Patient-by-Patient Deep Transfer Learning for Drug-Response Profiling Using Confocal Fluorescence Microscopy of Pediatric Patient-Derived Tumor-Cell Spheroids

Yannick Berker, Dina ElHarouni, Heike Peterziel, Petra Fiesel, Olaf Witt, Ina Oehme, Matthias Schlesner, and Sina Oppermann

***Abstract*— Image-based phenotypic drug profiling is receiving increasing attention in drug discovery and precision medicine. Compared to classical end-point measurements quantifying drug response, image-based profiling enables both the quantification of drug response and characterization of disease entities and drug-induced cell-death phenotypes. Here, we aim to quantify image-based drug responses in patient-derived 3D spheroid tumor cell cultures, tackling the challenges of a lack of single-cell-segmentation methods and limited patient-derived material. Therefore, we investigate deep transfer learning with patient-by-patient fine-tuning for cell-viability quantification. We fine-tune a convolutional neural network (pre-trained on ImageNet) with 210 control images specific to a single training cell line and 54 additional screen-specific assay control images. This method of image-based drug profiling is validated on 6 cell lines with known drug sensitivities, and further tested with primary patient-derived samples in a medium-throughput setting. Network outputs at different drug concentrations are used for drug-sensitivity scoring, and dense-layer activations are used in t-distributed stochastic neighbor embeddings of drugs to visualize groups of drugs with similar cell-death phenotypes. Image-based cell-line experiments show strong correlation to metabolic results ($R \approx 0.7$) and confirm expected hits, indicating the predictive power of deep learning to identify drug-hit candidates for individual patients. In patient-derived samples, combining drug sensitivity scoring with phenotypic analysis may provide opportunities for complementary combination treatments. Deep transfer learning with patient-by-patient fine-tuning is a promising, segmentation-free image-analysis approach for precision medicine and drug discovery.**

***Index Terms*— Molecular and cellular imaging, microscopy, pattern recognition and classification, neural network, spheroid, phenotypic drug profiling, translation.**

I. Introduction

FUNCTIONAL drug testing is an important step in drug discovery and drug repurposing [1], [2], [3], [4] which also receives increasing attention in precision medicine [5], [6], [7]. In particular, testing of primary patient-derived material using high-throughput technologies for cell seeding, drug treatment, and readouts is a promising approach to further improving treatment outcomes [6], [7], [8], [9], [10]. Functional drug profiling appears to be an obvious addition to precision-medicine strategies based on next-generation sequencing (NGS)[1] [11], providing confirmatory or complementary information on functionally active disease pathways as well as quantitative information on sensitivity to drug treatment such as dose-response curves (DRCs) and metrics. However, the technical set up of and, more so, data analysis in functional drug profiling remain challenging. We address four challenges through image-based drug profiling [12] using a novel, data-driven approach to image quantification.

A. Four Challenges

First, the current trend towards using three-dimensional (3D) cultures with increased tissue complexity (including co-cultures and organoid models) for drug testing in precision medicine likely mandates high-content analysis instead of scalar readouts [13]. In fact, it has already been shown using two-dimensional (2D) cultures that image-based phenotypic drug profiling can provide much higher-dimensional information [14] than classical end-point readouts, such as on single cells, cell/cell interactions, and tissue architecture, all of which may impact drug response [15], [16], [17], [18]. Specifically, drug responses of hypoxic and euoxic cells may differ significantly, motivating the use of 3D spheroids large enough to

Manuscript received 17 June 2022; accepted 23 August 2022. Date of publication 13 September 2022; date of current version 2 December 2022. This work was supported in part by the Federal Ministry of Education and Research (BMBF) under the frame of ERA PerMed (COMPASS, ERAPERMED2018-121) under Grant 01KU1903; in part by KirstinsWeg – Verein zur Förderung der Krebsmedizin e.V. and the Kirstin Diehl-Stiftung; and in part by the German Cancer Consortium (DKTK), BMBF, and the Scheu Family. The work of Ina Oehme was supported by the Deutsche Krebshilfe under Grant 70113843. *(Corresponding author: Sina Oppermann.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the Medical Faculty of Heidelberg under Approval No. S-502/2013, and performed in line with the Declaration of Helsinki.

Please see the Acknowledgment section of this article for the author affiliations.

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2022.3205554, provided by the authors.

Digital Object Identifier 10.1109/TMI.2022.3205554

[1] See Table S-I for a list of abbreviations (figures and tables with *S-* prefix refer to content in the unreviewed supplemental document on *IEEE Xplore*).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

establish hypoxic cores in precision oncology directed at solid tumors.

Image-based phenotypic features such as cell, nucleus and organoid texture or morphology enable sophisticated functional interpretation (e.g., mitochondrial outer membrane potential as an indicator for cell health). In addition, imaging not only enables advanced quality control (QC) and better suppression of background signals, but also, well or tissue-sample contaminations and undesired tissue-sample properties (such as polyclonal tumor origins or blood and other non-tumor cells in primary tumor tissue) can be accounted for [19]. Furthermore, using adequate controls, various drug mechanisms can be differentiated based on phenotypic features: this includes cell death (apoptosis, autophagy, stress) as well as effects on cell-cycle phases or organoid structure and function.

Second, advances in automated microcroscopy and decreasing costs for image storage have enabled rapid imaging of many treatment conditions. In fact, in each screen within this study, 75 drugs are tested on a single tissue sample within only 3 h image-acquisition time for one dye. In addition, as spheroid cultures of solid tumor material require imaging of several planes to capture the entire 3D structure, large amounts of imaging data are being acquired (dozens of image planes, scores of drugs, a concentrations series, and replicates, totaling hundreds of GB) for each tissue sample. In essence, large amounts of data mandate fully-automated image analysis.

Third, inter-sample heterogeneity of primary tissue samples implies that automated image analysis may require adaptation to each patient's tissue sample (e.g., when they represent different tumor entities). The often limited amount of patient material, and consequently low numbers of training controls, render automated adaptation challenging, and inter-plate variance may intensify this challenge – preventing turn-around within time frames compatible within clinical translation.

Fourth, and finally, classical image analysis is complicated by the fact that single-cell segmentation is still challenging in 3D spheroids. Classical image analysis usually combines image segmentation and feature extraction with supervised or unsupervised machine learning for feature selection [20], [21], [22]. In this setting, image features are extracted from a-priori (or iteratively determined [23]) regions of interest, such as whole images or supervoxels, segmented spheroids, or single cells. Unfortunately, each of these alternatives comes with its own challenges, especially when applied to 3D cultures: first, per-well features (based on whole images) may provide little additional information over metabolic readouts and are sensitive to well contaminations (Fig. 1, green arrow). Second, spheroid segmentation requires a spheroid to be established in the first place, which cannot be guaranteed for all tissue samples or drugs (in particular, those effecting cell viablity or spheroid formation). And third, while routine in 2D cultures, instance segmentation of single cells (or cell nuclei) from 3D fluorescence microscopy images is still challenging due to the tight packing of cells, resulting in insufficient spatial separation especially along the optical axis. This usually requires time-consuming fine-tuning of classical image-segmentation algorithms, or creation of large amounts of manual or semi-manual annotations for deep-learning segmentation methods,

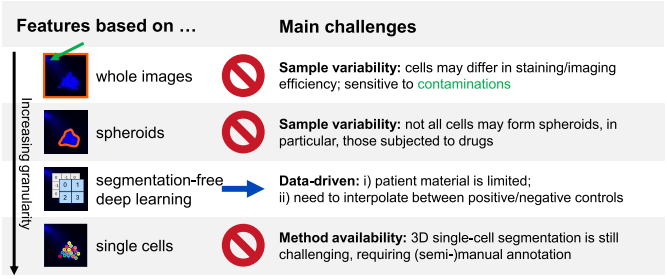

Fig. 1. Challenges met in 3D image analysis based on different approaches to feature extraction. Using features based on *whole images* or *spheroids* faces roadblocks such as contaminations (green arrow) or differences in spheroid formation (compare Fig. 2) and staining efficiency, in particular, with tissue samples originating from various patients and entities. Compared to *segmentation-free deep-learning* image analysis, for which we see particular potential (blue arrow), *single cells* will yield most granular image information as soon as reliable 3D single-cell instance-segmentation methods are available, which is still a major roadblock.

and still does not guarantee that cells in cores of larger spheroids, which may be dead or dying in hypoxic cores or difficult to stain or image in others [24], are fully considered.

In this study, we propose image-based drug profiling for patient-derived material obtained from pediatric solid tumors, focusing on segmentation-free deep-learning image analysis. This approach not only allows rapid and automated analysis of drug screens leading to quantitative information on drug sensitivity, but also provides information on drug-induced cell-death phenotypes (CDPs) on a patient level. The method tackles the aforementioned challenges of limited patient-derived material, analysis of 3D spheroid cultures without single-cell seqmentation, and the limitation of mechanism-specific end-point measurements and hence can be used to confirm, correct and to complement classical end-point readouts to improve precision-medicine and drug-discovery programs.

B. Related Work

Considering the tremendous progress that deep learning has brought to many areas of image analysis, its application to drug profiling seems only logical. However, despite a vast number of reports in the literature, we are not aware of any approach tackling the complete set of challenges introduced in the previous section. Looking at examples for single-cell analyses, the recently reported Deep-SeSMo [25] approach uses a convolutional neural network (CNN) to compute senescence probabilities for images acquired from cells undergoing drug treatment, after training using images of positive (H_2O_2) and negative controls. Notable differences to our settings include the use of 2D cell cultures (and hence ease of single-cell segmentation) and the use of cell lines (and hence an abundance of cell material). Similarly, Tox_CNN [26], which predicts drug toxicity based on training with controls such as Caspase 3/7, uses 2D cell cultures and cell lines. More generally, the practicability of single-cell feature embeddings, which have been proposed for 2D microscopic imaging [27], remains to be studied in 3D imaging.

Regarding analyses of 3D cell cultures, counting living cells [28] can be seen as a classical way of tackling the lack

of single-cell segmentation methods. More importantly, deep learning has been proposed for segmentation of organoids [29] and spheroids [30] as well as for classification of spheroid polarity [31]: yet so far, issues such as low amounts and inter-patient variability of biopsy material (and hence training data) in a pediatric precision-oncology platform have not been addressed to our knowledge.

C. Proposed Approach

As a proof of concept, this study was applied to tissue samples derived from pediatric patients with solid tumors, enrolled in the INFORM registry study ("INdividualized Therapy FOr Relapsed Malignancies in Childhood"). We aim to show the potential of image-based drug profiling in (pediatric) precision oncology programs, here specifically, to overcome the need of finding adequate therapy options for relapsed patients circumventing potential resistances induced by previous treatments. Our work is part of the COMPASS multi-center study ("Clinical implementation Of Multidimensional PhenotypicAl drug SenSitivities in paediatric precision oncology") which includes pediatric brain-tumor, sarcoma and neuroblastoma patient-tissue samples from several European registry studies (namely, "INdividualized Therapy FOr Relapsed Malignancies in Childhood" (INFORM), "Individualized Therapies for Children with Relapsed/Refractory Malignancies using Molecular Profiling" (iTHER), and "MoleculAr Profiling for Pediatric and Young Adult Cancer Treatment Stratification" (MAPPYACTS)) for functional drug profiling using metabolic and image-based readout techniques to quantify inhibition of cell viability.

To tackle all aforementioned challenges, we investigate patient-by-patient (in particular, screen-by-screen) deep transfer learning as an easily implementable solution for image-based drug profiling without complex image preprocessing or segmentation. Deep learning has produced outstanding results both in image analysis [35] and beyond, e.g., predicting drug response from genomic profiles and chemical structure [36]. In contrast to the three-step approach of image segmentation, feature extraction and machine learning, deep learning uses the available data to not only optimize selection of existing features, but also to guide extraction of optimized, potentially cell-phenotypic features without explicit single-cell segmentation [37]. Given sufficient training data, this end-to-end learning process can be carried out for each individual tissue sample.

Two main challenges of this end-to-end data-driven approach need to be addressed: the aforementioned limited patient-derived material and low number of assay controls, and the need to interpolate between positive and negative assay controls to compute drug-sensitivity scores (DSSs) over the tested drug concentration range. Transfer learning allows addressing the limited amount of screen-specific assay controls through pre-training with additional controls derived from more readily available cell lines, as well as from natural images; while label smoothing improves the interpolation between controls. Thus, the main technical novelty of our work is a training scheme using multiple stages of pre-training to account for low amounts of training data, including the use of ImageNet-trained network weights, pretraining using a cell line (Phase I), patient-specific fine-tuning (Phase II), and optionally employing images from technical replicates for training of a maximally robust CNN.

