bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Multidimensional Profiling of Drug-Treated Cells by Imaging Mass Cytometry

Alexandre Bouzekri, Amanda Esch, Olga Ornatsky

Fluidigm Canada Inc., 1380 Rodick Road, Markham, Ontario, L3R 4G5, Canada

Proteomics R&D Department

Running Title: Multidimensional drug profiling by Imaging Mass Cytometry

Corresponding author: Alexandre Bouzekri

Email: alexandre.bouzekri@fluidigm.com

Phone: 905-943-0133

Key words: Imaging Mass Cytometry, cell-based screening, image-based computational analysis, anti-cancer drug treatment, multiparametric similarities measurements

Abbreviations: IMC, Imaging Mass Cytometry; CellProfiler, CP; histoCAT, histology topography cytometry analysis toolbox; NCI-60, National Cancer Institute 60; MCF-7, Michigan Cancer Foundation-7; SNE, stochastic neighbor embedding; EGF, epidermal growth factor; 4E-BP1, eukaryotic initiation factor 4E-binding protein.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

In pharmaceutical research, high-content screening is an integral part of lead candidate development. Drug response *in vitro* over 40 parameters including biomarkers, signaling molecules, cell morphological changes, proliferation indexes and toxicity in a single sample could significantly enhance discovery of new therapeutics. As a proof of concept, we present a workflow for multidimensional Imaging Mass Cytometry™ (IMC™) and data processing with open source computational tools. CellProfiler was used to identify single cells through establishing cellular boundaries, followed by histoCAT™ (histology topography cytometry analysis toolbox) for extracting single-cell quantitative information visualized as t-SNE plots and heatmaps. Human breast cancer-derived cell lines SKBR3, HCC1143 and MCF-7 were screened for expression of cellular markers to generate digital images with a resolution comparable to conventional fluorescence microscopy. Predicted pharmacodynamic effects were measured in MCF-7 cells dosed with three target-specific compounds: growth stimulatory EGF, microtubule depolymerization agent nocodazole and genotoxic chemotherapeutic drug etoposide. We show strong pairwise correlation between nuclear markers pHistone3^{S28}, Ki-67 and p4E-BP1$^{T37/T46}$ in classified mitotic cells and anti-correlation with cell surface markers. Our study demonstrates that IMC data expands the number of measured parameters in single cells and brings higher-dimension analysis to the field of cell-based screening in early lead compound discovery.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Discovery of new treatments in oncology research relies extensively on the use of human-derived cell culture models [1]. High-content cell-based screens are widely applied in pharmaceutical drug development to prioritize lead molecules for animal testing [2]. These assays rely on the use of primary and cancer cell lines and mostly monitor cytotoxicity and proliferation. With the advent of sophisticated genomics and proteomics technologies for soluble proteins and mass cytometry for single cells, it is becoming possible to increase the multidimensionality of *in vitro* screens. Databases of genotypic and phenotypic profiles [3, 4, 5] across cancer drug panels have provided a comparative analysis between *in vitro* studies and clinical therapeutic responses *in vivo* [6]. One of these datasets developed and shared by the National Cancer Institute 60 (NCI-60) in the late 1980s was the first *in vitro* discovery screening tool for pharmacologic compounds inducing growth inhibition across 60 cancer cell lines [7]. Over the years several concerns were raised regarding the use of *in vitro* established human tumor-derived cell lines for drug testing. Concerns included genetic instability and dedifferentiation. However, the study of genetic mutations arising in immortalized cells remained a field of interest in drug discovery, providing knowledge on the dysregulation of cellular signaling pathways and the effects small-molecule inhibitors have on human tumor cell lines [8].

High-content imaging of cancer cell lines in response to drug treatment is a standard assay applied in preclinical studies for identification of different mechanisms of drug action [9, 10, 11, 12]. The main criteria in every morphophenotypic screen for early drug discovery is the selection of biomarkers and detection modalities (antibodies, chemical and enzymatic probes, reporter detection tags). In the field of fluorescent cellular imaging microscopy, biomarker analysis in single cells is limited to 4–6 due to the overlapping spectra of fluorescent dyes [13]. Repetitive

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rounds of staining using the same biological sample are required to achieve multiplexing [14]. New imaging technologies are needed to significantly increase multiplicity in a single sample preparation experiment and perform replicate analyses [15]. High-content imaging assays applied to screening drug perturbations in heterogeneous cancer cell models could uncover additional modes of drug action and biomarkers for clinical trials [16].

Imaging Mass Cytometry is an emerging and transformative technique in the field of digital histopathology applied to complex tissue sections [17, 18, 19]. The Hyperion™ Imaging System (Fluidigm®) can measure up to 40 parameters simultaneously in formalin-fixed, paraffin-embedded (FFPE) and frozen human tissue sections with subcellular resolution [20]. Sample preparation is very similar to standard immunohistochemical protocols [17], where the tissue section on a slide is first deparaffinized and treated with an antigen retrieval buffer for antigen epitope exposure in the case of FFPE, then stained once with a mixture of isotopically pure metal-labeled antibodies specific to structural and cell type biomarkers. After drying and inserting the slide into the ablation chamber, regions of interest (ROIs) are chosen and recorded by a camera integrated with the Hyperion Tissue Imager. These ROIs are then ablated in 1 μm steps as the slide moves under the laser. Each laser pulse vaporizes a 1 μm^2 area of the tissue and generates a plume of particles. The resulting particles are carried by a stream of helium/argon mixture into a CyTOF® instrument, an inductively coupled plasma time-of-flight mass cytometer, described elsewhere [18]. Data are stored as raw binary files of ion counts for each mass channel (parameter) per pixel. These files can be converted into a stack of single-parameter matrices (grayscale images), which are used to create merged multiparametric pseudocolor images and perform quantitative pixel- or segmentation-based analysis. Bodenmiller et al. are developing multivariate computational tools to visualize and analyze multiplexed images of

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

human tissue sections generated by IMC [21]. However, there is no single software package or analysis workflow that could currently be applied to answer specific biological questions.

In this proof-of-principle study, we set out to develop a comprehensive workflow for IMC data analysis based on recent advances in imaging algorithmic methods to visualize and measure multiple biomarkers in model cell lines cultured in chamber slides **(Fig. 1)**. Several human breast cancer-derived cell lines, SKBR3, HCC1143 and MCF-7, were screened for expression of surface and intracellular markers. Predicted pharmacodynamic effects were studied in MCF-7 cells dosed with three target-specific compounds: growth stimulatory epidermal growth factor (EGF), microtubule depolymerization agent nocodazole and genotoxic chemotherapeutic drug etoposide [22, 23]. Our method analysis workflow demonstrates the high multiplexing capability of IMC for *in vitro* research and offers quantitation of multiple biomarkers at subcellular resolution. Future improvements of the technology toward a higher acquisition speed of multiple samples will expand IMC applications as an imaging platform for *in vitro* cell-based drug screening.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Material and Methods

Cell culture

SKBR3 (HTB-30™), HCC1143 (CRL-2321™) and MCF-7 (HTB-22™) were purchased, tested for mycoplasma contamination and authenticated with short tandem repeat DNA profiling by American Tissue Culture Collection (ATCC®). Cells were cultured within 15 passages in their corresponding growth media supplemented with 10% bovine serum (HyClone™ Cosmic Calf™ Serum, Cat. No. SH30087.04) and penicillin/streptomycin (Thermo Fisher Scientific Cat. No. 15140-122): McCoy's 5a (ATCC 30-2007™) for SKBR3, RPMI-1640 (ATCC 30-2001™) for HCC1143 and DMEM (ATCC 30-2002™) for MCF-7 with 0.01 mg/mL insulin (Sigma-Aldrich® Cat. No. I9278). Cells at 70–80% monolayer confluency were washed with Versene (Thermo Fisher Scientific Cat. No. 15040-066), detached with Trypsin-EDTA 0.25% (Thermo Fisher Scientific Cat. No. 25200-056), collected and counted. For each cell line, chamber slides (BioCoat™ Collagen I 8-Well Culture Slide, Corning® Cat. No. 354630) were used to seed cells at a concentration of 0.1e6/mL in 0.5 mL per chamber and grown in 37 °C, 5% CO2, 100% humidity for 48 hours prior to drug compound addition. Each breast cancer cell line was seeded using individual chamber slides to avoid any cell cross-contamination, and each IMC experiment was carried out independently. In chronological order, seeding of HCC1143 was the first sample preparation, then we run the second experiment using SKBR3, and last we used MCF-7 for the drug-treated cells in vitro study.

In vitro drug treatment

All chemical compounds were dissolved and aliquoted at the supplier's recommended concentrations for long-term storage in 100% DMSO (Sigma-Aldrich Cat. No. 276855). Final

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

concentrations for each compound were selected from previously published data [32, 53, 54] by premixing the required initial drug stock volume with full growth media, then transferring 0.5 mL volume to corresponding cell chambers (**see Supplementary Table S1**): 10 µM of etoposide (Cell Signaling Technology® Cat. No. 2200S); 0.5 µM of nocodazole (Sigma-Aldrich Cat. No. M1404); 10 ng/mL of human EGF (Sigma-Aldrich Cat. No. E9644); 2% DMSO in growth media. One chamber was left with fresh growth media as non-treated control. Cells were exposed to compounds for 48 hours at 37 °C, 5% CO2 and 100% humidity. Preparation of additional eight chamber sample slides to replicate our experiment for IMC analysis was not required due to the known *in vitro* predicted effect of the drugs we tested on this cell line.

Immunocytochemistry and antibody validation

Immunostaining of live cells was performed in chamber slides. Following drug treatment, fresh media in all chambers was supplemented with Cell-ID™ Intercalator-103Rh (Fluidigm Cat. No. 201103A) at 1:500 for 15 minutes at 37 ºC, 5% CO2 for dead cell identification. The chamber slides were washed with DPBS (Thermo Fisher Scientific Cat. No. 14190-144) at room temperature (RT) for 5 minutes. Surface metal-labeled antibody cocktail (**see Supplementary Table S2**) was made in 0.5% BSA/DPBS from antibody stocks at a dilution of 1:100 for each. Cell monolayers were stained with 250 µL antibody mix for 120 minutes, RT. Following a washing step, cells were fixed with 250 µL freshly made 1.6% formaldehyde (Thermo Fisher Scientific Cat. No. 28906) in DPBS for 15 minutes at RT. After fixative removal, cells were permeabilized with fresh 0.3% Saponin/DPBS (Sigma-Aldrich Cat. No. S7900-25G) for 30 minutes at RT, then blocked with 1% BSA/DPBS for 1 hour at 37 ºC. After blocking, cells were incubated overnight at 4 ºC in 250 µL freshly prepared cocktail of metal-labeled antibodies

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

targeting intracellular markers (**see Supplementary Table S3**) diluted 1:50 in 0.3% Saponin/DBPS. The following day cells were washed once with 0.5% BSA and stained with 250 µL of nuclear Cell-ID Intercalator-Ir for 30 minutes (1:1000), RT. After the last wash, each chamber was filled with 250 µL deionized water for 5 minutes for salt removal. Chambers were detached from slides and left to dry at RT until IMC analysis. The commercial antibodies of our panels have been previously validated by immunocytochemistry applied to immunofluorescence analysis of adherent cells. A similar sample preparation method was applied to the validation of these antibodies labeled with metals regarding their antigen specificity detected by IMC. The selection of titers tested for surface markers (1:100) and intracellular markers (1:50) refer to rigorous Quality Control antibody validation functional assays using single-cell suspension Mass Cytometry, which has the same single isotopic detection modalities of antigens as IMC. The list of antibodies used for each breast cancer cell line experiment is cross-referenced in Supplementary Tables S2 and S3. A total number of 14 antibodies were tested on HCC1143 cells, 20 markers for SKBR3 and 25 for MCF-7. The markers from each panel were selected based on their specificity to surface membrane, cytoplasmic and nuclear components of single cells, their previous applications in the context of breast cancer cell line studies, and their commercial availability. Expanding a panel to 37 markers requires additional custom labeling of target specific monoclonal antibodies with non-used metal isotopes.

