Cell. Cardiol.* *52*, 13–20.
32. Zhou, P., and Pu, W.T. (2016). Recounting Cardiac Cellular Composition. *Circ. Res.* *118*, 368–370.
33. Schiattarella, G.G., and Hill, J.A. (2015). Inhibition of hypertrophy is a good therapeutic strategy in ventricular pressure overload. *Circulation* *131*, 1435–1447.
34. Sutcliffe, M.D., Tan, P.M., Fernandez-Perez, A., Nam, Y.J., Munshi, N.V., and Saucerman, J.J. (2018). High content analysis identifies unique morphological features of reprogrammed cardiomyocytes. *Sci. Rep.* *8*, 1258.
35. Peter, A.K., Bjerke, M.A., and Leinwand, L.A. (2016). Biology of the cardiac myocyte in heart disease. *Mol. Biol. Cell* *27*, 2149–2160.
36. Sanna, B., Bueno, O.F., Dai, Y.S., Wilkins, B.J., and Molkentin, J.D. (2005). Direct and indirect interactions between calcineurin-NFAT and MEK1-extracellular signal-regulated kinase 1/2 signaling pathways regulate cardiac gene expression and cellular growth. *Mol. Cell. Biol.* *25*, 865–878.
37. MacDonnell, S.M., Weisser-Thomas, J., Kubo, H., Hanscome, M., Liu, Q., Jaleel, N., Berretta, R., Chen, X., Brown, J.H., Sabri, A.K., et al. (2009). CaMKII negatively regulates calcineurin-NFAT signaling in cardiac myocytes. *Circ. Res.* *105*, 316–325.
38. Konstantin, M.H., Völkers, M., Collins, B., Quijada, P., Quintana, M., De La Torre, A., Ormachea, L., Din, S., Gude, N., Toko, H., and Sussman, M.A. (2013). Fibronectin contributes to pathological cardiac hypertrophy but not physiological growth. *Basic Res. Cardiol.* *108*, 375.
39. Chin, E.R., Olson, E.N., Richardson, J.A., Yang, Q., Humphries, C., Shelton, J.M., Wu, H., Zhu, W., Bassel-Duby, R., and Williams, R.S. (1998). A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes Dev.* *12*, 2499–2509.
40. Bruder, J.T., and Kovesdi, I. (1997). Adenovirus infection stimulates the Raf/MAPK signaling pathway and induces interleukin-8 expression. *J. Virol.* *71*, 398–404.
41. Linderman, G.C., Rachh, M., Hoskins, J.G., Steinerberger, S., and Kluger, Y. (2019). Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* *16*, 243–245.
42. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*, arXiv:1802.03426. <https://arxiv.org/abs/1802.03426>.
43. Bates, D., Machler, M., Bolker, B.M., and Walker, S.C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* *67*, 1–48.
44. Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D.; R Core Team (2017). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131. <https://cran.r-project.org/web/packages/nlme/index.html>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal Anti-Desmin antibody, rabbit	abcam	Cat# ab15200, RRID: AB_301744
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Life Technologies, Thermo Fisher	Cat# A-11012, RRID: AB_2534079
Monoclonal Anti-troponin T antibody, rabbit	abcam	ab209813
Bacterial and virus strains		
AdNFATc3-GFP adenovirus	Sanna et al., ³⁶ MacDonnell et al. ³⁷	N/A
Biological samples		
Aortic stenosis patient blood plasma	Department of Cardiology, University Hospital Heidelberg	Protocol S-587/2019
Chemicals, peptides, and recombinant proteins		
Cesium chloride	Carl Roth	7878.2
Suprarenin/Epinephrin, 1mg/ml	Sanofi	6053210
(R)-(-)-Phenylephrine hydrochloride	Sigma	A9525
Endothelin 1 97% (HPLC), powder	sigma	E7764
Insulin Insuman rapid 40 ie	Sanofi	1843315
(-) Isoproterenol hydrochloride	Sigma	I6504
Arterenol 1mg/ml	Sanofi	3870227
Tricibine Akt V Inhibitor	Sigma	124038
ERK/MEK Inhibitor	Promega	U0126
PF 573228 FAK inhibitor	Tocris	3239
BIO GSK3b inhibitor	Tocris	3194
Ly294002, PI3K Inhibitor	Millipore	440202
Laminin	Sigma	L2020-1MG
Paraformaldehyd, 20% solution	Electron Microscopy	15713-S
HEPES	Carl Roth	9105.2
Medium 199	Sigma	M7528-500ML
FBS	Life Technologies	10500064
DAPI	Life Technologies	D1306
DNase II Type V from Bovine spleen	Sigma	D8764-300KU
Experimental models: Cell lines		
Neonatal rat cardiomyocytes from rattus norvegicus	Janvier	Rattus norvegicus, wistar
Software and algorithms		
INCellAnalyzer 2200 Acquisition Software	GE	N/A
CellProfiler	The Broad Institute	N/A
R 4.0	R Core Team	https://www.r-project.org
cmoRe	This paper	https://github.com/mknoll/cmoRe
Other		
96 Well Black with Clear Flat Bottom TC-Treated Imaging Plate with Lid	corning	353219

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jennifer Furkel (Jennifer.furkel@med.uni-heidelberg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at github and is publicly available as of the date of publication (<https://github.com/mknoll/cmRe>). CellProfiler pipelines are available as online supplemental files ([Data S1](#) and [S2](#)).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

NRCM isolation and treatment

NRCMs were obtained from hearts of 1-2 days old neonatal rats using a trypsin based enzymatic digestion standard protocol as previously described.³⁸ To enrich cardiomyocytes, a percoll gradient centrifugation was performed after digestion. Cells were then counted and plated onto imaging compatible 96-well plates (corning, 353219), which were coated with Laminin (sigma, L2020, diluted to 1mg/l PBS) for 2h beforehand. 20 000 cells per well were seeded in 200μl high serum medium (10% FBS in M199) for the first 24h.

24h after isolation, NRCMs were washed with PBS and incubated in 80μl of low serum medium (0.5% FCS in M199) with NFAT-GFP adenovirus for 12h. The following 48h cells were incubated with hypertrophy inducing stimuli (phenylephrine (PE), adrenaline (A), noradrenaline (NA), isoproterenol (ISO), insulin (INS), endothelin (ET), angiotensin II (AT), all reagents were diluted in water and PBS) and an unstimulated PBS control (CTRL). For the liquid biopsy/ aortic stenosis experiments NRCMs were treated for 48h with blood plasma of healthy controls or patients showing a high-grade stenosis of the aortic valve undergoing transcatheter aortic valve implantation (before TAVR, within one week after TAVR). For the inhibitor screening experiment cells were for 48h concomitantly incubated with PE and an inhibitor of major pathways involved in cardiac hypertrophy (inhibitors of protein kinase B (AKT), extracellular signal-regulated kinase (ERK), focal adhesion kinase (FAK), glycogensynthase kinase (GSK) or phosphoinositid-3-kinase (PI3K)). All inhibitors were diluted in Dimethylsulfoxid (DMSO), adequate DMSO solvent controls were performed. Detailed information of the reagents used and concentrations can be found in [Table S2](#).

hiPSC-CMs

hiPSC generation and differentiation into hiPSC-CMs was performed by and as previously described by Seeger et al.³⁰ The identical plates utilized in the study by Seeger et al. were kindly provided to us, permeabilized and stained with anti-Troponin T.

METHOD DETAILS

AdNFATc3-GFP adenovirus amplification and purification

The adNFATc3-GFP adenovirus was kindly provided by Dr. Mark Sussmann. The AdNFATc3-GFP adenovirus, as described previously was propagated on HEK293 cells.^{36,37,39} In brief, cells were harvested and the pellet was washed with PBS (10 minutes, 300 g, 4°C). We performed four freeze-thaw cycles to lyse the cells (thawing at room temperature, freezing on dry ice), incubated the solution with DNase I (10μg/mL, 30min, 37°C). Debris was pelleted at 750 g for 10 minutes at 4°C. The supernatant was layered onto a Cesium Chloride (CsCl) gradient (CsCl cushions of 1.4 g/ml and 1.2 g/ml). We performed one centrifugation for 3h, and two subsequent centrifugations of 18h (Beckman Coulter SW41 rotor, 22500 rpm, 4°C temperature). The purified virus was dialyzed against dialysis buffer (3% sucrose, 10mM Tris (pH 7.8 at room temperature), 150mM NaCl, 10mM MgCl₂), as described previously.⁴⁰ Storage temperature of the purified virus was –80°C.

Immunocytochemistry

In all NRCM experiments cells were fixed with 4% paraformaldehyde (PFA), permeabilized with 0.2% Triton X-100, then incubated in blocking solution (10% FBS in PBS) for one hour at room temperature. In all NRCM experiments cytoskeletal staining using a desmin antibody (1:800) was applied on the plates overnight at 4°C. After washing with PBS the secondary antibody (1:600) was applied for 1h. The hiPSC-CMs were stained with troponin T antibody and secondary antibody (1:400 and 1:600). Nuclear DNA was stained in all experiments using 4',6-Diamidin-2-phenylindol (DAPI).

Image acquisition (IN Cell Analyzer 2200)

Images were acquired in an automated fashion using the IN Cell Analyzer 2200. For all NRCM experiments a centered square arrangement of a 4x4 grid of imaging fields per well was chosen and three channels were imaged (DNA - DAPI, desmin – TexasRed, NFAT- GFP native fluorescence /FITC), as outlined in [Figures 1A](#) and [1B](#). For hiPSC-CMs, a circular arrangement of 96 imaging fields for maximum coverage of the well was chosen and two fluorescent channels for each field were imaged (DNA - DAPI, troponin T – TexasRed). Images had a resolution of 2048x2048 pixel.

QUANTIFICATION AND STATISTICAL ANALYSIS

Image processing, cell segmentation and feature extraction

Image features were computed from raw images using CellProfiler v 3.1.8.⁵ Upon loading images into the software, images of the three fluorescent channels blue (DAPI), TexasRed (Desmin/TroponinT), GFP/FITC (NFAT) are matched by file name and image meta-data such as well number and image number are extracted from file names. Then nuclei are segmented based on an intensity-based algorithm on the DAPI image (single steps: [Figure S1A](#)). Using these nuclei as seeding points, cells are segmented by the propagation algorithm, which identifies cell borders based on intensity changes ([Figures S1B](#) and [S1D](#)). The occurrence of double nucleated cells is taken into account by assigning nuclei located within 10 pixels from each other to one cell ([Figure S1C](#)). This cutoff was validated by comparing unstimulated and PE treated cells to be optimal for discrimination of binucleated cells from single nucleated cells in close proximity. Cells touching the image borders are excluded from further analysis. Nuclei and cell segmentations are used to measure multiple primary morphological features such as shape, intensity, and texture in all acquired images. Single cell measurements are stored in .txt files (one row per cell). CellProfiler pipelines are provided as [Data S1](#) and [S2](#).