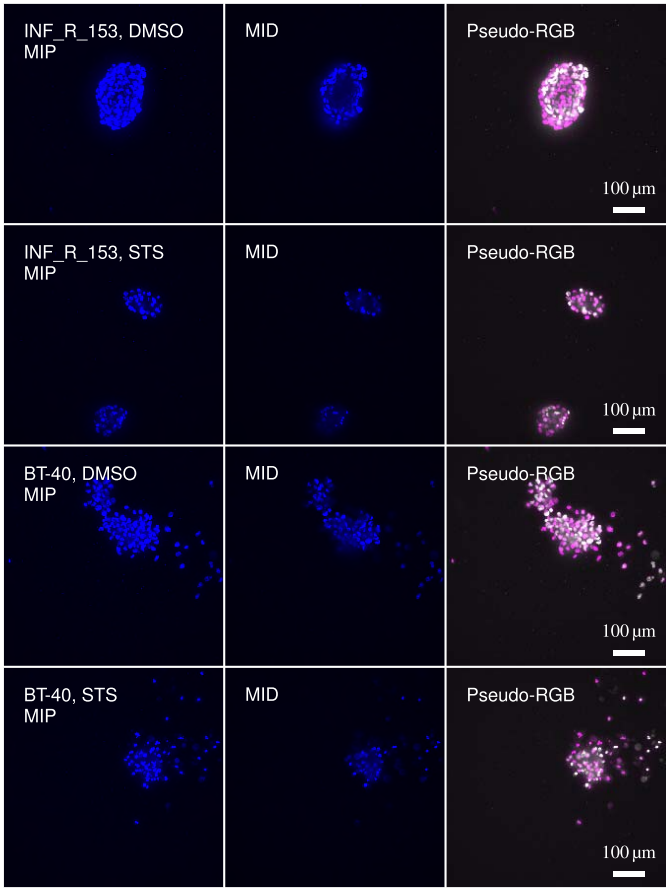

Fig. 2. Wavelength-colored MIP (left) and MID images (middle), and pseudo-RGB-colored MIP-MID-MIP composite images (right) of two tumor cell lines at 224 × 224 resolution before channel-wise normalization of image intensity. In the RGB image, white color indicates MID/MIP intensity, while pink color indicates MIP-only intensity (above or below the middle slice). These negative (DMSO) and positive control (STS) images for two cell lines illustrate differing affinities of cells to forming compact, potentially hypoxic-core spheroids, complicating application of classical image-analysis criteria to a wide range of cell lines and patient-derived cells. The INF_R_153 cell line confirms the targeted spheroid size range of 100–200 μm. While not part of routine screening of PDSs, Cas-3/7 staining in the control experiment with the INF_R_153 cell line resulted in some signal intensity (Fig. S-1), indicating a mix of either dead or hard-to-image cells in the inner core of the spheroid. For the BzCl positive control, see Fig. S-8.

To our knowledge, this is the first study applying such screen-by-screen deep transfer learning to image-based drug profiling. In the following, we present our proposed image-analysis approach using CNNs (Section II), the image data (Section III), and evaluation methods and results using image-based and metabolic readouts (Section IV). The significance of the results is discussed in Section V.

II. Image-Based Phenotypic Drug Profiling

Here, we describe the processing part of the proposed image-based drug-profiling pipeline. In summary, a pre-trained CNN was fine-tuned using control images generated using

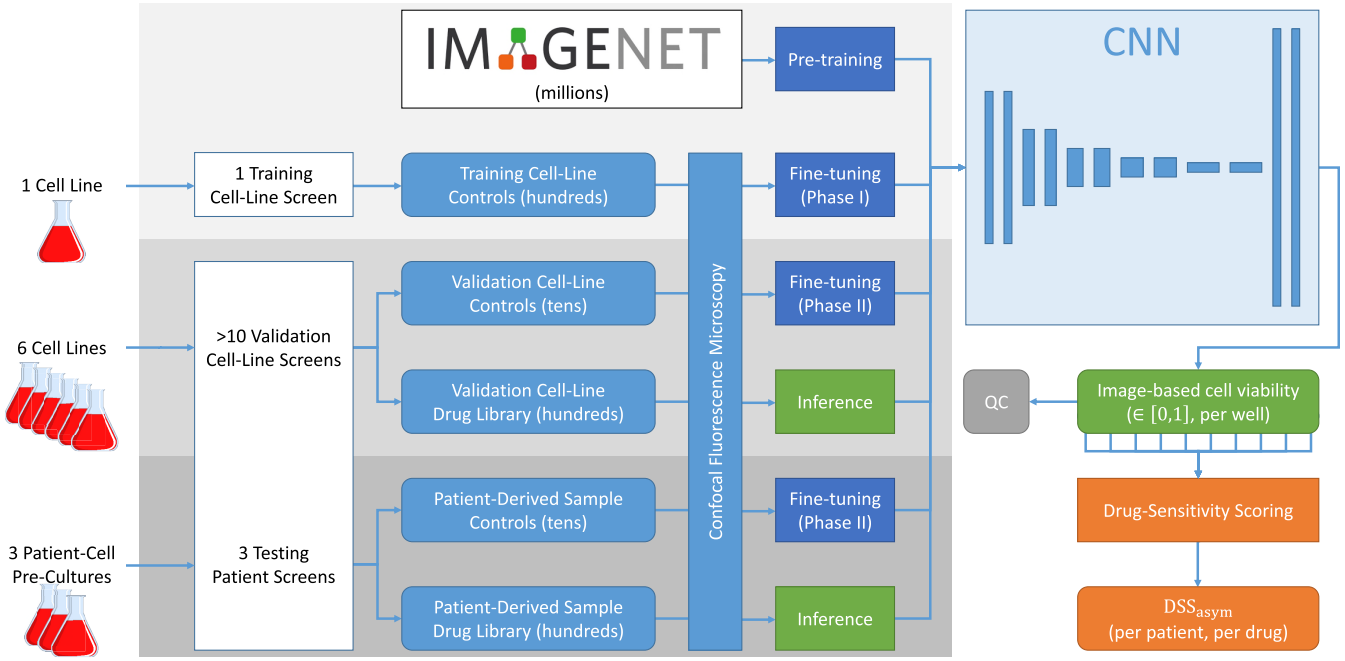

Fig. 3. A CNN pre-trained using the ImageNet database is fine-tuned using positive and negative controls from a controls-only training cell-line screen (top). For each screen of interest (validation cell-line screens, middle, and testing patient screens, bottom), in-plate controls are used to further fine-tune the network. Image-based cell viabilities are then inferred for each plate well holding library drugs, and following QC, combined in the form of one drug-response curve and summarized in one asymmetric DSS (see Section II-C) per patient and drug.

a cell line to input (preprocessed) images and output cell viability; further fine-tuned, for each screen, using screen-specific control images; and finally, applied to all of a screen's images.

A. Image Preprocessing

Available image data comprised one 3D image stack per well, consisting of 20–40 individual 2D image planes and a 2D maximum intensity projection (MIP) image.

1) Image Intensity Normalization: All used 2D images were downsampled from 2048 × 2048 to 224 × 224 pixels, meanwhile converting 16-bit integer image intensities to floating-point numbers $\in [0, 1]$. Maximum image intensities rarely exceeded 0.5, and an offset of 0.5 was added to obtain mean image intensities comparable to natural images, compensating for the dominance of low-intensity background pixels.

2) Pseudo-RGB Image Composition: For use of pre-trained established CNNs which input 2D red-green-blue (RGB) images, we composed 2D pseudo-RGB images for each well (Fig. 2). Therefore, after initial experimentation, we arranged the MIP and the mid-z plane (MID) image as three channels in the order MIP-MID-MIP.

3) Data Augmentation: During fine-tuning of the CNN, random data augmentation was applied, namely horizontal and vertical flips, image rotation (up to 30°), image shifts (up to 10%), image scaling ($\in [0.9, 1.1]$) and intensity scaling ($\in [0.5, 1.5]$). In particular, as we noticed that mean of stack (MoS) values can be used to separate controls in the control experiment, but performed poorly in the drug screens, we used intensity scaling to prevent the CNN from focusing on mean image intensity and thus to maximize generalization.

B. Neural Network Training

CNNs were trained in four steps: pre-training, phase-I and phase-II fine-tuning (Fig. 3), and optional phase-III robustness training (Fig. 4). We used a Linux workstation with a 10-core i9-7900X (Intel) processor, 32GB random-access memory (RAM), and a Quadro RTX 6000 (NVIDIA) graphics processing unit (GPU). All CNN computations were run in mixed precision using TensorFlow v2.8 and its Keras frontend in Python v3.10.

1) Image Labels: To quantify *image-based* cell viability, we assigned values of 1 and 0 to negative and positive cell-death controls, respectively. We used dimethyl sulfoxide (DMSO) and staurosporine (STS), respectively—see Section III-C.3 for details.

2) Network Architecture and Pre-Training: After comparing the performance of the VGG16 and the VGG19 image-classification networks [38], due to their large capacity and ease of availability, we chose VGG19 for its slightly improved performance on average. We used a pre-trained CNN by loading the network bottom[2] with weights pre-trained on the ImageNet image database [39] from the TensorFlow Keras *Application* module (TFKA), and added the network top[2] from the original architecture with random initialization. As in binary classification, the dense output layer had 2 outputs (one-hot encoding followed by softmax activation), of which we interpreted the first as *raw* image-based cell viability. This network has the same structure as in [38, Tables 1D and 2]

[2]In line with TFKA notation, in particular, its `include_top` parameter, we use the terms network *bottom* for its—usually convolutional—layers close to the input of the network, and network *top* for the remaining—usually fully-connected (dense)—and output layers, respectively.

Fig. 4. Illustrating Phase-II (left) and phase-III training (right) using percentage cell viabiltities. While phase II splits controls (−, +) between training and validation, phase III trains using all available controls as well as cell viabilities of duplicates inferred in phase II; the latter are exchanged between duplicates D^A and D^B (at the same concentration) of some drug D. Before training, label smoothing is applied (only) to controls; cell viabilities of drug wells are re-used directly. After inference, labels are un-smoothed by normalization in the iTReX web app [40], after which raw cell viabilitites 0.80, 0.62 ,0.59, 0.56, 0.50 and 0.20 correspond to percentage cell viabilities 100, 70, 65, 60, 50 and 0%, respectively. Colors indicate different groups of wells (green: negative controls; red: positive controls; blue, orange: two duplicates of drug D), while arrows indicate the way cell viabilities are used (◀: training label backpropagated into the network; ▶: value inferred by the network; ⇔: validation label).

except for the final dense layer (FC-2 instead of FC-1000) and hence the number of parameters (134 M instead of 138 M).

3) Phase-I Global Fine-Tuning Using Cell-Line Controls: The CNN was then fine-tuned using control images acquired from a cell line [88 and 102 wells for training; 2 × 10 wells for validation (Section III-D)]. The CNN globally fine-tuned in this way was used as a basis for further fine-tuning (Fig. 3).

We used an Adam optimizer with a learning rate of 1×10^{-7} to minimize the categorical cross-entropy loss function with a batch size of 1 (compare Section V-A.2) over 200 epochs, where the chosen learning rate was found heuristically based on the validation-loss curve's ability to approach the optimum, and the number of epochs was set to ensure convergence of said curve for most cell lines (compare Section IV-A). Importantly, we used label smoothing (using a heuristically determined value $\alpha = 0.4$) to prevent the network from becoming overly confident, as we prepared the network to encounter images between positive and negative controls for which it should infer meaningful intermediate values.

4) Phase-II Screen-Wise Fine-Tuning Using Screen Controls: For each screen independently, the globally fine-tuned CNN was fine-tuned further using the screen-specific controls across all 3 microplates (3 × 7 STS and 3 × 9 DMSO for training, 3 × 1 STS and DMSO for validation, respectively). Training parameters were the same as in phase I except for the learning rate (3×10^{-7}).

5) Optional Phase-III Screen-Wise Robustness Training: A potential weakness of the proposed method is low robustness, as measured by the reproducibility between duplicates (compare Section IV-C.2), supposedly due to using only negative and positive controls during network training. Considering also the low number of these controls, it may be advantageous to add drug wells to the training data. While these images are unlabeled a-priori, we are in the unique position of a) knowing that we expect their viabilities to lie between those of positive and negative controls, b) having duplicates available for all drugs and concentrations, and c) knowing that duplicate metabolic cell viabilities are highly correlated.

TABLE I
NUMBERS OF WELLS AND EFFECTIVE RAW CELL VIABILITY VALUES (AFTER LABEL SMOOTHING) OF SEVERAL WELL TYPES IN ALL PHASES

Phase	Operation	Wells	Raw cell viabilities
I	Training	190 control wells (88 neg, 102 pos)	88×0.80, 102×0.20
	Validation	20 control wells (10 neg, 10 pos)	10×0.80, 10×0.20
II	Training	48 control wells (27 neg, 21 pos)	27×0.80, 21×0.20
	Validation	6 control wells (3 neg, 3 pos)	3×0.80, 3×0.20
	Inference	750 drug wells (375 pairs of duplicates)	e.g., (0.62, 0.50) [for one example pair], …
III	Training	54 control wells (30 neg, 24 pos)	30×0.80, 24×0.20
		750 drug wells (375 pairs of duplicates)	e.g., (0.50, 0.62) [from the example pair in II], …
	Inference	750 drug wells (375 pairs of duplicates)	e.g., (0.59, 0.56) [for the same example pair], …

Having duplicates is a similar situation as in the SimCLR framework for contrastive learning [41], which uses two-fold, independent data augmentation on individual unlabeled images to generate image pairs expected to generate similar representations. However, instead of contrasting between all unlabeled images and thereby risking the introduction of unwanted differences in representations of similar drugs, we group images with expected similar viabilities.

Here, we use the raw image-based cell viabilities inferred by the CNN after phase II as training labels for an optional phase-III robustness training. Since training a network with its own output usually results in little training progress, we exchange the labels between the duplicates in each pair (Fig. 4).

In addition, we skip validation and use all control images of the screen for training in this phase. The resulting numbers of training and validation images are summarized with their raw cell viabilities in Table I.