Data acquisition by IMC

Samples were analyzed with the Hyperion Imaging System (Fluidigm). The dried slide was loaded into the imaging module, where an optical previewing of the ROIs was recorded for laser ablation. Areas of dimension 1000 x 1000 µm were acquired for HCC1143 and 1400 x 1400 µm

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for SKBR3 cell lines experiments. Replicate ROIs of 1500 x 1500 µm size were collected for MCF-7 exposed to chemical compounds, and a single ROI was ablated for non-treated control and DMSO-treated cells. Each ROI sample acquisition took 4.5 hours at an ablation frequency of 200Hz. The resulting data files were stored in MCD binary format. Multicolor images were generated with open source ImageJ 1.51 software [55]. Zoom-in regions for each multicolor image were made by selecting and cropping areas of dimension 400 x 400 µm from the original picture without any change of the pixel size. The resolution of each IMC image shown is 1 µm for 1 pixel.

Image analysis software tools

For each recorded ROI, stacks of 16-bit and 32-bit single-channel TIFF files were exported from MCD binary files using MCD™ Viewer 1.0 (Fluidigm). Cell-based morphological segmentation was carried out with two image processing pipelines for 16-bit or 32-bit TIFF files and CellProfiler 2.2.0 (CP), a widely adopted software in the open source image analysis community which has been continually improved since its availability in 2005 to read and analyze cell images using advanced algorithms [56]. The 16-bit TIFF unstacked image format was used as input data in CellProfiler to run a segmentation pipeline with a set of sequential modules to generate and save an unsigned 16-bit integer single cell mask TIFF image. Inputs of 16-bit TIFF images with their corresponding segmentation mask were uploaded in histoCAT to open a session data analysis. Due to the high number of 29 single isotopic channel TIFF images for each replicate and drug treatment condition, histoCAT analysis was generated faster with a 16-bit format instead of a 32-bit image format which is bigger in size. The 32-bit TIFF unstacked images were loaded in CellProfiler to run a segmentation and mean intensity multiparametric

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

measurement analysis of individual cells. The resulting outputs were exported and saved as SQLite single-cell object measurements database files directly compatible with CellProfiler Analyst for downstream computational analysis [37]**.** Spatial distribution maps, dimensionality reduction and unsupervised clustering for 16-bit single images were performed using the histoCAT 1.73 open source toolbox [21]. Supervised analysis by support vector machine learning classification on 32-bit datasets was processed with CellProfiler Analyst 2.2.1 [38]. Hierarchical clustering and correlation heat maps of classified cell populations were prepared using the web-based tool Morpheus (GenePattern, Broad Institute) [39]. Clustering and visualization of different types of markers in classified populations of cells across all controls and drug treatments were performed with the free software environment R 3.5.3, the packages pheatmap 1.0.12, igraph 1.2.4 and edgebundleR 0.1.4 [58].

Statistical analysis

Data from non-treated control, DMSO-only and drug-treated cells were analyzed using a non-parametric Mann-Whitney test with a two-tailed P value for every channel independently. Statistical significance was defined as $P < 0.05$. All statistical analysis was performed with GraphPad™ Prism® 7.04 software [57].

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Multiparametric characterization of breast cancer cell lines by IMC

IMC analysis revealed heterogeneity in the expression of surface, intracellular and nuclear markers in human breast cancer cell lines cultured in chamber slides. As an example, we selected the SKBR3 cell line, which is characterized by an invasive phenotype and increased proliferation *in vitro* and is used as a model to delineate mechanisms of resistance to ErbB2-targeted clinical therapies [24, 25]. As seen in **Fig. 2**, the culture consists of mostly small and medium-size rounded epithelial cells and occasional large multinucleated cells. Most cells show high levels of the cytoskeletal marker pan-keratin. Human epidermal growth factor receptor 2 (HER2) and EGFR are seen at the surface membrane, while tumor suppressor and transcription factor p53 is localized in the nuclei. Presence of lysosomal organelles is outlined in the cytoplasm with an antibody targeting the lysosome-associated membrane protein 2 (LAMP-2), known as CD107b. Image resolution is similar to fluorescence microscopy. Ductal breast carcinoma-derived triple-negative HCC1143 cell line was used to demonstrate mesenchymal morphology (**Fig. 3**) with high vimentin detection in the cytoplasm, nuclear localization of p53 and cytoplasmic distribution of the basal-like breast cancer marker cytokeratin 5 [26, 27]. The proliferation marker Ki-67 identifies cells in the active state of the cell cycle. Combination of multiple markers in an *in vitro* cell screening assay allows more precise analysis of cellular states of differentiation and invasiveness characteristic for breast cancer tumors.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In vitro drug effect profiling by IMC

To evaluate how IMC may be translated into future applications in the field of cell-based drug screening, we initiated a limited drug treatment study using one model breast cancer cell line. The phenotypic and functional responses of wild-type p53-expressing MCF-7 cells to etoposide, nocodazole or EGF compounds were investigated. In preclinical *in vitro* drug screening, this model is considered to be reproducible, fast and inexpensive despite the lack of clinical correlation to drug response *in vivo* due to breast tumor heterogeneity [28]. To identify how these chemical compounds induce cytostatic and cytotoxic effects on this noninvasive luminal breast cancer cell line, we generated multiplexed pseudocolor images of affected cellular compartments (**Fig. 4**). For example, nuclear size and proliferative activity of each cell were analyzed using the nuclear intercalator-Ir stain (Cell-ID Intercalator-Ir), cell cycle-S phase marker (Cell-ID IdU), proliferation marker Ki-67, cell cycle regulatory proteins cyclin B1 and D3 and estrogen receptor alpha (ERα). Antibody against pH2A.X^{S139} was used as a nuclear marker for DNA double-strand breaks and pHistone3^{S28} for mitotic cells. Cytoskeletal markers pan-keratin and cytokeratin 19 were used to follow changes in morphology and cell size. Drug response of MCF-7 morphology, adhesion and cell-to-cell interaction were visualized with surface membrane proteins, such as the membrane-bound mucin marker MUC-1 overexpressed in breast carcinoma [29], cell adhesion molecule EpCAM, integrin CD29, tetraspanin CD81, the low-expressed epidermal growth factor receptor EGFR, regulators of integrin-mediated cell adhesion CD98 and CD47 [30], and E-cadherin as epithelial interaction protein (**Supplementary Table S2**).

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Long-term exposure of MCF-7 cells to EGF induced significant morphological changes and an increase in cell size. Lineage shift from a classic epithelial cobblestone monolayer to a more invasive mesenchymal type was observed. The presence of large spindle-shaped cells that seemed to lose contact with other cells and migrate into empty spaces may be related to an increase in cell motility (**Fig. 4A**). Activation of the EGFR signal transduction pathway by EGF is known to play a role in cellular motility and size increase in the tumor aggressiveness [31].

Screening of MCF-7 cells treated with etoposide, nocodazole or EGF shows multivariate phenotypic response profiles compared to non-treated and DMSO-treated controls (**Supplementary Fig. S1**).

Treatment of MCF-7 cells with the antineoplastic agent nocodazole resulted in an increase in the population of pHistone3^{S28}-positive cells with a mitotic index three-fold higher compared to DMSO-treated or non-treated control (**Supplementary Table S4**). Another cytostatic phenotype was observed after prolonged drug exposure: subpopulations of cyclin D3-positive and tetraploid cells exiting mitosis without cytokinesis as a result of microtubule disruption by nocodazole (**Fig. 4B**). Presence of cyclin D3-positive and tetraploid cells in the G1/S interphase confirms nocodazole effect on MCF-7 cell cycle [32].

Incubation of MCF-7 with the anti-cancer agent etoposide shows inhibition of cell proliferation, elevated numbers of cells positive for nuclear pH2A.X^{S139} protein and compaction of the monolayer compared to EGF- or nocodazole-treated samples (**Fig. 4C**). Exposure of tumor-derived cell lines to this inhibitor induces the formation of a stable covalent topoisomerase II-cleaved DNA complex, which causes multiple breaks in double-stranded DNA. Etoposide treatment of MCF-7 activates phosphorylation of H2A.X, which is required for checkpoint-mediated cell cycle arrest, promotes DNA repair and maintains genomic stability.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Thus, pH2A.X^{S139} is a sensitive biomarker for measuring levels of drug-induced double-stranded DNA breaks in cancer cells [33].

High-dimensional phenotype analysis of *in vitro* drug response

Multivariate phenotypic responses of MCF-7 cells following chemical perturbations were quantified with a set of available software tools. Unfortunately, there is no single software package for this type of analysis. Therefore, we employed a workflow that includes a wide variety of tools designed for cellular imaging. Hopefully, this work will underscore the need for software specifically developed for analysis of multiparametric data acquired by IMC.

First, we used CellProfiler 2.2.0 (CP) to identify single cells through establishing nuclear and cellular boundaries [34]. Accuracy and robustness of cell segmentation was assessed by using Cell-ID Intercalator-Ir to identify cell nuclei and pan-keratin for cytoplasm (**Fig. 5A, Supplementary Fig. S2A and S3**). The segmentation masks were generated for each ROI, then exported for further downstream analysis of changes in MCF-7 cellular phenotypes in response to drug treatment.

Second, to visualize and analyze data we used the open source platform histoCAT, which enables highly multiplexed, quantitative and detailed analysis of cell phenotypes, microenvironment interactions and tissue architecture [21]. The workflow of histoCAT consists of overlaying the segmentation masks previously generated by CP to extract single-cell level information.

To evaluate the diversity of phenotypic profiles collected from drug-treated MCF-7 cells, we then selected and generated Z-score normalized mean intensities of pHistone3^{S28} and pH2AX^{S139} as drug-sensitive protein markers. Spatial heat maps of both channels were displayed for various drug-treated MCF-7 ROIs using a 99th percentile cut-off (**Figs 5B and 5C, Supplementary Fig.**

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

S2B). This quantitative and visual approach allows us to compare at the same scale the spatial distribution of these parameters between ROIs. For example, pH2A.X^{S139} nuclear expression level in etoposide-treated MCF-7 cells is higher than in nocodazole or EGF-treated cells on average. Mitotic cell marker pHistone3^{S28} expression level in nocodazole-treated MCF-7 is two-fold higher than those identified for other drugs. Related single-cell Z-score mean intensity values for nuclear markers (pHistone3^{S28}, pH2A.X^{S139}, Ki-67, p4EBP1$^{T37/T46}$, p53, cyclin D3), surface markers (CD98, CD81, CD29, CD49e, CD47) and cell size parameters (area, perimeter, major and minor axis lengths) were exported from each drug treatment dataset to assess statistically significant changes of protein expression levels and cellular morphology (**Fig. 5D and Supplementary Fig. S4)**. We can see differences between both controls and drug-treated samples in the expression of nuclear markers, with a co-increase of pH2A.X^{S139} and Ki-67 levels, and phosphorylation of the translation repression protein 4E-BP1 under etoposide treatment. This suggests the presence of a non-quiescent population of MCF-7 undergoing DNA repair and maintaining high proliferative activity. For nocodazole and EGF treatments, low Ki-67 levels may be related to the disruption of cell cycle and lineage shift induced in response to respective compound [35]. EGF treatment shows increased cell size and expression of adhesion surface markers compared to other conditions. These quantitative results are in agreement with multicolor images of treated cells.