Preprocessing

Cell and feature processing was performed as described in the main text using functions of the custom written R package cmoRe, publicly available on github (<https://github.com/mknoll/cmoRe>). Its general functionality is outlined in [Figure 2](#), a detailed outline with function names is provided in [Figure S2](#), a full handbook of all package functions and an example are provided in the [Methods S1](#) and with the package vignette. The NFAT score was calculated by automated thresholding on the median GFP intensity of the nucleus ([Figure S4F](#)). It reflects translocation of NFAT-GFP from the cytoplasm compartment into the nuclear compartment upon activation of the Calcineurin-NFAT pathway. For cell cycle analysis we used the integrated DAPI intensity of the nucleus, for identification of non-attached cells we used the cellular/nuclear area ratio and for identification of non-cardiomyocytes we used the median DAPI intensity of the nucleus ([Figure S3](#)).

Single cell phenotyping

Filtered, non-aggregated data of highest concentrations of tested substances were analyzed with t-SNE using the FI-tSNE implementation⁴¹ and UMAP,⁴² similarity between substance induced phenotypes were assessed by binning the two-dimensional t-SNE data ($n = 50$ bins), normalizing data (rates) and calculating pairwise differences. 5% and 95% quantiles of combined unique pairwise combination differences were used to indicate significance ([Figure 3E](#)).

Population-level phenotyping

All analyses were performed in R, version 3.4.1. Mixed effect models were computed using lme4⁴³ or nlme.⁴⁴ Significance level was set to $\alpha = 0.05$ (two-sided).

Cell Reports Medicine, Volume 2

Supplemental information

**C-MORE: A high-content single-cell morphology
recognition methodology for liquid biopsies
toward personalized cardiovascular medicine**

Jennifer Furkel, Maximilian Knoll, Shabana Din, Nicolai V. Bogert, Timon Seeger, Norbert Frey, Amir Abdollahi, Hugo A. Katus, and Mathias H. Konstandin

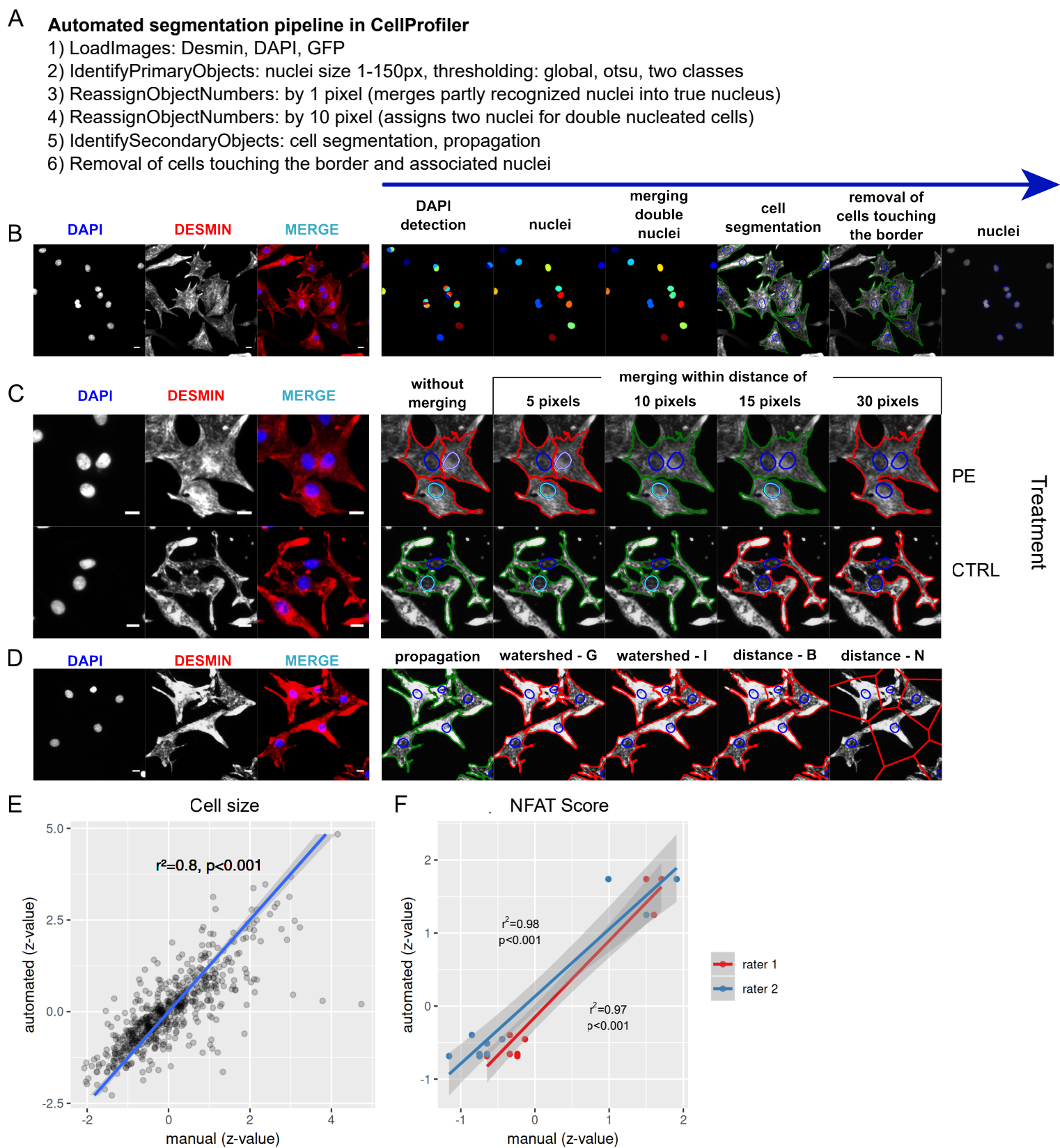


Figure S1. CellProfiler pipeline functionality and validity assessment of automated measurements. Related to Figure 1 and STAR methods. (A) Overview of the CellProfiler analysis steps and important settings. (B) Overview of the CellProfiler analysis step with an image example. (C) shows the detection of double nucleated cells by distance of two nuclei. The distance of 10 pixels was validated manually to be optimal to identify true multi-nucleated cardiomyocytes, while minimizing misidentification of nearby single nucleated cells. Unstimulated and hypertrophic cardiomyocytes were used as ground truth. We show representative recognition on images of unstimulated (CTRL) and phenylephrine (PE) stimulated cardiomyocytes. (D) shows a comparison of cell segmentations performed by different algorithms available in CellProfiler. For our pipeline we chose the propagation algorithm. (E) Correlation between manual and automatically determined cell sizes (z-transformation, Spearman correlation, $n=612$ cells, p -value: likelihood ratio test between linear models). (F) Numbers of nuclear GFP positive cells as determined automatically (density thresholding) and manually by two raters. Metrics are calculated as in E. Images were cropped for better visualization.

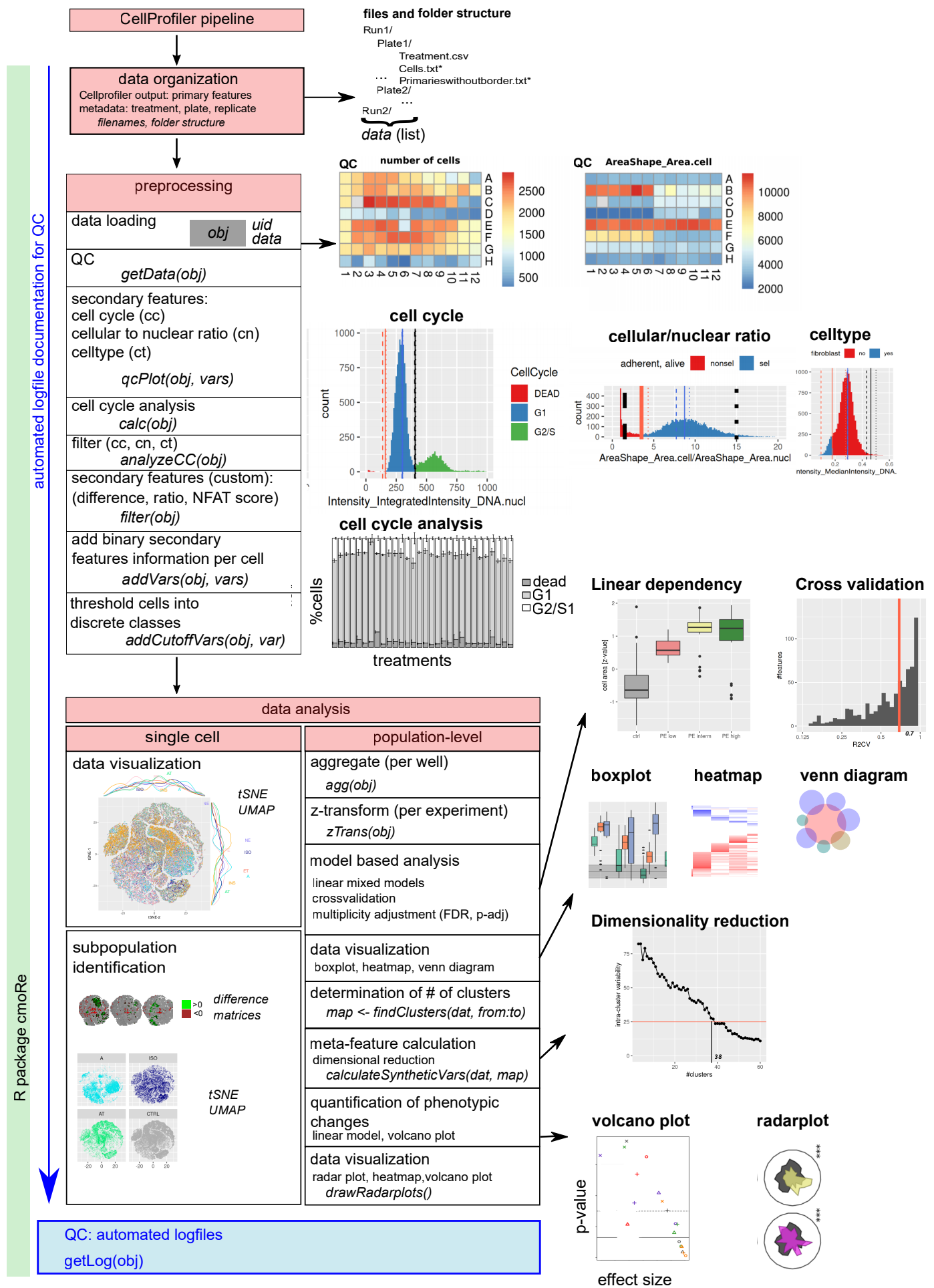


Figure S2. Detailed overview of the R package cmoRe and the proposed workflow. Related to Figure 2 and STAR methods. The analysis workflow comprises four main parts: 1) Loading data matrices calculated by CellProfiler using a convenient file-folder structure, 2) Preprocessing includes quality control using median feature distribution over the plate (e.g. cell number or cell area). Secondary features are calculated using the cmoRe thresholding function. Specific secondary features - cell cycle (cc), cellular to nuclear ratio for detection of non-attached cells (cn) and the celltype (ct) - are used to filter for vital and properly attached cardiomyocytes for downstream analyses. Custom secondary features, e.g. when custom reporters as the NFAT-GFP reporter are used can be calculated additionally on the curated data set. For data analysis cmoRe offers 3) a single cell level phenotyping with functions to quantify patterns for subpopulation identification. And 4) a population-level phenotyping aggregating single cell data per well. cmoRe functions to be used for the respective analysis step are indicated in italic font. In-depth documentation, a handbook of all package functions and an example are provided in Methods S1 and the package vignette online.