C. Inference and Drug Scoring

The CNNs fine-tuned using screen-specific controls (phases II and III, respectively) were subsequently applied to images acquired from all wells of that screen. Drug-sensitivity scoring used a prerelease version of the iTReX web application for drug-sensitivity scoring ([40], v1.1.1.9000, functionally identical to release v1.2.0), run in R v4.1. Due to label smoothing ($\alpha = 0.4$), raw image-based cell viabilities generally range around the interval $[0.2, 0.8]$ and were therefore normalized using positive (STS, set to 0%) and negative (DMSO, 100%) controls to yield *percentage* image-based cell viabilities.

For each drug, 10 percentage image-based cell viabilities (5 concentrations, duplicates) were converted to percentage inhibition values and used to fit a curve model based on the Hill equation (namely, a generalized 5-parameter logistic curve model in the log-concentration domain [42] having an asymmetry coefficient). The curve model was then integrated in the log-concentration domain to yield the which was then converted to an in-house-developed DSS

(DSS_{asym}, [40]) inspired by [43, Fig. 1], using an automated drug-scoring pipeline such as previously shown by others [43], [44], [45], [46].

Raw metabolic cell viabilities aquired after imaging (compare Section III-C.6, Fig. S-8) underwent similar normalization using readouts from benzethonium chloride (BzCl) and DMSO wells, followed by otherwise identical drug scoring.

III. Materials

Materials used in this work include a total of 17 screens of 6 cell lines (each screened between 1 and 4 times) and 3 patient-derived samples (PDSs) (screened once). All cells underwent our standard 3D spheroid drug-profiling assay including fluorescence microscopy (*imaging*; inspired by [8] and described in detail in Section III-C.5) followed by luminescence-based *metabolic* readouts (compare Section III-C.6 and [40], [47]); one cell line underwent additional imaging (Section III-D) of assay controls in larger quantities for CNN training (phase I).

A. Cell Lines

Details of 6 cell lines used are summarized in Table II. The INF_R_153 cell line (established, growing in fetal bovine serum (FBS), passage 40+) was generated from a primary tumor biopsy obtained from a soft tissue sarcoma patient enrolled in the INFORM registry study.

Cells were cultured according to cell-line-specific culture protocols largely identical to those described in [40]. The BT-40, HD-MB03, and INF_R_153 cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium containing L-glutamine (Gibco, Life Technologies, Thermo Fisher Scientific) supplemented with 10% FBS and 1× non-essential amino acid (NEAA) solution. The NCI-H3122 cells were cultured in RPMI 1640 medium (American Type Culture Collection) supplemented with 10% FBS and 2% penicillin-streptomycin (P/S). The SJ-GBM2 and SMS-KCNR cells were cultured in high-glucose Dulbecco's Modified Eagle's medium (DMEM), supplemented with 10% FBS and 1× NEAA.

TABLE II
Overview of Cell Lines and Patient Samples Used

ID§	Entity & Characterization
INF_R_153	Inflammatory myofibroblastic tumor (IMT) • ETV6-NTRK3 fusion • Expected to be sensitive to TRKi [48] – TRKi: Entrectinib, Larotrectinib, Merestinib
BT-40	Pleomorphic xanthoastrocytoma (PXA) • $BRAF^{V600E}$ mutation, CDKN2A/B loss • Sensitive to BRAFi, MEKi [49] – BRAFi: Vemurafenib, Dabrafenib – MEKi: Trametinib, Cobimetinib, Selumetinib
HD-MB03 (CVCL_S506)	Medulloblastoma group 3 (MB_G3) • MYC amplification • Sensitive to HDACi [50], PLKi [51] – HDACi: Entinostat, Vorinostat†, Panobinostat‡
NCI-H3122 (CVCL_5160)	Lung adenocarcinoma (LUAD) • EML4-ALK fusion • Sensitive to ALKi [52] – ALKi: Lorlatinib
SJ-GBM2 (CVCL_M141)	Glioblastoma (GBM), H3 wild type • MET fusion, MYC amplification, TP53 mutation • Expected to be sensitive to METi, HDACi – METi: Cabozantinib, Foretinib – HDACi: Entinostat, Vorinostat†, Panobinostat‡
SMS-KCNR (CVCL_7134)	Neuroblastoma (NBL) • ALK-F1174L mutation, MYCN amplification, BCL2 overexpression • Sensitive to ALKi [53], [54], BCL2i [55] – ALKi: Lorlatinib – BCL2i: Venetoclax
INF_R_1021 _relapse1_V1	Cortical brain metastasis of clear cell sarcoma (CCS)
INF_R_1025 _primary_V2	Fibromatosis-like soft-tissue sarcoma (STS)
INF_R_1123 _primary_V1	High-grade neuroepithelial tumor with BCOR alteration (HGNET-BCOR), PLAGL2 amplification

§ in parentheses: research resource identifier (RRID)
† modest clinical efficacy, ‡ high non-specific toxicity.

B. Patient-Derived Tissue Samples

After implementation of the method in cell lines, PDSs were used for additional validation. Tumor biopsies collected during surgical tumor resections in pediatric and young adult (<21 yrs.) patients enrolled in the INFORM registry study (German Clinical Trials Register ID: DRKS00007623) were submitted by the respective Society for Paediatric Oncology and Haematology (GPOH) study centers for functional drug testing within 24 h of surgery.

Written informed consent covering the use of tumor material for drug profiling and drug profiling analysis was obtained by the local study centers. Ethics committee approval for performing drug profiling, the use of consent forms and scientific evaluation of the data were obtained from the Ethics Committee of the Medical Faculty of Heidelberg (S-502/2013).

PDSs from tumor biopsies, of which we selected 3 for this work based on good metabolic QC and strict adherence to the requirements of the imaging protocol (Table II), were dissociated according to entity-specific protocols: tumor-tissue samples INF_R_1021_relapse1_V1 and INF_R_1123_primary_V1 were dissociated according to a protocol adapted from [56] and pre-cultured as short-term cultures under defined serum-free conditions in complete tumor stem medium (TSM) with 4 mM L-glutamine and 2% penicillin-streptomycin up to 7 d before drug screening; tumor-tissue sample INF_R_1025_primary_V2 was dissociated by the group of E. Koscielniak, cultured in RPMI, 10% FBS, and 1% NEAA, and screened more than 6 weeks after surgery.

C. Drug-Profiling Assay

1) Nomenclature: Over the course of several months, all cell lines in Table II underwent one or multiple rounds of metabolic screening in the framework of the COMPASS project at our

institution, and we included all screens for which imaging data were also acquired. In screen identifiers, separate rounds of screening are indicated by _V1, _V2 suffixes, while _DS1, _DS2 suffixes indicate repeated drug screens within the same round. For PDSs, _V1, _V2 suffixes indicate *V*ital cells followed by an order number designating different deliveries of samples from the INFORM study center, while _DS1, _DS2 indicate different drug screens of the same delivered tissue sample.

2) Drug-Microplate Sets: Each drug in a library of 75–76 oncology drugs (Table S-II) was profiled in duplicates, using 5 concentrations in a 10-fold serial dilution. The drug library and additional assay controls were pre-printed (High Throughput Biomedicine core unit, Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland) onto sets of 3 round-bottom 384-well microplates (3830, Corning), each set being designed to be used for one screen [57].

3) Assay Controls: Each plate set contained positive as well as negative assay controls: positive controls comprise 3 × 8 wells with 250 nM of the pan-kinase inhibitor STS as a 0%-cell-viability control for all image-based measurements. The advantage of STS lies in its numerous cell-death mechanisms [58], [59] which explains its common use as a positive control for cytotoxic and anti-proliferative effects on cancer cell lines and primary cancer cells [8]. 3 × 5 wells with 100 μM BzCl were used as a 0%-cell-viability control for metabolic measurements. Negative controls include 3 × 10 wells with DMSO, set to 100% cell viability (despite basal levels of cell death) for image- and metabolics-based readouts, and additionally untreated cells in medium (3 × 15 wells mostly in top and bottom rows, and an extra 30 wells in the third plate).

4) Cell Seeding: We aimed for a spheroid size of 100–200 μm to allow imaging dyes (and light) to penetrate into (and out of) the spheroid. 25μL cell suspension (20 000 and 40 000 cells/mL for cell lines and PDSs, respectively) were seeded using automated liquid handling (Multidrop Combi Reagent Dispenser, Thermo Fisher Scientific) onto the pre-printed drug plates and incubated at 37 °C and 5% CO_2 for 72 h before image-based and metabolic readouts of the same plates. The first and last columns (32wells) were left empty (i.e., wells were filled with medium, but without cells).

5) Image-Based Readout: Cells were stained with Hoechst 33342 (H3570, Thermo Fisher Scientific) dye following a no wash-protocol (final dilution factor, 1:2000). Plates were incubated for 30 min after staining. Image stacks were acquired at 37 °C using an ImageXpress Micro Confocal (Molecular Devices, San Jose, CA) high-content microscope equipped with light-emitting diode (LED) excitation, a 60 μm pinhole disk, a CFI Plan Apo Lambda 20X (Nikon) objective (numerical aperture, 0.75), a DAPI filter set (excitation, 377 nm; emission, 447 nm) and laser-assisted focusing. With an exposure time of 50 ms, Hoechst image acquisition took approximately 60 min per plate and resulted in 100 GB of data stored as 16-bit unsigned integers in uncompressed tagged image file format (TIFF).

We imaged a single, central site in all inner 22 × 14 wells. Images comprised 2048 × 2048 pixels, with a calibrated pixel size of 0.343 μm, and each stack contained 20–40 images with a distance along z of 2.5 μm.

6) Metabolic Readout: After imaging, a metabolic luminescence readout was carried out in the same plates using CellTiter-Glo 2.0 (CTG) and either a FLUOstar OPTIMA (BMG LABTECH) or a Spark multimode (Tecan) microplate reader as described before [40]. These measurements are referred to as raw *metabolic* cell viabilities.

D. Cell-Line Control Experiment

We acquired one 384-well microplate with negative (DMSO, 98 wells) and positive controls (250 nM STS, 112 wells) holding the INF_R_153 cell line (500 cells/well) for phase-I CNN fine-tuning. This control experiment (INF_R_153_CE) was carried out 4 mos before the first drug screens of the same cell line, allowing for potential changes of the assay over time (such as through using different batches of plates or storage of preprinted drug plates).

IV. Evaluation Methods & Results

We evaluated the proposed approach using raw cell viabilities, QC metrics and drug-sensitivity scores, including data from metabolic screens and the proposed CNN-based approach after phase II and III, respectively. In addition, we considered a naïve image-based approach averaging image intensities across each well's 3D image stack (mean of stack, MoS).

A. Network Training

Network training performance metrics (categorical cross-entropy loss, accuracy, and area under the receiver operator curve (AUROC)) for cell-line and PDS screens are detailed in Table S-III. Of note, the minimum achievable loss is 0.5004024235381879 due to label smoothing ($\alpha = 0.4$). Training accuracy is close to ideal (1.0) except for BT-40_V3_DS2 and INF_R_1123 (1 error among 48 training images, respectively); AUROC values are $>= 0.99$. Validation accuracy is 6/6 for 11 screens, 5/6 for 5 screens, and 4/6 for the remaining screen. AUROC values are $>= 0.97$ for all but two images (BT-40_V3_DS1 and BT-40_V3_DS2). Training times amount to approximately 45 min, 15 min, and 2.5–3 h in CNN phases I, II and III, respectively.

To demonstrate the added value of transfer learning, we repeated Phase-II training for the INF_R_153_V2_DS1 drug screen while skipping different parts of pre-training. Results shown in Fig. 5 indicate the importance of carrying out both Phase-I and Phase-II training, as fewer Phase-I iterations lead to higher Phase-II validation losses than cannot be compensated for by longer Phase-II training.

B. Quality Control

We quantified assay quality (using high-throughput screening (HTS) QC metrics such as the Z'-factor [60], where $Z' \in (-\infty, 1]$, and assays with $Z' > 0.5$ are usually considered excellent) and reproducibility (using correlation coefficients between duplicate wells). Robust computation was used for Z'-factors by using median instead of mean values, and

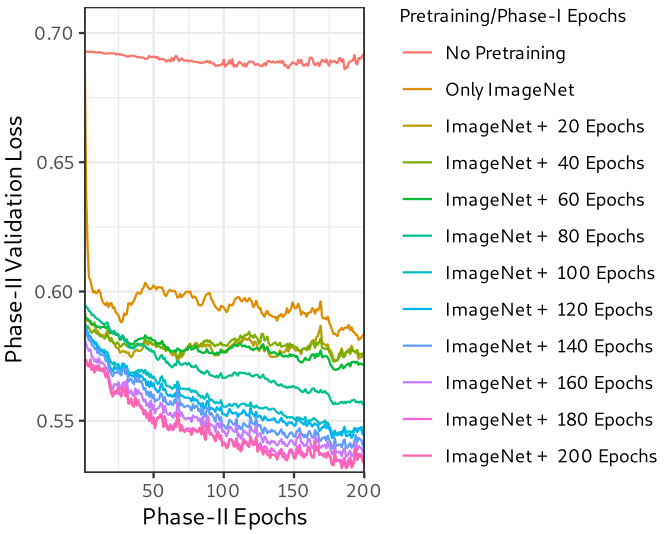

Fig. 5. Phase-II validation loss (lower is better) for INF_R_153_V2_DS1 as a function of Phase-II epochs and various pretraining configurations (using ImageNet and INF_R_153_CE). Longer Phase-I pretraining consistently improves the validation loss achievable in Phase II, indicating the importance of transfer learning.