Finally, the heterogeneity of MCF-7 cell phenotypes was analyzed using the t-distributed stochastic neighbor embedding (t-SNE) algorithm, a data dimensionality reduction method based on the similarity of selected markers or cell types [36]. We processed replicates of ROI for each compound using Z-score-normalized protein markers and cell size parameters to generate two-dimensional t-SNE plots. This visualization tool shows signal distribution of phosphoproteins,

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

size parameters and nuclear and cell surface membrane markers over the different ROIs (**Fig. 6A and Supplementary Fig. S4**). For example, the population of mitotic cells was grouped in a single region on the map with high expression levels of p4EBP1$^{T37/T46}$, pHistone3^{S28} and Ki-67. The high level of pH2A.X^{S139} was detected in the region of etoposide-treated cells and correlates with the strong Ir-intercalator signal due to compaction of cellular DNA in response to the drug. Area plot reveals the presence of larger-size cells in EGF and nocodazole treatments, in line with EGF-induced lineage shift and cytokinesis inhibition by nocodazole (**Fig. 6B**).

Classification of drug-treated MCF-7 phenotypic changes by support vector machine learning

Here we applied the support vector machine learning classification ruler termed Fast Gentle Boosting for accurate IMC dataset multiparametric analysis of nuclear states of compound-treated MCF-7 cells [37]. This classifier was integrated into CellProfiler Analyst (CPA) and trained [38] to identify five different nuclear classes of MCF-7 cells: pHistone3^{S28}-positive mitotic cells, pH2A.X^{S139} for cells undergoing DNA repair, p53 as tumor suppressor activation, cyclin D3 for G1/S transition in interphase and Ki-67 for proliferating cells. A composite image of all nuclear markers is shown in **Fig. 7A**. These single-cellular objects were then scored and classified for a supervised analysis of every protein marker and cell size parameter per specified nuclear subpopulation. Accuracy of image classification on each dataset of MCF-7 drug compound was evaluated by cross-validation between the true labeled cells as classification training set and the image full scoring as predicted label set for each nuclear class. Classification reports were displayed as confusion matrices where each value corresponds to average cross-validation metrics between the trained (true label) and scored (predicted label) single-cell objects

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for each nuclear class on a performance scale from 0 to 1 (**Fig. 7B and Supplementary Fig. S6**). As shown, classification of each nuclear class per drug compound dataset shows relatively good performance between 0.5 and 1**.**

Raw mean intensity values of protein expression for each of the five nuclear markers were exported per classified MCF-7 population and compared to evaluate statistical significance (**Fig. 7C and Supplementary Fig. S7**). Bar charts show various levels of protein markers among the five predicted classes of nuclear states, with higher protein detection in the corresponding class, except for the Ki-67 proliferative marker, which shows stronger detection in pHistone3^{S28} mitotic cells.

Multiparametric MCF-7 drug compound data pre-labeled by nuclear classification were reprocessed using hierarchical clustering similarity heat maps. First, we computed pairwise similarities for each class of cells per drug exposure with Pearson correlation coefficient between each measured parameter. This method is applicable to non-normalized data and can be used when data measurements vary between samples. Similarity Pearson correlation heat maps were computed using the browser-based tool Morpheus [39]. The core interface in Morpheus is a graphical color heat map representing protein marker and size parameter measurements applied to each single-cell object. Determination of the Pearson correlation coefficient between each measured parameter generates a reorganized heat map of multiparametric similarity relationships for each classified MCF-7 drug dataset [40]. Using these hierarchical similarity clustering heat maps, we compared the grouping pattern of mitotic pHistone3^{S28}-positive cells across the multiple ROIs of compound-treated MCF-7 (**Fig. 8 and Supplementary Fig. S8**). Naturally occurring isotopes of iridium, ^{191}Ir and ^{193}Ir, are present in the nuclear staining reagent Cell-ID Intercalator-Ir and are grouped consistently across all heat maps. The two cytoskeleton

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

biomarkers pan-keratin and cytokeratin 19 strongly correlate as well, since anti-pan-keratin and anti-CK19 antibodies identify cytokeratin isoforms highly expressed in MCF-7. Size parameters including perimeter, area and axis length of each mitotic cell are grouped together after hierarchical clustering across all drug treatments.

Cell surface protein markers (CD29, CD98, CD81, CD47, EpCAM and CD49e) generated a cluster with high correlation and redundancy in the pHistone3^{S28} mitotic cell subclass. However, EGF treatment influenced expression of these markers, and some level of dissimilarity occurred between the surface membrane cluster and two integrin-mediated regulators, CD98 and CD47. This is consistent with the fact that MCF-7 cells switch to a more mesenchymal phenotype after EGF treatment and higher motility, with a concomitant increase in the expression level of some adhesion proteins.

The intracellular and nuclear biomarkers pHistone3^{S28}, p4EBP1$^{T37/T46}$ and Ki-67 for mitotic cells cluster together in all conditions, showing a specific phenotypic profile. By looking at the correlation between the mitotic marker pHistone3^{S28} and other markers, we see that p53, cyclin B1 and cleaved poly (ADP-ribose) (cPARP) show variations in their correlation similarities depending on the drug treatment. p53 and cPARP exhibit close relationships with other nuclear markers, since both bind to nuclear DNA to be functional. Cyclin B1 shows strong correlation with pHistone3^{S28} in cells from non-treated control and EGF-treated ROIs. This correlation is less pronounced in the case of etoposide and nocodazole.

The clustering of different types of markers in the class of pHistone3^{S28} mitotic cells was measured by comparing the similarity level of each target-specific cluster across all controls and drug treatments. First, we reprocessed cell surface membrane CD markers for each treatment as a new hierarchical similarity heatmap by using Pearson correlation as distance metric and

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

complete linkage method (**Fig. 9A**). As expected, each individual cluster is aligned and identified around the matrix diagonal component. Both non-treated and DMSO controls group of protein markers show close similarities to each other, and low or negative correlation with drug-treated markers. We used a hierarchical visualization edge bundle circle plot diagram to highlight the positive intra and inter-similarities across surface membrane protein parameters (**Fig. 9B**). Visualization of the high pairwise correlations within each cluster is shown as edge connecting nodes, and bundled lines correspond to close inter-relationships between both control conditions. Drug-treated clusters of proteins show poor or no interdependency. A similar visual workflow was applied to the multivariate analysis by pairwise Pearson correlation of specific nuclear markers and cytoplasmic markers types (**Supplementary Fig. S9**).

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

The goal of this study was to demonstrate a multidimensional IMC workflow and quantitative data processing using open source computational tools for *in vitro* drug response research. The multiplexed measurements acquired from a single image may provide a more informative and reliable screening of drug leads for further clinical development [41]. In this work, phenotypic features of tumor-derived cell lines such as SKBR3, HCC1143 and MCF-7, which recapitulate different subtypes found in breast tumors, were characterized. The wild-type p53-expressing MCF-7 epithelial cell line, a model of luminal breast adenocarcinoma, was selected for *in vitro* drug treatment and exposed for 48 hours to three classes of bioactive compounds: EGF, nocodazole and etoposide. The cells were stained with a mixture of metal-tagged antibodies against cell membrane, cytoplasmic and nuclei markers and subjected to IMC. The panel of metal tag antibodies used in our study was designed to minimize any potential spill over cross-interference between neighbor isotopic mass channels [42, 43]. When analyzing our IMC data, we did not encounter strong and significant channel contamination from one metal labeled antibody to another. The collected data allowed us to develop a workflow for cellular imaging analysis, which relies on established computational methods [44]. Multiplexed images of cell compartments give a detailed and broad visualization of biomarker localization and cell size

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

parameter changes. Cell-based morphological segmentation of generated digital images is a prerequisite to identify and export multiple features of interest at single-cell resolution. Exploratory downstream analysis was performed with the histoCAT platform to generate, normalize, measure and compare multiple images according to their multidimensional content in the form of spatial distribution heat maps, dimensionality reduction and unsupervised clustering. In this report, we also describe cell classification by machine learning to pre-label cells before measuring their multivariate similarities. Prior knowledge of different cellular states is important for the calculation of distance and similarity metrics between each marker in the evaluation of the heterogeneity of cells and the effects of different compounds on cancer cells [45]. In the example of pHistone3^{S28}-positive mitotic cells, we generated a multivariate high-dimensional heat map that allows a comparative study of mechanisms of action of multiple drugs. Additional nuclear classes such as pH2A.X^{S139}, p53, cyclin D3 and Ki-67 were analyzed using similar multiparametric clustering approaches (data not shown). The strategy could enable new ways to perform deep proteomic drug screening on different models for translational medicine and preclinical trials [46, 47].

Other *in vitro* assays using stem cells, or co-cultures of different cell types, could be studied by IMC as an additional approach to designing high-content analysis to investigate cell-to-cell interactions and signaling. A recent targeted therapy against cancer oncogenes described a cytotoxic small-molecule compound that activated the steroid receptor signaling pathways and led to cell stress and death [48]. IMC may assist researchers to better understand perturbations in this and other pathways through high-dimensional analysis. Three-dimensional tumor spheroids grown *in vitro* are more representative of the complexity of cancer tissue than two-dimensional cell cultures and can be analyzed by IMC through the preparation of sections of FFPE pelleted

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spheroids [49]. We summarized in **Table 1** several advantages and challenges during preparation on slide, staining procedure, acquisition by IMC and software data analysis for three different models of sample. IMC was useful in the biodistribution study of the platinum-containing anti-cancer drug cisplatin in pancreatic adenocarcinoma xenograft model and could be applied to the emerging metal-containing chemotherapies and photodynamic therapies, which are considered as alternative approaches to classic compounds [50, 51]. Therefore, the goal for IMC in preclinical screening of metallodrugs will be to efficiently and effectively identify cellular responses of *in vitro* tumor-derived cell lines for lead optimization in target-based drug discovery [52]. A comparison of IMC performance metrics to other high content imaging system technologies in the field of single cell analysis is shown in **Table 2**.

In conclusion, we demonstrate a comprehensive image analysis workflow for cell-based IMC data analysis of surface and intracellular markers in drug-treated cells. The correlation coefficient for pairs of multiple parameters and the similar distances among a collection of treatment profiles facilitate downstream analysis and allow for direct data visualization. Image-based cell profiling studies compute statistical estimates of the likelihood of equivalence between two drug-induced profiles. Similarity measurements quantify proximity between profiles, since they detect deviations from one sample to another regardless of the absolute magnitude. This procedure is useful in finding relations and groups of samples that share common properties in high-content screening. Predicted pharmacodynamic effects were visualized and quantified in MCF-7 cells dosed with three target-specific compounds. Strong pairwise correlation between nuclear markers pHistone3^{S28}, Ki-67 and p4E-BP1$^{T37/T46}$ in mitotic cells and anti-correlation with cell surface markers CD29, CD98, CD81, CD47 and EpCAM was demonstrated.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

The authors thank the employees of Fluidigm Canada Inc. (Markham, Ontario) for their support of this research.

Funding

This research received no specific grant from any funding agency.

Associated Content

Supporting Information

Additional tables and figures and all the experimental details are mentioned in the supporting text.

Disclosures

A.B., A.E. and O.O. are employees of, and receive remuneration from, Fluidigm Corporation.

Fluidigm, Cell-ID, CyTOF, Imaging Mass Cytometry and IMC are trademarks and/or registered trademarks of Fluidigm Corporation in the United States and/or other countries. All other trademarks are the sole property of their respective owners.

For Research Use Only. Not for use in diagnostic procedures.