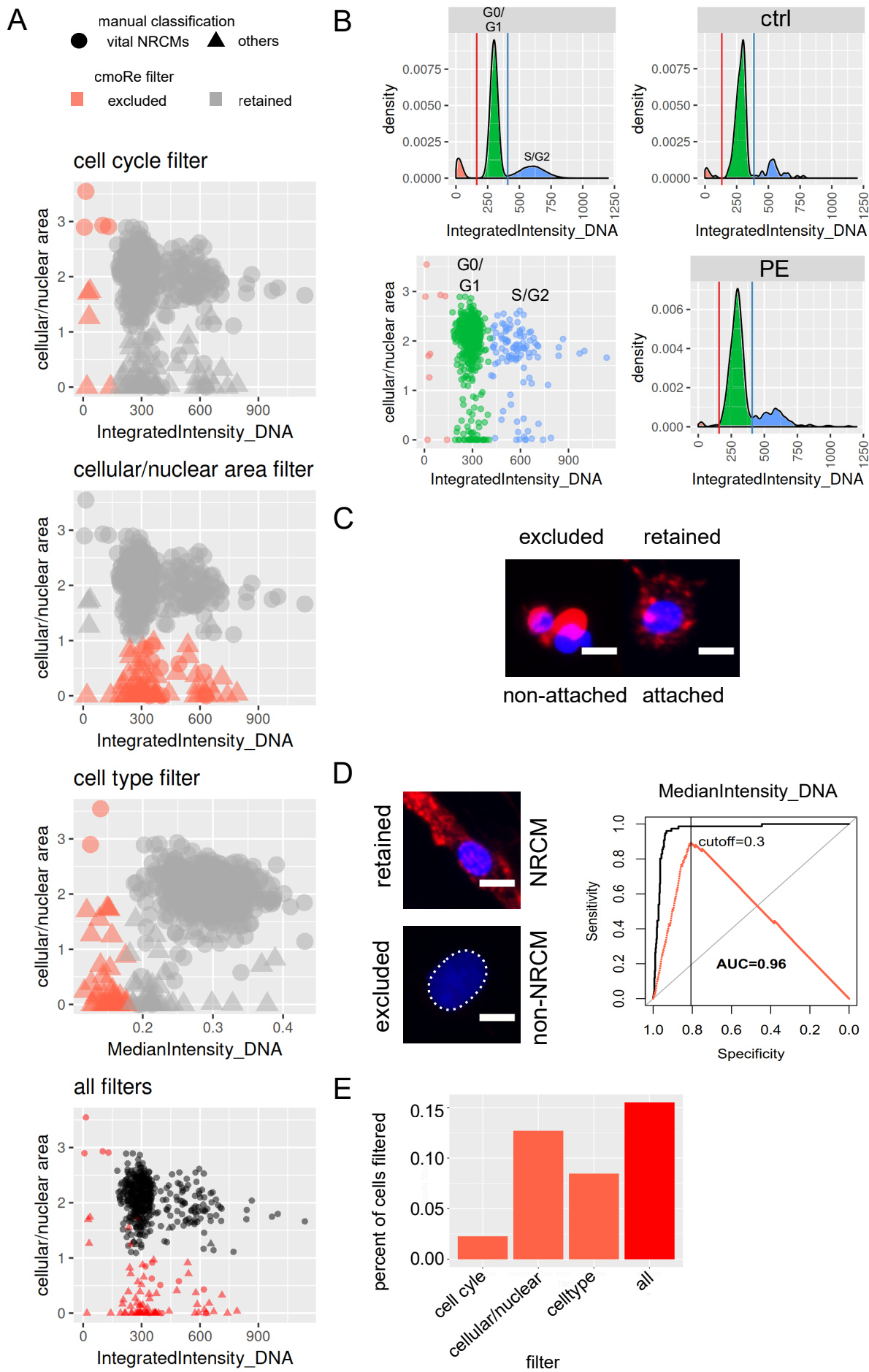


Figure S3. Filters implemented in the cmoRe workflow validated against a manually curated data set. Related to Figure 2 and STAR methods. (A) Scatter plots of the respective secondary features used for filtering via the cmoRe automated thresholding function. (B) Representative plots of nuclear integrated DNA intensity and cmoRe calculated thresholds for automated cell cycle analysis analogously to usage e.g. in flow cytometry (FACS). (C) Filter for non-attached cells: A representative non-attached cell is shown which is characterized by a very small cellular to nuclear ratio and excluded by the filter. In contrast, we show a very small, but properly attached cell which is correctly retained by the filter. (D) Left: Filter for non-cardiomyocytes: Non-cardiomyocytes are characterized by their low intense nucleus. A representative retained NRCM cell, as well as an excluded non-NRCM nucleus (dotted line) is shown. Right: The feature was chosen based on its high sensitivity and specificity in separating NRCMs and non-NRCMs as depicted on the right (receiver operating characteristic (ROC); data: manually annotated dataset, black: ROC curve, red: Youden index). (E) Fraction of cells excluded with the respective filter step and fraction of overall excluded cells.

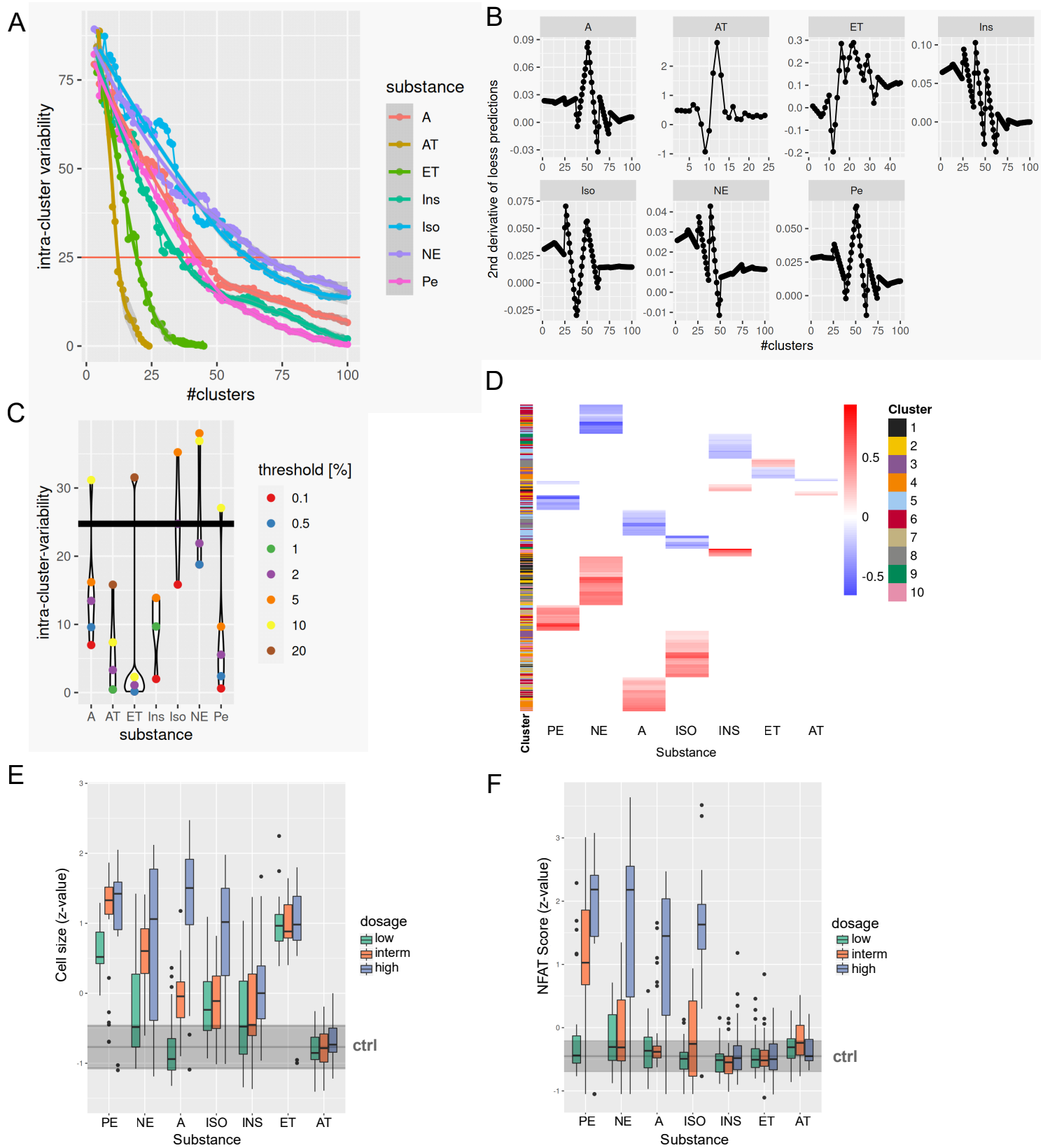


Figure S4. Selection of the intracluster variability threshold, stimulus specific feature changes and differences in cell sizes and nuclear NFAT positive cell fraction. Related to Figure 2,3, and STAR methods. (A) Intracluster variability for increasing number of clusters for all substances. (B) Second derivative of loess fit predictions trained on data from A. The cluster number corresponding to the value which surpasses its predecessor value for more than a fraction of the observed range for the respective substance is retained for multiple fractions (see C, thresholds). (C) Intracluster variability per identified threshold and median value of all thresholds (black horizontal line). (D) Heatmap of canonical hypertrophic stimulus specific feature alterations. Red indicates a dose-dependent increase, blue indicates a dose-dependent decrease. (E) Cell size for all stimuli and concentrations (z-values). (F) Nuclear GFP (NFAT) positive fraction of cells (z-score) for all stimuli and concentrations. For control condition the median and median absolute deviation is depicted in dark grey in E and F.

