Quantitative Phase Microscopy Spatial Signatures of Cancer Cells

Darina Roitshtain,1 Lauren Wolbromsky,1 Evgeny Bal,1 Hayit Greenspan,1 Lisa L. Satterwhite,2 Natan T. Shaked1*

Abstract

We present cytometric classification of live healthy and cancerous cells by using the spatial morphological and textural information found in the label-free quantitative phase images of the cells. We compare both healthy cells to primary tumor cells and primary tumor cells to metastatic cancer cells, where tumor biopsies and normal tissues were isolated from the same individuals. To mimic analysis of liquid biopsies by flow cytometry, the cells were imaged while unattached to the substrate. We used low-coherence off-axis interferometric phase microscopy setup, which allows a single-exposure acquisition mode, and thus is suitable for quantitative imaging of dynamic cells during flow. After acquisition, the optical path delay maps of the cells were extracted and then used to calculate 15 parameters derived from the cellular 3D morphology and texture. Upon analyzing tens of cells in each group, we found high statistical significance in the difference between the groups in most of the parameters calculated, with the same trends for all statistically significant parameters. Furthermore, a specially designed machine learning algorithm, implemented on the phase map extracted features, classified the correct cell type (healthy/cancer/metastatic) with 81–93% sensitivity and 81–99% specificity. The quantitative phase imaging approach for liquid biopsies presented in this paper could be the basis for advanced techniques of staging freshly isolated live cancer cells in imaging flow cytometers.

Key terms
cytometry; digital holographic microscopy; interferometric imaging; machine learning; quantitative phase microscopy

Finding cancer in its early curable stages is a clear and critical unmet need. Late stage metastatic forms of cancer are almost always fatal (1,2). Many cancers are found in early stages only due to incidental or proactive screening, especially in high risk groups (3,4). However, some cancers, including pancreatic cancer and colon cancer, can develop asymptotically and thus escape early detection (5,6). The primary basis for diagnosis of cancer is evaluation of morphological changes in a tissue biopsy by a trained pathologist, a process with inherent subjectivity, performed usually after the location of the suspect tissue is known (7).

Flow cytometry of body fluids obtained by routine medical tests can identify circulating tumor cells after their separation from the other fluid contents (8,9). Interest in using circulating tumor cells found in liquid biopsies to diagnose solid tumors has exploded recently, especially to address difficulties in analyzing tumor biopsies due to intractable anatomical placement or to determine if metastatic disease is present (10–13). In 2004, a landmark study found that the number of circulating tumor cells in blood could predict survival in metastatic breast cancer (14). Cancer-specific signatures in liquid biopsies include DNA sequence, composition of cancer cell exosomes, and unique systemic response reflected by components in blood, tears, saliva or urine assayed by genomics, epigenetics, proteomics and...
metabolomics (4,15). Increasingly, liquid biopsies have been suggested as a viable and affordable mechanism to safeguard good health and to detect cancer in its asymptomatic early curable stages through proactive monitoring, especially in high risk groups.

Detection of circulating tumor cells in liquid biopsies requires a highly sensitive method to identify a small number of diseased cells in a large cell population. Isolation of these cancer cells is laborious and typically yields uniformly round cells, which are hard to stage without advanced methods (4).

During the progression of a healthy cell to immortal cancerous cell and later to metastatic cell, the biophysical and morphological phenotypes of the cell change (16). Many scientific studies have focused on cytometry of cancer cells with the goal of revealing the unique biophysical properties of these cells for cancer prognosis, which includes measuring the mechanical (17–21) or optical properties of the cells by label-based (22) or label-free (23–29) techniques. For example, multiphoton laser tomography combined with fluorescence lifetime imaging has been used to generate 30 optical parameters that are able to distinguish normal nevi from melanoma in situ with high sensitivity and specificity (30).

Although markers provide biological assays with a high degree of specificity, using markers might cause cytotoxicity by perturbing the cell environment, by influencing its behavior over time and its viability, and eventually damaging the accuracy of the test or prohibiting further clinical use of the isolated cells (31).

In flow cytometry for cell sorting, one evaluates cellular features through fluorescence markers and purifies the heterogeneous cell suspension into fractions containing a single cell type (32). However, in addition to possible cytotoxicity, suitable markers might be not available for certain cell types, and some markers might be difficult to use (33). Specifically, fluorescent markers tend to photobleach, which damages the image contrast and the prognosis results (34).