Author Contributions

AB designed the study and drafted the manuscript. AB, AE and OO took part to the manuscript writing. All authors have read and approved the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Masters JR. Human cancer cell lines: fact and fantasy. Nat Rev Mol Cell Biol 2000;1:233–236.
2. An WF, Tolliday N. Cell-based assays for high-throughput screening. Mol Biotechnol 2010;45:180–186.
3. Wilding JL, Bodmer WF. Cancer cell lines for drug discovery and development. Cancer Res 2014;74:2377–2384.
4. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Gonçalves E, Barthorpe S, Lightfoot H, Cokelaer T, Greninger P, van Dyk E, Chang H, de Silva H, Heyn H, Deng X, Egan RK, Liu Q, Mironenko T, Mitropoulos X, Richardson L, Wang J, Zhang T, Moran S, Sayols S, Soleimani M, Tamborero D, Lopez-Bigas N, Ross-Macdonald P, Esteller M, Gray NS, Haber DA, Stratton MR, Benes CH, Wessels LFA, Saez-Rodriguez J, McDermott U, Garnett MJ. A Landscape of Pharmacogenomic Interactions in Cancer. Cell 2016;166:740–754.
5. Haverty PM, Lin E, Tan J, Yu Y, Lam B, Lianoglou S, Neve RM, Martin S, Settleman J, Yauch RL, Bourgon R. Reproducible pharmacogenomic profiling of cancer cell line panels. Nature 2016;533:333–337.
6. Goodspeed A, Heiser LM, Gray JW, Costello JC. Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. Mol Cancer Res 2016;14:3–13.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

7. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 2006;6:813-823.
8. Sharma SV, Haber DA, Settleman J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nat Rev Cancer 2010;10:241–253.
9. Lang P, Yeow K, Nichols A, Scheer A. Cellular imaging in drug discovery. Nat Rev Drug Discov 2006;5:343–356.
10. Kang J, Hsu CH, Wu Q, Liu S, Coster AD, Posner BA, Altschuler SJ, Wu LF. Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. Nat Biotechnol 2016;34:70–77.
11. Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM, Petri Seiler K, Carrel HA, Golub TR, Schreiber SL, Clemons PA, Carpenter AE, Shamji AF. Multiplex cytological profiling assay to measure diverse cellular states. PLoS One 2013 doi: 10.1371/journal.pone.0080999.
12. Wu PH, Phillip JM, Khatau SB, Chen WC, Stirman J, Rosseel S, Tschudi K, Van Patten J, Wong M, Gupta S, Baras AS, Leek JT, Maitra A, Wirtz D. Evolution of cellular morpho-phenotypes in cancer metastasis. Sci Rep 2015 doi:10.1038/srep18437.
13. Boutros M, Heigwer F, Laufer C. Microscopy-Based High-Content Screening. Cell 2015;163:1314-1325.
14. Lin JR, Izar B, Wang S, Yapp C, Mei S, Shah PM, Santagata S, Sorger PK. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. eLife 2018; 7:e31657.
15. Wawer MJ, Li K, Gustafsdottir SM, Ljosa, V, Bodycombe NE, Marton MA, Sokolnicki KL, Bray MA, Kemp MM, Winchester E, Taylor B, Grant GB, Hon CS, Duvall JR,

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wilson JA, Bittker JA, Dančík V, Narayan R, Subramanian A, Winckler W, Golub TR, Carpenter AE, Shamji AF, Schreiber SL, Clemons PA. Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. Proc Natl Acad Sci U S A 2014;111:10911–10916.

16. Wagner BK, Schreiber SL. The Power of Sophisticated Phenotypic Screening and Modern Mechanism-of-Action Methods. Cell Chem Biol 2016;23:3–9.
17. Chang Q, Ornatsky OI, Siddiqui I, Loboda A, Baranov VI, Hedley DW. Imaging Mass Cytometry. Cytometry A 2017;91:160–169.
18. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, Pavlov S, Vorobiev S, Dick JE, Tanner SD. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. Anal Chem 2009;81:6813–6822.
19. Lou X, Zhang G, Herrera I, Kinach R, Ornatsky O, Baranov V, Nitz M, Winnik MA. Polymer-based elemental tags for sensitive bioassays. Angew Chem Int Ed Engl 2007;46:6111–6114.
20. Giesen C, Wang HA, Schapiro D, Zivanovic, N, Jacobs A, Hattendorf, B, Schüffler PJ, Grolimund D, Buhmann JM, Brandt S, Varga Z, Wild PJ, Günther D, Bodenmiller B. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods 2014;11:417–422.
21. Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanotelli VRT, Schulz D, Giesen C, Catena R, Varga Z, Bodenmiller B. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat Methods 2017;14:873–876.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S, Roberts ME, Jones TR, Singh S, Genovesio A, Clemons PA, Carragher NO, Carpenter AE. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J Biomol Screen 2013;18:1321–1329.
23. Bougen-Zhukov N, Loh SY, Lee HK, Loo LH. Large-scale image-based screening and profiling of cellular phenotypes. Cytometry A 2017;91:115–125.
24. Henjes F, Bender C, von der Heyde S, Braun L, Mannsperger HA, Schmidt C, Wiemann S, Hasmann M, Aulmann S, Beissbarth T, Korf U. Strong EGFR signaling in cell line models of ERBB2-amplified breast cancer attenuates response towards ERBB2-targeting drugs. Oncogenesis doi:10.1038/oncsis.2012.16.
25. Holliday DL, Speirs V. Choosing the right cell line for breast cancer research. Breast Cancer Res 2011;13:215–224.
26. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, Perou CM. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast Cancer Res Treat 2013;142:237–255.
27. Lacroix M, Leclercq G. Relevance of breast cancer cell lines as models for breast tumors: an update. Breast Cancer Res Treat 2004;83:249–289.
28. Caie PD, Walls RE, Ingleston-Orme A, Daya S, Houslay T, Eagle R, Roberts ME, Carragher NO. High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol Cancer Ther 2010;9:1913–1926.
29. Bafna S, Kaur S, Batra SK. Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. Oncogene 2010;29:2893–2904.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

30. Kim SM, Hahn JH. CD98 activation increases surface expression and clustering of beta1 integrins in MCF-7 cells through FAK/Src- and cytoskeleton-independent mechanisms. Exp Mol Med 2008;40:261–270.

31. Garcia R, Franklin RA, McCubrey JA. EGF induces cell motility and multi-drug resistance gene expression in breast cancer cells. Cell Cycle 2006;23:2820–2826.

32. Blajeski AL, Phan VA, Kottke TJ, Kaufmann SH. G(1) and G(2) cell-cycle arrest following microtubule depolymerization in human breast cancer cells. J Clin Invest 2002;110:91–99.

33. Ji J, Zhang Y, Redon CE, Reinhold WC, Chen AP, Fogli LK, Holbeck SL, Parchment RE, Hollingshead M, Tomaszewski JE, Dudon Q, Pommier Y, Doroshow JH, Bonner WM. Phosphorylated fraction of H2AX as a measurement for DNA damage in cancer cells and potential applications of a novel assay. PLoS One 2017 doi: 10.1371/journal.pone.0171582.

34. Molnar C, Jermyn IH, Kato Z, Rahkama V, Östling P, Mikkonen P, Pietiäinen V, Horvath P. Accurate Morphology Preserving Segmentation of Overlapping Cells based on Active Contours. Sci Rep 2016 doi: 10.1038/srep32412.

35. Sobecki M, Mrouj K, Colinge J, Gerbe F, Jay P, Krasinska L, Dulic V, Fisher D. Cell-Cycle Regulation Accounts for Variability in Ki-67 Expression Levels. Cancer Res. 2017;10:2722-2734.

36. Van der Maaten LJP, Hinton GE. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 2008;9:2579–2605.

37. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P, Sabatini DM. Scoring diverse cellular morphologies in

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

image-based screens with iterative feedback and machine learning. Proc Natl Acad Sci U S A. 2009 Feb 10;6:1826-31.

38. Dao D, Fraser AN, Hung J, Ljosa V, Singh S, Carpenter AE. CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. Bioinformatics 2016;32:3210–3212.
39. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, Lahr DL, Hirschman JE, Liu Z, Donahue M, Julian B, Khan M, Wadden D, Smith IC, Lam D, Liberzon A, Toder C, Bagul M, Orzechowski M, Enache OM, Piccioni F, Johnson SA, Lyons NJ, Berger AH, Shamji AF, Brooks AN, Vrcic A, Flynn C, Rosains J, Takeda DY, Hu R, Davison D, Lamb J, Ardlie K, Hogstrom L, Greenside P, Gray NS, Clemons PA, Silver S, Wu X, Zhao WN, Read-Button W, Wu X, Haggarty SJ, Ronco LV, Boehm JS, Schreiber SL, Doench JG, Bittker JA, Root DE, Wong B, Golub TR. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 2017; 171:1437–1452.
40. Zhou JX, Isik Z, Xiao C, Rubin I, Kauffman SA, Schroeder M, Huang S. Systematic drug perturbations on cancer cells reveal diverse exit paths from proliferative state. Oncotarget 2016;7:7415–7425.
41. Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nat Rev Drug Discov 2017;16:531–543.
42. Takahashi C, Au-Yeung A, Fuh F, Ramirez-Montagut T, Bolen C, Mathews W, O'Gorman WE. Mass cytometry panel optimization through the designed distribution of signal interference. Cytometry A. 2017;91:39-47.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

43. Chevrier S, Crowell HL, Zanotelli VRT, Engler S, Robinson MD, Bodenmiller B. Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. Cell Syst. 2018;6:612-620.

44. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, Wawer M, Paavolainen L, Herrmann MD, Rohban M, Hung J, Hennig H, Concannon J, Smith I, Clemons PA, Singh S, Rees P, Horvath P, Linington RG, Carpenter AE. Data-analysis strategies for image-based cell profiling. Nat Methods 2017;14:849–863.

45. Ljosa V, Caie PD, Ter Horst R, Sokolnicki KL, Jenkins EL, Daya S. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J Biomol Screen 2013;18:1321–1329.

46. Doan M, Vorobjev I, Rees P, Filby A, Wolkenhauer O, Goldfeld AE, Lieberman J, Barteneva N, Carpenter AE, Hennig H. Diagnostic Potential of Imaging Flow Cytometry. Trends Biotechnol 2018 doi: 10.1016/j.tibtech.2017.12.008.

47. Rahmoune H, Guest PC. Application of Multiplex Biomarker Approaches to Accelerate Drug Discovery and Development. In: Guest PC, editor. Multiplex Biomarker Techniques. Methods Mol Biol; 2017. p.3–17.

48. Wang L, Yu Y, Chow DC, Yan F, Hsu CC, Stossi F, Mancini MA, Palzkill T, Liao L, Zhou S, Xu J, Lonard DM, O'Malley BW. Characterization of a Steroid Receptor Coactivator Small Molecule Stimulator that Overstimulates Cancer Cells and Leads to Cell Stress and Death Cancer. Cell 2015;28:240–252.

49. Stock K, Estrada MF, Vidic S, Gjerde K, Rudisch A, Santo VE, Barbier M, Blom S, Arundkar SC, Selvam I, Osswald A, Stein Y, Gruenewald S, Brito C, van Weerden W,

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Rotter V, Boghaert E, Oren M, Sommergruber W, Chong Y, de Hoogt R, Graeser R. Capturing tumor complexity *in vitro*: Comparative analysis of 2D and 3D tumor models for drug discovery. *Sci Rep* 2016 doi: 10.1038/srep28951.