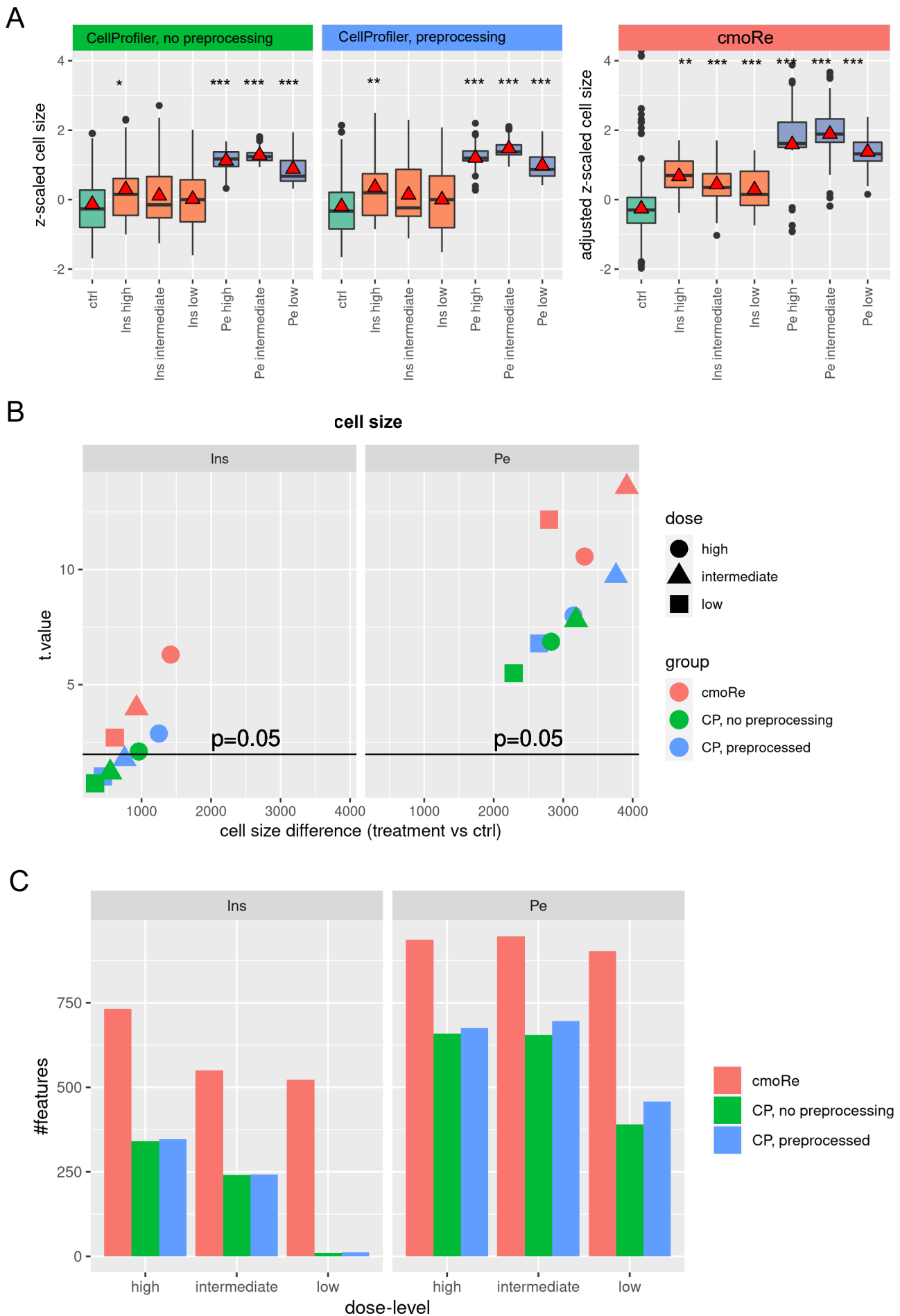


Figure S5. Benchmarking of cmoRe against CellProfiler raw data implemented with naïve linear models. Related to Figure 2. (A) Cell size as reported by CellProfiler (left) and adjusted with linear mixed models as implemented in cmoRe (right). Red triangles: mean, preprocessed: filtered data (fibroblasts, dead cells). z-transformed values. (B) Pairwise differences in cell size (ctrl vs treatment/dose) for compared analysis methods. (C) Number of significant features for different analysis methods with FDR < 0.05. P-values were calculated with linear mixed effect models; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$

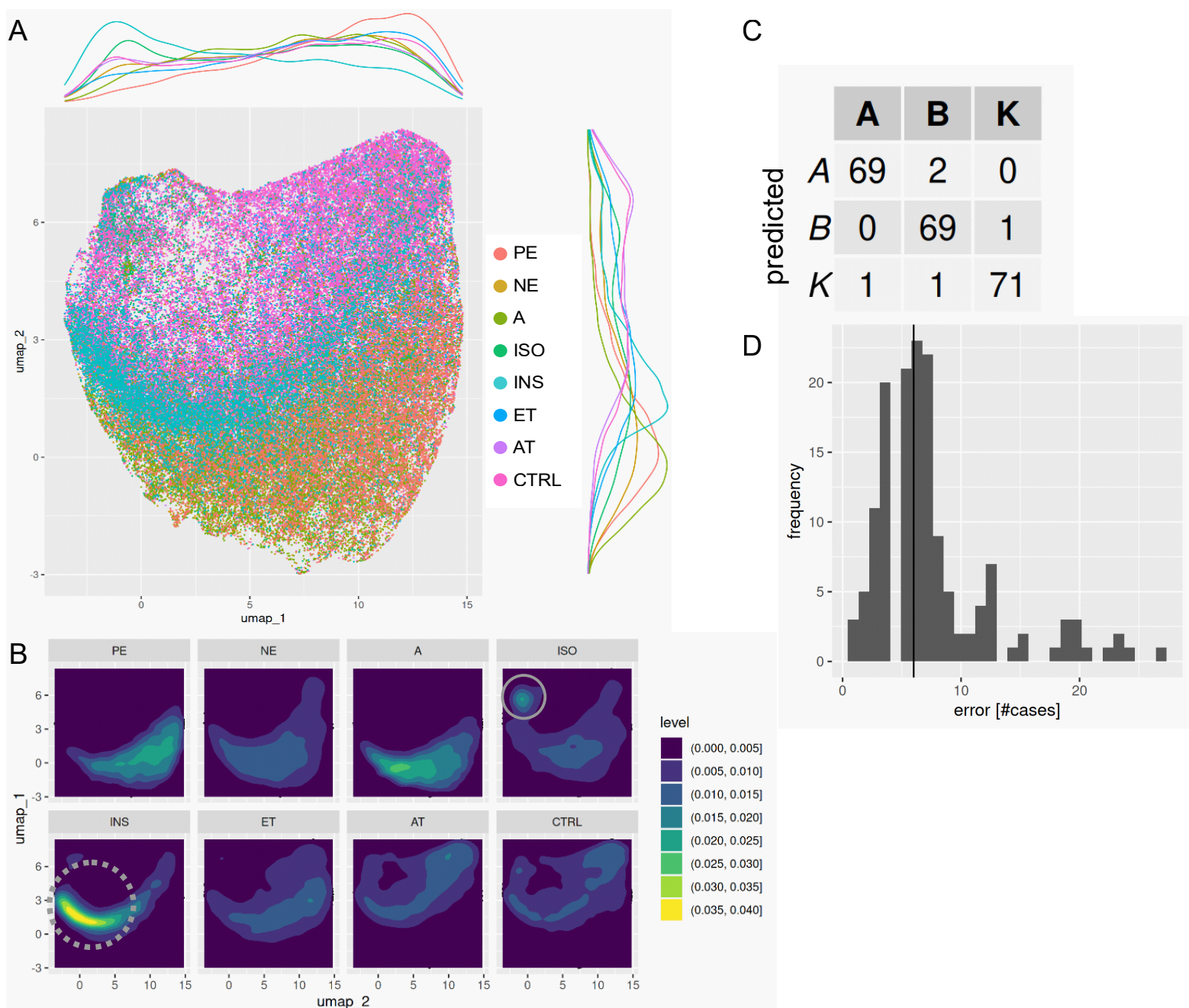


Figure S6. UMAP representation of single cell data of all canonical hypertrophic stimuli and random forest classification of aortic stenosis dataset. Related to Figure 3 and 5 and STAR methods. (A) UMAP of all canonical hypertrophic stimuli, with marginal distributions. (B) Cellular densities per stimulus color-coded on the UMAP plot. Identified substance specific subpopulations are circled for ISO, for INS circled with dotted line. (C) Accuracy of random forest classification of preTAVR (A), postTAVR (B) and healthy controls (K) per well: The table shows true vs. predicted classes. (D) Error for all permutations of random forest classification as shown in C, training and test set with fixed group sizes. Median number of misclassified wells are 5. For A and B: Single cell data of $n=1$ experiment is shown. For C and D: Training set consisted of patients 1,2,3, K1,K2,K3,K4; test set consisted of patients 4,5, K5,K6 for C. In D all permutations were used.

References

- Bakal C, Aach J, Church G, Perrimon N (2007) Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* 316: 1753-6
- Garvey CM, Spiller E, Lindsay D, Chiang C-T, Choi NC, Agus DB, Mallick P, Foo J, Mumenthaler SM (2016) A high-content image-based method for quantitatively studying context-dependent cell population dynamics. *Scientific reports* 6: 29752-29752
- Jentzsch C, Leierseder S, Loyer X, Floherschütz I, Sassi Y, Hartmann D, Thum T, Lagerbauer B, Engelhardt S (2012) A phenotypic screen to identify hypertrophy-modulating microRNAs in primary cardiomyocytes. *Journal of Molecular and Cellular Cardiology* 52: 13-20
- Manzella G, Schreck LD, Breunis WB, Molenaar J, Merks H, Barr FG, Sun W, Römmele M, Zhang L, Tchinda J, Ngo QA, Bode P, Delattre O, Surdez D, Rekhi B, Niggli FK, Schäfer BW, Wachtel M (2020) Phenotypic profiling with a living biobank of primary rhabdomyosarcoma unravels disease heterogeneity and AKT sensitivity. *Nature Communications* 11: 4629
- Ryall KA, Bezzerides VJ, Rosenzweig A, Saucerman JJ (2014) Phenotypic screen quantifying differential regulation of cardiac myocyte hypertrophy identifies CITED4 regulation of myocyte elongation. *J Mol Cell Cardiol* 72: 74-84
- Sero JE, Sailem HZ, Ardy RC, Almuttaqi H, Zhang T, Bakal C (2015) Cell shape and the microenvironment regulate nuclear translocation of NF- κ B in breast epithelial and tumor cells. *Mol Syst Biol* 11: 790
- Slack MD, Martinez ED, Wu LF, Altschuler SJ (2008) Characterizing heterogeneous cellular responses to perturbations. 105: 19306-19311
- Snijder B, Vladimer GI, Krall N, Miura K, Schmolke A-S, Kornauth C, Lopez de la Fuente O, Choi H-S, van der Kouwe E, Gültekin S, Kazianka L, Bigenzahn JW, Hoermann G, Prutsch N, Merkel O, Ringler A, Sabler M, Jeryczynski G, Mayerhoefer ME, Simonitsch-Klupp I et al. (2017) Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label, pilot study. *The Lancet Haematology* 4: e595-e606
- Woo LA, Tkachenko S, Ding M, Plowright AT, Engkvist O, Andersson H, Drowley L, Barrett I, Firth M, Akerblad P, Wolf MJ, Bekiranov S, Brautigan DL, Wang QD, Saucerman JJ (2019) High-content phenotypic assay for proliferation of human iPSC-derived cardiomyocytes identifies L-type calcium channels as targets. *J Mol Cell Cardiol* 127: 204-214