The internal morphology and texture of cancer cells changes during oncogenesis (35–37). Specifically, the intrinsic refractive index of live cells can indicate abnormal cell morphology. The cell refractive index is related to the optical interaction of the light field with cellular organelles and chemical composition, can be potentially used for quantitative monitoring and diagnosis of the cellular phenotypes (28), and indicate abnormal cell morphology. In addition, dry mass of cancer cells has been recognized as a possible diagnostic and monitoring marker (38,39).

These cellular changes in cancer cells can be potentially detected by label-free imaging techniques. Without staining, however, biological cells are nearly transparent under bright-field microscopy, as their absorption differs only slightly from that of their surroundings, resulting in a low image contrast. An internal contrast mechanism that can be used when imaging cells without staining is their refractive index. The light beam passing through the imaged cells is delayed, since the cells have a slightly higher refractive index compared with their surroundings. Conventional intensity-based detectors are not fast enough to record this delay directly. Phase imaging methods, on the other hand, use optical interference to record the delays of light passing through the sample, and thus they yield label-free contrast in the image (40). In contrast to qualitative phase contrast methods, such as Zernike’s phase contrast and differential interference contrast (DIC) microscopies, quantitative phase microscopy yields the optical thickness map or optical path delay (OPD) map of the cell, so that on each x–y point of this map, OPD is equal to the integral of the refractive-index values across the cell thickness.

Quantitative phase imaging techniques, interferometric based (23,28,41–43) and non-interferometric based (24,29), have been used to analyze cell features for red blood cells (41), stem cells (42), cancer tissues (23), and cancer cells (24,28,29,43), and have shown the ability to differentiate between various conditions of cells and tissues, based on the average OPD values (23,42,43), or other parameters that are derived from the OPD map and the cell visible morphology (24,29), as well as their spatial-frequency content (41).

Machine learning of cytometric data from circulating tumor cells enables automatic analysis of a large number of cells with good classification abilities (44). Specific to interferometry, after extraction of various cellular features from the cellular OPD maps, machine learning techniques can assign weights to these features for classification. Several machine learning techniques have been previously applied on reconstructed digital interferograms starting from 2005 to identify filamentous microorganisms (45); to classify stained and unstained cells, and quantify cell viability and concentration (46); and to grade red blood cells infected by the malaria parasite (47).

In the present study, we compared the quantitative phase imaging-based features of healthy and cancer cells and of primary cancer and metastatic cancer cells. When performing these comparisons, we have chosen pairs of cell lines taken from the same individual to avoid differences that are related to changes between people’s organs and disease expression. In order to obtain data for significant statistics, cell imaging techniques should have high throughput capabilities, but still be affordable (16). Acquiring the cells during fast flow can enable high throughput. In our research, the cells were alive and in a nonadherent state to allow analyzing a large number of cells during flow, in contrast to previous studies that used adherent cells (28,43), fixed cells (24,29), or tissue samples (23), which were limited in the amount of recorded data. In addition, we have used off-axis interferometry, which requires only a single camera exposure, and thus is suitable for acquisition of rapid dynamics, such as those occurring during flow.

After acquiring off-axis interferograms of the cells and extracting their OPD maps, we calculated cellular features based on these quantitative maps, and applied machine learning approaches for cell group classification. Our quantitative phase imaging approach is expected to yield an automated tool to distinguish oncogenic progression and metastasis based on label-free cancer cell cytometry.
cells was chosen to compare healthy cells versus cancer cells; the other two pairs of cells were chosen to compare primary cancer cells versus metastatic cells. Each of the cell line comparisons is from the same individual.

The complete growth medium used for the Hs cell pair is Dulbecco’s Modified Eagle’s Medium (DMEM) (ATCC, SN. 30-2002) supplemented with 10% Fetal Bovine Serum (FBS) (BI, SN. 04-007-1A).

The complete growth medium used for the WM cell pair is DMEM (BI, SN. 01-055-1A) supplemented with 10% FBS (BI, SN. 04-007-1A) and 2 mM L-glutamine (BI, SN. 03-020-1B) (48,49).

The complete growth medium used for the SW cell pair is B1 Roswell Park Memorial Institute (RPMI) 1640 Medium without L-glutamine (BI, SN. 01-104-1A) supplemented with 10% FBS (BI, SN. 04-007-1A) and 2 mM L-glutamine (BI, SN. 03-020-1B) (50–55).