50. Chang Q, Ornatsky OI, Siddiqui I, Straus R, Baranov VI, Hedley DW. Biodistribution of cisplatin revealed by imaging mass cytometry identifies extensive collagen binding in tumor and normal tissues. *Sci Rep*. 2016 doi: 10.1038/srep36641.
51. Mjos KD, Orvig C. Metallodrugs in medicinal inorganic chemistry. Chem Rev 2014;114:4540–4563.
52. Gillet JP, Varma S, Gottesman MM. The Clinical Relevance of Cancer Cell Lines. J Natl Cancer Inst 2013;105:452–458.
53. Osborne CK, Hamilton B, Titus G, Livingston RB. Epidermal growth factor stimulation of human breast cancer cells in culture. Cancer Res 1980;40:2361–2366.
54. Hau A, Ceppi P, Peter ME. CD95 is part of a let-7/p53/miR-34 regulatory network. PLoS One 2012 doi: 10.1371/journal.pone.0049636.
55. Schulz D, Zanotelli VRT, Fischer JR, Schapiro D, Engler S, Lun XK, Jackson HW, Bodenmiller B. Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. Cell Syst 2018;6:25–36.
56. Bray MA, Vokes MS, Carpenter AE. Using CellProfiler for Automatic Identification and Measurement of Biological Objects in Images. Curr Protoc Mol Biol 2015;109:1–13.
57. Swift ML. GraphPad Prism, Data Analysis, and Scientific Graphing. J Chem Inf Comput Sci 1997;37:411–412.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

58. Holten D. Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans Vis Comput Graph. 2006 ;12:741-748

59. Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, Levenson RM, Lowe JB, Liu SD, Zhao S, Natkunam Y, Nolan GP. Multiplexed ion beam imaging of human breast tumors. Nat Med. 2014;20:436-442.

60. Pomerantz AK, Sari-Sarraf F, Grove KJ, Pedro L, Rudewicz PJ, Fathman JW, Krucker T. Enabling drug discovery and development through single-cell imaging. Expert Opin Drug Discov. 2019;14:115-125.

61. Beghin A, Kechkar A, Butler C, Levet F, Cabillic M, Rossier O, Giannone G, Galland R, Choquet D, Sibarita JB. Localization-based super-resolution imaging meets high-content screening. Nat Methods. 2017;14:1184-1190.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. IMC multiplex image of MCF-7 non-treated and DMSO-treated control cells.

Figure S2. Cell segmentation and single-marker distribution of MCF-7 non-treated and DMSO-treated controls.

Figure S3. Segmentation accuracy of nuclei and cytoplasmic contours of identified MCF-7 cells drug-treated using binary image overlap.

Figure S4. Comparison of average Z-score expression level of p53, cyclin D3, CD47, CD49e and Cell-ID Intercalator-Ir between each control and drug treatment condition.

Figure S5. Unsupervised cluster identification by t-SNE of control and compound-treated MCF-7.

Figure S6. Fast Gentle Boosting classification matrices of stitched replicates of MCF-7 ROIs per control and drug compound.

Figure S7. Comparative expression levels of protein markers Ki-67, cyclin D3, pH2A.X^{S139}, p53 and pHistone3^{S28} (rows) per nuclear class and drug treatment condition (column) of MCF-7 cells.

Figure S8. Heat maps of Pearson correlation coefficients of multiple parameters for non-treated and DMSO-treated controls ROIs on mitotic pHistone3^{S28} cells.

Figure S9. Graphical representations of Pearson correlation coefficients for nuclei markers (A) and cytoplasmic markers (B) across all controls and drug treatments in the classified population of mitotic pHistone3^{S28} MCF-7 cells.

Table S1. List of chemical compounds used for MCF-7 cell treatment, 48 h exposure.

Table S2. Metal-labeled antibody panel for cell surface markers.

Table S3. Metal-labeled antibody panel for intracellular markers.

Table S4. Mitotic index of compound-treated MCF-7 generated by support vector machine classification with Fast Gentle Boosting ruler.

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. Image-based IMC data analysis workflow with open source software.

Figure 2. IMC images of cellular and nuclear markers expressed by SKBR3 with zoom-in area. Cell-ID: nuclei labeled with Ir-intercalator. Zoom-in images are rendered with multiple combinations of protein markers detected in different cellular compartments. Scale bar = 100 µm.

Figure 3. IMC images of cellular and nuclear markers expressed by HCC1143 with zoom-in areas (two ROI). Cell-ID: nuclei labeled with Ir-intercalator. Each zoom-in composite image is rendered with a selection of different markers. Scale bar = 100 µm.

Figure 4. Multiplexed IMC images of MCF-7 cells treated with different compounds. (A) EGF, (B) nocodazole, (C) etoposide. Composite images of 1500 x 1500 µm size (scale bar = 160 µm) with 400 x 400 µm zoom-in areas (scale bar = 100 µm). Columns images correspond to the same combination of proteins detected per drug treatment (row).

Figure 5. Image processing by cell-based segmentation and spatial parametric distribution across ROIs of drug-treated MCF-7 cells. (A) Cell and nuclear masks overlaid on color IMC images. Scale bar = 100 µm. Zoom-in areas of selected parameters visualized by Z-score normalized color heat maps for (B) pHistone3^{S28} (C) pH2A.X^{S139}. Scale bar = 10 µm. (C) Z-score mean intensity levels of nuclear markers (top), surface markers (middle) and cellular size parameters (bottom) for control and drug treatment conditions. Data are presented as mean ± SEM. ****P < 0.0001 (unpaired t-test).

bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6. Unsupervised cluster identification by t-SNE of control and compound-treated MCF-7. (A) Combined colored t-SNE maps of individual cells (EGF, green; etoposide, black; nocodazole, purple), DMSO (red) and non-treated control (blue). (B) t-SNE maps colored according to the expression level of nuclear markers and cell size (area).

Figure 7. Support vector machine learning classification and multiparametric hierarchical similarity clustering of drug-treated MCF-7.

(A) 6-plex color image of nocodazole-treated cells generated by CPA Image Viewer. Scale bar = 160 µm. (B) Normalized confusion matrix of six different cell nuclear phenotypes. (C) Protein expression levels of nuclear markers used for cell classification of MCF-7 treated with nocodazole. Graphs represent the non-normalized mean intensity (±SD) of nuclear proteins within each assigned category. $P < 0.05$ and $P < 0.0001$ (unpaired t-test) are included.

Figure 8. Hierarchical clustering heat maps of Pearson coefficients of multiple parameters for drug-treated pHistone3^{S28}-positive cells. Clusters of cells are highlighted according to cells size parameters (black), CD surface markers (blue), Cell-ID Intercalator-Ir isotopes (green), pan-keratin and CK19 markers (orange) and nuclear markers (yellow).

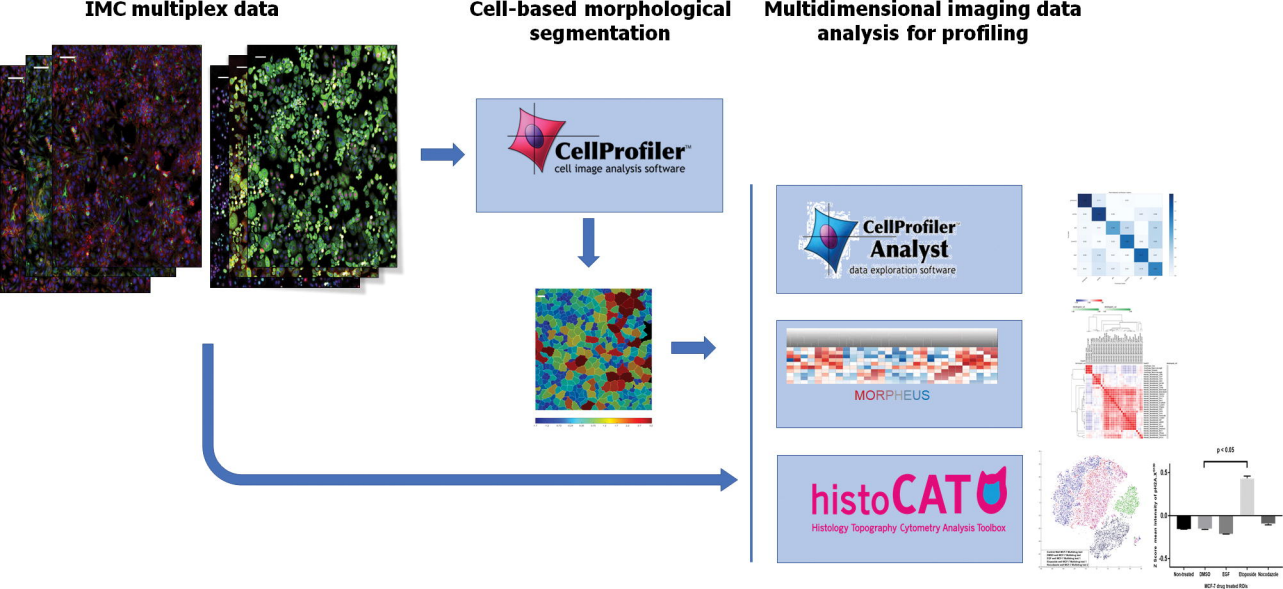
Figure 9. Comparative analysis of multiparametric pairwise correlations between surface membrane markers expressed by MCF-7 pHistone3^{S28} positive mitotic cells across controls and drug treatments. (A) Hierarchical similarity heatmap of CD98, CD29, EpCAM, CD81, CD47 and CD49e protein markers. Individual clusters are highlighted around the diagonal component for each condition (Non-treated, blue; DMSO, red; EGF, green; etoposide,

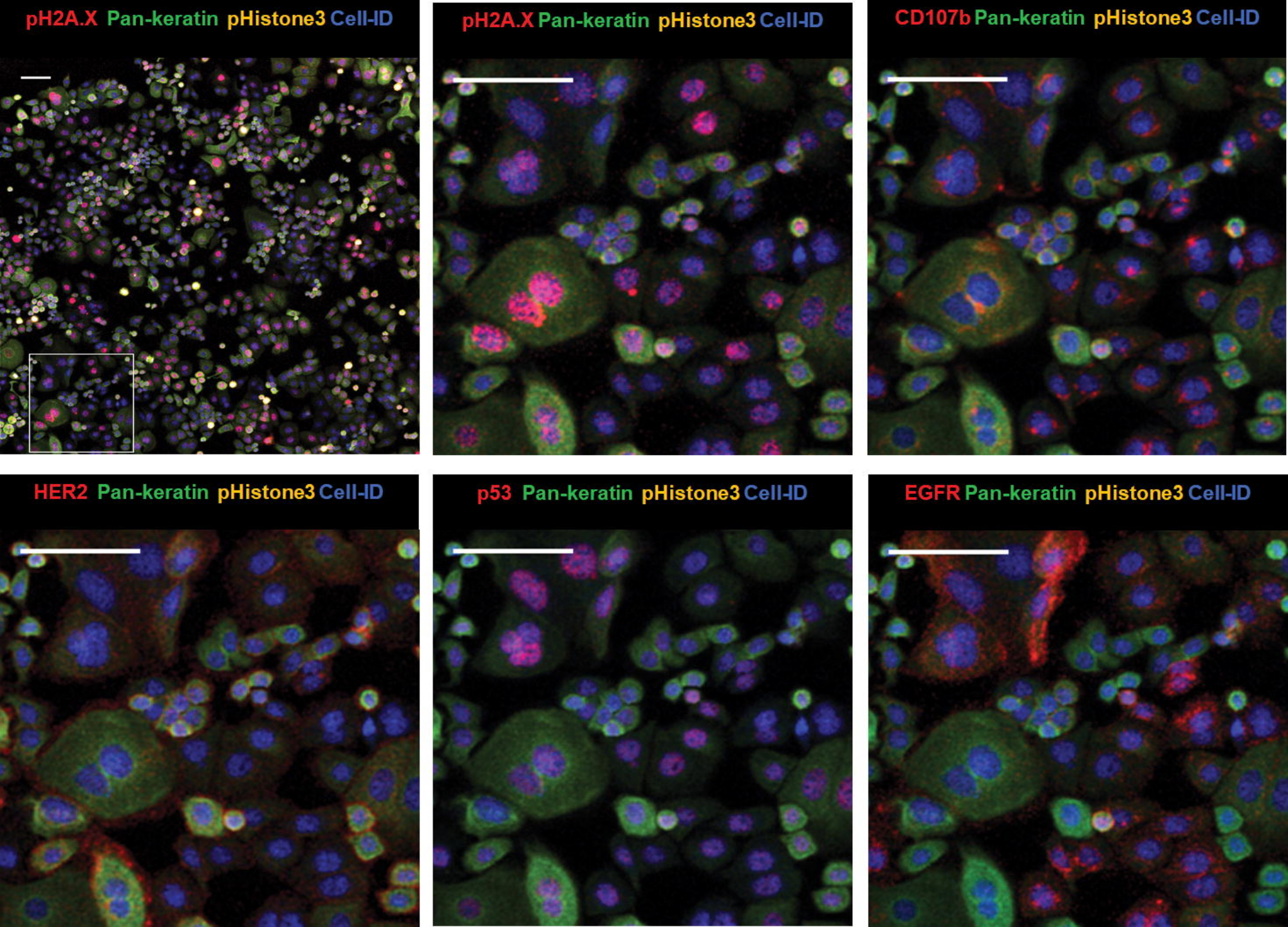
bioRxiv preprint doi: https://doi.org/10.1101/549592; this version posted May 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

