

Cite this: *Lab Chip*, 2020, **20**, 3696

Deep-learning-assisted biophysical imaging cytometry at massive throughput delineates cell population heterogeneity†

Dickson M. D. Siu,^a Kelvin C. M. Lee,^a Michelle C. K. Lo,^a Shobana V. Stassen,^a Maolin Wang,^a Iris Z. Q. Zhang,^a Hayden K. H. So,^a Godfrey C. F. Chan,^b Kathryn S. E. Cheah,^c Kenneth K. Y. Wong,^a Michael K. Y. Hsin,^d James C. M. Ho^e and Kevin K. Tsia^{a*}

The association of the intrinsic optical and biophysical properties of cells to homeostasis and pathogenesis has long been acknowledged. Defining these label-free cellular features obviates the need for costly and time-consuming labelling protocols that perturb the living cells. However, wide-ranging applicability of such label-free cell-based assays requires sufficient throughput, statistical power and sensitivity that are unattainable with current technologies. To close this gap, we present a large-scale, integrative imaging flow cytometry platform and strategy that allows hierarchical analysis of intrinsic morphological descriptors of single-cell optical and mass density within a population of millions of cells. The optofluidic cytometry system also enables the synchronous single-cell acquisition of and correlation with fluorescently labeled biochemical markers. Combined with deep neural network and transfer learning, this massive single-cell profiling strategy demonstrates the label-free power to delineate the biophysical signatures of the cancer subtypes, to detect rare populations of cells in the heterogeneous samples (10–5), and to assess the efficacy of targeted therapeutics. This technique could spearhead the development of optofluidic imaging cell-based assays that stratify the underlying physiological and pathological processes based on the information-rich biophysical cellular phenotypes.

Received 26th May 2020,
Accepted 8th September 2020

DOI: 10.1039/d0lc00542h

rsc.li/loc

Image-based profiling has become a new thrust in biological research for census of individual cells in a large population. By extracting a myriad of spatial features from the single-cell image, it is possible to create a “fingerprint-like” profile that defines cellular properties and identity.¹ While current profiling methods are predominantly based on the molecular-specific image contrast given by fluorescence labels,^{2,3} the intrinsic image contrast of single cells employed in microscopy for centuries (*e.g.*, bright-field, phase-contrast), can also be indicative of many biological processes. However,

the bright-field image contrast of an unlabeled cell is poor and is thus broadly conceived to be practical only for probing bulk single-cell biophysical and mechanical properties, *e.g.*, cell size and deformation.^{4,5} Yet, what has commonly been overlooked is that the subtly observable bright-field texture (or optical density distribution) within cells could indeed be an effective biomarker. For example, shapes and textures of cells assessed in bright-field have been used for the classification of immune cell types,⁶ cell cycle analysis⁷ and disease-specific phenotypes for blood analysis.⁸

The label-free phenotypic content can further be enriched by employing quantitative phase imaging (QPI).⁹ Not only can QPI overcome the contrast limitation of bright-field imaging, but it also quantifies cellular dry mass density that can be derived from the optical phase at high precision.^{10–12} Cell mass is contributed primarily by intracellular proteins, lipids, metabolites, and nucleic acids.¹³ Regulation of cell mass is thus tightly linked to underlying molecular pathways, from protein folding equilibrium^{14,15} to single-cell transcriptomic signatures.¹⁶ Indeed, a growing body of evidence shows that the label-free assessment of cell mass density is as effective, or even more accurate, descriptor of cellular heterogeneity, compared to the conventional fluorescence markers.^{17–20}

^a Department of Electrical and Electronic Engineering, Choi Yei Ching Building, The University of Hong Kong, Pokfulam Road, Pokfulam, Hong Kong.

E-mail: tsia@hku.hk

^b Department of Pediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong

^c School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong

^d Department of Cardiothoracic Surgery, Queen Mary Hospital, Pokfulam Road, Hong Kong

^e Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0lc00542h

Although the association between the molecular genetic landscape that instructs expression of different proteins and macromolecules in single-cells must be intrinsically linked with their morphological features, systematic demonstration of the relationship and practical application to resolve population heterogeneity has not been shown. The main limitation lies in the lack of high-throughput quantitative label-free imaging tool that offers sufficient statistical power to reveal the heterogeneity of large populations of cells, and thus provides maximal information to link to the molecular signatures and responses.

To address this gap, we established a systematic strategy to profile the spatial characteristics of optical and mass densities of single cells, which are here referred to as optophysical phenotypes. Given the limitations of current QPI in terms of throughput (10–1000's cells per s) and phenotypic content,^{21–24} here we employ an ultrahigh-throughput (>10 000 cells per s) QPI flow cytometry platform, called multi-ATOM,²⁵ to perform deep single-cell optophysical phenotyping. In contrast to the prior work on QPI (including multi-ATOM), here we adopt a hierarchical approach to exploit not only the bulk features²⁶ (*e.g.*, cell size, shape, and mass), but most importantly, their global and local textural information within a cell. Systematic hierarchical phenotyping has not been applied for assessing the statistical power of the optophysical properties of cells. Enabling this effort could facilitate better semantic interpretation of optical and mass density of cells and their link to the established biological knowledge.

We show that this label-free image-based profiling method, combined with deep learning (including a neural network model and transfer learning), allows us to robustly delineate the spatial optophysical profiles of different lung cancer subtypes within large populations of cells (beyond millions). Going beyond the previous multi-ATOM platform, we further advance the system by integrating the fluorescence cytometry capability such that the single-cell optophysical phenotypes and molecular-specific fluorescent markers can be simultaneously captured without sacrificing the throughput. Not only can this system sensitively detect rare lung cancer cell population in human peripheral blood in a label-free manner, but also correlate the single-cell optophysical markers to the fluorescent surface-marker information. This correlation is an important yet underexploited knowledge that could allow us to identify new cost-effective surrogate biomarkers derived from optophysical phenotypes. We finally demonstrate that this hierarchical spatial optophysical profiling method is also sensitive to predict targeted drug response of non-small cell lung cancer (NSCLC), showing this potential in cost-effective label-free drug screening.

Results

General workflow

The phenotyping workflow begins with ultrahigh-throughput single-cell imaging using multi-ATOM in the microfluidic

flow configuration (Fig. 1 and see Fig. S1†). Based upon optical time-stretch,^{27–29} multi-ATOM enables ultrafast QPI without the need for dedicated interferometry, which largely limits the speed of most of the QPI modalities, at an imaging line scan rate beyond 10 MHz.^{24,25} To enable robust single-cell multi-ATOM, we fabricated a custom microfluidic chip which focused the flowing cells in a single stream in a channel by inertial focusing.³⁰ The serpentine structure of the microfluidic chip first focused the cells to form an aligned stream through balancing of inertial lift and Dean force. With the following long straight, tapered channel, it further confined the focal spots of cells with various sizes, achieving a tightly focused cell flow. 2D snapshots of individual flowing cells were reconstructed and processed in real-time at a throughput >10 000 cells per s by digitally stacking the line-scans, which are orthogonal to the cell flow (Fig. 1). Here, we used a real-time field programmable gate array (FPGA) based signal processing system on which custom logic was implemented to automatically detect and segment cells from the digitized data stream, all in real-time. Based upon the intensity-only measurements, multi-ATOM retrieves the complex-field information of light transmitting through the cell and yields two image contrasts at subcellular resolution: bright-field (BF: amplitude of the complex-field) which displays the distribution of light attenuation (or optical density) within the cell; and quantitative phase (ϕ : phase of the complex field), which relates to the refractive index and thus the mass density distribution within the cells.³¹ We experimentally validated that the complex fields (*i.e.*, amplitude and phase) of the microbeads measured by multi-ATOM agreed well with both classical Fourier optics model, and the Mie-scattering theory³² (Fig. S3†).

We further integrated a synchronized module that simultaneously detects the fluorescence signal (channel) of the cell (Fig. S1†). To simplify the detection hardware and make it scalable in the number of fluorescence channels, we implemented a frequency multiplexing scheme in which the two fluorescence signals (excited by two continuous wave (CW) lasers at the wavelengths: 488 nm and 532 nm) were modulated at different frequencies before detection; and were digitally separated by frequency demodulation (Fig. S2†). The fluorescence signals were synchronized with the multi-ATOM signals such that it allows linked measurement of fluorescence-labeled biochemical phenotypes and the optophysical phenotypes of the same cell at high throughput. A detailed workflow and information can be found in Methods and materials.