Table S2. Stimulus and inhibitor information. Related to Figure 3 and 4.

hypertrophic stimulus	low concentration	medium concentration	high concentration	source	identifier
Adrenaline (Suprarenin/Epinephrin, 1mg/ml)	0,1 μ M	1 μ M	10 μ M	Sanofi	6053210
Angiotensin II ((R)-(-)-Phenylephrine hydrochloride)	0,1 μ M	1 μ M	10 μ M	Sigma	A9525
Endothelin-1 (Endothelin 1 97% (HPLC), powder)	10 nM	0,1 μ M	1 μ M	sigma	E7764
Insulin (Insulin Insuman rapid 40 ie)	4*10 ⁻⁵ IE/ml	4*10 ⁻⁴ IE/ml	4*10 ⁻³ IE/ml	Sanofi	1843315
Isoproterenol ((-) Isoproterenol hydrochloride)	0,1 μ M	1 μ M	10 μ M	Sigma	I6504
Noradrenaline (Aterenol 1mg/ml)	0,1 μ M	1 μ M	10 μ M	Sanofi	3870227
inhibitor	low concentration	medium concentration	high concentration	source	identifier
AKT (Tricibine Akt V Inhibitor)	0,156 μ M	1,56 μ M	15,6 μ M	Sigma	124038
ERK (ERK/MEK Inhibitor)	0,1 μ M	1 μ M	10 μ M	Promega	U0126
FAK (PF 573228 FAK inhibitor)	10 μ M	0,1 mM	1mM	Tocris	3239
GSK (BIO GSK3b inhibitor)	5 nM	50 nM	500 nM	Tocris	3194
PI3K (Ly294002, PI3K Inhibitor)	650 nM	6,5 μ M	65 μ M	Millipore	440202

Methods S1. Detailed description and manual for C-MORE functions. Related to Figure 1,2,3 and STAR methods

1. Calculation of intracluster variability

Intra-cluster variability was computed using Euclidean distance and ward.D2 clustering method. Intra-cluster variability values for different numbers of clusters (2 to 100) were used to fit loess models. Second derivative of loess fit model predictions were evaluated to determine an intra-cluster variability cutoff. Starting from the highest number of clusters, the change between this value and its predecessor was calculated as fraction of the maximal observed difference per substance. The maximum intra-cluster variability detected in any of the substances for a given threshold was retained (0.1, 0.5, 1, 2, 5, 10 and 20%), the median of all values was 24.75. Thus a cutoff of 25 was used for analysis.

2. Benchmarking of cmoRe against CellProfiler raw data implemented with naïve linear models

To evaluate cmoRe performance, we tested differences in NRCM cell size data obtained with CellProfiler (CP) between non-treated and INS/PE treated cells by omitting single cmoRe (pre-)processing steps.

Therefore, we tested for differences between control (non-treated, ctrl) and INS/PE treated cells (per dose level, i.e. ctrl vs INS high data, ctrl vs. PE intermediate) on well aggregated data.

First, we assessed non-filtered data (no exclusion of dead cells and fibroblasts, Supplementary Figure 6A, left [CellProfiler, no preprocessing]) using linear models. For the comparison ctrl vs INS, we only found a significant difference ($p < 0.05$) between highest dose of INS and ctrl, and for all comparisons between PE and ctrl. Next, we used data from which dead cells and fibroblasts were filtered prior to aggregation per well (Supplementary Figure 6A, preprocessed). Again, only the comparison ctrl vs INS high was the only significant finding within the comparisons of ctrl and INS treated cells. T-values (p-values), corresponding to the respective tests, however, were larger (smaller) (Supplementary Figure 6B).

Next, we assessed the effect of using linear mixed models for analyses (labeled cmoRe), to adjust for variation between repetitions of experiments. All comparisons between ctrl and INS irrespective of dose level yielded significant results. For visualization, z-transformed residuals $y_i - \hat{y}_i$ (y_i : measured data in well i , \hat{y}_i : predicted data with the mixed model formula $y \sim 1 + (1|experiment/plate)$) are shown.

Finally, we tested the number of significant features from all CP features. cmoRe yielded the highest number of significant results ($FDR < 0.05$), much less features were obtained when using linear models instead of linear mixed models for analysis (CP, no preprocessing and CP, preprocessed). The latter, however, showed slightly higher numbers of features, highlighting the beneficial effect of filtering fibroblasts and dead cells in detecting differences in NRCM morphology induced by differential treatment.

3. Manual for our R-package cmoRe

C-MORE: A high content single cell morphology
assay for cardiovascular medicine:
The cmoRe R package

July 26, 2021

Document version 0.3

Package version 0.3

Contents

1 Introduction	4
2 Cellprofiler data preparation	4
2.1 Cellprofiler output	4
3 The R package <i>cmoRe</i>	4
3.1 The <i>imgExp</i> class	6
3.2 Data loading	6
3.3 QC plots	6
3.4 Filtering of single cells	6
3.4.1 Detected numbers per well	7
3.4.2 Morphology based	7
3.4.3 Fibroblast filter (<i>fb</i>)	8
3.4.4 Cell-cycle assignment (<i>cc</i>)	9
3.4.5 Debris/detached cells filtering (<i>nc</i>)	10
3.5 Threshold identification	10
3.5.1 Cutoff identification	12
3.5.2 Imputation of missing data	12
3.5.3 Assignment of single cells	12
3.6 Cell-cycle analysis	12
3.7 Additional features	13
3.8 Aggregation	14
3.9 Z-transformation	14
3.10 Removal of single cell data	15
4 Single cell data analysis	15
5 Well-aggregated data analysis	15
5.1 Dose dependent specific alterations	15
5.2 Inhibitors	16
5.2.1 Positive control vs inhibitor	16
5.2.2 Negative control and inhibitor	17
5.3 TAVI	17
5.3.1 Concentration dependent regulation	17
5.3.2 Concentration independent regulation	19
5.3.3 Crossvalidation	19
5.4 IPS cells	19
6 Crossvalidation	20
7 Data prediction	20
8 Feature selection	20
9 Metafeature calculation	20

10 Phenotype visualization

22

1 Introduction

The *cmoRe* R package contains a collection of functions for preprocessing and analysis of morphological analysis obtained with *CellProfiler* (CP) on a single cell basis.

Main functionality of the package spans:

- Preprocessing
 - Quality control
 - Filtering for vital cardiomyocytes
 - Exclusion of non-cardiomyocytes (fibroblasts)
 - Cell-cycle assignment
 - Automatic threshold selection of multimodal distributions¹
- Data analysis
 - Feature selection
 - Meta-feature calculation
 - Visualization

Parts of the analysis steps (modell based testing of single features without crossvalidation) are used in a more general form using the *dataAnalysisMisc* package.

2 Cellprofiler data preparation

2.1 Cellprofiler output

Cellprofiler output files are stored in a standardized folder structure (Fig. [1](#)).

Different experimental runs are stored in a folder (*run_n*), with subfolders for each measured 96-well plate (*plate_m*). Within each subfolder, cell profiler output data (*Cells.txt*, *Primarieswithoutboder.txt*, *Cytoplasm.txt*) as well as metadata information (*Treatment.csv*) file are stored. The latter contains information about treatments applied to each well, see. Tbl. [1](#).

3 The R package *cmoRe*

The following sections give a short overview of the functionality implemented in the *cmoRe* package. Not all parameters are outlined in detail, please refer to the package vignette (`vignette(package="cmoRe")`) and documentation (`?fun`) for further information.

¹Currently only implemented as dichotomization, see below.

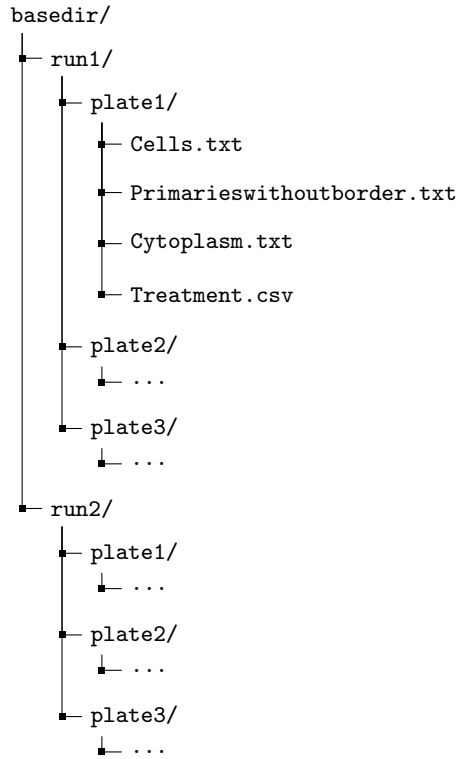


Figure 1: Required structure of *CellProfiler* output data and metadata files.

Well	Treatment	Konzentration
A1	ctrl	1
⋮	⋮	⋮
H12	PE	3

Table 1: Representative information of a *Treatment.csv* file containing the metadata for each plate layout.

Suffix	CP file
.cell	Cells.txt
.nucl	Primarieswithoutborder.txt
.cyto	Cytoplasm.txt

Table 2: Suffixes of features for CellProfiler output files.

3.1 The *imgExp* class

For a given analysis, an *imgExp* class is instantiated (*obj*). Its constructor expects a *list* containing paths to the different experimental runs with the corresponding plates in subfolders and a user specified unique ID (*uid*) (Fig. 1 and Lst. 1).

```

1 # Where is the CP data stored?
  data <- list()
3 data[[1]] <- paste0("run1/", c("plate1/", "plate2/", "plate3/"))
  data[[1]] <- paste0("run2/", c("plate1/", "plate2/", "plate3/"))
5
  # Instantiate imgExp class
7 obj <- new("imgExp", data, "UID1")

```

Listing 1: Instantiating an *imgExp* object.

3.2 Data loading

`getData(obj)` checks completeness, loads and stores data in the `@data` slot of *obj*.

Measurements from different files are merged by *Metadata.Well*, *ImageNumber* and *ObjectNumber* (parameter *mrg* in *loadData()*) per experimental run and plate. Suffixes are added to feature names to denote their origin (Tbl. 2). File-names are specified by the *fn* parameter (CP output files) and *treatF* (metadata file, *Treatment.csv*) in *loadData()*.

3.3 QC plots

Initial quality control plots can be obtained with the `qcPlots(obj)` function to visualize the number of CP recognized cells per well or the distribution of any selected calculated feature per plate, e.g. median cell size (Lst. 2 Fig. 2).

```

1 # Creates a pdf file in folder
  qcPlots(obj, folder="/tmp/", var="AreaShape.Area.cell", fun=median)

```

Listing 2: Representative QC plot for the distribution of median cell size.

3.4 Filtering of single cells

To retain mostly vital, adherent cardiomyocytes with a minimum number of cells per well for subsequent analyses, different filtering steps are implemented.