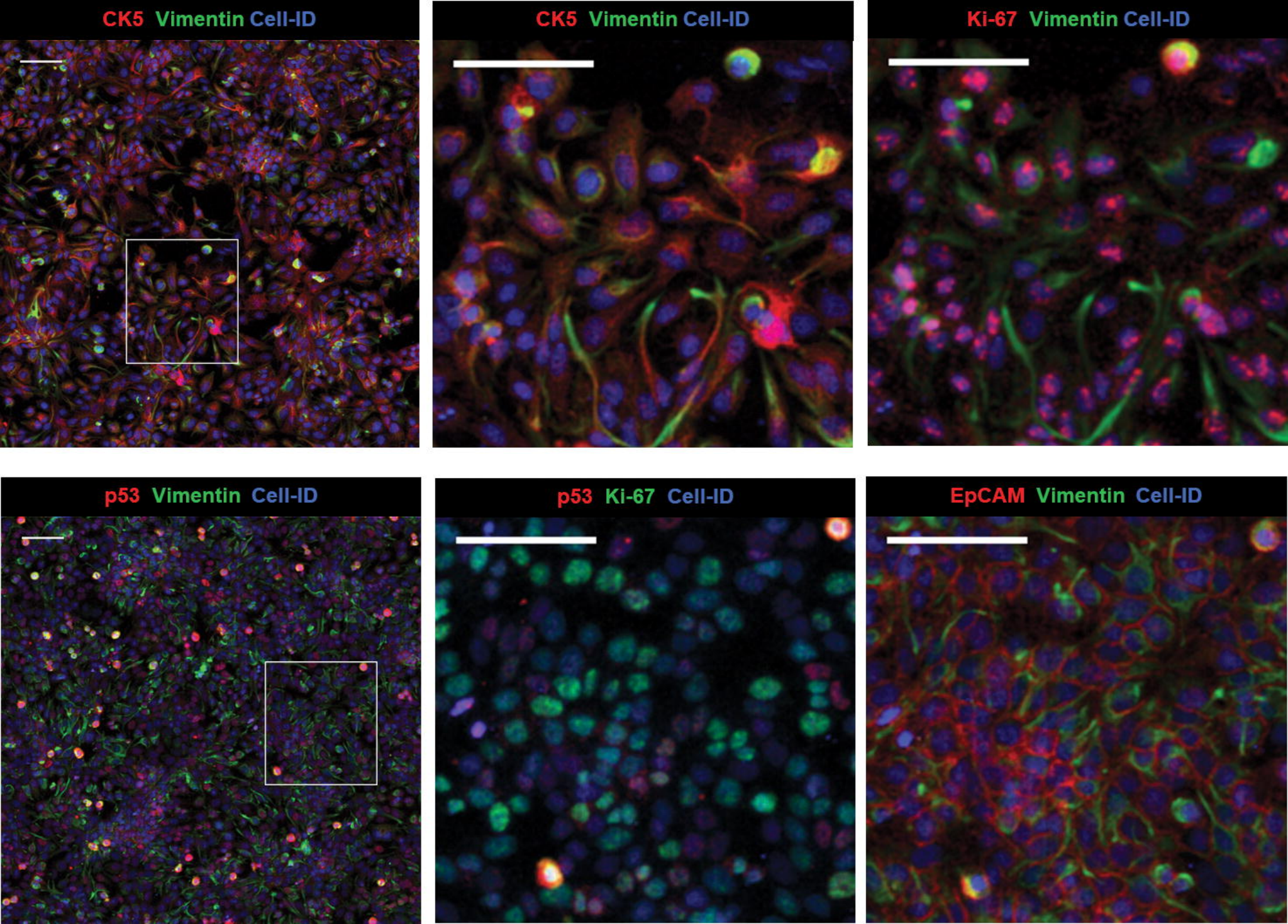
black; nocodazole, purple). (B) Hierarchical edge bundle visual graph showing connecting edges between parametric nodes. Bundled lines between nodes correspond to positive pairwise correlations with a cut-off value > 0.3. The bundled interconnections show similarities of surface membrane phenotyping profiles between non-treated and DMSO controls.

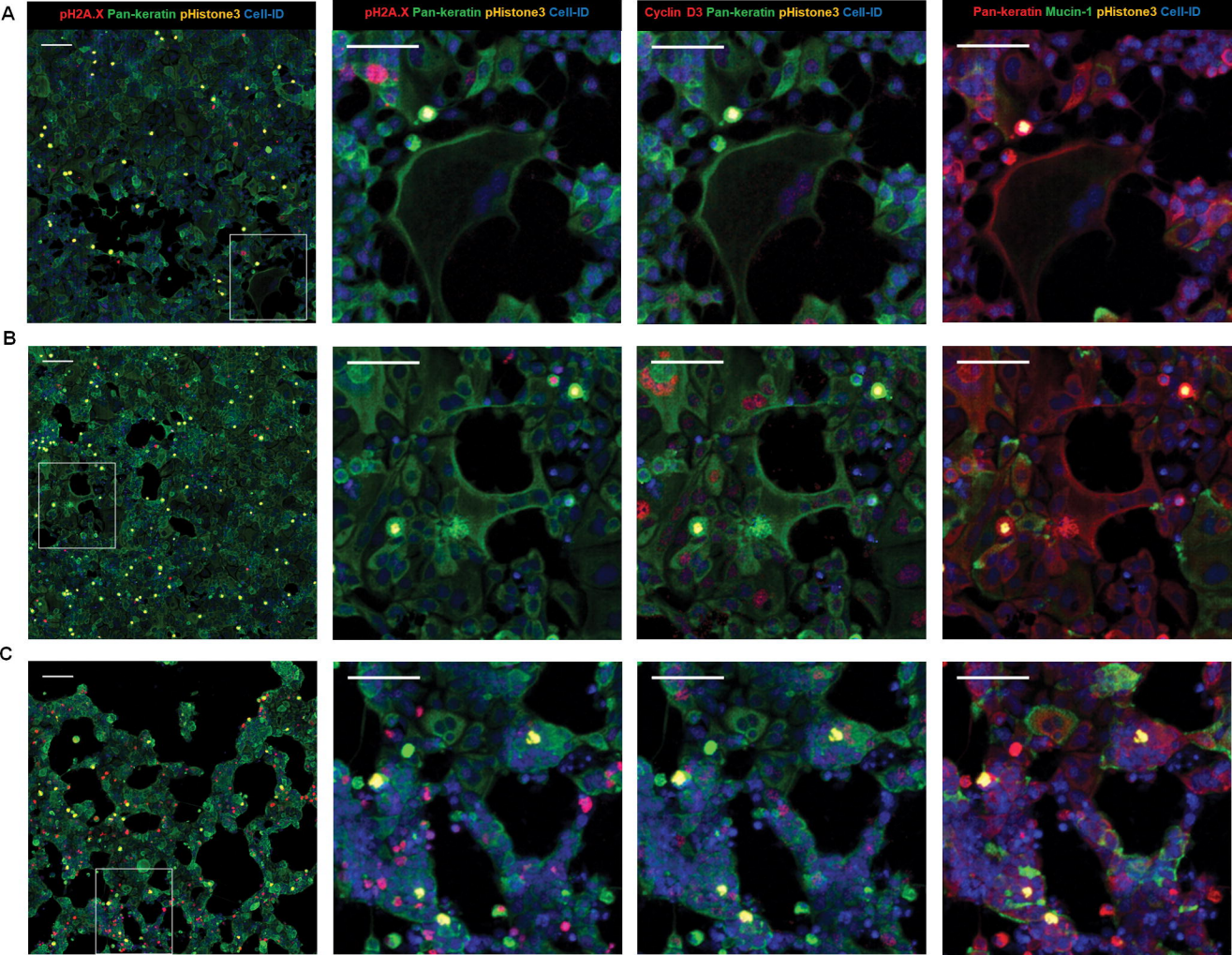
Table 1. Key similarities and differences of biological samples processing used for IMC.