The cell lines were incubated under standard cell culture conditions at 37°C and 5% CO2 in a humidified incubator until 80% confluence was achieved.

Preparation for Imaging

Prior to the imaging experiment, the cells were trypsinized for suspension, supplemented with a suitable medium, and inserted into an adhesive chamber (Grace Bio-Labs SecureSeal adhesive chamber, volume 18 μL, 13 mm diameter × 0.15 mm thickness, 1.5 mm ports diameter, Sigma Aldrich SN. GBL611101) attached to a cover slip. This chamber induced a constant thickness value on the entire imaged sample, which is important for the flatness of the final phase map. Another adhesive chamber was filled with the suitable medium and placed in the reference beam propagation path. Then, all cell lines were quantitatively imaged without labeling using the low-coherence IPM system shown in Figure 1.

Data Analysis

To extract the quantitative phase map from the acquired off-axis image interferograms, we used the off-axis interferometry Fourier-based algorithm (56), which includes a 2D Fourier transform, filtering one of the cross-correlation terms, and an inverse 2D Fourier transform, where the argument of the resulting matrix is the wrapped phase of the sample. To compensate for stationary aberrations and field curvatures, we subtracted from the wrapped phase map of the sample, a phase map which is extracted from an interferogram acquired with no sample. We then applied the unweighted least squares phase unwrapping algorithm to resolve 2π phase ambiguities (57). The resulting unwrapped phase map is multiplied by the wavelength and divided by 2π to obtain the quantitative OPD map of the sample, which is defined as follows:

\[
OPD_n(x, y) = [\pi_c(x, y) - n_0] \times h_i(x, y),
\]

where \(n_0\) is the refractive index of the medium, \(h_i\) is the thickness profile of the cell, and \(\pi_c\) is the cell integral refractive index, which is defined as follows (58):
\[ \pi_c(x, y) = \frac{1}{h_k} \int_0^h r_c(x, y, z) \, dz. \] (2)

To separate single cells from the background and be able to process only the OPD related to the cells, we used the normalized cut algorithm as an edge detector (59). This method is based on a graph formulation, wherein the nodes of the graph are the points in the feature space with a similarity function weight connecting them. The goal is to partition the vertices into disjoint sets \( V_1, V_2, \ldots, V_m \), whereby the similarity within a set \( V_i \) is high, and across different sets is low. After implementation of this algorithm, a morphological opening operator was applied in order to connect the gradient lines that were detected. Next, a morphological dilation operator was applied in order to expand the connected lines. At last, a global threshold was applied in order to remove background pixels that erroneously appeared in the cell area.

Using the above described methods, we could create a data set containing the OPD information of the cell areas only, and calculate the following parameters that are based directly on the OPD map defined in Eq. (1), without decoupling the cellular thickness profile from the refractive index as a prior stage:

1. Mean and median of OPDc.
2. Dry mass: This parameter quantifies the mass of the non-aqueous material of the cell, yielding information about cell growth (60,61).
   \[ m = \frac{1}{S_c} \int S_c(x, y) \, ds = \frac{S_c}{x} \times \langle \text{OPD}_c \rangle, \] (3)
   where \( x \) is the refractive increment and approximated as 0.18–0.21 mL/g, \( S_c \) is the projected cell area on the \( x-y \) plane, and \( \langle \text{OPD}_c \rangle \) is the averaged OPD over the cell area. We used \( x = 0.2 \).
3. Dry mass averaged density: This parameter can be calculated using the cell dry mass as follows (62):
   \[ \sigma_M = \frac{M}{S_c} \] (4)
4. Phase volume: This is not the actual cell volume but only the equivalent of the cell volume that is based on OPDc directly and takes into consideration refractive-index variations inside the cell (62).
   \[ V_\phi = \int S_c \, \text{OPD}_c(x, y) \, ds. \] (5)
   As can be seen, the phase volume is proportional to the dry mass of the cell.
5. Phase surface area: This parameter can be calculated as the sum of the upper surface area of the phase profile and the projected area, as follows:
   \[ SA_\phi = \frac{1}{S_c} \int dA + S_c = \int \frac{(1 + \delta h_x^2 + \delta h_y^2)^{1/2}}{S_c} \, dx \, dy + S_c, \] (6)
   where \( dA \) is the discrete cell surface area as projected over a single camera pixel, \( \delta h_x \) and \( \delta h_y \) are the gradients along the \( x \) and \( y \) directions of the cell OPD map (63,64).
6. Phase surface area to volume ratio: This parameter is a generalized version of the physical surface area to volume parameter (65,66), but again it takes into consideration phase changes in the cell:
   \[ SAV = \frac{SA_\phi}{V_\phi}. \] (7)
7. Phase surface area to dry mass ratio: This parameter is defined as follows (62):
   \[ SDM = \frac{SA_\phi}{M}. \] (8)
   The last two parameters can quantify cell metabolism and describe how much material a surface unit transfers to one volume unit or mass unit.
8. Projected area to volume ratio: This parameter describes the flatness of the cell (62).
   \[ PAV = \frac{S_c}{V_\phi}. \] (9)
9. Phase sphericity index: This parameter quantifies the degree of cell roundness. The sphericity of an object is the ratio of object volume and the surface area. Round shape is a value that may imply on cell abnormality (62,67–69). It is a dimensionless constant with values ranging from zero for a laminar disk to unity for a sphere (68).
   \[ \psi = \pi^{1/3} \times \frac{(6 \times V_\phi)^{2/3}}{SA_\phi}. \] (10)
10. Phase statistical parameters: These parameters describe the dry mass or volume distribution in the cell. They are based on changes in phase values and thus react to structural alternations of cell organelles and factors. To use these parameters, the phase values over the projected cell area need to be written as a single vector containing \( K \) values. Then, the following statistical parameters can be defined (62):
   a. Phase variance: This parameter measures how a set of the cell OPD values is spread out.
      \[ \sigma_\phi = \frac{1}{K-1} \sum_{i=1}^{K} (\text{OPD}_c(i) - \mu_{\text{OPD}})^2, \] (11)
      where \( \mu_{\text{OPD}} \) is the mean of the OPD of the cell.
   b. Phase kurtosis: This parameter measures whether the cell OPD distribution is peaked or flat.
Kurtosis$_{\phi} = \frac{1}{\sigma_{\phi}^4} \sum_{i=1}^{K} \left( \frac{\text{OPD}_c(i) - \mu_{\text{OPD}_c}}{\sigma_{\phi}} \right)^4. \tag{12}
$

\text{c. Phase skewness: This parameter measures the lack of symmetry of the cell OPD values from the mean value.}

\text{Skewness}_{\phi} = \frac{1}{\sigma_{\phi}^3} \sum_{i=1}^{K} \left( \frac{\text{OPD}_c(i) - \mu_{\text{OPD}_c}}{\sigma_{\phi}} \right)^3. \tag{13}
$

11. Energy: This parameter characterizes the cell texture (70).

\[ E = \sum_{i=1}^{K} \text{OPD}_c(i)^2. \tag{14} \]

It is important to mention that the OPD values have coupling between the cellular refractive index and the physical thickness (see Eq. (1)). The entire analysis was performed directly on the OPD value without decoupling these parameters to allow single exposure mode per each instance of the sample, which is suitable for acquiring cells during fast flow. Note also that all 15 parameters presented above are based on the quantitative OPD map and thus cannot be calculated based on simple bright-field microscopy or fluorescent microscopy (recording the intensity of light), which is the basis of conventional flow cytometry, or based on non-quantitative phase techniques such as Zernike’s phase contrast microscopy and differential interference contrast (DIC) microscopy.

**Statistical Analysis**

In total, we acquired 106, 97, 71, 102, 118, and 163 OPD maps for cell lines Hs 895.Sk, Hs 895.T, WM 115, WM 266-4, SW 480, and SW 620, respectively. To evaluate statistical difference amongst the cell groups, for each cell line pair and for each of the calculated parameters, we used the two-sample test for the $P$ values of the data shown in Table 1. In addition, we also implemented the Mann–Whitney test, which yielded similar or even greater statistical difference between the groups.

**Machine Learning**

Following the OPD-map-based feature extraction, our objective was to obtain classification decisions. For this goal, we used state-of-the-art machine learning techniques to perform three separate analysis for each pair of cell lines: Hs 895.Sk versus Hs 895.T, WM 115 versus WM 266-4, SW 480, and SW 620.

In general, a classification solution comprises of three main stages: 1) selection of features for the specific task, 2) a dimensionality reduction (and noise removal), and 3) the final classification using a selected classifier. In this work, we used the proposed analysis of the descriptors to select the features.