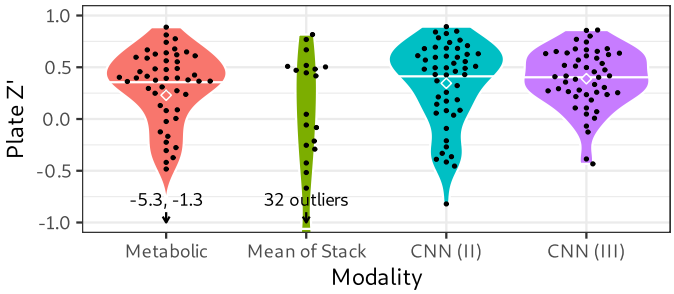

Fig. 6. Z'-factors for four methods across all screens (3 plates per screen, cell-line as well as patient screens). White diamonds, mean values of data points; white lines, violin medians; bottom annotations, number or numerical values of outliers.

mean absolute deviations instead of standard deviations. Both metrics were compared to the same metrics in the metabolic assay using the same plates.

Per-plate Z'-factors are summarized in Fig. 6, indicating comparable QC metrics for metabolic and CNN-based quantification, compared to drastically lower values for MoS (mean and median Z', < -1.0). Table S-III also shows individual Z'-factors for phase II: negative Z'-factors for 9/51 plates in most cases coincide with non-ideal (<1.00) training accuracy (1/9 plates) or validation accuracy (5/9 plates): this is in line with the facts that a) Z' values are computed from cell viabilities of all (training and validation) control images, and b) Z' values are particularly prone to misclassification errors, in the sense that a perfect Z' value computed from classification-based cell viabilities of 10 positive and 8 negative control images can become negative under the misclassification of a single control image.

C. Well Quantification

1) Uniformity: For drug-treated wells, to ensure the desired smoothness of DRCs (as opposed to binary classification of

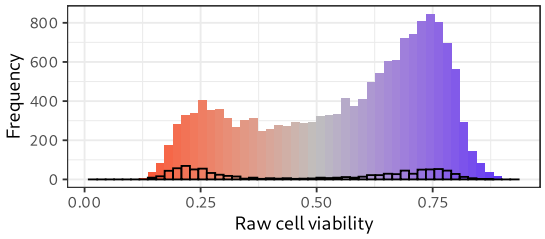

Fig. 7. Red to blue: image-based raw cell viabilities of all wells after CNN phase II. Black: control wells only, with values expected at 0.2 (STS) and 0.8 (DMSO) due to label smoothing.

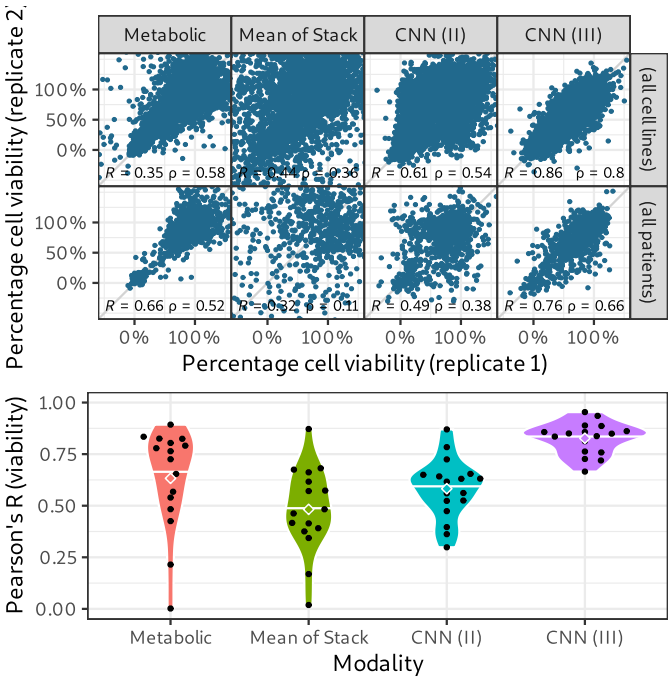

Fig. 8. Top: percentage cell viabilities estimated for duplicate drug wells using four methods, summarized for cell-line and patient screens. R, Pearson's correlation coefficient; ρ, Spearman's rank correlation coefficient; per-screen results in Fig. S-2. Bottom: Pearson's R per screen.

drug wells into positive or negative), we verified that network outputs (across all drugs and concentrations) were distributed uniformly in the range between negative and positive controls. Fig. 7 confirms a continuous distribution of CNN-based raw cell viabilities for controls and drugs after CNN phase II.

2) Reproducibility: Complementary to Z'-factors, Fig. 8 summarizes the reproducibility of network outputs for drug wells within each plate (excluding any control wells) by scatter plots as well as the Pearson correlation coefficents. Scatter plots (Fig. 8, top) show that metabolic data is more variable for high cell viabilities, indicating a noise amplitude proportional to the luminescence readout. Smaller extensions of MoS point clouds indicate high background image intensity that can lead to mean values similar or larger than standard deviations. By contrast, CNN phase II is highly robust for both high and low cell viabilities (close to the training controls), showing high variability for intermediate cell viabilities. CNN phase III is able to reduce variability in the intermediate range in particular.

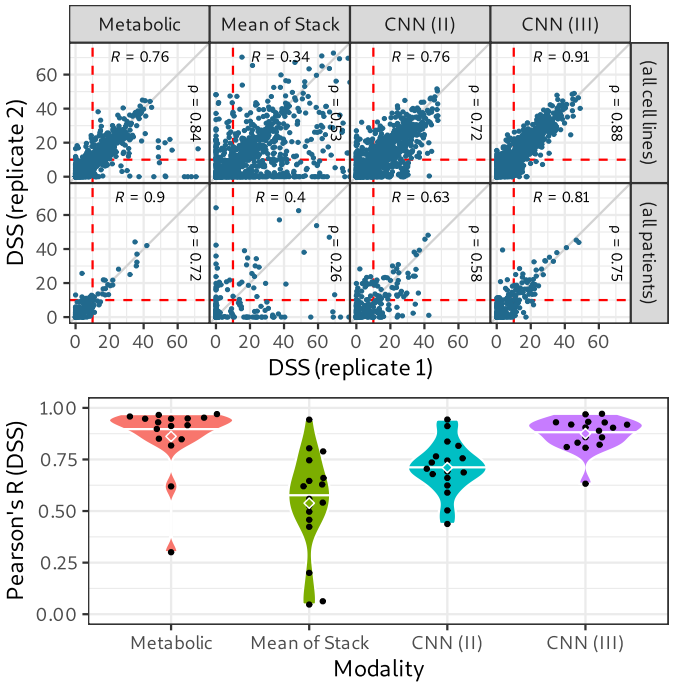

Fig. 9. Top: image-based DSSs computed using duplicate dilution series (5 conc. each) based on cell viabilities in Fig. 8. R, ρ as in Fig. 8; per-screen results in Fig. S-2. Bottom: Pearson's R per screen.

In line with Z'-factors, reproducibility as measured by correlation coefficients (Fig. 8, bottom) is worse for MoS than for metabolic data. Reproducibility for CNN phase II is also slightly lower, while that for CNN phase III is higher, confirming the effect of robustness training using duplicates.

D. Drug Scoring

1) Reproducibility: The trends observed for cell viabilities from different modalitities are confirmed by DSSs computed based on these cell viabilities (Fig. 9), with unusable MoS results and somewhat reduced (CNN phase II) and highest (CNN phase III) correlations between duplicates, compared to correlations within metabolic data.

2) Modalities: Heatmaps of DSSs across all modalitites, drugs, and screens indicate severe differences between MoS and all other modalities (Fig. 12). In particular, MoS yields a number of large outlier values compared to both metabolic and CNN-based results, and clustering of screens is also improved in CNN-based over MoS results.

3) Expected Hits: On a biological level, we verified that expected *hit* compounds (drugs expected to be efficient in killing cells according to their molecular characterization; see Table II for a list of molecular characteristics and related references) ranked high also according to their drug scores (using a DSS of 10 as a threshold). Image-based DSSs of expected hits in cell lines (Fig. 10) confirm most drug hits seen in metabolic data. Image-based DSSs appear to be slightly higher, which is not concerning considering different normalizations of image-based (STS) and metabolic (BzCl) cell viabilities.

4) Imaging–Metabolics Correlation: Finally, through scatter plots of image-based and metabolic-based drug scores, we compared image-based DSSs to their metabolic counter-parts

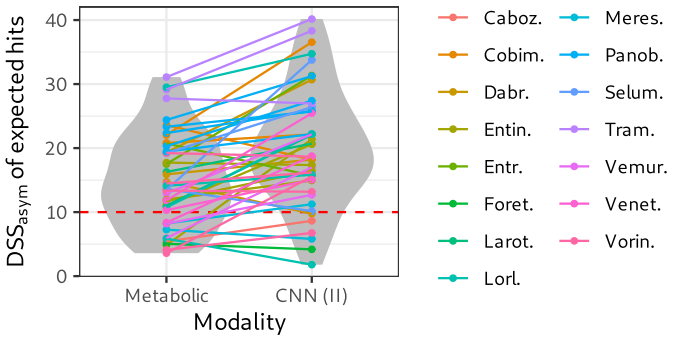

Fig. 10. DSSs of expected hits in all cell line screens for metabolic data as well as after CNN phase II. See Table S-IV for numerical results.

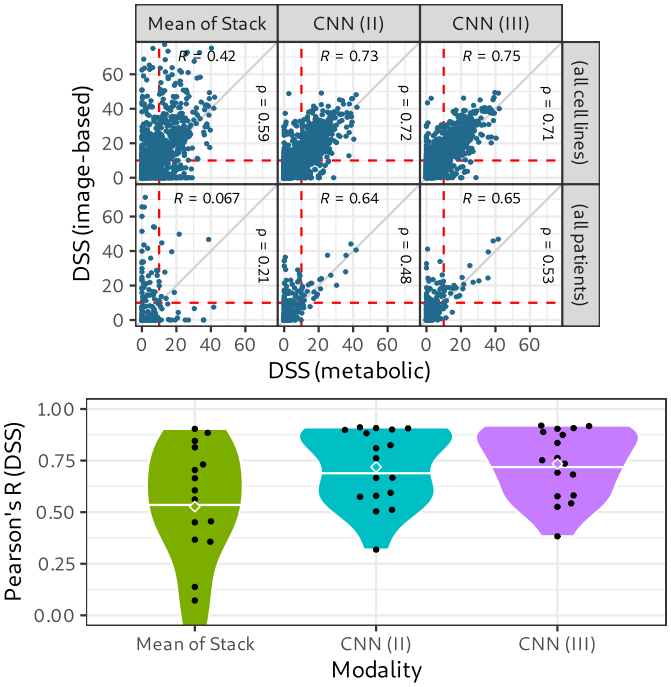

Fig. 11. Top: image-based vs. metabolic DSSs (using 2 × 5 conc. each), summarized for cell-line and patient screens. R, ρ as in Fig. 8; per-screen results in Fig. S-2. Bottom: Pearson's R per screen.

(Fig. 11). Note that contrary to earlier scatter plots, perfect correlation between metabolic and imaging results is neither intended nor expected. Nonetheless, CNN phases II and III behave similarly in terms of DSS, and with higher correlation to metabolics than MoS. Applying the hit threshold to image-based and metabolic DSSs, we identified mismatches (hits in one, but not the other) in 26% of scores for MoS, compared to only 20 and 21% for CNN phases II and III.

E. Beyond Metabolics: Imaging-Exclusive Benefits

1) Imaging–Metabolics Mismatches: While the output of metabolic readouts consist of only a single number per well, multi-dimensional imaging readouts enable additional information to be extracted from the assays, including those allowing to rule out certain mistakes in sample preparation such as missing cells, non-viable cells, lack of spheroid formation, lack of imaging dye, etc. We exemplify the benefit of having this information available for user interpretation through one