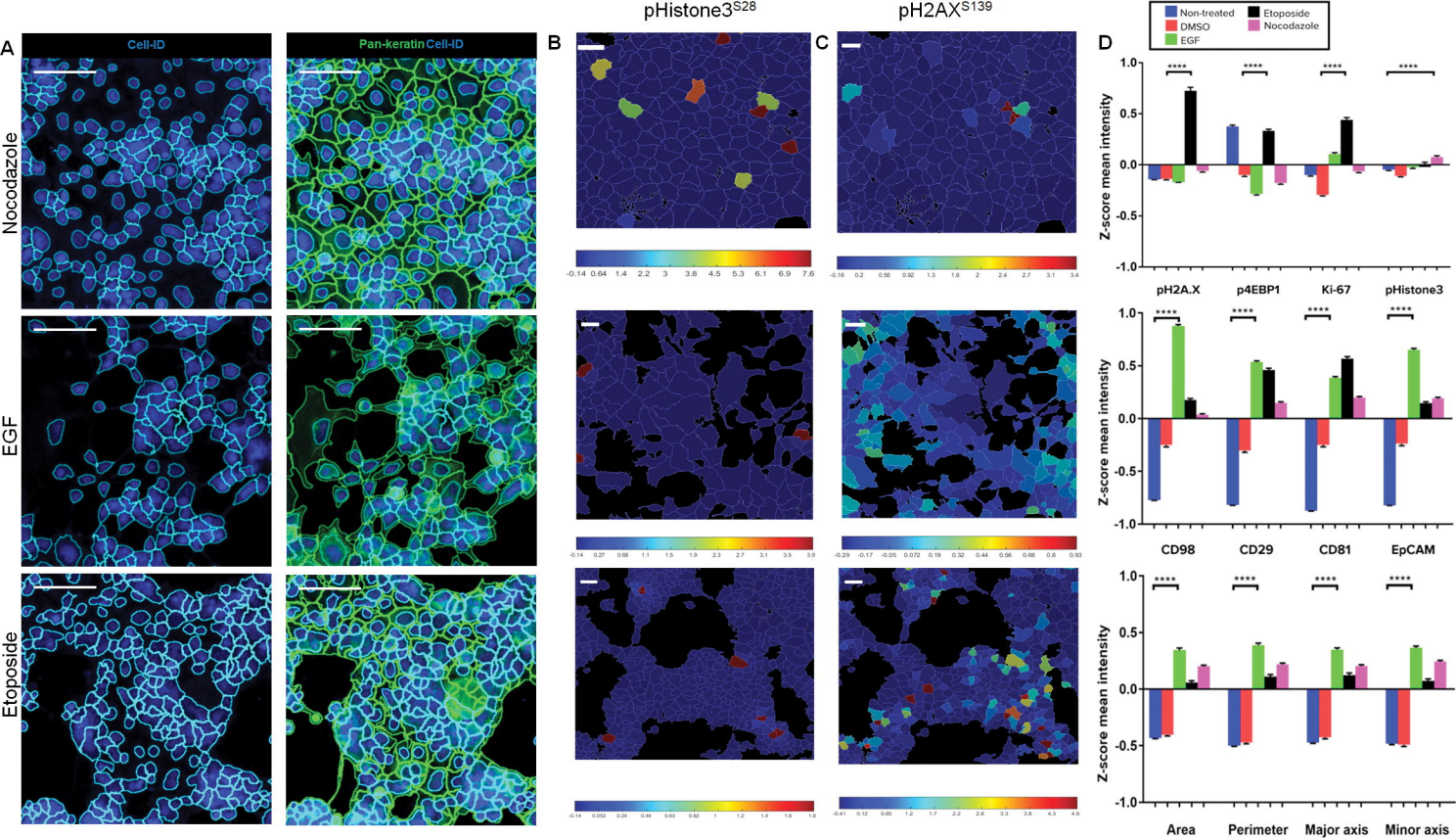
Table 2. Performance comparison of IMC to other high-content single cell imaging technologies.

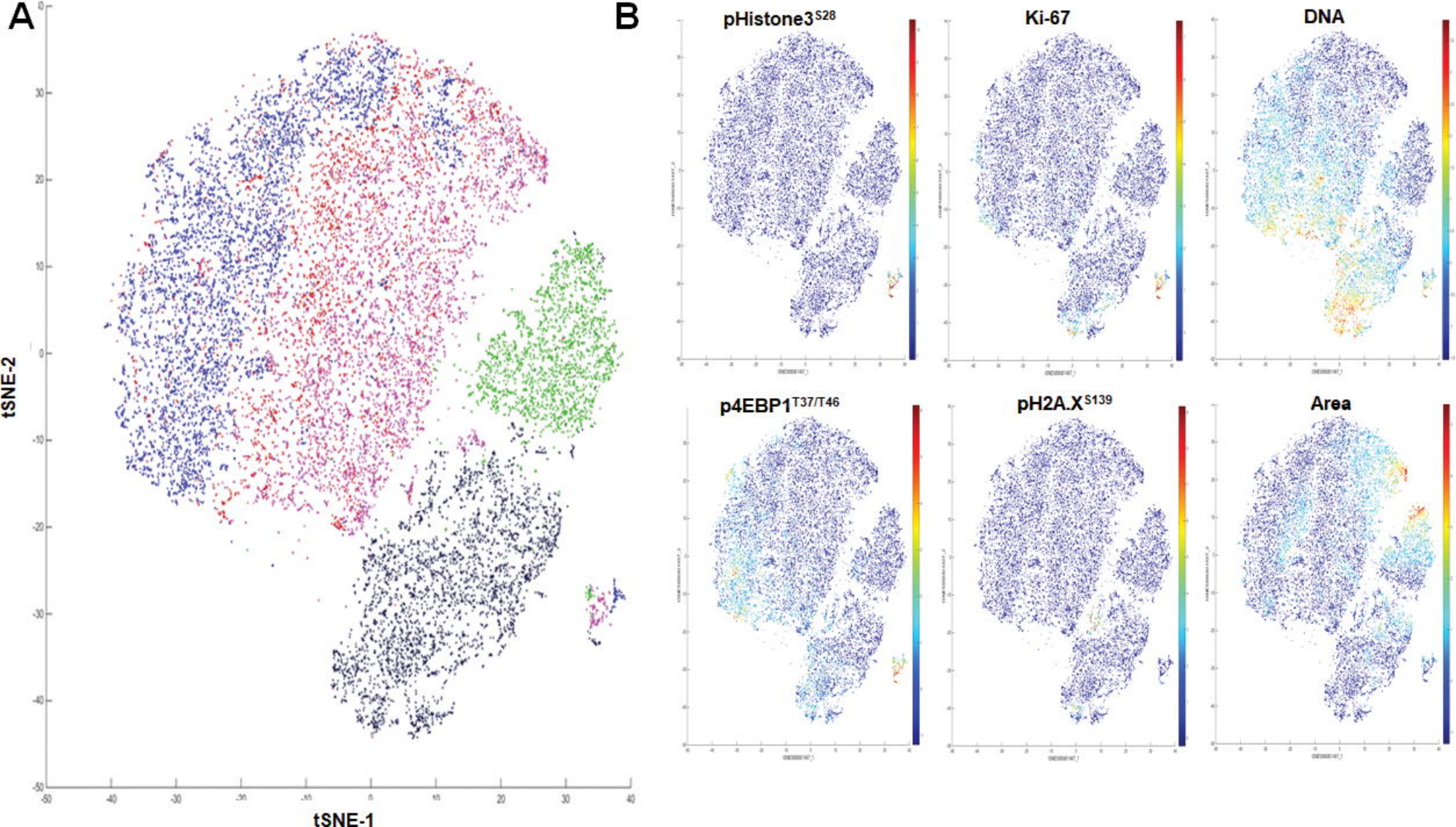

IMC multiplex data
Cell-based morphological segmentation
Multidimensional imaging data analysis for profiling
CellProfiler™
cell image analysis software
CellProfiler
Analyst
data exploration software
MORPHEUS
histoCAT
Histology Topography Cytometry Analysis Toolbox
p < 0.05
Nontreated
DMSO
EGF
Etoposide
Nocodazole
MCF-7 drug treated RDs

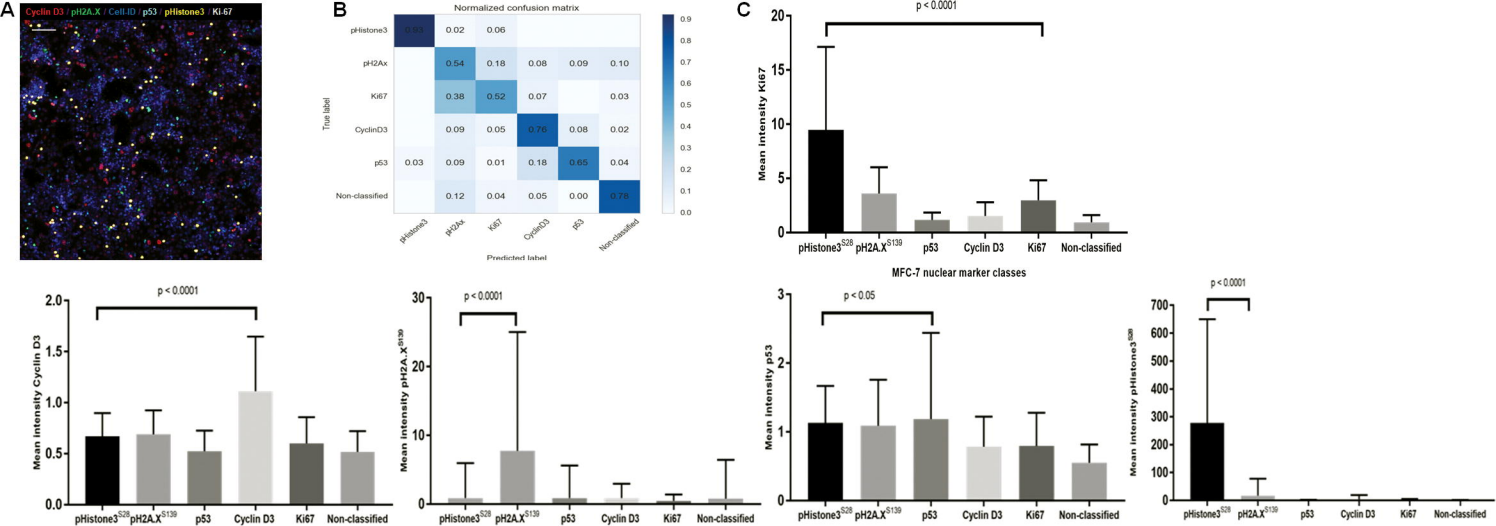

pH2A.X Pan-keratin pHistone3 Cell-ID
pH2A.X Pan-keratin pHistone3 Cell-ID
CD107b Pan-keratin pHistone3 Cell-ID
HER2 Pan-keratin pHistone3 Cell-ID
p53 Pan-keratin pHistone3 Cell-ID
EGFR Pan-keratin pHistone3 Cell-ID

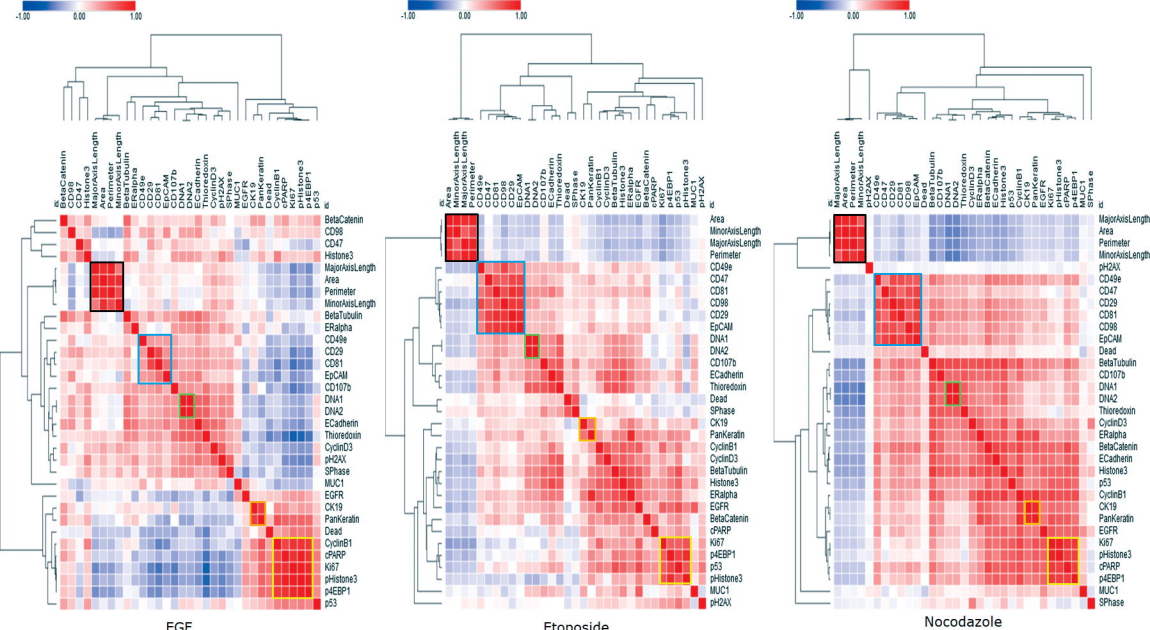

CK5 Vimentin Cell-ID
CK5 Vimentin Cell-ID
Ki-67 Vimentin Cell-ID
p53 Vimentin Cell-ID
p53 Ki-67 Cell-ID
EpCAM Vimentin Cell-ID