Our phenotypic strategy defines the spatial characteristics of both optical and mass density of a cell in a hierarchical manner, *i.e.*, from the bulk features to the local subcellular textures (Fig. 1). At the bulk level, we quantified the cell size, dry mass density, and cell shape. At the global texture level, we parameterized the holistic textural characteristics of optical density and mass density across the cell by their multi-order moment statistics. We note that analyzing higher-order moments (3rd and 4th) is known to provide a

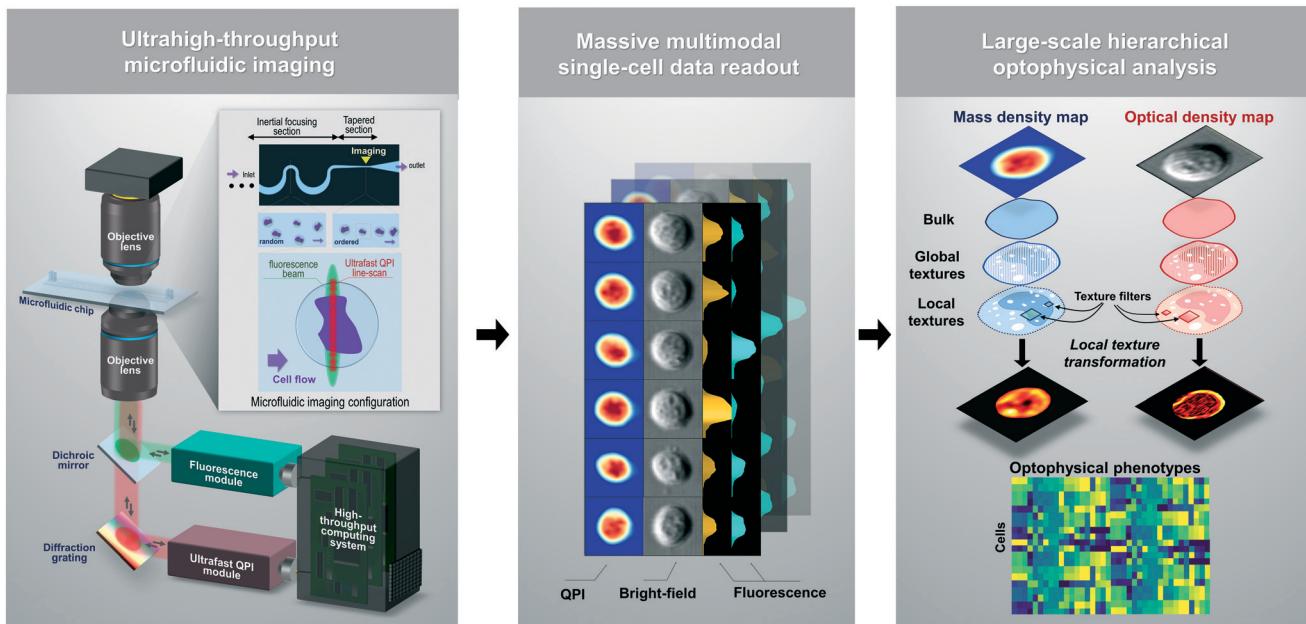


Fig. 1 General workflow of single-cell hierarchical spatial optophysical cytometry. First, a microfluidic-based imaging flow cytometer is employed to synchronously capture the quantitative phase images (QPI, equivalent to the mass density map), bright-field images (equivalent to the optical density maps) and fluorescence signals of the individual cells, at a real-time throughput of $\sim 10\,000$ cells per s. The system mainly consists of two integrated parts: a fluorescence module and an ultrafast QPI module (*i.e.* multi-ATOM for QPI and bright-field image acquisition). The microfluidic chip is designed to align the cells in a single stream by inertial focusing. The key structure includes a serpentine structure followed by a straight, tapered channel. 2D snapshots of individual flowing cells were reconstructed and processed in real-time at a throughput $>10\,000$ cells per s by digitally stacking the line-scans. Based on the optical density and mass density map, the spatial optophysical phenotypes across various hierarchical layers are then extracted to generate a profile of spatially-resolved optophysical properties. Which is used for cell-type classification and correlative analysis.

better statistic for detecting subpopulations.³³ However, this is only valid for a large sample size, which is otherwise limited in the existing label-free phenotyping techniques. We further applied a set of texture filters at the different kernel sizes to quantify the local textural characteristics at both the coarse and fine scale (*e.g.*, local variation of mass density (text contrast in Fig. 2a and 3a), its higher-order statistics). This hierarchical phenotyping approach allowed us to establish high-dimensional spatial profiles employing more than 80 optophysical features, which were normalized based on the z-score (see ESI Note S1†).

Label-free lung cancer subtype classification with transfer learning

We first evaluated the specificity of our label-free phenotyping strategy by imaging and profiling seven lung cancer cell lines comprising a total population of 2.3 million cells. Specifically, we sought to ask if this label-free method could delineate three major histologically differentiated subtypes of lung cancer amongst seven cell lines, *i.e.*, two subtypes of NSCLC (adenocarcinoma, squamous cell carcinoma) and one for SCLC.³⁴ First of all, we observed that it is not effective to distinguish the subtypes based on either visual inspection of the single-cell images (Fig. 2a) or the common label-free biophysical

features (*e.g.* cell size, the averaged mass/optical density) extracted from the images (Fig. 2b). To this end, we applied the neural network, combined with transfer learning, to classify the three main subtypes by using either the common biophysical features (*i.e.* volume, attenuation density and dry mass density) or the high-dimensional optophysical phenotypic profile as the network inputs (Fig. 2c). To ensure practical optophysical single-cell analysis, we took into account the batch variation effect in our phenotyping pipeline. Batch effect accounts for the technical sources of data variation due to measurement in different setups or/and at different times, which is known to compromise genuine data interpretation and analysis in cytometry. Here we adopt a transfer-learning approach that could offer a two-fold advantage: (1) reducing the training data amount;³⁵ (2) alleviating the batch-to-batch variation problem.³⁶ Compared to the use of the common biophysical features, the neural network model using the full set of optophysical phenotypes yields a significant improvement in the prediction accuracy (from 76% to $>91\%$) (Fig. S6†). In addition, we found that transfer learning significantly improved the predicted probability for each lung cancer subtype (Fig. 2d). Overall, after transfer learning had been included, the neural network consistently predicted each subtype from the unseen (test) datasets at high accuracy of 91% to 95%. This is in contrast to the same neural network

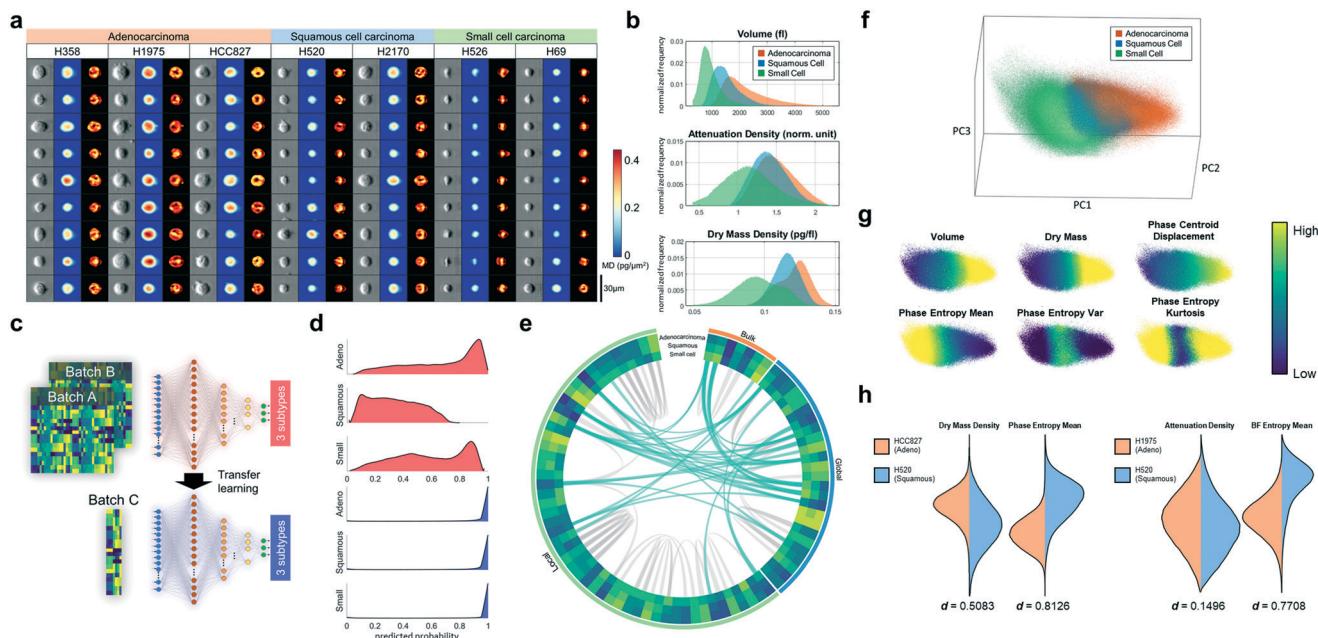


Fig. 2 Massive (~2.3 million cells) label-free lung cancer subtypes classification by deep learning. (a) Randomly selected label-free single-cell images of the seven lung cancer cell lines, which belong to three histologically differentiated subtypes of lung cancer. From left to right in each cell line, the spatial maps are optical density, mass density and local texture transformation of mass density. The colorbar shows the scale of the mass density (MD). (b) The distributions of cell volume, averaged optical density and averaged mass density of the three lung cancer subtypes. (c) General workflow of the label-free lung-cancer subtype classification by transfer-learning-assisted deep neural network. (d) The predicted probability distribution of each subtype before (red) and after (blue) transfer learning. (e) A circular plot summarizing the phenotypic heatmap of the lung-cancer subtypes and the correlations between the optophysical phenotypes (see Fig. S9† for detailed feature labels). The outermost layer (orange, blue and green) indicates the phenotypic groups belonging to the bulk, global and local features. The three inner rings represent the mean phenotypic profiles of the three lung cancer subtypes. The lines in the centre connect the phenotypes with correlation coefficients >0.7 . Gray lines represent the correlations within the same phenotypic group whereas green lines represent correlations across different groups. (f) A three-dimensional PCA visualization of the 2.3 millions cells based on the spatial optophysical profiles. The first 3 principal components are shown. (g) The same PCA maps as (f), color-coded with the expression levels of six different optophysical phenotypes, normalized by the z-scores. (h) Classification performance comparisons between the global and local spatial optophysical phenotypes. d stands for Cliff's delta.