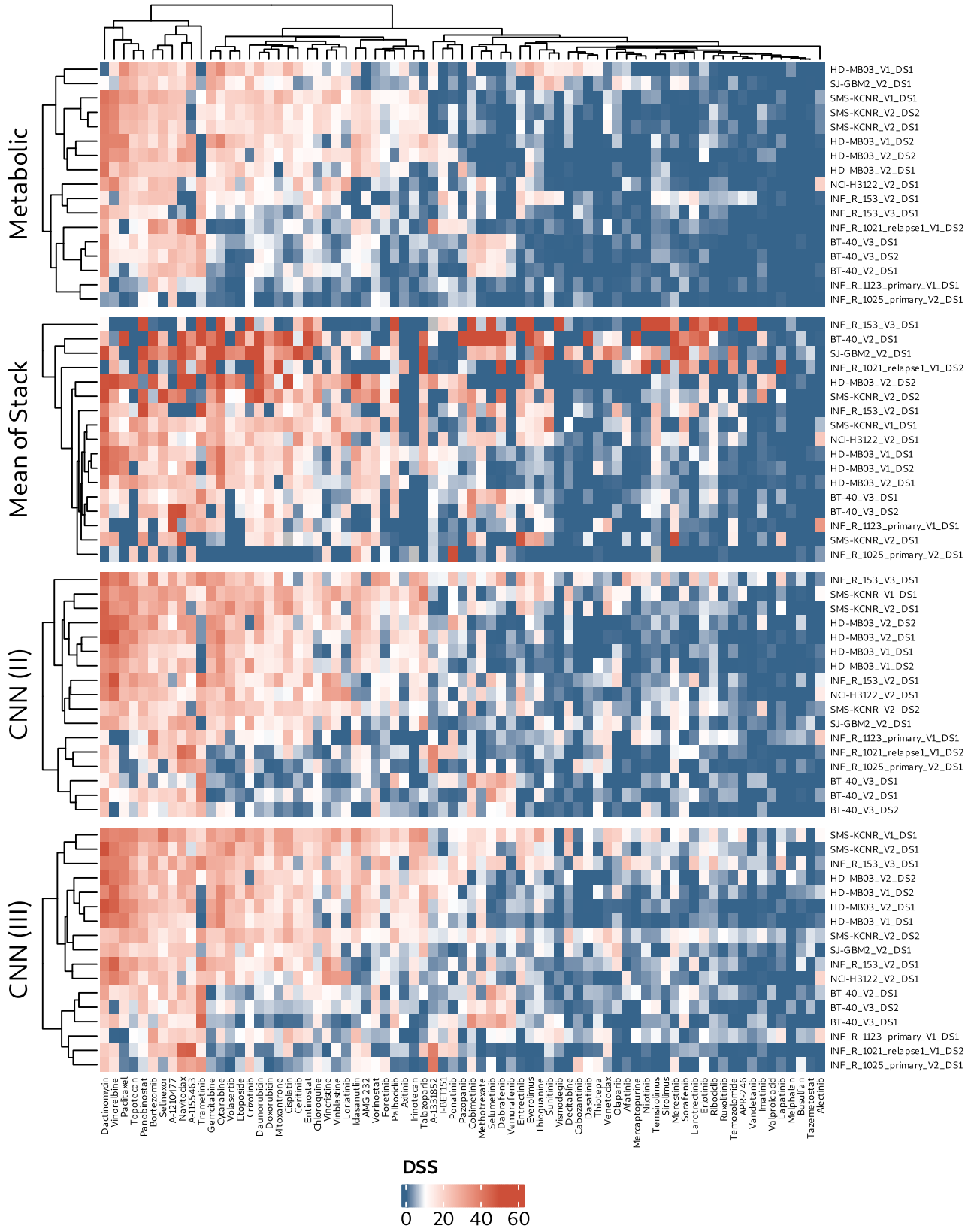

Fig. 12. Heatmaps of DSS values for all modalitites (vertical, labels on the left), drugs (vertical, labels on the right), and screens (horizontal).

case of mismatch between image-based and metabolics-based cell viabilities in the same well.

In an extreme case of imaging–metabolics mismatch (Vinorelbine in screen INF_R_153_V2_DS1, Fig. S-3), DRCs show almost maximum effect in imaging (DSS, 48.8), but almost no effect in metabolics (DSS, 2.7). Qualitative comparison of images suggests correctness of the image-based DRCs. An independent imaging–metabolics screen of the

same cell line (INF_R_153_V3_DS1; image-based DSS, 38.8; metabolics, 22.4) as well as metabolic screens carried out at 3 COMPASS partner sites, namely DKFZ (DSS, 28.7), Institut Curie (23.4), and FIMM (20.7), support the hypothesis that this drug's metabolic DSS value in INF_R_153_V2_DS1 is an outlier. In conclusion, while imaging–metabolics DSS differences persist (not unexpectedly, due to normalization to different controls), Vinorelbine is detected as a hit in all screens except one, in which quantitative and qualitative image analysis overrule metabolic findings.

2) Phenotypic Analysis: In addition, imaging information can be used for technical (detecting contaminations and empty wells; compare Fig. S-9) as well as biological purposes (identifying similarities and differences between CDPs). In this section, we visualize groups of drugs with similar CDPs of known drugs using CNN features, namely, 4096 layer activations of the penultimate fully-connected layers of the trained CNN after principal component analysis (PCA) to reduce the number of features to those representing at least 95% of the variance.

Figs. 13 and 14 and Figs. S-4 to S-7 show 2D t-SNE plots of the CNN layer activations after PCA. Features shown are for the highest concentration of each drug and control, for all three patients, and for phases II and III, respectively. Two distinct clusters can be observed in all cases, presumably representing non-effective and effective drugs, respectively, as indicated by the presence of untreated wells (u) and DMSO in the former and STS, BzCl and apoptotic modifiers in the latter. As images of BzCl (used primarily for metabolic analysis) and STS appear strikingly different, STS and BzCl wells tend to cluster in separate, disjoint sub-regions, while DMSO and untreated wells intermingle with each other and other non-effective drugs. It is notable how closely duplicates of many effective drugs cluster in phase II. In phase-III t-SNE plots (Fig. 14 and Figs. S-5 and S-7), we note the same proximity in replicates of ineffective drugs. Finally, in one example, Fig. 15, we show the t-SNE plot annotated with a selection of the respective microscopy images to visualize the different CDPs across the plot.

3) Explainability: Finally, Figs. S-8 and S-9 show three different class-activation maps (CAMs) for controls and a Foretinib dilution series, respectively. CAMs [61], [62], [63] illustrate that the CNN directs its attention towards the spheroid if one can be seen (Fig. S-8); in Fig. S-9, they visualize how the CNN is potentially confused by a well contamination for 10 nM of Foretinib.[3]

F. Proposed Patient Workflow

Fig. 16 summarizes analysis of a PDS analysis using a waterfall plot of drug scores to select top hits, and a t-SNE plot of CNN features to select complementary phenotypes for combinatorial treatments. Drug DRCs and images of drugs and controls (DMSO, STS) help visually confirm sensitivity and complementarity of this drug combination.

[3]It should be noted that while the VGG16 network, in an earlier version of this manuscript, returned an incorrectly low cell viability for this well, VGG19 was not impacted in the same way.

V. Discussion

In this work, we have achieved image-based quantification of cell viability using fully-automated CNN fine-tuning for individual patients, especially for the use with limited primary material. We have validated image-based cell viability using metabolic cell viability. In addition, image-based functional drug profiling offers unique potential to control, correct and supplement metabolic analyses. Deep-learning based features shed light on drug CDPs in individual cell lines or PDSs, allowing comparison of CDPs and selection of complementary drugs for combination therapies. This method is currently being used in "Clinical implementation Of Multidimensional PhenotypicAl drug SenSitivities in paediatric precision oncology" (COMPASS) for future improvement of treatment stratification in precision-medicine platforms such as INFORM, especially for cases with few genetic alterations [64].

A. Confirmation of Specific Aims

1) Versatility: A major aim of this approch was broad applicability which we ensure by screen-by-screen fine-tuning, enabling us to employ the same workflow for 3D cell cultures of different structural phenotypes without any manual recalibration. This includes tissue samples forming spheroids with hypoxic or hard-to-image cores as well as less-densely packed cells (Figs. 2, S-8 and S-9).

Another example is screen HD-MB03_V1_DS2, which suffered from side-illumination issues during microscopic imaging (Fig. S-10). Despite rather low Z'-factors (0.21–0.57) in CNN phase II and negative Z'-factors in metabolics (Table S-III), the CNN yields acceptable within-screen reproducibility ($R = 0.84$) for both raw cell viabilities as well derived DSS without any dedicated image preprocessing or spheroid segmentation. Performance in patients (INF_R_dddd) clearly favors the CNN-based approaches over MoS (Figs. 8, 9 and 11).

Our definition of "image-based cell viability" through positive controls is highly flexible, as it can be varied by choosing different positive (and potentially, negative) control conditions. In fact, our choice of STS as a positive control implies an *STS-likeness* interpretation of cell viability, one of many possible. Finally, study conditions need not be limited to drug treatments: for example, images from ex-vivo irradiation assays could also be used, as could multi-color image acquisitions (Fig. S-1) to provide additional distinction between phenotypes.

2) Limited Patient Material: Another aim being the ability to work with limited patient material, and thus a very low number of control images, we train the network using a minimal training batch size of 1 for maximum generalization, inspired by [65]. The resulting loss of compute parallelism is irrelevant to our application due to the low number of training images and, consequently, short training durations. We also employed regularization through label smoothing to prevent the CNN from overfitting on the few training controls, as seen in the broad distribution of raw cell viabilities for control wells (Fig. 7). Additional regularization approaches, such as

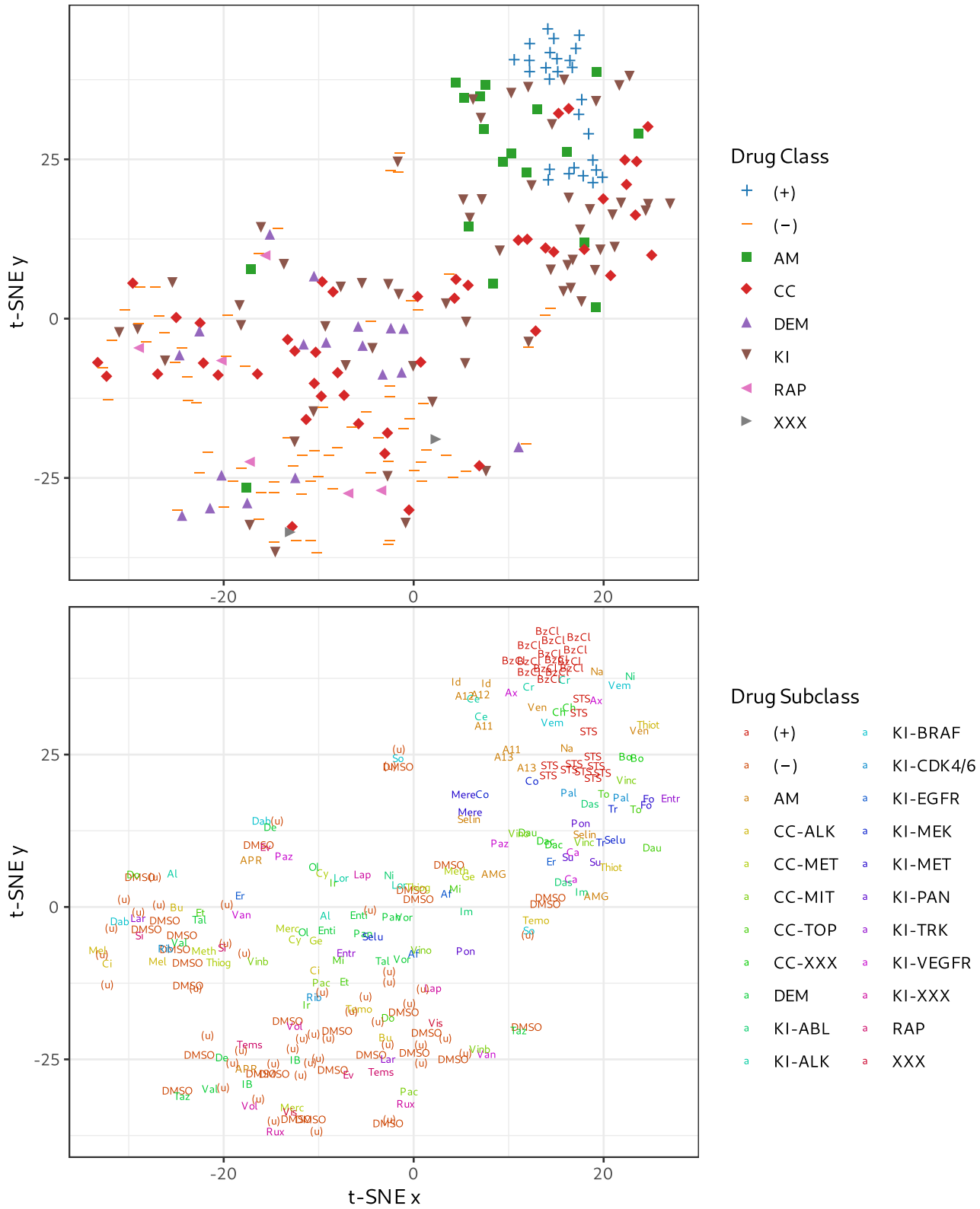

Fig. 13. t-SNE plots of CNN penultimate-dense-layer activations for wells holding highest concentrations of drugs and assay controls for patient INF_R_1025 after CNN phase II. Top: wells color- and shape-coded by drug class. Bottom: wells color-coded by drug subclass and labeled by drug name abbreviation (compare Table S-II, second column), allowing separating drugs within subclasses for biological interpretation and matching wells holding drug duplicates for technical validation. t-SNE based on 131 principal components representing 95% of the variability of the 4096 features.

mixup [66] and CutMix [67], shall be mentioned here for the sake of completeness, but have not been studied in this work.