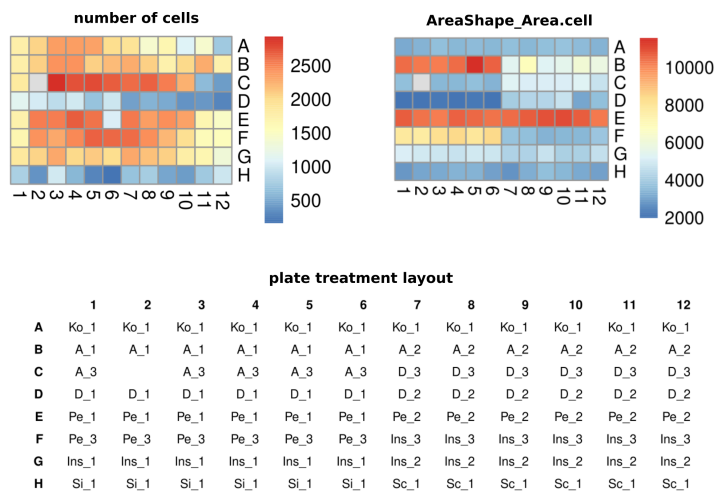


Figure 2: QC plots for numbers of cells, cell size per well (median) and plate treatment layouts.

3.4.1 Detected numbers per well

```
obj <- checkFilter(obj, filter=100)
```

Listing 3: Remove wells with <100 cells per well.

Wells/treatments with only very few identified cells (e.g. for highly cytotoxic substances and too high concentrations) can be removed with the `checkFilter()` function. The `filter` parameter removes all treatment-experimental run-plate combinations containing less than the specified number of cells. Defaults to `NULL` and prints the respective numbers in this case without filtering (Lst. 3).

Furthermore, rows (corresponding to single cells) with unspecified treatments (`is.na(obj@data$TREATMENT) == TRUE`) are removed.

3.4.2 Morphology based

```
1 # Calculate cutoffs
  obj <- calc(obj)
3 # Alternative: calculate only cell cycle cutoffs
  #obj <- calc(obj, fun="cc")
5
7 # Apply cutoffs and filter data
  obj <- filter(obj)
```

Listing 4: Identification of thresholds and assignment per cell.

An alternative, fast method with allows to calculate cutoffs without the need to previously load all data into RAM, is available with the following lines:

```

1 # create new object for demonstration purposes
  obj0 <- new("imageExp", "uid2", data)
3
4 # calculate cutoffs
5 cutoffs <- calcCutoffs(obj0, fun="nc")
6
7 # load data
  obj0 <- getData(obj0)
9
10 #and use the precalculated cutoffs
11 obj0 <- addCutoffs(obj0, cutoffs)

```

Listing 5: Fast identification of thresholds (full CP data) and separated assignment per cell.

Vital, adherent cardiomyocytes are selected by applying three filters on each single cell.

- Differentiation between fibroblasts and cardiomyocytes (fibroblast filter, `fun="fb"`), Fig. 3
- Selection of adherent cells and removal of debris (nuclear-to-cellular ration, `fun="nb"`), Fig. 5
- Selection of vital cardiomyocytes (Assignment of cell-cycle, `fun="cc"`), Fig. 4

The latter two filters rely on the analysis value distributions for identification of cutoffs with given constraints. The general approach is shown in Sec. 3.5.

3.4.3 Fibroblast filter (*fb*)

- feature: *Intensity-MedianIntensity-DNA.nucl*
- transformation: *identity*
- method: identify minimum left of global maximum
- constraints
 - *xMinGlobMax* := 0.2
 - *xMaxGlobMax* := 0.6

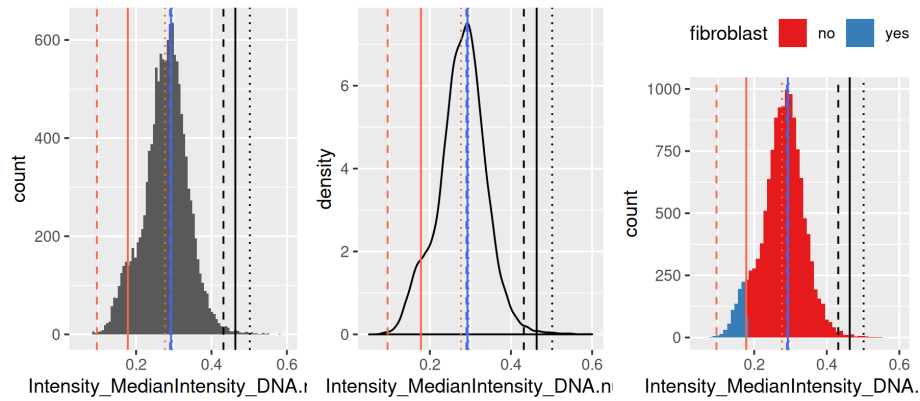


Figure 3: Threshold identification for the differentiation between fibroblasts and cardiomyocytes. Left: histogram with calculated cutoffs (see Sec. 3.5), corresponding density (middle) and histogram with group assignments (right). Solid (median), dashed (10%) and dotted (90%) quantiles. Bold lines: Constraints on cutoff identification ($xCut$, $xMinGlobMax$).

3.4.4 Cell-cycle assignment (cc)

- feature: $Intensity_IntegratedIntensity_DNA.nucl$
- transformation: log
- method: identify global maximum (G1 peak), minimum left: cutoff for dead cells, minimum right: G2 cells
- constraints
 - $xMinGlobMax := log(100)$

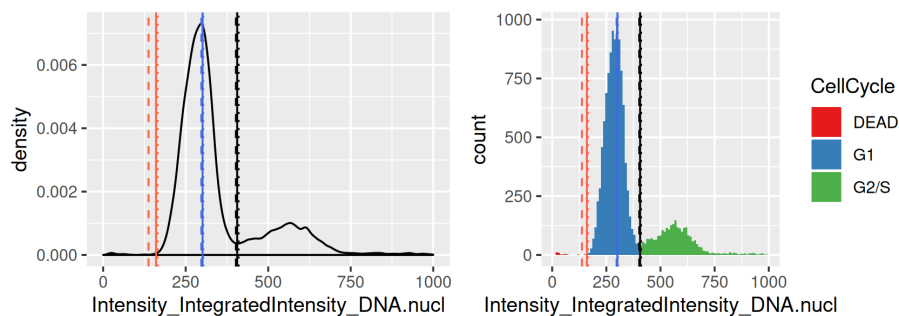


Figure 4: Threshold identification for cell cycle assignment. Left: density with calculated cutoffs (see Sec. 3.5); solid (median), dashed (10%) and dotted (90%) quantiles and histogram of cell-cycle assigned cells (right).

3.4.5 Debris/detached cells filtering (*nc*)

- feature: *AreaShape_Area.cell*, *AreaShape_Area.nucl*
- transformation: *identity*
- method: identify minimum left of global maximum within the interval (*xMinGlobMax*, *xCut*)
- constraints
 - *bw* := 0.1
 - *xCut* := 15
 - *xMinGlobMax* := 1.5

3.5 Threshold identification

The general approach for threshold selection is based on a density analysis of the respective value distribution with given constraints (e.g. cutoff within a pre-specified interval), transformations (*log*, *identity*) and a specific method (e.g. identifying the minimum left of the global maximum)

To account for variability, cutoffs are calculated (and applied) for each experimental run, plate and treatment separately. Robustness of cutoffs is assured by repeatedly performing analyses on resampled data and aggregating the obtained results. 10, 50 (median) and 90% quantiles are calculated. If no cutoffs could be identified for a given combination, data is imputed based on the remaining information.

An overview of utilized constraints is shown in Tbl. 3.

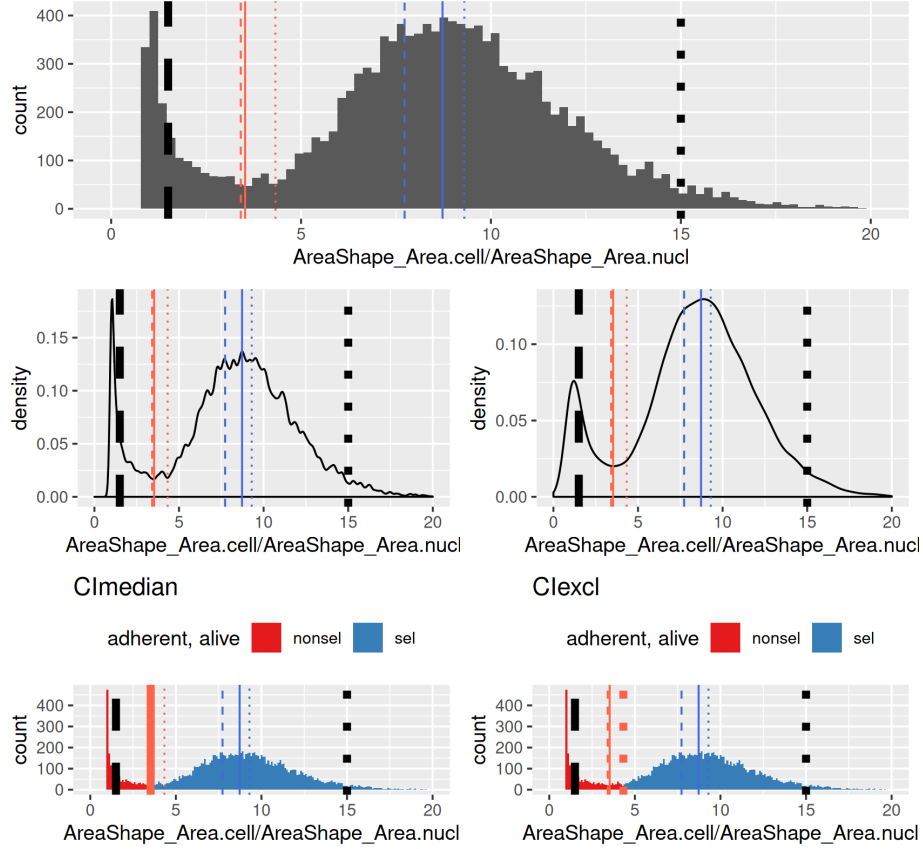


Figure 5: Threshold identification for identification of attached, vital cardiomyocytes. Upper row: histogram of ratio distribution. Middle: different bandwidths for density estimation (left: 0.1, right: auto). Bottom: Differently selected cutoffs (left: median CI_{median} , right: 90% CI_{excl} quantile) of the left minimum. Solid (median), dashed (10%) and dotted (90%) quantiles.

Parameter	Description
bw	Bandwith for R's density function
$xCut$	Cutoff to remove all values above $xCut$ to focus on the relevant data below this threshold,
$xMinGlobMax$	Lower border of the interval in which the global maximum is expected.
$xMaxGlobMax$	Upper border of the interval in which the global maximum is expected.

Table 3: Parameters and constraints for threshold identification.

3.5.1 Cutoff identification

Cutoffs are identified by repeated (e.g. $n=100$) sampling from the available data (with replacement, number of observations equals number of drawings).

First, the density of the resulting data is calculated with default parameters. Obtaining the first and second derivative leads to identification of extremal points from which all maxima and minima are retrieved and the global maximum is selected.