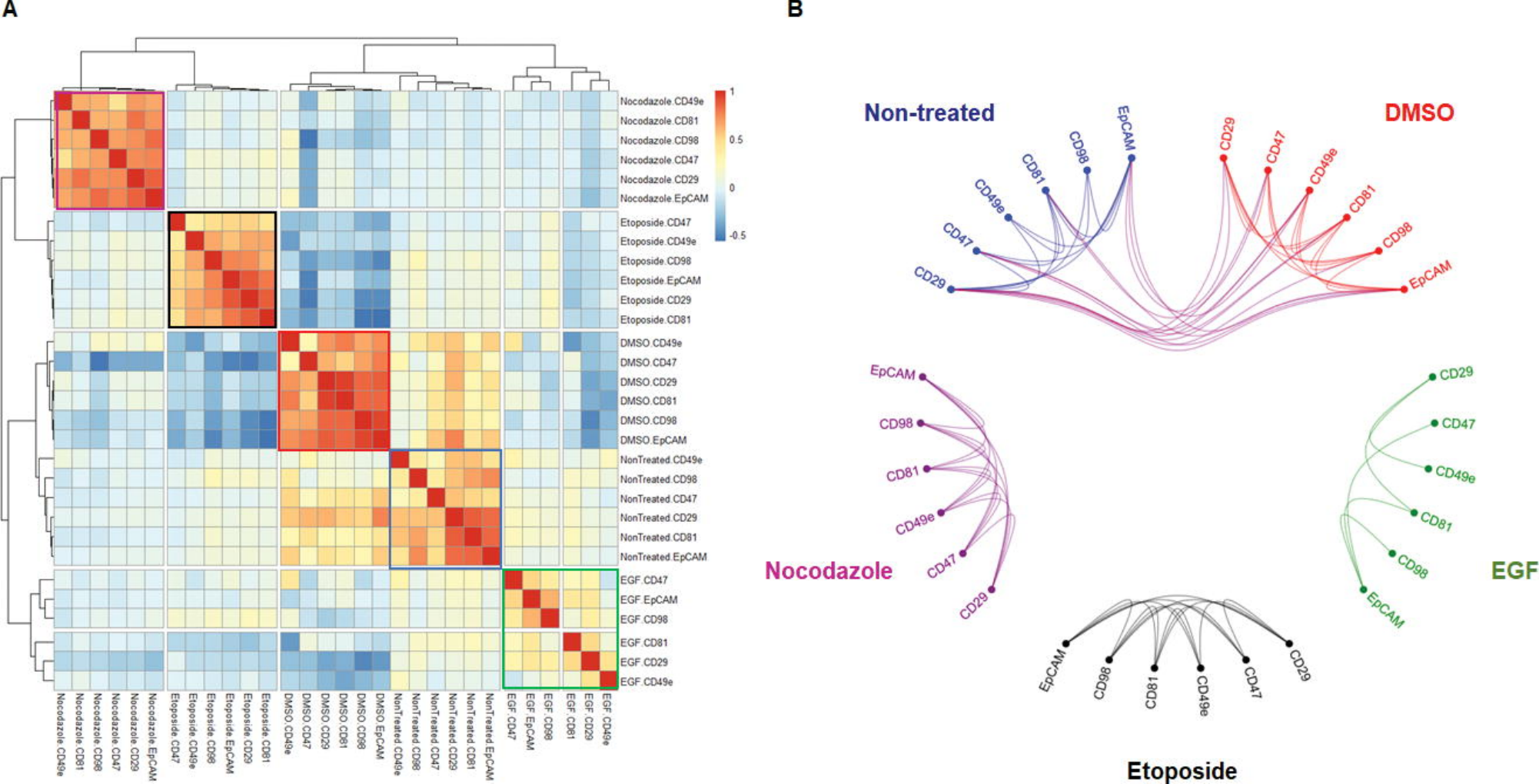

A
pH2A.X Pan-keratin pHistone3 Cell-ID
pH2A.X Pan-keratin pHistone3 Cell-ID
Cyclin D3 Pan-keratin pHistone3 Cell-ID
Pan-keratin Mucin-1 pHistone3 Cell-ID
B
C


A
Cell-ID
Pan-keratin Cell-ID
Nocodazole
EGF
Etoposide
B
pHistone3S28
C
pH2AXS139
D
Non-treated
DMSO
EGF
Etoposide
Nocodazole
Z-score mean intensity
pH2A.X
p4EBP1
Ki-67
pHistone3
CD98
CD29
CD81
EpCAM
Area
Perimeter
Major axis
Minor axis


A
B
tSNE-1
tSNE-2
pHistone3S28
Ki-67
DNA
p4EBP1T37/T46
pH2A.XS139
Area


A
Cyclin D3 / pH2A.X / Cell ID / p53 / pHistone3 / Ki-67
B
Normalized confusion matrix
True label
Predicted label
pHistone3
pH2Ax
Ki67
CyclinD3
p53
Non-classified
0.93 0.02 0.06
0.54 0.18 0.08 0.09 0.10
0.38 0.52 0.07 0.03
0.09 0.05 0.76 0.08 0.02
0.03 0.09 0.01 0.18 0.65 0.04
0.12 0.04 0.05 0.00 0.78
C
p < 0.0001
Mean intensity Ki67
pHistone3S28
pH2A.XS139
p53
Cyclin D3
Ki67
Non-classified
MFC-7 nuclear marker classes
Mean intensity Cyclin D3
Mean intensity pH2A.XS139
p < 0.05
Mean intensity p53
Mean intensity pHistone3S28


-1.00
0.00
1.00
id
BetaCatenin
CD98
CD47
Histone3
MajorAxisLength
Area
Perimeter
MinorAxisLength
BetaTubulin
ERalpha
CD49e
CD29
CD81
EpCAM
CD107b
DNA1
DNA2
ECadherin
Thioredoxin
CyclinD3
pH2AX
SPhase
MUC1
EGFR
CK19
PanKeratin
Dead
CyclinB1
cPARP
Ki67
pHistone3
p4EBP1
p53
EGF
Etoposide
Nocodazole


A
B
Nocodazole.CD49e
Nocodazole.CD81
Nocodazole.CD98
Nocodazole.CD47
Nocodazole.CD29
Nocodazole.EpCAM
Etoposide.CD47
Etoposide.CD49e
Etoposide.CD98
Etoposide.EpCAM
Etoposide.CD29
Etoposide.CD81
DMSO.CD49e
DMSO.CD47
DMSO.CD29
DMSO.CD81
DMSO.CD98
DMSO.EpCAM
NonTreated.CD49e
NonTreated.CD98
NonTreated.CD47
NonTreated.CD29
NonTreated.CD81
NonTreated.EpCAM
EGF.CD47
EGF.EpCAM
EGF.CD98
EGF.CD81
EGF.CD29
EGF.CD49e
1
0.5
0
-0.5
Non-treated
DMSO
Nocodazole
EGF
Etoposide
EpCAM
CD98
CD81
CD49e
CD47
CD29

Sample type	Tissue sections (Frozen and FFPE)	Cell smear on microscopic slide (Liquid biopsy)	Adherent cell lines grown in chamber slides
Sample quality and integrity criteria	• Thin thickness of the cut tissue slice (5µm). • Effectiveness of tissue fixation for FFPE (temperature, fixing chemicals, pH, time). • Storage conditions for tissue preservation (temperature, humidity, light exposure).	• Cell monolayer smear on slide by surface tension. • Fixation method. • Low temperature storage conditions. • Pre-identification of cells by immunofluorescence staining and scanning.	• High cell viability and proliferation rate. • Adherence efficiency of growing cells to the surface of the slide. • Optimal seeding density on slide and growth conditions per cell line tested.
Sample staining critical steps	• Optimization of dewaxing and hydration procedures for FFPE tissue. • Removal of tissue freezing matrix for frozen tissue. • Antigen retrieval and blocking conditions. • Metal isotope labeling of validated antibody clones for tissue staining. • Titration range testing of metal tag antibodies and DNA metallo-intercalators.	• Use of hydrophobic circle barrier around cells in specific regions for liquid staining on slide. • Blocking conditions. • Preparation of metal tag antibodies liquid cocktail. • Washing and drying of the sample slide before acquisition.	• Metal tag antibodies panel specific to surface membrane, cytoplasm and nucleus of adherent cells. • Chamber well liquid volume measurements for cell staining procedure. • Fixation and permeabilization conditions. • Removal of chamber wells from the slide.
Sample acquisition by IMC	• Simultaneous acquisition of all markers. • Whole ablation of tissue regions. • No autofluorescence.	• Plotting of two dimensional coordinates of cells pre-identified by immunofluorescence. • Laser ablation of areas with good cellular density and morphology.	• Regions of interest with intact cells. • Ablation of areas with even cell monolayers rather than densely packed overlapping cells.
Data analysis challenges	• Pathology expertise required for IMC imaging data visualization. • Annotation of structural and cellular regions. • Segmentation of individual and distinct cells within tissue samples require nuclei and membrane markers.	• Single-cell segmentation of small-sized cells. • Identification of nuclei with DNA metal-intercalator. • Use of membrane stain for detection of cellular boundaries. • Classification of cellular subpopulations.	• Single-cell morphological segmentation analysis. • Nuclei segmentation with DNA metal-intercalator. • Whole cell detection using cytoplasmic marker. • Classification of different cell states.
Multiplexing biological applications	• Immuno-oncology • Solid tumor oncology • Quantitative pathology	• Identification of rare circulating tumor metastatic cells • Hematology • Immunology	• Drug discovery • Cell lines authentication and characterization • Single-cell systems biology

High-Content Imaging Technology	Imaging Mass Cytometry	Imaging Flow Cytometry	Imaging Mass Spectrometry (SIMS MIBI, MALDI)	Single Cell Fluorescence Microscopy	Super-resolution Microscopy
Reagents used	• Metal tag antibodies • DNA metal-intercalators	• Fluorescent tag antibodies • DNA dyes	• Metal tag antibodies (SIMS MIBI) • Label free, matrix solutions (MALDI)	• Fluorescent tag antibodies • DNA dyes	• Fluorescent probes
Sample format	• Chamber slide (2, 4 ,8, 16-well format) • Tissue on slide • Cell smear on slide	• Multi-well plates • Tubes	• Tissue on slide • Cell smear on slide	• Multi-well plates (96, 384, 1536-well format)	• Single coverslip • Multi-well plates
Sample preparation time	6 hours	6 hours	6 hours	6 hours per staining cycle	6 hours
Spatial resolution	1 μm	1 μm	• < 1 μm (SIMS MIBI) • 5-200 μm (MALDI)	1–100 μm	< 1 μm
Throughput time	200 pixels per second	1000-5000 cells per second	• 10 000 pixels per second (SIMS MIBI) • 100 pixels per second (MALDI)	100-300 cells per second	10 cells per hour
Number of cells analyzed	10^3-10^4	10^4-10^5	10^3-10^4 (SIMS MIBI)	10^3-10^5	10-10^2
Sample types	• Fixed adherent and suspension cells • FFPE and frozen tissue sections	• Cells in suspension	• Fixed adherent and suspension cells • FFPE and frozen tissue sections	• Live and fixed cells	• Live and fixed cells
Number of targets	> 37 markers	12 markers	• 7 markers per scan (SIMS MIBI) • 0-100 000 Dalton m/z spectra range of molecules (MALDI)	6 markers per scan	4 markers per scan
References	17, 20	46	59, 60	14, 60	60, 61