before transfer learning that performs poorly in distinguishing the subtypes (Fig. S7†). The transfer-learning-assisted models also demonstrated an improved accuracy ranging from 3% to 7%, compared with the models being trained and tested on the same batch of data. This improvement could be attributable to the ability to transfer the knowledge of the batch effects (including systematic image focus conditions, system drift, or variations in laser power and photodetector sensitivity) to new classification tasks. By inspecting the averaged optophysical phenotypic profiles of all the 2.3 million cells, we observed three distinct signatures corresponding to the three lung-cancer subtypes (Fig. 2e and S9†). When we grouped the single-cell features into three main classes covering the hierarchy: bulk, global and local textures, we found that the highly correlated features (Pearson's correlation coefficient >0.7) generally occur more within the same class (gray lines in Fig. 2e), but less common across different classes (green lines in Fig. 2e). It signifies the low-redundancy of this hierarchical phenotyping approach. We further studied the phenotypic correlations among the bulk, optical density, and mass density features and observed a similar low-redundancy among these categories of features (Fig. S10†).

We also performed principal component analysis (PCA) to visualize the single-cell optophysical data in the low-dimensional representation by using the first three principal components, which represent 74% of the total variance (Fig. 2f). To interpret the phenotypic differences among the subtypes, we visualized some general trends in the feature expressions on this 3D PCA map (at the bulk, local texture, and global texture levels) (Fig. 2g and S11†). For instance, comparing the two NSCLC subtypes, we found that the adenocarcinoma population generally exhibits larger cell size, higher averaged mass density, but lower values in local textures of mass density (e.g., phase entropy variance), than the squamous cell carcinoma. The significance of such a hierarchy can also be evident by comparing the local texture features and global texture features in classifying different lung cancer subtypes (Fig. 2h). We showed that the local mass density feature (QP entropy mean) had a larger effect size ($d = 0.8126$ using Cliff's delta statistics) than its global equivalent (dry mass density, $d = 0.5083$) in classifying adenocarcinoma (HCC827) and squamous cell carcinoma (H520). Likewise, the local optical-density feature (BF entropy mean) showed a larger effect size ($d = 0.7708$) than the global equivalent (attenuation density, $d = 0.1496$). Note that we

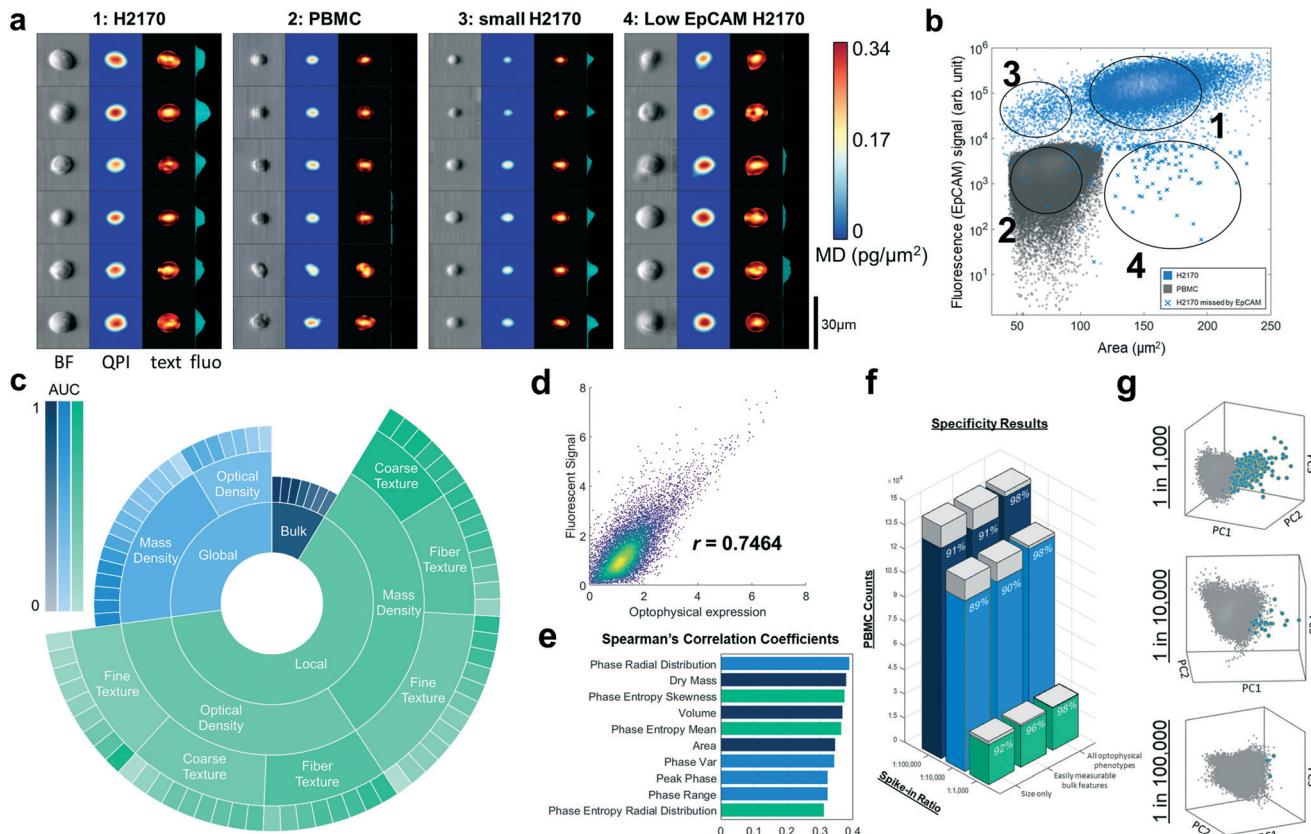


Fig. 3 Spatial optophysical profile detects rare cell populations and correlates with the molecular markers. (a) Examples of single-cell images synchronously captured with the fluorescence signals (EpCAM) of the H2170 and PBMC. The colorbar shows the scale of the mass density (MD). (b) Measured fluorescence (EpCAM) signal and cell size of the separate populations of H2170 and PBMC. Four regions were identified, and the corresponding spatial profiles are shown in (a). Cross markers indicate the H2170 subpopulation with low EpCAM expression level. This subpopulation is commonly missed by EpCAM-based fluorescence detection but could be identified by optophysical phenotypes (Fig. S13†). (c) A sunburst chart showing the significance of the hierarchical spatial optophysical phenotypes in classification of H2710 and PBMC. The outermost layer displays the individual phenotypes. The brightness of color in each grid of the chart represents the significance of the optophysical phenotype (quantified by the area under curve (AUC)). The AUC values were calculated directly by plotting the distributions of H2170 and PBMCs on each feature axis. A classifier was then directly scanned on this axis to produce the ROC curve. The AUC was finally obtained from the ROC curve. Thus, an AUC value could be computed from each individual feature. The significance is ranked in clockwise (see Fig. S12† for detailed feature labels). (d) Correlation between the optophysical phenotypes and EpCAM expression. A linear regression model was implemented to fit all the optophysical phenotypes to the EpCAM expression and thus show the integral correlative relationship between optophysical phenotypes and EpCAM expression. r denotes the Spearman's correlation coefficient. (e) Spearman's correlation between individual optophysical phenotypes (top 10 most correlated) and fluorescent signal (EpCAM expression). The bar color, following the same color scheme as (c), indicates the hierarchical level to which the phenotype belongs. (f) Specificity improvement of using the spatial optophysical phenotypes versus using easily measurable bulk features. The gray stacks in the bar chart denote the false-negative cases and the colored stacks denote the true negative cases. The value of the specificity in each case is specified. (g) Three-dimensional PCA visualizations of three spike-in ratios (1:1000, 1:10 000 and 1:100 000). The first 3 principal components are used to visualize the data. Blue dots denote the H2170 and gray dots denote the PBMCs.

used effect size, which is independent of sample size, instead of the commonly used p -value because of our large sample size ($>100\,000$ cells),³⁷ which is otherwise challenging to attain in other techniques.