If no value could be identified, *NA* is saved for this iteration, and a warning is printed. Quantiles (default: 10, 50 and 90%) of all relevant extremal points (usually global maximum, left and right minimum) are calculated (while removing *NAs*) and returned.

3.5.2 Imputation of missing data

If no cutoffs could be detected in the previous step, missing data is imputed by calculating the median of all additional thresholds detected for the respective analysis.

3.5.3 Assignment of single cells

In the faster cutoff calculations using `calcCutoffs()` and `addCutoffs()`, only a simplified assignment method is currently provided (*CI*median, see below).

Two methods to assign single cells to their respective category are implemented, *CI*median (default) and *CI*excl. The former assigns each cell based on the 50% quantile (median) cutoff, the latter uses the boundaries of the calculated interval and marks values within the intervals with *NA*.

3.6 Cell-cycle analysis

If we used `calc()` to compute cutoffs, we can use the following code.

```
1 # Plot cell cycle distributions per plate
  analyzeCC(obj)
3
4 # Plot cell cycle distributions aggregated per treatment
5 analyzeCC(obj, agg=T)
```

Listing 6: Obtain cell cycle information.

Alternatively, if we used `calcCutoffs()` and `addCutoffs()`, the relevant fractions are calculated as follows (Lst. [7](#)):

```
1 cc <- sumAgg(obj0)
```

Listing 7: Calculation of cell cycle fractions.

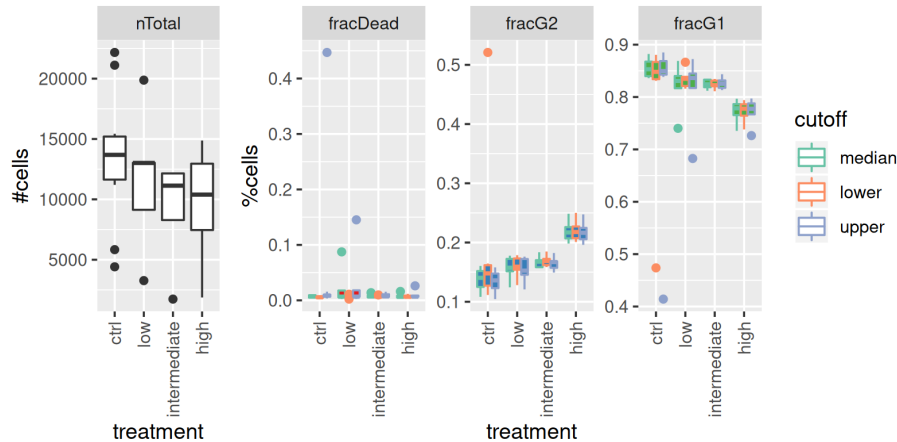


Figure 6: Cell cycle distribution for increasing concentrations for a given treatment. Left: Numbers of cells, right: fractions of cells in the respective cellcycles with different cutoffs used to separate populations.

Cell cycle classification is performed as described in Sec. 3.4.4. All cells, which are included in the `obj@data` *data.frame* at the time of running the `calc(obj)` function² are evaluated. Aggregated visualization and values can be obtained with the `analyzeCC()` function³. The latter returns a *data.frame* with the the total number/fraction of cells per as signed cell-cycle state (three values each, estimate [median], lower and upper interval bounds) (Lst. 6), `sumAgg` a *data.frame* with the fraction of dead, G1 and G2/S cells per well (Fig. 6).

3.7 Additional features

```
1 # vars: vector of feautres (colnames in obj@data)
  obj <- addCutoffVars(obj, vars, fun=log, nBoot=100)
```

Listing 8: Automatic threshold identification for additional features.

The previously applied method to identify cutoffs can be applied to any feature, with an arbitrary transformation (often log). The `addCutoffVars()` function is used to detect cutoffs similarly to the cell cycle analysis.

Currently, only dichotomization based on the identification of the first minimum right of the global maximum is implemented. For identification of the left minimum multiply the respective data by -1. A function for transformation is specified with the `fun` parameter (Lst. 8).

²and all cells present in the CellProfiler output files for `calcCutoffs()` and `addCutoffs()`

³only in combination with `calc()`

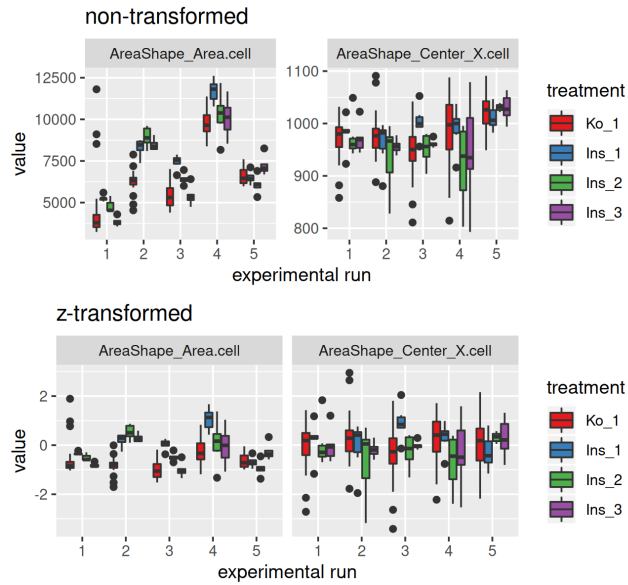


Figure 7: Non- and z-transformed data.

3.8 Aggregation

```

1 # median aggregation
2 obj@dataAgg <- medianAgg(obj)

```

Listing 9: Aggregation of single cell data per well

The `medianAgg()` function aggregates data by well (median, Lst. 9). Fractions of binary features can be calculated and added to the aggregated data as shown in Lst. 10.

```

1 #add fractions of binary features
2 exp@dataAgg <- cbind(exp@dataAgg, fractAgg(exp))

```

Listing 10: Calculation of dichotomized feature fractions.

3.9 Z-transformation

```

1 obj <- zTrans(obj)

```

Listing 11: Z-transformation of data.

The `zTrans()` function performs an experimental-run wise z -transformation of data from treatments present in all treatments and for aggregated data (Lst. 11). A minimum of two experimental runs is required (Fig. 7).

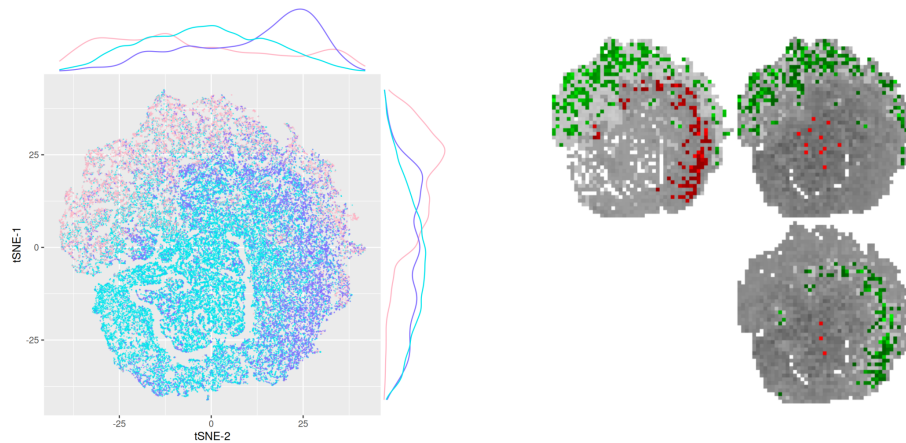


Figure 8: Representative single-cell analysis results. Left: t-SNE of data treated with three treatments. Right: difference maps with 2.5% and 97.5% quantiles used for color code thresholding to highlight differences in population composition.

3.10 Removal of single cell data

```
1 obj <- removeRawData(obj)
```

Listing 12: Remove single cell data.

If no further single cell data analysis is planned, data can be removed to keep the object sizes with the calculated metadata and aggregated data small (Lst. [12](#)), Fig. [8](#)

4 Single cell data analysis

Single cell data can be accessed by the `obj@data` slot as `data.frame` and might be used e.g. in a variety of dimension reduction methods (t-SNE, umap).

5 Well-aggregated data analysis

Analysis of well-aggregated data by mixed model analysis, following feature selection and phenotype visualization is depicted.

5.1 Dose dependent specific alterations

For the assessment of effects for varying dose treatments of NRCMs, a linear mixed effect model was utilized. Three doses were assessed together with non-treated data and a linear relationship was assumed (Tbl. [4](#) Lst. [13](#)).

Dose	Comment
0	control
1	low concentration
2	intermediate concentration
3	high concentration

Table 4: Recoding of concentrations for equidistant assumptions to test for differences.

Dose was evaluated as fixed effect, *treatment* nested under plate (*plate*) and experimental run (*expRun*) were included as random effects (see Lst. 13) using the *lme4* package.

The effect of dose as independent variable on each morphological feature was assessed by likelihood ratio tests of the null and full (differing in *dose* as covariate) model.

Crossvalidation (see Sec. 6) and the testing method are implemented in the `analyze()` function.

treatment: substance, e.g. PE.

dose: assumed equidistant dose concentrations, $dose \in \{0, 1, 2, 3\}$

expRun: experimental runs, $expRun \in \{1, 2, 3, 4, 5\}$

plate: plate per experimental run, $plate \in \{1, 2\}$

val: measurement of a given morphological feature.

```
1 val ~ dose + (1|expRun/plate/treatment)
```

Listing 13: Formula used to identify dose dependent differences.

5.2 Inhibitors

The same approach as outlined in Sec. 5.1 was used to calculate substance specific differences (positive control).

5.2.1 Positive control vs inhibitor

Differences between positive controls and inhibitor data was calculated from meta-features (see Sec. 9) as follows.

i : index of meta features, $i \in \mathbb{N}^+$

m_i^{ctrl} : median aggregated meta feature i value (positive control).

m_i^{tr} : values of meta feature i for treatment tr

The differences d between meta features of treated and control data are defined as follows:

$$d_i^{tr,ctrl} := m_i^{tr} - m_i^{ctrl} \quad (1)$$

d^2 values are then evaluated using a linear mixed effects model to test for differences between controls and inhibitor treated measurements separate for each concentration while adjusting for the respective meta-features (Lst. 5.2.1). Experimental run and well are included as random effects.

meta_feature: meta-feature label (e.g. $\{MF1, MF2, MF3\}$ if three meta features were selected / calculated).

treat: treatment factors to be evaluated for a significant effect on differences (e.g. $\{ins_1, pe_2\}$).

```

1 # fixed effect formula
  d^2 ~ meta_feature + treat
3 # random effect formula
  ~ 1 | expRun/well

```

5.2.2 Negative control and inhibitor

In addition to the the testing for differences between positive control vs positive control substance + inhibitor combinations, the similarity/ equivalence between negative controls and positive control substance + inhibitor was evaluated as follows: Per treatment and dose, a mixed model was fitted with (negative) control data and data from treated cells (substance + inhibitor in a specific dose) with a random effects *experimentalRun* and *treatment* (multiple wells per treatment, Lst. 14) for each meta feature (*SYN_VAR_n / VAL*, Fig. 10).

```
lmer(VAL~1+treatment+(1|experimentalRun/treatment), data=data)
```

Listing 14: Model estimates for controls and substance + inhibitor treated data.