Label smoothing also affords some detectability of out-of-distribution cases, such as contaminations, through filtering of extreme CNN outputs (raw image-based cell

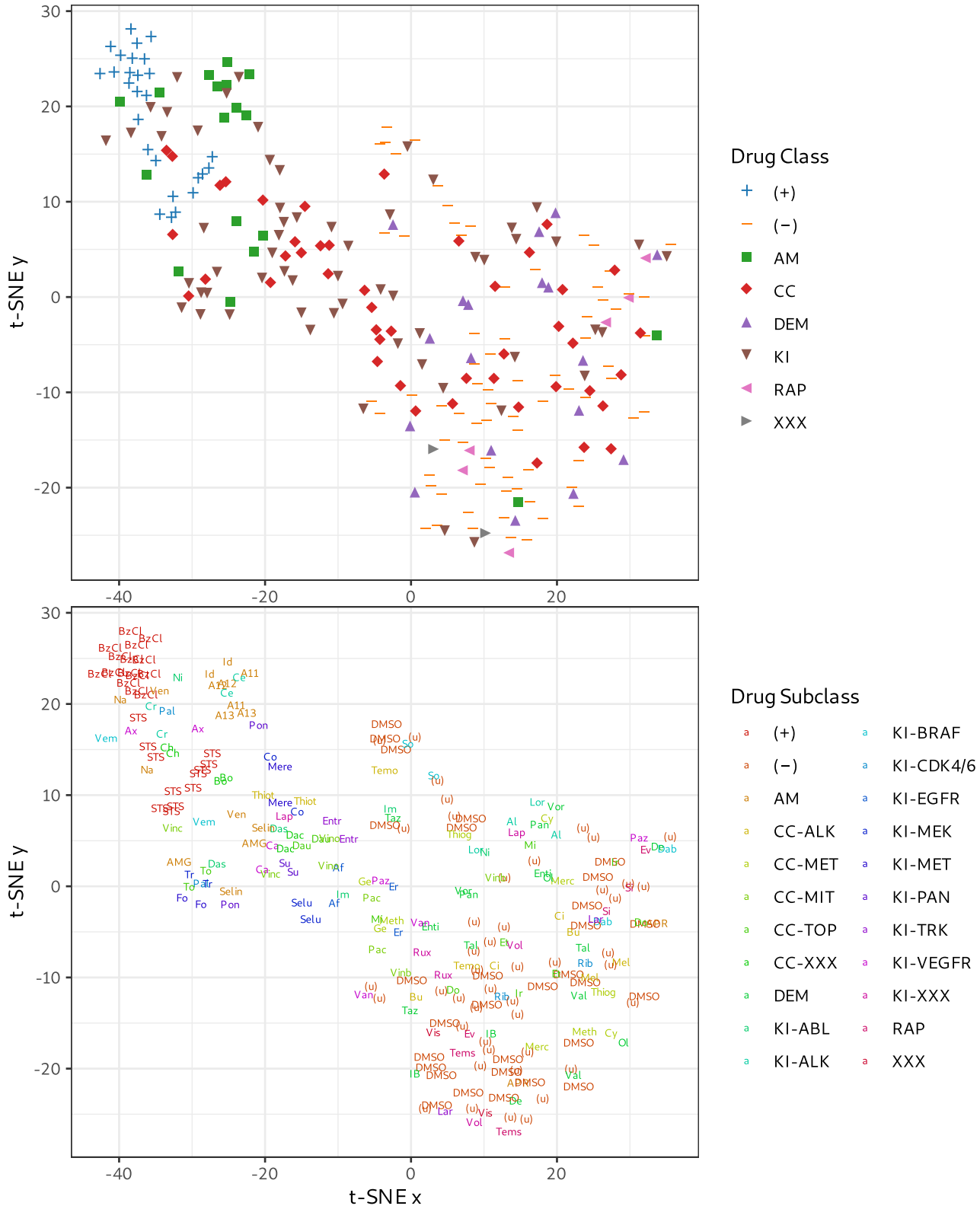

Fig. 14. t-SNE plots as in Fig. 13 after CNN phase III (157 principal components). See Fig. 15 for image annotations.

viabilities outside of $[0.2, 0.8]$) which are not normally attained (Fig. S-10).

To further optimize the use of limited patient material, we use all available controls in final (phase III) CNN training, as well as duplicate drug wells with phase-II outputs switched. This can be interpreted as an extra regularization step, as we do not introduce any new information into the CNN, but force it to consider the given information in a way designed to improve

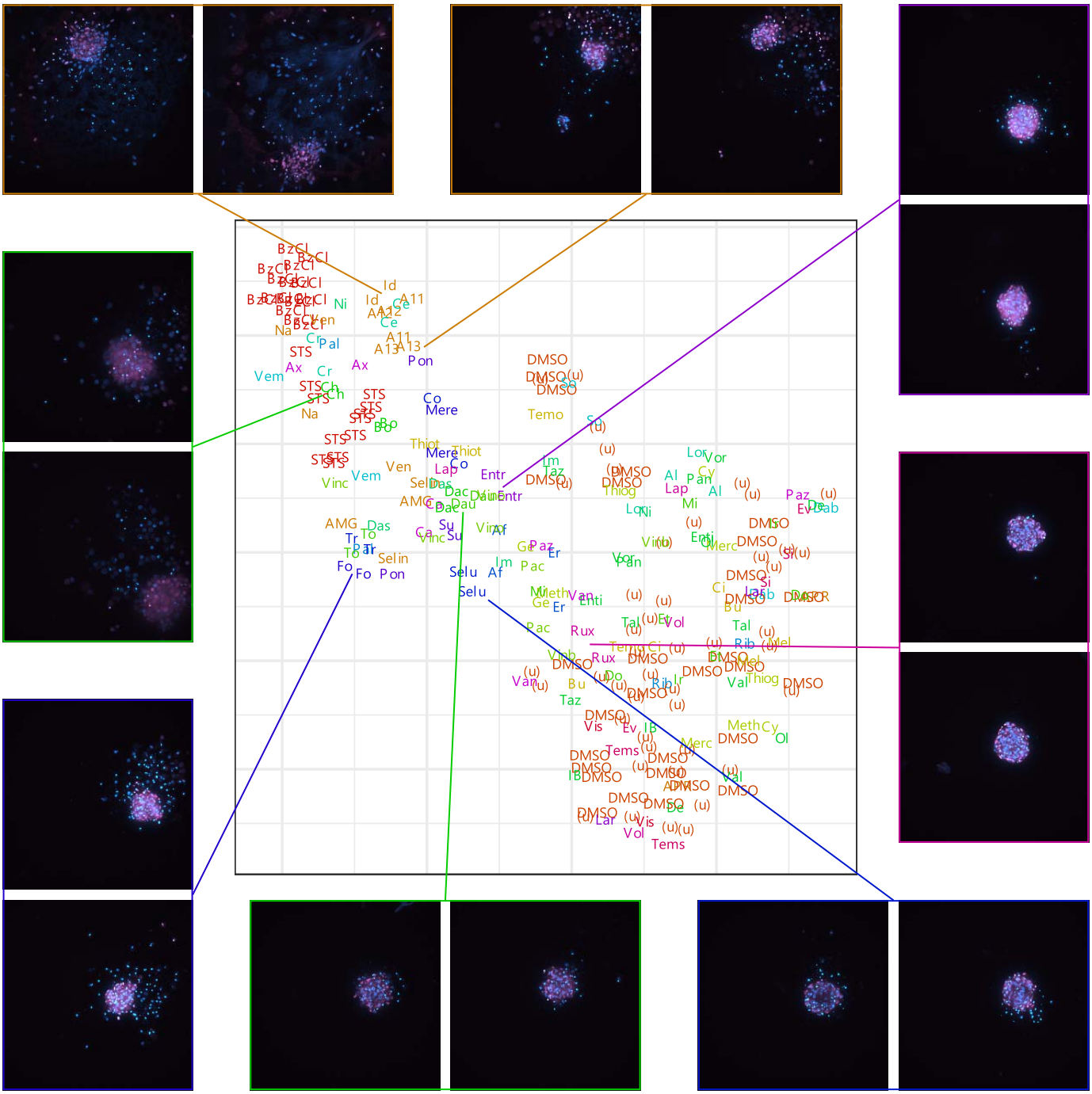

Fig. 15. Image-annotated t-SNE plot from Fig. 14 (patient INF_R_1025, phase III), including images for (clockwise from top left) Idasanutlin (Id), A-1331852 (A13), Entrectinib (Entr), Ruxolitinib (Rux), Selumetinib (Selu), Dactinomycin (Dac), Foretinib (Fo), and Chloroquine (Ch).

robustness. If the CNN were only overfitting in this step, we would expect the correlation plots between duplicate wells (Fig. 8) to flip, leaving the correlation coefficient between duplicates mostly unchanged. Therefore, we can safely evaluate the success of this step by the increase of correlation coefficients from phase II to phase III. It should be noted, however, that phase-III training relies on the existence of duplicate drug wells, which is not generally guaranteed – especially with limited patient material, in which case operators can face a decision between maximizing the number of drugs tested (without duplicates) or including duplicates (at the expense of testing a lower number of drugs). It is for this reason that we term phase III *optional*.

B. Beyond Metabolic Findings

We have already presented the largest imaging–metabolics mismatch (DSS 48.8 vs. 2.7, Fig. S-3) above, with the image-based DRCs being backed by qualitative interpretation of the images and metabolics-based DRC being contradicted by other screens of the same cell line. Beyond this anecdotal

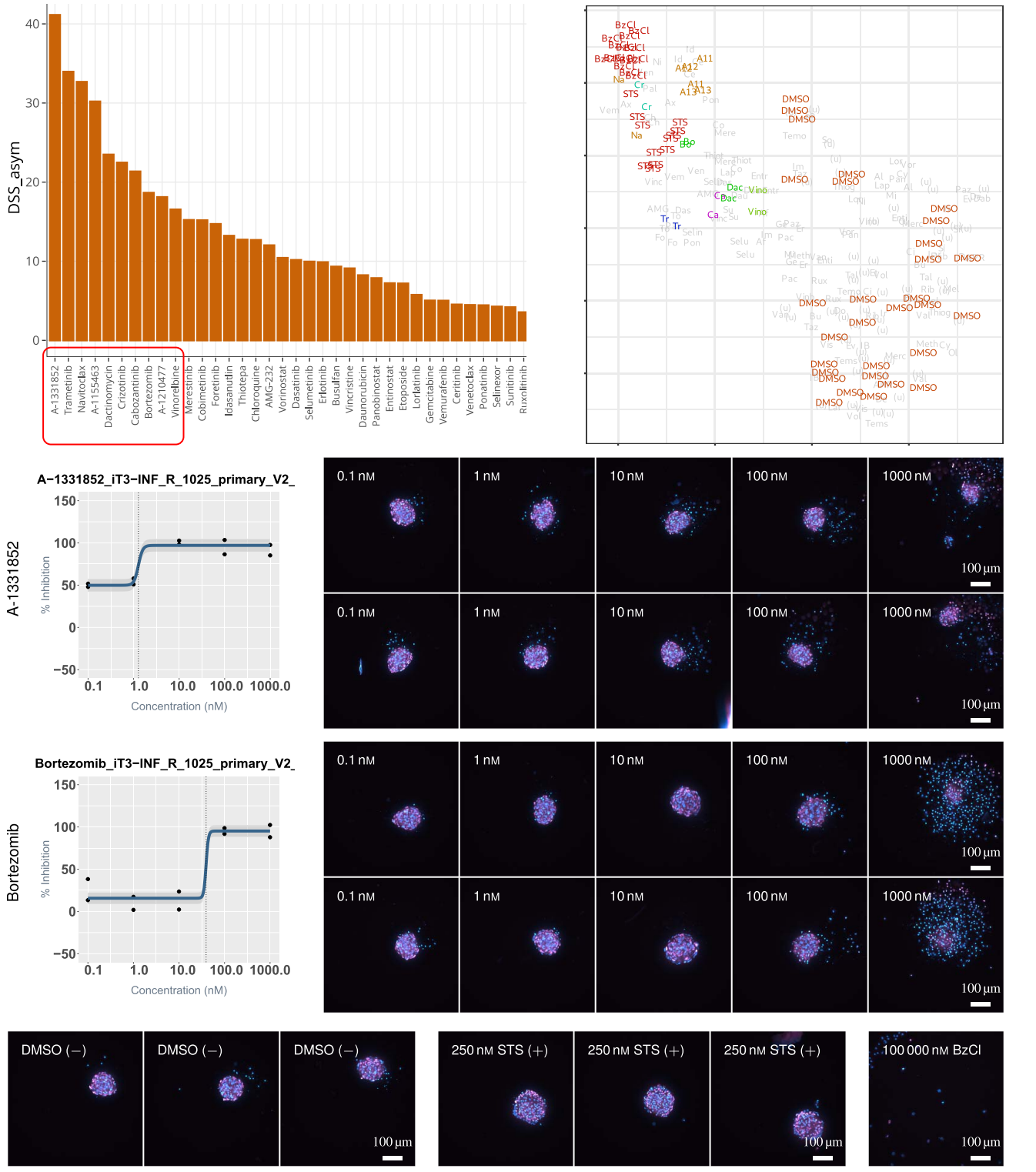

Fig. 16. Summary of possible workflow for patient INF_R_1025. Top left: waterfall plot of DSS values per drug, indicating **A-1331852**, Trametinib, Navitoclax, A-1155463, Dactinomycin, Crizotinib, Cabozantinib, **Bortezomib**, A-1210477, and Vinorelbine as top-10 drug hits. Top right: t-SNE plot of CNN features (compare Fig. 14), indicating different phenotypes between wells with highest concentrations of A-1155463/A-1210477/**A-1331852**/Ceritinib/Idasanutlin (A-11/A12/**A-13**/Ce/Id) and **Bortezomib** (**Bo**). Middle: DRCs and pseudo-RGB images of A-1331852 and Bortezomib dilution series, highlighting particularly different phenotypes at highest concentrations. Bottom: pseudo-RGB images of assay controls (3 negative imaging controls, DMSO; 3 positive imaging controls, STS; and 1 positive metabolic control, BzCl). We expect this analysis workflow uncovering complementary CDPs to be relevant for drug discovery as well as for personalized oncology.

evidence, image-based, even segmentation-free CNN-based, cell-viability quantification offers two additional, unique advantages.