Spatial optophysical profile detects rare populations and correlates with molecular markers

Using the single-cell spatial optophysical profiles, we next investigated whether our method could have the sensitivity to detect rare cell populations. To establish the basic understanding of the role of optophysical phenotypes in this

task, we first analyzed the pure populations of human peripheral blood mononuclear cells (PBMC) and a NSCLC cell line (H2170: squamous cell carcinoma) fluorescently labeled with antibodies against the cell surface protein epithelial cell adhesion molecule (EpCAM) – a common marker used for detection and enrichment of NSCLC in liquid biopsy.^{38–40} The multi-ATOM platform allows simultaneous measurements of both spatial optophysical phenotypes and the fluorescent EpCAM expression, at single-cell precision. In general, using either the cell size or EpCAM expression alone is not always effective to distinguish the NSCLC from PBMC, especially the small populations of small EpCAM+ NSCLC

(cluster 3) and large EpCAM- NSCLC (cluster 4) (Fig. 3a and b). This essentially reflects the main limitations in many popular circulating tumor cell (CTC) detection methods, notably the loss of sensitivity due to the biased prior knowledge primarily toward epithelial-specific cell-surface markers, and the low specificity given by the size-based detection methods, which are often difficult to capture due to the heterogeneity of CTC size.^{41–43} Indeed, by visualizing the ranked optophysical phenotypes in a hierarchical structure, we observed the indispensable role of both the global and local features in detecting NSCLC (Fig. 3c and S12†).

We again used the deep neural network model based upon the single-cell spatial optophysical phenotypes, but modified it for predicting the NSCLC identity. Our analysis showed that the spatial optophysical profile, together with the EpCAM information, identify the additional rare populations (0.61% of the original H2170 population), that were not recognized by either EpCAM expression, cell size or the combination of common biophysical features with EpCAM expression, and thereby further reduce the false-negative count (Fig. 3b and S13†). Indeed, compared to the use of EpCAM expression alone, our approach of using both optophysical phenotypes and EpCAM expression increases the number of true positives detected by 119, which resulted in improved sensitivity and accuracy in spite of a slight compromise in specificity (<0.1%) (Fig. S13†). Leaving out the fluorescence information, we found that the performance of NSCLC detection by using the label-free optophysical phenotypes alone (accuracy of 98.5%) is on par with that solely based on the EpCAM expression (accuracy of 99.7%) (Fig. S13†). This suggests that the optophysical phenotypes, to a certain degree, correlate with the EpCAM expression, as also reflected from a linear regression analysis (Spearman's correlation coefficient of $r = 0.7464$) (Fig. 3d). Ranking the individual features in terms of their correlations to the EpCAM expression, we found that 7 of the top 10 features relate to the single-cell global and local textures of mass density, whereas the remaining are the bulk features, namely cell size and mass (Fig. 3e). It signifies the importance of subcellular mass density textures in correlating with the EpCAM expression.

To further evaluate the detection sensitivity, we performed a series of spike-in tests in which anti-EpCAM-labeled-H2170 cells are spiked in human PBMC from healthy donors at three different proportions of 1:1000, 1:10 000 and 1:100 000 (Fig. 3f). Here we modified the deep neural network, which was trained on the pure populations of PBMC and H2170, to detect the rare H2170 cells in the three spike-in cases. In general, the label-free optophysical phenotypic profiling allowed us to faithfully distinguish the rare populations of H2170 in all 3 cases (see the visualizations by PCA in Fig. 3g). We also identified that the local and global optical and mass density textures are the crucial features for classifying H2170 and PBMC, apart from the commonly used cell size or the easily measurable biophysical features

(Fig. 3f). Indeed, our results show that the specificity of H2170 detection was greatly improved when the neural network used the entire optophysical profile compared to the case using the cell-size only (increased from 89% to 98% for the spike ratio of 1:10 000; and from 91% to 98% for the spike ratio of 1:100 000) or the easily measurable features (increased from 90% to 98% for the spike ratio of 1:10 000; and from 91% to 98% for the spike ratio of 1:100 000). Notably, the high-dimensional single-cell optophysical profile reduces ~10 000 false positives in the case of 1:100 000 ratio, as compared to the size-based-only detection (Fig. 3f).

Single-cell optophysical spatial profile predicts sensitivity to targeted drug treatment

We further tested whether label-free single-cell spatial optophysical phenotyping offers sufficient molecular sensitivity to assess efficacy of targeted drug response. Specifically, we characterized the optophysical response of the NSCLC (H1975, which is known to harbor T790M mutation) treated with osimertinib, the third-generation epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) for NSCLC patients with EGFR T790M mutation (Fig. 4a).^{44–46} In parallel with the multi-ATOM imaging experiments (Fig. 4b), we validated the dose response of H1975 to osimertinib in terms of the inhibition of expression of phosphorylated EGFR (pEGFR) by osimertinib across the concentration from 1 nM to 1000 nM (Fig. 4c). It was consistent with the previously reported experiments and its known mechanism of action.^{44,45}

Based on the single-cell multi-ATOM images captured under different drug concentrations, H1975 cells however did not show any observable change in morphology in both the single-cell optical density and mass density (Fig. 4b). We thus turned to investigate the significance of the spatial optophysical features, derived from the multi-ATOM images, based on the Spearman's correlations with the different concentrations. We visualized the high-dimensional label-free single-cell data (from all different concentrations) in a 2D plot using a dimensionality-reduction method, called UMAP (Fig. 4d).^{47,48} We observed that the cells generally migrated from the right to the left side of the UMAP plot according to the increasing concentration of osimertinib. It indicated the change in the single-cell optophysical profiles when the cells were exposed to this targeted drug for only 6 hours. We note that the effect on the label-free single-cell optophysical spatial profile preceded the common drug assays based on proliferation measurements (72 hours) (see the cytotoxicity assay in Fig. S14†),⁴⁵ and generally occurs as soon as the others using the molecular readouts, *e.g.* EGFR phosphorylation (2–6 hours).⁴⁴

Having profiled the single-cell spatial optophysical features, we found that both local and global features are among the top 10 features most correlated to the drug response (*i.e.*, 7 are local texture features whereas 3 are global texture features of single-cell optical and mass density) (Fig. 4e and S15†). We further