Model estimates for controls and substance+inhibitor treated cells were evaluated, using their 95% confidence intervals. Two qualitative metrics were used to evaluate similarity between model estimates: the number of meta features which show distinct estimate distributions (*different*) and the number of metafeatures for which the treatment estimates lay within the confidence interval of the control estimate (*estim in ref CI*), see Fig. 9. Representative data is shown in Fig. 10.

5.3 TAVI

5.3.1 Concentration dependent regulation

For the TAVI/aortic stenosis experiment, multiple serum concentrations were tested. As it was unknown which serum concentration might yield appropriate results (and an appropriate concentration might differ between features), it was tested for a significant interaction term between the two factors *concentration* and *treatment* (Lst. 15).

treatment $\in \{control, aorticStenosis, postTAVI\}$
concentration $\in \{0.025, 0.5, 1.0, 2.0\}$ *val*: measurement of a given morphologi-

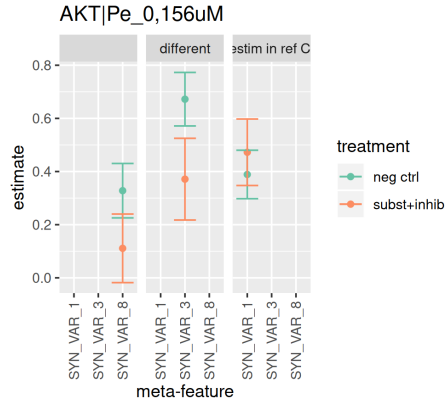


Figure 9: Qualitatively different metrics for the evaluation of similarity between negative controls and substance+inhibitor treated cells on meta-feature level.

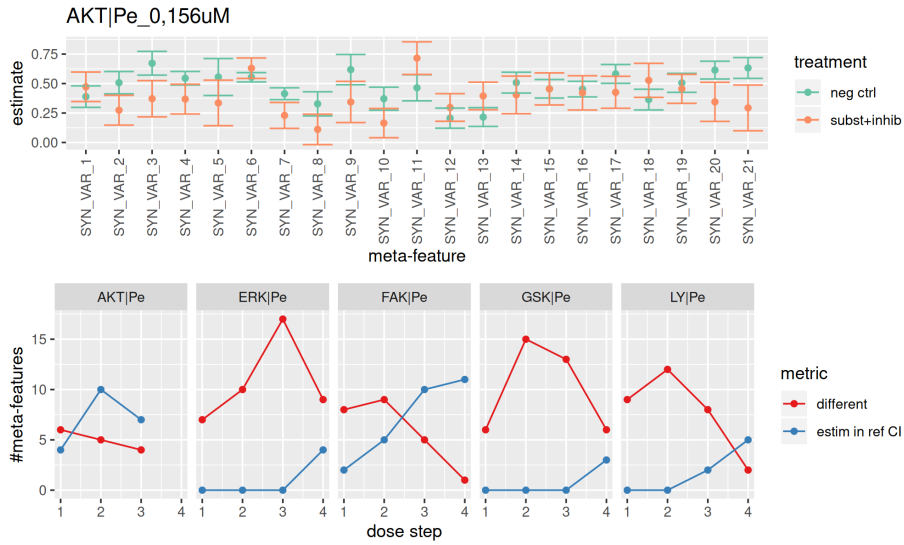


Figure 10: Representative data showing the quantification of similarity between negative controls and substance + inhibitor treated cells for meta-features.

State	Category
0	healthy
1	heterozygous
2	homozygous

Table 5: Encoding of evaluated IPS types for an analysis analogously to the dose-dependent substance approach.

cal feature.

```
1 val ~ treatment*concentration
```

Listing 15: Formula to test for differences in TAVI data.

For the evaluation, features showing a significant interaction term *treatment : concentration* for the comparison *ctrl : concentration* vs *aorticStenosis : concentration* were selected as being specific for aortic stenosis. Meta-features with significant interaction effects *aorticStenosis : concentration* vs *postTAVI : concentration* and inverted effect sized were considered reversible after TAVI.

5.3.2 Concentration independent regulation

Features not showing a significant interaction effect, and thus being regulated independently of the tested concentration, were also included into the evaluation. Analysis was performed as described in Sec. ??, but excluding all features with a non-adjusted interaction p-value above 0.05 or 0.1.

5.3.3 Crossvalidation

Crossvalidation (see Sec. 6) and the testing method are implemented in the `analyzeTAVI()` function.

5.4 IPS cells

For the analysis of (genetically altered) IPS cells, equidistant differences were assumed and differences were assessed with a linear mixed effect model using the *dataAnalysisMisc* package (Tbl. 5, Lst. 16).

```
1 #Patient: patient ID
  res <- randEffAnalysis(dat[, , drop=F], pheno[, ],
3                       frm0=as.formula(VAL~1+(1|Patient)),
                       frm=as.formula(VAL~state+(1|Patient)))
```

Listing 16: Testing for differences in the IPS dataset.

For visualization of phenotypic changes, batch adjusted data was used (see Sec. 7).

6 Crossvalidation

Crossvalidation was performed as follows: measurements (aggregated per well) were left out per treatment and experimental run to train a model, the left out data was predicted with the latter.

\mathbf{v}^{full} : predicted values from the model trained with all data.
 $i \in \mathbb{N}^+$: numbers of models fitted with incomplete data.
 \mathbf{v}_i^{sub} : predicted values of the left out data not used for model training.

$$R_{cv}^2 := 1 - \sum_i (\mathbf{v}^{full} - \mathbf{v}_i^{sub})^2 \cdot \frac{\text{Var}(\mathbf{v}_i^{sub})}{i} \quad (2)$$

$$R_{cv}^2 = \begin{cases} R_{cv}^2 & \text{if } R_{cv}^2 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

7 Data prediction

Model based feature selection utilized mixed effects model to estimate random intercepts for specific factors as e.g. different batches. For visualization of meta-feature calculation, non-adjusted data might be utilized or predicted data. The latter was performed by fitting a mixed effects model for each feature with all observed data, and using this model to predict values.

8 Feature selection

Features were selected with a given significance threshold after adjustment for multiplicity of likelihood-ratio test p-values p^* (Bonferroni, Benjamini-Hochberg), R_{cv}^2 and effect size Δ cutoffs.

9 Metafeature calculation

Similar features were aggregated to meta features mf_i based on a hierarchical cluster analysis and the selection of a maximal intracluster variability (Fig. 11).

For metafeature calculation, either directly measured, filtered and transformed data (as described above) was used. Alternatively, for data with even higher variability (TAVI, IPS cells) linear mixed models were fitted for data of each feature, and predicted values were used for following analysis (cluster selection, meta-feature calculation, visualization in radarplots, see Lst. 17 for IPS analyses).

```
fit <- lmer(VAL~Mut_num + (1|Batch), data=data)
prd <- predict(fit, newdata=data)
```

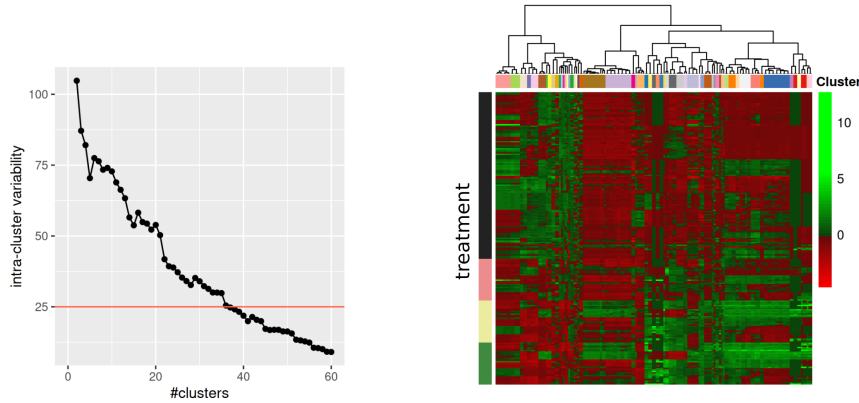


Figure 11: Intra-cluster variability for increasing numbers of clusters (left) and heatmap of selected features with a representative clustering.

Listing 17: Prediction of data per feature.

Using the `findClusters()` function, a range of potential clusters can be evaluated. The latter returns a intra-cluster variability value as a function of assumed clusters (Lst. 18):

$\mathbf{v}_i^{F_j}$: measurement i of feature set F_j corresponding to cluster j .
 $mad(\mathbf{x})$: median absolute deviance of \mathbf{x} .
 F : set of all selected features.
 k : number of clusters with $k := \{i | i \in \mathbb{N}^+ 1 \leq k \leq |F|\}$.

For each number of clusters k to be tested (`nClust` parameter), the following metric m_k is calculated:

$$m_k := \frac{\sum_{j=1}^k mad(\mathbf{v}_i^{F_j})}{k} \quad (3)$$

Ward.D2 clustering (`clustering_method` parameter) and Euclidean distance are used for hierarchical cluster analysis to derive the respective number of clusters (Fig. 11).

```
##calculate metric for 2 to 60 clusters
2 cluster <- findClusters(clusterDat, nClust = c(2,60))
```

Listing 18: Calculation of intracuster variability.

After selection of a desired intra-cluster variability cutoff, such defined similar features are aggregated to meta-features by median aggregation (Lst. 19).

```
## select intra-cluster variability cutoff
```

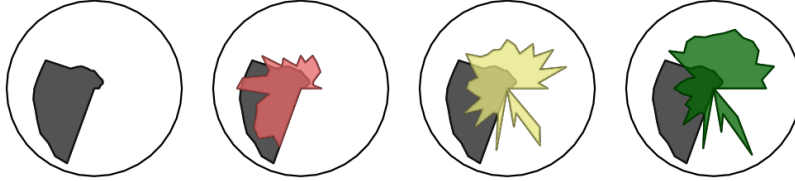



Figure 12: Radarplots of representative data showing meta-features for increasing substance concentrations.

```

2 cutoff <- 100
4 cm <- cluster[order(unlist(cluster[, "dispSum"])),]
# list with feature names and cluster assignment
6 map <- cm[which(cm[, "dispSum"] > cutoff),,drop=F][1,,drop=F][1,"pm.
  clust"][[1]]
8 # calculate meta features
synthMetr <- calculateSyntheticVars(clusterDat, map)

```

Listing 19: Calculation of metafeatures for a specific intra-cluster variability cutoff.

10 Phenotype visualization

Changes in meta-features can be visualized using radar plots (Lst. [20](#), Fig. [12](#)).

```

1 drawRadarplots(plotDat, vars=names(synthMetr$anno)[order(ret
  [[1]][[1]])], labels = F, ctrlLevel="Ko.1", pTest=NULL, agg
  =mean, col=col, main="")

```

Listing 20: Visualization of phenotype changes using radarplots.