1) Explainability of Image-Based Viabilities: A common point of criticism against deep-learning approaches concerns a lack of interpretability, which it shares with other machine-learning

approaches such as random forests. In terms of explainability, fine-tuning all layers of the CNN for each patient impedes computation of features that can be compared across patients. However, CNNs still offer unique possibilities for visualization of network mechanics such as CAMs (Fig. S-9): classical machine-learning approaches based on whole-well features would be challenged localizing such influences.

2) Visualization of Drug CDPs: t-SNE plots of CNN features enable further imaging-exclusive benefits to be realized, such as visualizing CDPs. Our interpretation is that t-SNE plots highlight distinctly different phenotypes between different positive and negative controls and, by extension, different drugs, which we support in three ways:

a) Confirming drug clusters through drug (sub)classes: We observe several noticeable patterns patterns in Figs. 13 and 14 and Figs. S-4 to S-7. On the level of drug classes (top panel of each plot), positive and negative controls are well separated in all plots. Similarly, apoptotic modifiers (AM) usually cluster around or close to positive controls. On the level of drug subclasses (bottom panel of each plot), BzCl and STS positive controls usually form separate clusters. On that same level, e.g., for patient INF_R_1025 (Fig. 13), all three MEK inhibitors (KI-MEK: Cobimetinib [Co], Selumetinib [Selu], and Trametinib [Tr]) cluster in the same region. Similarly, the two kinase inhibitors whose main target is the BCR-ABL kinase, Dasatinib (Das) and Pontatinib (Po), cluster together.

b) Confirming drug clusters through microscopy images: As seen in Fig. 15, microscopy images linked to individual data points in the t-SNE plot show distinctly different phenotypes between drugs at different locations in the plot, while images of neighboring drugs appear similar.

c) Confirming drug combinations through literature: Finally, several combinations of drugs that are effective yet distant from each other in the t-SNE plot (indicating complementary phenotypes) have been found to provide synergetic benefits in the literature. For example, the t-SNE plot for patient INF_R_1025 (Fig. 13) shows Venetoclax (Ven) in the top-right cluster and Ponatinib (Pon) south of it (with one replicate of the latter central in the supposedly effective cluster and the other closer to the supposedly ineffective cluster). In fact, venetoclax-ponatinib has been described as synergistic for several forms of leukemia [68], [69]. Another such example, from the same plot, is venetoclax-dasatinib, Dasatinib (Das) being found central and towards the south end of the effective cluster [68].

Considering that distances in t-SNE plots are not indicative of distances in the original features especially with different cluster sizes, the data do not allow the conclusion that negative controls exhibit higher inter-well variability. By contrast, the relative proximity of replicates despite variability in spheroid formation and staining suggests that different drug CDPs can be distinguished and could potentially be linked to drug mechanisms of actions.

C. Limitations and Potential for Improvement

1) Image Analysis: Detection of contaminations by non-cellular material (such as dirt, dust, or lint) could be improved using advanced training, for example, using a third network output class and inputting either images with extreme output values or even manually annotated images collected across patients. After detection, such data points could be discarded before drug-sensitivity scoring as each drug is described by 10 data points in total (5 concentrations in duplicates). Alternatively, contaminations not overlapping any spheroid could be segmented and removed, and a cleaned-up images fed into the cell-viability quantification network. We did not follow any of these approaches in this work due to rare occurrence of contaminations (Fig. S-10).

Despite acquisition and storage of huge amounts of 3D imaging data, the proposed approach does not make use of the full 3D information, as the network inputs only MID and MIP images. However, the added benefit of designing larger 3D CNNs is not obvious and would need to be demonstrated. Unfortunately, pre-trained 3D image-classification networks are much less widely available, and since all training phases appear critically important for screen-by-screen learning with limited patient data, this would require a re-design of all aspects of the proposed training pipeline.

Similarly, the network operates on low-resolution, 224 × 224-pixel images due to GPU memory constraints and does not use the full image resolution available. As before, the advantage of using the full image resolution is not guaranteed and would need to be demonstrated using, for example, lower-complexity architectures such as MobileNet [70] for which pre-trained networks are available.

The amount of training data could be augmented by using not only DMSO, but also untreated wells as negative controls. However, this would increase the risk of class imbalance between negative and positive controls. BzCl wells could additionally be considered as positive controls, but t-SNE plots imply that while DMSO and untreated wells are phenotypically similar, the same is not true for STS and BzCl wells. By contrast, t-SNE plots may reveal drugs similar enough to STS in CDP and effectiveness to be used as additional positive controls on a patient-by-patient basis.

Finally, it would be helpful make CDPs accessible to human interpretation; however, the grouping of drugs with similar CDPs is based on CNN features which are complex, non-linear combinations of input pixel values, rarely coincide with classical morphological properties, and are thus inherently challenging to describe in human terms. In addition, since a CNN is fine-tuned for every single screen, features represent *different* pixel-value combinations for different screens and are thus not suitable for comparison across screens. Translating similarities in CNN features into similarities of human-perceived morphology therefore requires further study.

2) Imaging Reproducibility: As we have seen, reproducibility of image-based cell viabilities is lower compared to that in metabolics, especially in phase II. While we are not able to attribute this effect to either the imaging workflow (in particular, nuclear staining and imaging) or to the image-quantification method, there is some notable potential for the former:

Variabilities may for example be introduced into the imaging workflow by spheroids outside of our imaging field of view, which covers only a central site rather than the

whole well. If wells known to be empty (free of cells) were imaged, these could be added to the training data as another class to detect wells where no cells can be seen which could then be discarded. Beyond such approaches, the proposed phase-III training does appear to improve reproducibility.

3) QC Failure: Despite the demonstrated good versatility, the proposed approach is not universal. For example, one PDS excluded from this work due to using only 566 cells/well did not establish spheroids and failed image-based QC despite good QC of the metabolic readout. However, no comparable case was available to verify if the imaging failure was related to the number of cells, the disease entity, or other factors.

4) Extended QC Analysis: In addition to Z' values computed from cell viabilities of negative (DMSO) and positive (250 nM STS) control images, STS concentration series (0.1–1000 nM) are routinely acquired as shown in Fig. S-11 and could be used for additional, automated QC analysis, using the assumption that continuously increasing STS concentrations should lead to continuously increasing inhibition values.

5) Choice of Pre-Training Data: Regarding phase-I pre-training with the INF_R_153 cell line, it is especially relevant to evaluate the need for pre-training and the impact of the choice of the training cell lines. Since pooling controls across patients screens yielded training accuracies and AUROCs around unacceptable 75%, supposedly due to large variability between PDSs, patient-specific fine-tuning seems to be mandatory. We have also seen in earlier experiments that patient-specific controls alone are insufficient for training, requiring use of a dedicated control plate for pre-training. Using another cell line for pre-training remains to be investigated.

6) Lack of Comparison Methods: This manuscript demonstrates the value of patient-by-patient deep transfer learning, in particular when compared to a simple MoS method. We have considered implementing additional comparison methods; however, due to the difficulties mentioned in Section I-A (in particular, different degrees of spheroid formation between different patients and a lack of reliable 3D single-cell segmentation methods), a classical method based on spheroid or single-cell segmentation was out of reach in this work.

VI. Outlook and Future Work

As a next step, we will apply the proposed cell-viability quantification method to a larger cohort of PDS from the INFORM study for exploration of potentially novel drug-response patterns, investigation of phentotypic clustering accross disease entities and drug mechanisms of action, and ultimately clinical translation. Lacking a unique gold standard for image quantification, clinical translation will focus on independent confirmation of metabolic findings (similar to expected hits in cell lines, Fig. 10) and genomic alterations. We further aim to train CNNs across patients to distinguish between entities, e.g., for verification of histological diagnoses.

In the future, we will compare the proposed deep-learning approach with single-cell-segmentation-based approaches, including the opportunity to differentiate cell types within one tissue sample (e.g., tumor from non-tumor cells, using a patient-specific healthy tissue sample). For illustration, Fig. S-12 shows the example of red blood cells (RBCs) making up the majority of a PDS during imaging. While RBCs are unaffected by Hoechst staining due to their lack of nuclei, bright-field (BF) imaging demonstrates how other (blood) cells centrifuged to the well bottom could be mistaken for a healthy tumor-cell spheroid in simple image-based readouts, potentially leading to over- or underestimation of measured cell viabilities, respectively.

VII. Conclusion

Deep transfer learning with patient-specific training via screen-by-screen fine-tuning allows quick implementation of high-throughput microscopy-based image-quantification pipelines for functional drug response profiling. Two-phase fine-tuning addresses limited amounts of patient material while label smoothing supports interpolation between positive and negative controls. CNN features may be used for mapping the landscape of effective and ineffective drugs by their cell-death phenotypes. Future work will employ this pipeline for discovering image-based drug-response patterns across different drugs and tumor entities, as well as correlation with NGS data.

Acknowledgment

The authors would like to thank the GPOH study centers for providing fresh vital tumor cells; Till Milde (DKFZ) for support in the Translational Drug Screening Unit and providing HD-MB03 cells; Ewa Koscielniak from Klinikum Stuttgart, Stuttgart, Germany, for providing the INF_R_153 cell line and INF_R_1025 cultured cells; Peter J. Houghton from the University of Texas Health Science Center at San Antonio, San Antonio, TX, USA, for providing cells; Emmy Dolman from Prinses Máxima Centrum, Utrecht, the Netherlands, for providing SMS-KCNR cells; the High Throughput Biomedicine Core Unit (Biocenter Finland/Institute for Molecular Medicine Finland (FIMM)), HiLIFE, University of Helsinki, Helsinki, Finland, in particular, Laura Turunen and Jani Saarela, for printing drug plates; Clarissa Holitsch and Aileen Mangang (DKFZ) for carrying out drug screens; Olli Kallioniemi and Vilja from FIMM and Gudrun Schleiermacher and Elaine Del Nery from Institut Curie for providing metabolic screen data for the INF_R_153 cell line; and Gregory P. Way from the University of Colorado Anschutz, as well as the numerous other, anonymous reviewers for their remarks leading to improvements of the performance of the method.

Yannick Berker was with the Hopp Children's Cancer Center Heidelberg (KiTZ), 69120 Heidelberg, Germany, and also with the German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Clinical Cooperation Unit Pediatric Oncology, 69120 Heidelberg, Germany. He is now with Siemens Healthineers, Molecular Imaging, 91301 Forchheim, Germany (e-mail: yannick.berker@alumni.dkfz.de).

Dina ElHarouni was with the German Cancer Research Center (DKFZ), Bioinformatics and Omics Data Analytics, 69120 Heidelberg, Germany. She is now with the Hopp Children's Cancer Center Heidelberg (KiTZ), 69120 Heidelberg, Germany, also with the Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), 69120 Heidelberg, Germany, and also with the Faculty of Biosciences, Heidelberg University, 69117 Heidelberg, Germany (e-mail: d.elharouni@kitz-heidelberg.de).

Heike Peterziel, Ina Oehme, and Sina Oppermann are with the Hopp Children's Cancer Center Heidelberg (KiTZ), 91301 Heidelberg, Germany, and also with the German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Clinical Cooperation Unit Pediatric Oncology, 69120 Heidelberg, Germany (e-mail: h.peterziel@kitz-heidelberg.de; i.oehme@kitz-heidelberg.de; sina.oppermann@kitz-heidelberg.de).

Petra Fiesel is with the Hopp Children's Cancer Center Heidelberg (KiTZ), 69120 Heidelberg, Germany, and also with the German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Clinical Cooperation Unit Neuropathology, 69120 Heidelberg, Germany (e-mail: p.fiesel@kitz-heidelberg.de).

Olaf Witt is with the Hopp Children's Cancer Center Heidelberg (KiTZ), 69120 Heidelberg, Germany, also with the German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Clinical Cooperation Unit Pediatric Oncology, 69120 Heidelberg, Germany, and also with the Department of Pediatric Oncology, Hematology, Immunology and Pulmonology, University Hospital Heidelberg, 69120 Heidelberg, Germany (e-mail: o.witt@kitz-heidelberg.de).

Matthias Schlesner was with the German Cancer Research Center (DKFZ), Bioinformatics and Omics Data Analytics, 69120 Heidelberg, Germany. He is now with the Chair of Biomedical Informatics, Data Mining and Data Analytics, University of Augsburg, 86159 Augsburg, Germany (e-mail: matthias.schlesner@informatik.uni-augsburg.de).

Author Contributions

Yannick Berker: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing: original draft, writing: review & editing, and visualization; Dina ElHarouni: software, and writing: review & editing; Heike Peterziel: resources, and writing: review & editing; Petra Fiesel: resources; Olaf Witt: supervision, writing: review & editing, and funding acquisition; Ina Oehme: supervision, and writing: review & editing; Matthias Schlesner: supervision, and writing: review & editing; Sina Oppermann: conceptualization, methodology, investigation, supervision, writing: original draft, writing: review & editing, project administration, and funding acquisition.

Code and Data Availability

The code used for training the CNNs, driving the iTReX web app, and generating the plots is available on GitHub at https://github.com/yannickberker/iTRADE, the metabolic and image data used here are available on Zenodo at https://doi.org/10.5281/zenodo.5885481

References

[1] R. Macarron *et al.*, "Impact of high-throughput screening in biomedical research," *Nat. Rev. Drug Discov.*, vol. 10, no. 3, pp. 188–195, Mar. 2011.

[2] Y. Feng, T. J. Mitchison, A. Bender, D. W. Young, and J. A. Tallarico, "Multi-parameter phenotypic profiling: Using cellular effects to characterize small-molecule compounds," *Nature Rev. Drug Discov.*, vol. 8, no. 7, pp. 567–578, Jul. 2009.