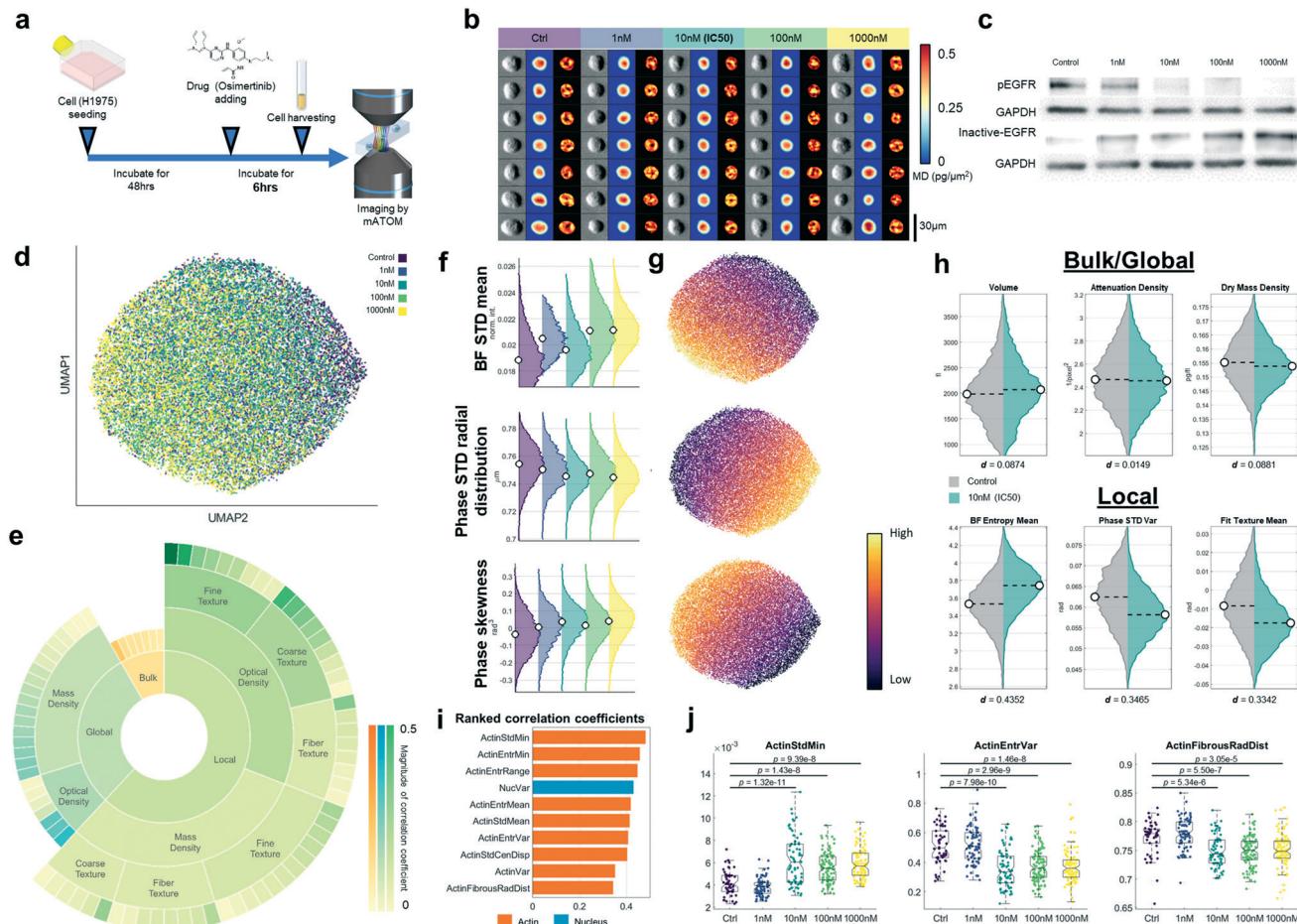


Fig. 4 Single-cell spatial optophysical profile is sensitive to targeted drug treatment. (a) Experimental workflow. (b) Single-cell spatial maps captured by multi-ATOM at different concentrations of osimertinib treatment. (c) Immunoblotting results obtained from osimertinib treatment for 72 hours. The top row is immunoblotted for phosphorylated-EGFR. The 3rd row is immunoblotted for inactive-EGFR. The rows below them (GAPDH) are their respective loading controls. (d) UMAP visualization based on the optophysical phenotypes. (e) A sunburst chart showing the significance of the hierarchical spatial optophysical phenotypes in label-free assay of targeted drug treatment. The ranking arrangement is the same as Fig. 3c (see Fig. S15† for detailed feature labels). (f) Changes in the distributions of the spatial optophysical phenotypes in response to osimertinib treatment. The colors of the distributions follow the indication in (d). (g) The same UMAP as in (d), color-coded with the expression levels of selected phenotypes shown in (f). (h) Performance comparison among bulk/global texture (top) and local texture features (bottom) in distinguishing the drugged cells (at 10 nM (IC_{50})) from control. d stands for Cliff's delta used in the effect size analysis. (i) Top 10 spatial features (i.e., actin, nucleus) obtained from fluorescence images (Fig. S16†) that are most correlated to the drug treatment (see also Fig. S17†). (j) The distributions of selected fluorescent images-based features according to different drug treatment conditions. Mann-Whitney U -test was performed between conditions.

identified the clear trends of these texture features (both global and local textures of optical and mass density, *i.e.*, BF STD Mean, DMD Radial Distribution, and Phase skewness) on the UMAP plot that can be mapped to different concentrations of osimertinib (Fig. 4f and g).

To further gauge the significance of these single-cell optophysical textural features, we compared the feature distributions under the control condition and treatment at 10 nM, which is near the IC_{50} (half maximal inhibitory concentration) determined by the viability assay (Fig. S14†). We observed the sizeable changes in the texture-based features (*e.g.*, BF entropy mean, phase STD var and fit texture mean) by the drug treatment (mean effect size of 0.372). By contrast, osimertinib had no significant effect on the

commonly adopted bulk features,^{49,50} *i.e.*, cell size, mass and optical density (mean effect size of 0.063) (Fig. 4h).

Furthermore, by adopting the fluorescence image-based profiling strategy,⁵¹ we investigated whether changes in the subcellular organization, such as nucleus and cytoskeleton, in response to the same drug treatment could be detected (Fig. S16†). We applied the same hierarchical spatial profiling strategy on the fluorescently-labelled nucleus and actin cytoskeleton images and identified that the changes in the subcellular textures of mass density correlate more closely with the textural change of actin cytoskeleton (9 of the top 10 features are related to actin cytoskeleton) (Fig. 4i and S17†). This result suggests the possibility of using spatial changes in optophysical properties (*e.g.*, mass density) in single-cells

to detect changes in subcellular organization *e.g.*, in the cytoskeleton in response to perturbations.

Conclusion

We have presented an important advance in the massive analysis of single-cell intrinsic optophysical properties at levels of sensitivity and specificity that have been inconceivable in the current cell-based assays. This is enabled by a label-free image-based profiling approach that creates a list of single-cell spatial optophysical descriptors in a hierarchical manner. A key factor to this achievement is the combination of ultrahigh-throughput single-cell imaging (enabled by multi-ATOM combined with synchronized fluorescence cytometry) and automated image analytics (powered by deep neural network and transfer learning). This integrative method enabled the high statistical power that robustly delineated the label-free signatures of the lung cancer cell subtypes from more than a million single-cells. Furthermore, this approach achieved high *label-free* sensitivity of detecting rare lung cancer cell populations (10^{-5}) in human PBMCs with implications for the detection of circulating cancer cells. Our findings could underpin the potential utility of this automated, and high-throughput optophysical single-cell profiling strategy in the liquid-biopsy-based detection and classification of lung cancer, minimizing laborious and costly sample preparations mandated in the staining-based histopathological and cytological examinations.^{52,53} We note that the strategy of transfer learning could also be used for correcting and aligning the phenotypic profiles from different batches. While this has been applied in other single-cell data types (notably single-cell RNA-sequencing data),⁵⁴ we anticipate that similar approach would potentially be applicable to the single-cell image data. In contrast to the common biophysical readouts using the bulk single-cell mass and size,^{50,55} we found that the subcellular optical density and mass density textures are more effective in assessing the drug sensitivity of single NSCLC cells. Its *label-free* sensitivity to predict targeted treatment could also inspire new functional assays for identification of new therapeutics (small-molecule and biologics) and assessment of drug efficacy. We also envision that the hierarchical optophysical phenotyping methods could potentially also be harnessed to interpret the black-box operation of the deep learning models, such as convolutional neural network (CNN)^{56,57} or multilayer perceptron (MLP)⁵⁸ (*e.g.* building the hierarchical graphs across layers in the network that are interpretable⁵⁹). We argue this is an important step to make deep learning a credible element in imaging cytometry used in biological research and clinical diagnosis.

We further demonstrated the compatibility of this image-based profiling workflow with fluorescence flow cytometry in which synchronous acquisition of optophysical and biochemical markers of the same cells. This is particularly useful for the exploration of the correlations between

optophysical and biochemical phenotypes. Linking these two types of single-cell data opens a new dimension for better defining specific cell types, assessing population heterogeneity, and comprehending the biophysical–biochemical-regulatory relationships in individual cells. This coupled single-cell information could also be applied in enrichment and detection of rare CTCs, which are regarded as a prognostic cancer marker for guiding treatment decisions.^{38–40} Despite of these, to fully exploit the potential of transfer learning for robust cytometry, especially making it adaptable to the clinical setting, it is necessary to conduct a more elaborated survey of diverse types of datasets (*e.g.* different cell types) so that it could accelerate training convergence and improve the generalization of the model. We also note that new algorithmic developments in this aspect could potentially be inspired by a variety of single-cell analysis methods (*e.g.* single-cell RNA-sequencing) (*e.g.* ref. 60 and 61). Overall, this work could spearhead the development of cost-effective label-free alternative of functional image-based cell assay to its fluorescence counterpart that could help identify the underexploited sets of optophysical biomarkers and their roles in homeostasis and pathogenesis of the complex biological systems.

Methods and materials

Multi-ATOM system with synchronized fluorescence detection

The optical system consists of two integrated parts: the multi-ATOM and fluorescence detection module (Fig. S1†). Without using dedicated interferometry, multi-ATOM acquires the complex-field image information at high speed by optical time-stretch combined with multiplexed differential phase-gradient contrast encoding. A custom fiber-based pulsed laser (bandwidth: ~10 nm; repetition rate: 11 MHz; centered wavelength: 1064 nm), followed by a dispersive optical fiber (group-velocity dispersion (GVD): 1.78 ns nm⁻¹) was employed to create an ultrafast wavelength-swept light source for the system. A diffraction grating was then used to transform the single-beam of wavelength-swept light into a 1D scanning line beam that was projected onto the cells flowing in a microfluidic channel, which was oriented orthogonal to the line beam. Having passed through the cells, the scanning beam was then transformed back to a single beam, which contained the phase-gradient information of the cells.^{62,63} Subsequently, the beam was directed to the photodetection path through a multiplexed phase-gradient encoding module. In this module, the beam is split into four beam paths, along which four knife edges are positioned respectively to partially block the beam from different directions (*i.e.*, up, down, left and right). This procedure effectively generates four different phase-gradient contrasts in each single line-scan for complex-field retrieval. The four beams were then combined and coupled into a fiber-based optical delay line, such that the four copies were time multiplexed and detected in real-time by a high-speed

single-pixel photodetector (electrical bandwidth = 12 GHz (Newport, US)). In the system backend, we used a real-time field programmable gate array (FPGA) based signal processing system (electrical bandwidth = 2 GHz, sampling rate = 4 GSa s⁻¹), on which custom logic was implemented to automatically detect and segment cells from the digitized data stream, at a processing throughput equivalent to >10 000 cell per s. All segmented cell images (four different gradient-contrast contrasts per cell) were sent through four 10G Ethernet links and were stored by four data storage nodes with a total memory capacity of over 800 GB. For each cell, the 2D complex-field information (*i.e.*, bright-field and quantitative phase) were retrieved from the four different phase-gradient contrasts based on an algorithm using complex Fourier integration. Detailed algorithm can be referred to ref. 25.