[3] M. Shen *et al.*, "Quantitative high-throughput phenotypic screening of pediatric cancer cell lines identifies multiple opportunities for drug repurposing," *Oncotarget*, vol. 9, no. 4, pp. 4758–4772, Jan. 2018.

[4] S. Li and M. Xia, "Review of high-content screening applications in toxicology," *Arch. Toxicol.*, vol. 93, no. 12, pp. 3387–3396, Dec. 2019.

[5] K. Gorshkov *et al.*, "Advancing precision medicine with personalized drug screening," *Drug Discov. Today*, vol. 24, no. 1, pp. 272–278, Jan. 2019.

[6] T. Pemovska *et al.*, "Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia," *Cancer Discov.*, vol. 3, no. 12, pp. 1416–1429, 2013.

[7] P. Barahona *et al.*, "Zero childhood cancer (ZERO): A comprehensive precision medicine platform for children with high-risk cancer," *Cancer Res.*, vol. 80, no. 14, p. A52, Jul. 2020.

[8] S. Oppermann *et al.*, "High-content screening identifies kinase inhibitors that overcome venetoclax resistance in activated CLL cells," *Blood*, vol. 128, no. 7, pp. 934–947, Aug. 2016.

[9] C. Pauli *et al.*, "Personalized *in vitro* and *in vivo* cancer models to guide precision medicine," *Cancer Discov.*, vol. 7, no. 5, pp. 462–477, May 2017.

[10] S. Dietrich *et al.*, "Drug-perturbation-based stratification of blood cancer," *J. Clin. Investig.*, vol. 128, no. 1, pp. 427–445, Jan. 2018.

[11] E. A. Ashley, "Towards precision medicine," *Nat. Rev. Genet.*, vol. 17, no. 9, pp. 507–522, Aug. 2016.

[12] S. N. Chandrasekaran, H. Ceulemans, J. D. Boyd, and A. E. Carpenter, "Image-based profiling for drug discovery: Due for a machine-learning upgrade?" *Nat. Rev. Drug Discov.*, vol. 20, no. 2, pp. 145–159, Feb. 2021.

[13] T. H. Booij, L. S. Price, and E. H. J. Danen, "3D cell-based assays for drug screens: Challenges in imaging, image analysis, and high-content analysis," *SLAS Discov.*, vol. 24, no. 6, pp. 615–627, Jul. 2019.

[14] M.-A. Bray *et al.*, "Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes," *Nat. Protoc.*, vol. 11, no. 9, pp. 1757–1774, 2016.

[15] Y. Imamura *et al.*, "Comparison of 2D- and 3D-culture models as drug-testing platforms in breast cancer," *Oncol. Rep.*, vol. 33, no. 4, pp. 1837–1843, Apr. 2015.

[16] B. Pinto, A. C. Henriques, P. M. A. Silva, and H. Bousbaa, "Three-dimensional spheroids as *in vitro* preclinical models for cancer research," *Pharmaceutics*, vol. 12, no. 12, p. 1186, Dec. 2020.

[17] V. Brancato, J. M. Oliveira, V. M. Correlo, R. L. Reis, and S. C. Kundu, "Could 3D models of cancer enhance drug screening?" *Biomaterials*, vol. 232, Feb. 2020, Art. no. 119744.

[18] J. Wardwell-Swanson *et al.*, "A framework for optimizing high-content imaging of 3D models for drug discovery," *SLAS Discov.*, vol. 25, no. 7, pp. 709–722, Aug. 2020.

[19] A. Arjonen *et al.*, "Image-based *ex vivo* drug screen to assess targeted therapies in recurrent thymoma," *Lung Cancer*, vol. 145, pp. 27–32, Jul. 2020.

[20] A. E. Carpenter *et al.*, "CellProfiler: Image analysis software for identifying and quantifying cell phenotypes," *Genome Biol.*, vol. 7, no. 10, p. R100, Oct. 2006.

[21] C. Scheeder, F. Heigwer, and M. Boutros, "Machine learning and image-based profiling in drug discovery," *Curr. Opin. Syst. Biol.*, vol. 10, pp. 43–52, Aug. 2018.

[22] S. Berg *et al.*, "ilastik: Interactive machine learning for (bio)image analysis," *Nature Methods*, vol. 16, no. 12, pp. 1226–1232, Dec. 2019.

[23] P. Mergenthaler, S. Hariharan, J. M. Pemberton, C. Lourenco, L. Z. Penn, and D. W. Andrews, "Rapid 3D phenotypic analysis of neurons and organoids using data-driven cell segmentation-free machine learning," *PLOS Comput. Biol.*, vol. 17, no. 2, Feb. 2021, Art. no. e1008630.

[24] S. Nath and G. R. Devi, "Three-dimensional culture systems in cancer research: Focus on tumor spheroid model," *Pharmacol. Therapeutics*, vol. 163, pp. 94–108, Jul. 2016.

[25] D. Kusumoto *et al.*, "Anti-senescent drug screening by deep learning-based morphology senescence scoring," *Nature Commun.*, vol. 12, no. 1, p. 257, Jan. 2021.

[26] D. Jimenez-Carretero *et al.*, "Tox_(R)CNN: Deep learning-based nuclei profiling tool for drug toxicity screening," *PLOS Comput. Biol.*, vol. 14, no. 11, Nov. 2018, Art. no. e1006238.

[27] J. C. Caicedo, C. McQuin, A. Goodman, S. Singh, and A. E. Carpenter, "Weakly supervised learning of single-cell feature embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 9309–9318.

[28] S. Kim *et al.*, "Comparison of cell and organoid-level analysis of patient-derived 3D organoids to evaluate tumor cell growth dynamics and drug response," *SLAS Discov.*, vol. 25, no. 7, pp. 744–754, Aug. 2020.

[29] J. Betge *et al.*, "The drug-induced phenotypic landscape of colorectal cancer organoids," *Nat. Commun.*, vol. 13, no. 1, Dec. 2022, Art. no. 660993.

[30] Z. Chen *et al.*, "Automated evaluation of tumor spheroid behavior in 3D culture using deep learning-based recognition," *Biomaterials*, vol. 272, May 2021, Art. no. 120770.

[31] B. Soetje, J. Fuellekrug, D. Haffner, and W. H. Ziegler, "Application and comparison of supervised learning strategies to classify polarity of epithelial cell spheroids in 3D culture," *Front. Genet.*, vol. 11, p. 248, Mar. 2020.

[32] C. M. van Tilburg *et al.*, "The pediatric precision oncology INFORM registry: Clinical outcome and benefit for patients with very high-evidence targets," *Cancer Discov.*, vol. 11, no. 11, pp. 2764–2779, Nov. 2021.

[33] K. Langenberg, E. Dolman, and J. Molenaar, "Abstract A40: Integration of high-throughput drug screening on patient-derived organdies into pediatric precision medicine programs: The future is now!" *Cancer Res.*, vol. 80, no. 14, p. A40, Jul. 2020.

[34] P. Berlanga *et al.*, "Pediatric precision medicine program in recurrent tumors: Results of the first 500 patients included in the European MAPPYACTS molecular profiling trial," *Cancer Res.*, vol. 79, no. 13, p. CT081, Jul. 2019.

[35] E. Meijering, "A bird's-eye view of deep learning in bioimage analysis," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 2312–2325, Aug. 2020.

[36] B. M. Kuenzi *et al.*, "Predicting drug response and synergy using a deep learning model of human cancer cells," *Cancer Cell*, vol. 38, no. 5, pp. 672–684, Nov. 2020.

[37] W. J. Godinez, I. Hossain, S. E. Lazic, J. W. Davies, and X. Zhang, "A multi-scale convolutional neural network for phenotyping high-content cellular images," *Bioinformatics*, vol. 33, no. 13, pp. 2010–2019, Feb. 2017.

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., May 2015, pp. 1–14.

[39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

[40] D. ElHarouni *et al.*, "iTReX: Interactive exploration of mono- and combination therapy dose response profiling data," *Pharmacol. Res.*, vol. 175, Jan. 2022, Art. no. 105996.

[41] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, H. Daumé, III and A. Singh, Eds., vol. 119, Jul. 2020, pp. 1597–1607.

[42] N. Bindslev, "Hill in hell," in *Drug-Acceptor Interactions: Modeling Theoretical Tools to Test and Evaluate Experimental Equilibrium Effects*. London, U.K.: CRC Press, Feb. 2017, ch. 10, pp. 257–281.

[43] B. Yadav *et al.*, "Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies," *Sci. Rep.*, vol. 4, no. 1, p. 5193, Jun. 2014.

[44] O. Pelz, M. Gilsdorf, and M. Boutros, "web cellHTS2: A web-application for the analysis of high-throughput screening data," *BMC Bioinf.*, vol. 11, no. 1, p. 185, Apr. 2010.

[45] N. A. Clark *et al.*, "GRcalculator: An online tool for calculating and mining dose–response data," *BMC Cancer*, vol. 17, no. 1, p. 698, Oct. 2017.

[46] S. Potdar *et al.*, "Breeze: An integrated quality control and data analysis application for high-throughput drug screening," *Bioinformatics*, vol. 36, no. 11, pp. 3602–3604, Mar. 2020.

[47] C. Gatzweiler *et al.*, "Functional therapeutic target validation using pediatric zebrafish xenograft models," *Cancers*, vol. 14, no. 3, p. 849, Feb. 2022.

[48] D. S. Hong *et al.*, "Larotrectinib in patients with TRK fusion-positive solid tumours: A pooled analysis of three phase 1/2 clinical trials," *Lancet Oncol.*, vol. 21, no. 4, pp. 531–540, Apr. 2020.

[49] H. K. Bid *et al.*, "Development, characterization, and reversal of acquired resistance to the MEK1 inhibitor selumetinib (AZD6244) in an *in vivo* model of childhood astrocytoma," *Clin. Cancer Res.*, vol. 19, no. 24, pp. 6716–6729, Dec. 2013.

[50] T. Milde *et al.*, "HD-MB03 is a novel group 3 medulloblastoma model demonstrating sensitivity to histone deacetylase inhibitor treatment," *J. Neuro.-Oncol.*, vol. 110, no. 3, pp. 335–348, Dec. 2012.

[51] G. Valinciute *et al.*, "MEDU-01. HDACi and PLK1i act synergistically in MYC-amplified medulloblastoma," *Neuro-Oncol.*, vol. 21, no. 2, p. 103, Apr. 2019.

[52] J. P. Koivunen *et al.*, "EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer," *Clin. Cancer Res.*, vol. 14, no. 13, pp. 4275–4283, Jul. 2008.

[53] I. Lambertz *et al.*, "Upregulation of MAPK negative feedback regulators and RET in mutant ALK neuroblastoma: Implications for targeted treatment," *Clin. Cancer Res.*, vol. 21, no. 14, pp. 3327–3339, Jul. 2015.

[54] J. Guan *et al.*, "The ALK inhibitor PF-06463922 is effective as a single agent in neuroblastoma driven by expression of ALK and MYCN," *Dis. Models Mech.*, vol. 9, no. 9, pp. 941–952, Sep. 2016.

[55] A. Van Goethem *et al.*, "Dual targeting of MDM2 and BCL2 as a therapeutic strategy in neuroblastoma," *Oncotarget*, vol. 8, no. 34, pp. 57047–57057, Aug. 2017.

[56] T. Eisemann *et al.*, "Podoplanin expression is a prognostic biomarker but may be dispensable for the malignancy of glioblastoma," *Neuro-Oncol.*, vol. 21, no. 3, pp. 326–336, Feb. 2019.

[57] E. Kulesskiy, J. Saarela, L. Turunen, and K. Wennerberg, "Precision cancer medicine in the acoustic dispensing era: *Ex vivo* primary cell drug sensitivity testing," *J. Lab. Autom.*, vol. 21, no. 1, pp. 27–36, Feb. 2016.

[58] C. A. Belmokhtar, J. Hillion, and E. Ségal-Bendirdjian, "Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms," *Oncogene*, vol. 20, no. 26, pp. 3354–3362, Jun. 2001.

[59] M. W. Karaman *et al.*, "A quantitative analysis of kinase inhibitor selectivity," *Nat. Biotechnol.*, vol. 26, no. 1, pp. 127–132, Jan. 2008.

[60] J.-H. Zhang, T. D. Y. Chung, and K. R. Oldenburg, "A simple statistical parameter for use in evaluation and validation of high throughput screening assays," *SLAS Discov.*, vol. 4, no. 2, pp. 67–73, Apr. 1999.

[61] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020.

[62] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 839–847.

[63] H. Wang *et al.*, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 24–25.

[64] G. Schleiermacher and O. Delattre, "Kids enter the MATCH," *J. Nat. Cancer Inst.*, vol. 109, no. 5, pp. 1–15, Dec. 2016.

[65] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," 2018, *arXiv:1804.07612*.

[66] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., Apr. 2017, pp. 1–13.

[67] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6023–6032.

[68] A. Maiti *et al.*, "Venetoclax and BCR-ABL tyrosine kinase inhibitor combinations: Outcome in patients with Philadelphia chromosome-positive advanced myeloid leukemias," *Acta Haematol.*, vol. 143, no. 6, pp. 567–573, Dec. 2020.

[69] H. Wang *et al.*, "Venetoclax-ponatinib for T315I/compound-mutated Ph+ acute lymphoblastic leukemia," *Blood Cancer J.*, vol. 12, no. 1, p. 20, Jan. 2022.

[70] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.