In the fluorescence detection module, two continuous wave (CW) lasers (wavelength: 488 nm and 532 nm) were employed to generate line-shaped fluorescence excitation, that were spatially overlapped with the multi-ATOM illumination (Fig. 1). The two epi-fluorescence signals were detected by two photomultiplier tubes (PMT) separately. In the analog electronics backend, we multiplexed the PMT-detected signals by frequency modulation (11.8 MHz and 35.4 MHz respectively, using a multichannel direct digital synthesizer). The multiplexed signals were then separated by digital demodulation and low-pass filtering (Fig. S2†). The same FPGA was configured to synchronously obtain the signal from multi-ATOM and fluorescence detection from each single cell at high-speed. As the required number of fluorescence channels (and thus PMTs) increases, only one analog-to-digital convertor (ADC) is needed for the frequency multiplexing implementation. This is important for cost-effective high-speed data acquisition as the high-bandwidth (>GHz) ADC module is not as scalable as photodetectors.

Microfluidic channel fabrication

Using the standard soft lithography technique, the microfluidic channel was fabricated by curing polydimethylsiloxane (PDMS) on a silicon wafer mold. We first used a spin coater (spinNXG-P1, Apex Instruments Co., India) to coat a layer of photoresist (SU-82025, MicroChem, US) on a silicon wafer, followed by a two-step soft-bake (at 65 °C for 3 minutes and then at 95 °C for 6 minutes). Upon cooling under the ambient temperature, we used a maskless soft lithography machine (SF-100 XCEL, Intelligent Micro Patterning, LLC, US) to pattern the photoresist with a computer-aided design (CAD)(exposure for 4 seconds). The exposed photoresist was then post-baked (for 1 minute at 65 °C and then 6 minutes at 95 °C) and developed with the SU-8 developer (MicroChem, US) for 5 minutes. After the rinsing and drying step, the PDMS precursor (SYLGARD® 184 Silicone Elastomer kit, Dow Corning, US) was mixed with the curing agent (with a ratio at 10:1) before pouring onto the silicon wafer. A custom-designed acrylic block was placed on

the silicon wafer to control the channel height of the imaging section. The channel was then cured in an oven at 65 °C for 2 hours before demolding. Two holes are then punched (Miltex 33-31 AA, Integra LifeSciences, US) for inserting plastic tubings (BB31695-PE/2, Scientific Commodities, Inc., US) to the inlet and outlet of the microfluidic channel. The channel and a glass slide were then bonded with oxygen plasma (PDC-002, Harrick Plasma, US), followed by an oven bake at 65 °C for 30 minutes to strengthen the bonding. Note that the serpentine structure (prior to the imaging section of the microfluidic channel in order to create inertial focusing effect, which optimizes the balance between the inertial lift force and the viscous drag force³⁰ – the key to achieving robust in-focus single-cell imaging under a fast microfluidic flow (>1 m s⁻¹). The serpentine structure consists of 8 repeated units, each of which has one small turn (channel width: 0.15 mm; radius of curvature: 0.4 mm) and one large turn (channel width: 0.42 mm; radius of curvature: 1 mm). The channel dimension at the imaging section is 30 mm × 60 mm (height × width) (yellow arrow in Fig. 1).

Automated label-free cell-type classification and transfer learning

We employed a deep neural network model, which has three fully-connected hidden layers of 100, 50 and 25 nodes respectively,⁶⁴ and used the hierarchical optophysical phenotypes as inputs (see ESI† Methods). A rectified unit function was used as the activation function between them, while the softmax function was used at the output layer. Cross-entropy function was selected as a loss function. To ensure robust optophysical single-cell analysis, we took into account the batch variation effect in our phenotyping pipeline. Batch effect accounts for the technical sources of data variation (*e.g.* image focus conditions, system drift, or variations in laser power and photodetector sensitivity) due to measurement in different setups or/and at different times, which is known to compromise genuine data interpretation and analysis in cytometry. To address batch variation effects in the single-cell optophysical data (Fig. S4†), we incorporated transfer learning in the neural network model. This was done by first pre-training the neural network model with the large datasets from 2 different batches. The model was then trained on additional but smaller datasets (5% of the first training dataset) from the third batch.

Cell lines and culture

Three adenocarcinoma cell lines (H358 (EGFR WT), HCC827 (EGFR exon 19 del) and H1975 (L858R and T790M)) two squamous cell carcinoma cell lines (H520 and H2170), two small cell lung cancer cell lines (H526 and H69) were obtained from American Type Culture Collection (ATCC) and authenticated using the Human STR profiling cell authentication service. They were expanded and cultured in the tissue culture flasks (surface area of 75 cm²) (TPP). The full culture medium was ATCC modified RPMI-1640 (Gibco)

supplemented with 10% fetal bovine serum (FBS) (Gibco) and 1% antibiotic–antimycotic (Gibco). The cells were placed in a CO₂ incubator at 37 °C and 5% CO₂. Passage or change of medium was done 2–3 times a week depending on cell confluence.

Preparation of PBMC

Human buffy coat was provided by the Hong Kong Red Cross. Written consents for clinical care and research purposes were all obtained from the donors. The research protocol was approved by the Institutional Review Board of the University of Hong Kong (IRB Reference No.: UW 17-219) and complied with the Declaration of Helsinki and acts in accordance to ICH GCP guidelines, local regulations and Hospital Authority and the University policies. Buffy coat, 1× PBS and Ficoll were pre-warmed to room temperature. 5 mL of buffy coat was then mixed with 5 mL of PBS in a 50 mL centrifuge tube. 5 mL of Ficoll was layered on top carefully to avoid mixing with the solution below. The solution was centrifuged under 400g for 20 minutes producing 5 distinct layers in the centrifuge tube. The second layer from the top which corresponds to PBMCs was then carefully extracted using a 1 mL pipette tip. Next, the extracted PBMCs were rinsed with 1× PBS once by centrifuging under 200g for 5 minutes and resuspended in fresh 1× PBS.

Live-cell immunofluorescence labeling

Suspended cell solution was centrifuged and washed with 1X PBS once. 1e6 cells were extracted and topped up to 1 mL with 1X PBS. Then, 5 µL of PE-conjugated anti-EpCAM monoclonal antibody (12-9326-42, eBioscience) was added to the 1 × 10⁶ cells. The cells were incubated at room temperature for 30 minutes. The cells were then washed with 1× PBS once and resuspended in 3 mL of full medium until the imaging experiments began.

Osimertinib treatment

Osimertinib (Axon Medchem) was dissolved in 1× PBS at 10 µM. It was stored in -20 °C and used within 2 weeks. The H1975 cell line was seeded in the 6-well plates with a density of 10⁵ cells per well and was cultured in full medium for 24 hours. After being replaced by the fresh medium. Each well was added with the desired amount of osimertinib and 1× PBS to attain the desired concentrations *i.e.*, 1 nM, 10 nM, 100 nM and 1000 nM. The control well was added with 1× PBS. The 6-well plate was then incubated for 6 hours until being harvested for imaging.

Western blot

We seeded the cells and cultured them for 48 hours. They were then treated with the selected concentrations of osimertinib for 72 hours. Then, the cells were harvested with trypsinization. After washing with ice-cold 1× PBS thrice, the cells were resuspended in a cocktail containing RIPA buffer

(Abcam), protease and phosphatase inhibitor (Thermo Scientific), and were then stored in ice for 4 hours for digestion. They were then centrifuged at 12 000 rpm to remove the cell debris and thus to obtain the solution containing the protein samples. Protein concentration of each sample was determined using the BCA assay kit (Pierce). The assay was performed according to the protocol specified by the manufacturer. The gel for SDS-PAGE consisted of the stacking part (4%) and the resolving part (10%). The stacking gel was prepared by mixing 3 mL of distilled water, 1.25 mL of 0.5 M Tris-HCl, 0.05 mL of 10% SDS, 0.67 mL of acrylamide, 0.027 mL of APS and 7 µL of TEMED. The resolving gel was prepared by mixing 3 mL of distilled water, 1.875 mL of 1.5 M Tris-HCl, 0.075 mL of 10% SDS, 2.5 mL of acrylamide, 0.04 mL of APS and 10 µL of TEMED. Equal amounts of protein were loaded into different lanes. SDS-PAGE and the transference of protein to polyvinylidene difluoride membrane were performed using the BIO-RAD system. Interested region of the membrane was cut out and put into 5% BSA for blocking. After that, appropriate part of the membrane was blotted with phospho-EGFR antibody (ab32430, Abcam) and EGFR antibody (ab52894, Abcam), GAPDH (MA5-15738, Thermo Fisher) was blotted as loading control. Finally, specific horseradish peroxidase (HRP)-conjugated secondary antibodies (W401B, W402B, Promega). The signals of the target proteins on the blots were detected using enhanced chemiluminescence (Azure Biosystems).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The work is partially supported by the Research Grants Council of the Hong Kong Special Administrative Region of China (HKU 17208918, 17209017, 17259316, C7047-16G), Innovation and Technology Commission - Hong Kong (ITS/204/18), and the University Development Funds of the University of Hong Kong. We would like to acknowledge Prof. Barbara P. Chan, Mr. Johnny Wong and Dr. Vincent Lau from the Mechanical engineering department of the University of Hong Kong for supporting related wet lab experiments. We would also like to acknowledge the contribution of our research assistants, particularly Ms. Maitraee Mistry in helping to maintain the cell culture.

References

- 1 J. Caiceo, S. Cooper and F. Heigwer, Data-analysis strategies for image-based cell profiling, *Nat. Methods*, 2017, **14**, 849–863.
- 2 P. J. Thul, L. Åkesson, M. Wiklund, D. Mahdessian, A. Geladaki, H. A. Blal, T. Alm, A. Asplund, L. Björk and L. M. Breckels, A subcellular map of the human proteome, *Science*, 2017, **356**(6340), eaal3321.

- 3 P.-H. Wu, J. M. Phillip, S. B. Khatau, W.-C. Chen, J. Stirman, S. Rosseel, K. Tschudi, J. Van Patten, M. Wong and S. Gupta, Evolution of cellular morpho-phenotypes in cancer metastasis, *Sci. Rep.*, 2015, **5**(1), 1–10.
- 4 O. Otto, P. Rosendahl, A. Mietke, S. Golfier, C. Herold, D. Klaue, S. Girardo, S. Pagliara, A. Ekpenyong and A. Jacobi, Real-time deformability cytometry: on-the-fly cell mechanical phenotyping, *Nat. Methods*, 2015, **12**(3), 199–202.
- 5 T. Henry, D. R. Gossett, Y. S. Moon, M. Masaeli, M. Sohsman, Y. Ying, K. Mislick, R. P. Adams, J. Rao and D. Di Carlo, Quantitative diagnosis of malignant pleural effusions by single-cell mechanophenotyping, *Sci. Transl. Med.*, 2013, **5**(212), 212ra163.
- 6 M. Lippeveld, C. Knill, E. Ladlow, A. Fuller, L. J. Michaelis, Y. Saeys, A. Filby and D. Peralta, Classification of Human White Blood Cells Using Machine Learning for Stain-Free Imaging Flow Cytometry, *Cytometry, Part A*, 2019, **97**(3), 308–319.
- 7 T. Blasi, H. Hennig, H. D. Summers, F. J. Theis, J. Cerveira, J. O. Patterson, D. Davies, A. Filby, A. E. Carpenter and P. Rees, Label-free cell cycle analysis for high-throughput imaging flow cytometry, *Nat. Commun.*, 2016, **7**(1), 1–9.
- 8 N. Toepfner, C. Herold, O. Otto, P. Rosendahl, A. Jacobi, M. Kräter, J. Stächele, L. Menschner, M. Herbig and L. Ciuffreda, Detection of human disease conditions by single-cell morpho-rheological phenotyping of blood, *Elife*, 2018, **7**, e29213.
- 9 Y. Park, C. Depersinge and G. Popescu, Quantitative phase imaging in biomedicine, *Nat. Photonics*, 2018, **12**(10), 578–589.
- 10 T. A. Zangle and M. A. Teitel, Live-cell mass profiling: an emerging approach in quantitative biophysics, *Nat. Methods*, 2014, **11**(12), 1221.
- 11 M. Mir, Z. Wang, Z. Shen, M. Bednarz, R. Bashir, I. Golding, S. G. Prasanth and G. Popescu, Optical measurement of cycle-dependent cell growth, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(32), 13124–13129.
- 12 G. Popescu, Y. Park, N. Lue, C. Best-Popescu, L. Deflores, R. R. Dasari, M. S. Feld and K. Badizadegan, Optical imaging of cell mass and growth dynamics, *Am. J. Physiol.*, 2008, **295**(2), C538–C544.
- 13 W. Palm and C. B. Thompson, Nutrient acquisition strategies of mammalian cells, *Nature*, 2017, **546**(7657), 234–242.
- 14 R. J. Ellis, Macromolecular crowding: obvious but underappreciated, *Trends Biochem. Sci.*, 2001, **26**(10), 597–604.
- 15 M. Al-Habibi, Macromolecular crowding and its role as intracellular signalling of cell volume regulation, *Int. J. Biochem. Cell Biol.*, 2001, **33**(9), 844–864.
- 16 R. J. Kimmerling, S. M. Prakadan, A. J. Gupta, N. L. Calistri, M. M. Stevens, S. Olcum, N. Cermak, R. S. Drake, K. Pelton and F. De Smet, Linking single-cell measurements of mass, growth rate, and gene expression, *Genome Biol.*, 2018, **19**(1), 1–13.
- 17 W. H. Grover, A. K. Bryan, M. Diez-Silva, S. Suresh, J. M. Higgins and S. R. Manalis, Measuring single-cell density, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(27), 10992–10996.
- 18 J.-M. Park, J.-Y. Lee, J.-G. Lee, H. Jeong, J.-M. Oh, Y. J. Kim, D. Park, M. S. Kim, H. J. Lee and J. H. Oh, Highly efficient assay of circulating tumor cells by selective sedimentation with a density gradient medium and microfiltration from whole blood, *Anal. Chem.*, 2012, **84**(17), 7400–7407.
- 19 F. Feijó Delgado, N. Cermak, V. C. Hecht, S. Son, Y. Li, S. M. Knudsen, S. Olcum, J. M. Higgins, J. Chen and W. H. Grover, Intracellular water exchange for measuring the dry mass, water mass and changes in chemical composition of living cells, *PloS One*, 2013, **8**(7), e67590.
- 20 S. Byun, V. C. Hecht and S. R. Manalis, Characterizing cellular biophysical responses to stress by relating density, deformability, and size, *Biophys. J.*, 2015, **109**(8), 1565–1573.
- 21 B. Guo, C. Lei, H. Kobayashi, T. Ito, Y. Yalikun, Y. Jiang, Y. Tanaka, Y. Ozeki and K. Goda, High-throughput, label-free, single-cell, microalgal lipid screening by machine-learning-equipped optofluidic time-stretch quantitative phase microscopy, *Cytometry, Part A*, 2017, **91**(5), 494–502.
- 22 D. Roitshtain, L. Wolbromsky, E. Bal, H. Greenspan, L. L. Satterwhite and N. T. Shaked, Quantitative phase microscopy spatial signatures of cancer cells, *Cytometry, Part A*, 2017, **91**(5), 482–493.
- 23 M. Ugele, M. Weniger, M. Stanzel, M. Bassler, S. W. Krause, O. Friedrich, O. Hayden and L. Richter, Label-Free High-Throughput Leukemia Detection by Holographic Microscopy, *Adv. Sci.*, 2018, **5**(12), 1800761.
- 24 K. C. Lee, M. Wang, K. S. Cheah, G. C. Chan, H. K. So, K. K. Wong and K. K. Tsia, Quantitative phase imaging flow cytometry for ultra-large-scale single-cell biophysical phenotyping, *Cytometry, Part A*, 2019, **95**(5), 510–520.
- 25 K. C. Lee, A. K. Lau, A. H. Tang, M. Wang, A. T. Mok, B. M. Chung, W. Yan, H. C. Shum, K. S. Cheah and G. C. Chan, Multi-ATOM: Ultrahigh-throughput single-cell quantitative phase imaging with subcellular resolution, *J. Biophotonics*, 2019, **12**(7), e201800479.
- 26 Y. Gu, A. C. Zhang, Y. Han, J. Li, C. Chen and Y.-H. Lo, Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification, *Cytometry, Part A*, 2019, **95**(5), 499–509.
- 27 K. Goda, K. Tsia and B. Jalali, Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena, *Nature*, 2009, **458**(7242), 1145–1149.
- 28 A. K. Lau, H. C. Shum, K. K. Wong and K. K. Tsia, Optofluidic time-stretch imaging—an emerging tool for high-throughput imaging flow cytometry, *Lab Chip*, 2016, **16**(10), 1743–1756.
- 29 A. Mahjoubfar, D. V. Churkin, S. Barland, N. Broderick, S. K. Turitsyn and B. Jalali, Time stretch and its applications, *Nat. Photonics*, 2017, **11**(6), 341.
- 30 J. M. Martel and M. Toner, Inertial focusing in microfluidics, *Annu. Rev. Biomed. Eng.*, 2014, **16**, 371–396.
- 31 R. Barer, Interference microscopy and mass determination, *Nature*, 1952, **169**(4296), 366–367.
- 32 H. Ding, Z. Wang, F. Nguyen, S. A. Boppart and G. Popescu, Fourier transform light scattering of inhomogeneous and dynamic structures, *Phys. Rev. Lett.*, 2008, **101**(23), 238102.
- 33 D. Hsu and S. M. Kakade, in Learning mixtures of spherical gaussians: moment methods and spectral decompositions, *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, 2013, pp. 11–20.

- 34 D. W. Travis, G. E. Brambilla, M. A. Nicholson, B. Y. Yatabe, R. J. H. Austin, B. M. Beasley, R. L. Chiriac, M. S. Dacic, A. E. Duhig, S. D. Flieder, S. K. Geisinger, S. F. Hirsch, S. Y. Ishikawa, S. K. Kerr, S. M. Noguchi, S. G. Pelosi, S. C. Powell, S. M. Tsao and S. I. Wistuba, The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification, *J. Thorac. Oncol.*, 2015, **10**(9), 1243–1260.
- 35 J. Yosinski, J. Clune, Y. Bengio and H. Lipson, in How transferable are features in deep neural networks?, *Advances in neural information processing systems*, 2014, pp. 3320–3328.
- 36 T. Wang, T. S. Johnson, W. Shao, Z. Lu, B. R. Helm, J. Zhang and K. Huang, BERMUDA: a novel deep transfer learning method for single-cell RNA sequencing batch correction reveals hidden high-resolution cellular subtypes, *Genome Biol.*, 2019, **20**(1), 1–15.
- 37 G. Sullivan and R. Feinn, Using effect size-or why the p value is not enough, *J. Grad. Med. Educ.*, 2012, **4**(3), 279–282.
- 38 M. G. Krebs, R. Sloane, L. Priest, L. Lancashire, J.-M. Hou, A. Greystoke, T. H. Ward, R. Ferraldeschi, A. Hughes, G. Clack, M. Ranson, C. Dive and F. H. Blackhall, Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer, *J. Clin. Oncol.*, 2011, **29**(12), 1556–1563.
- 39 A. Fiorelli, M. Accardo, E. Carelli, D. Angioletti, M. Santini and M. Di Domenico, Circulating tumor cells in diagnosing lung cancer: clinical and morphologic analysis, *Ann. Thorac. Surg.*, 2015, **99**(6), 1899–1905.
- 40 J. Kapeleris, A. Kulasinghe, M. E. Warkiani, I. Vela, L. Kenny, K. O'Byrne and C. Punyadeera, The prognostic role of circulating tumor cells(CTCs) in lung cancer, *Front. Oncol.*, 2018, **8**, 311.
- 41 C. Alix-Panabières and K. Pantel, Challenges in circulating tumour cell research, *Nat. Rev. Cancer*, 2014, **14**(9), 623–631.
- 42 E. E. van der Toom, J. E. Verdone, M. A. Gorin and K. J. Pienta, Technical challenges in the isolation and analysis of circulating tumor cells, *Oncotarget*, 2016, **7**(38), 62754.
- 43 E. Lin, T. Cao, S. Nagrath and M. R. King, Circulating Tumor Cells: Diagnostic and Therapeutic Applications, *Annu. Rev. Biomed. Eng.*, 2018, **20**(1), 329–352.
- 44 D. A. Cross, S. E. Ashton, S. Ghiorgiu, C. Eberlein, C. A. Nebhan, P. J. Spitzler, J. P. Orme, M. R. V. Finlay, R. A. Ward and M. J. Mellor, AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer, *Cancer Discovery*, 2014, **4**(9), 1046–1061.
- 45 Z.-H. Tang, X.-M. Jiang, X. Guo, C. M. V. Fong, X. Chen and J.-J. Lu, Characterization of osimertinib(AZD9291)-resistant non-small cell lung cancer NCI-H1975/OSIR cell line, *Oncotarget*, 2016, **7**(49), 81598.
- 46 J.-C. Soria, Y. Ohe, J. Vansteenkiste, T. Reungwetwattana, B. Chewaskulyong, K. H. Lee, A. Dechaphunkul, F. Imamura, N. Nogami and T. Kurata, Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer, *N. Engl. J. Med.*, 2018, **378**(2), 113–125.
- 47 L. McInnes, J. Healy and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, 2018, arXiv preprint arXiv:1802.03426.
- 48 E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux and E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP, *Nat. Biotechnol.*, 2019, **37**(1), 38.
- 49 M. Mir, A. Bergamaschi, B. S. Katzenellenbogen and G. Popescu, Highly sensitive quantitative imaging for monitoring single cancer cell growth kinetics and drug response, *PloS One*, 2014, **9**(2), E89000.
- 50 D. Huang, K. A. Leslie, D. Guest, O. Yeshcheulova, I. J. Roy, M. Piva, G. Moriceau, T. A. Zangle, R. S. Lo and M. A. Teitell, High-Speed Live-Cell Interferometry: A New Method for Quantifying Tumor Drug Resistance and Heterogeneity, *Anal. Chem.*, 2018, **90**(5), 3299–3306.
- 51 J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington and A. E. Carpenter, Data-analysis strategies for image-based cell profiling, *Nat. Methods*, 2017, **14**(9), 849–863.
- 52 C. Fitzmaurice, D. Dicker, A. Pain, H. Hamavid, M. Moradi-Lakeh, M. F. MacIntyre, C. Allen, G. Hansen, R. Woodbrook and C. Wolfe, The global burden of cancer 2013, *JAMA Oncol.*, 2015, **1**(4), 505–527.
- 53 F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *Ca-Cancer J. Clin.*, 2018, **68**(6), 394–424.
- 54 I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh and S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony, *Nat. Methods*, 2019, **16**(12), 1289–1296.
- 55 M. M. Stevens, C. L. Maire, N. Chou, M. A. Murakami, D. S. Knoff, Y. Kikuchi, R. J. Kimmerling, H. Liu, S. Haidar and N. L. Calistri, Drug sensitivity of single cancer cells is predicted by changes in mass accumulation rate, *Nat. Biotechnol.*, 2016, **34**(11), 1161–1167.
- 56 Y. Zhou, A. Yasumoto, C. Lei, C.-J. Huang, H. Kobayashi, Y. Wu, S. Yan, C.-W. Sun, Y. Yatomi and K. Goda, Intelligent classification of platelet aggregates by agonist type, *Elife*, 2020, **9**, e52938.
- 57 N. Nitta, T. Sugimura, A. Isozaki, H. Mikami, K. Hiraki, S. Sakuma, T. Iino, F. Arai, T. Endo and Y. Fujiwaki, Intelligent image-activated cell sorting, *Cell*, 2018, **175**(1), 266–276. e13.
- 58 A. A. Nawaz, M. Urbanska, M. Herbig, M. Nötzel, M. Kräter, P. Rosendahl, C. Herold, N. Toepfner, M. Kubánková, R. Goswami, S. Abuhattum, F. Reichel, P. Müller, A. Taubenberger, S. Girardo, A. Jacobi and J. Guck, Intelligent image-based deformation-assisted cell sorting with molecular specificity, *Nat. Methods*, 2020, **17**(6), 595–599.
- 59 Q. Zhang, R. Cao, F. Shi, Y. N. Wu and S.-C. Zhu, in Interpreting cnn knowledge via an explanatory graph, *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

- 60 A. Butler, P. Hoffman, P. Smibert, E. Papalexis and R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, *Nat. Biotechnol.*, 2018, **36**(5), 411–420.
- 61 L. Haghverdi, A. T. Lun, M. D. Morgan and J. C. Marioni, Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors, *Nat. Biotechnol.*, 2018, **36**(5), 421–427.
- 62 T. T. Wong, A. K. Lau, K. K. Ho, M. Y. Tang, J. D. Robles, X. Wei, A. C. Chan, A. H. Tang, E. Y. Lam and K. K. Wong, Asymmetric-detection time-stretch optical microscopy(ATOM) for ultrafast high-contrast cellular imaging in flow, *Sci. Rep.*, 2014, **4**, 3656.
- 63 A. H. L. Tang, Q. T. K. Lai, B. M. F. Chung, K. C. M. Lee, A. T. Y. Mok, G. K. Yip, A. H. C. Shum, K. K. Y. Wong and K. K. Tsia, Microfluidic Imaging Flow Cytometry by Asymmetric-detection Time-stretch Optical Microscopy(ATOM), *J. Visualized Exp.*, 2017(124), e55840.
- 64 F. Ma and M. Pellegrini, ACTINN: automated identification of cell types in single cell RNA sequencing, *Bioinformatics*, 2020, **36**(2), 533–538.