As for stage 1) of feature selection, for each cell OPD map, a set of 13 features, which were found to be statistically discriminative according to their $P$ values (mean, median, projected surface area, phase volume, dry mass, dry mass average density, energy, surface area, phase surface area to volume ratio, phase surface area to dry mass ratio, projected area to volume ratio, sphericity and phase variance), were concatenated into a feature vector of size 13. For each classification task, a feature matrix of size $(M + N) \times 13$ was built, where $M$ and $N$ are the numbers of cells in the matching cell line pair.

For stage 2) of dimensionality reduction, a standard principal component analysis (PCA) (71) was performed on the feature matrix. The resulting PCA feature matrix was column normalized to a mean of 0 and a standard deviation of 1. A set of the most informative components can be selected following the PCA stage (71). In this work, we selected the first six components to serve as the input representation.

Next, in stage 3) of classification, since we know all of the cell labels, a supervised learning method was selected. A state-of-the-art support vector machine (SVM) classifier was used for this task. SVM is widely used for pattern classification problems (72). A leave-one-out SVM classification was performed using LibSVM (73). The classification performance was evaluated using the area under curve (AUC) of the receiver operating characteristic (ROC) curve.

**Results**

We used the low-coherence IPM system presented in Figure 1 to acquire OPD maps of the six cell lines described above, with the goal of comparing healthy to cancer cells (Hs 895.Sk [skin] versus Hs 895.T [melanoma]), and primary cancer to metastatic cancer cells (WM 115 [melanoma] versus WM 266-4 [metastatic melanoma], and SW 480 [adenocarcinoma colon] versus SW 620 [metastatic adenocarcinoma colon]). Note that the cells were unattached and therefore had mostly round projected areas. The ability to analyze unattached isolated cells is important for flow cytometry via high-throughput quantitative imaging of cells during flow. Visualization 1 (Fig. 2) presents the OPD map of SW 480 cells flowing in a microfluidics channel (IBIDJ, SN. 80666, 1 mm width, 17 mm length, 0.1 mm height). In this situation, when the isolated cells are round and unattached, most of the cells look alike and subjective pathological examination cannot be performed. Indeed, Figure 3 shows one representative OPD map from each of the cell line groups, demonstrating that even when using quantitative phase imaging, a bare eye cannot see significant differences between each pair of cell lines. This is solved by the automatic machine-learning method analyzing the cell topological OPD maps. In contrast, nonquantitative imaging methods, such as label-free bright-field, Zernike’s phase contrast or DIC microscopies, which do not have access to these cell topological OPD maps, do not allow for quantitative automatic identification.

Prior to the analysis of these OPD maps, we applied the image segmentation procedure, described in the Methods section, to track the cell edges. Next, we applied Eqs. (3–14) to the cell OPD area selected by the segmentation process to calculate the following parameters: mean, median, projected...
Table 1. Averages and standard deviations for all 15 parameters calculated for the three cell line pairs

| Parameter                        | HS 895.SK | HS 895.T | WM 115 | WM 266-4 | SW 480 | SW 620 |
|----------------------------------|-----------|----------|--------|----------|--------|--------|
| Mean (nm)                        | 333.09    | 272.40   | 474.10 | 319.90   | 306.45 | 286.04 |
|                                  | 60.99     | 48.45    | 86.58  | 35.70    | 34.59  | 27.82  |
| Median (nm)                      | 348.88    | 282.53   | 502.78 | 342.90   | 335.89 | 312.34 |
|                                  | 72.37     | 52.90    | 93.88  | 39.04    | 40.07  | 32.95  |
| Projected surface area ($10^6$ nm$^2$) | 251.87   | 202.99   | 342.24 | 153.14   | 95.60  | 70.4   |
|                                  | 78.94     | 84.96    | 106.06 | 34.62    | 21.48  | 13.53  |
| Phase volume ($10^9$ nm$^3$)     | 86.34     | 57.67    | 165.93 | 49.75    | 29.74  | 20.34  |
|                                  | 38.76     | 33.01    | 69.02  | 14.87    | 9.29   | 5.35   |
| Dry mass ($10^{-11}$ g)          | 43.17     | 28.84    | 82.96  | 24.87    | 14.87  | 10.17  |
|                                  | 19.38     | 16.51    | 34.51  | 7.44     | 4.64   | 2.68   |
| Dry mass average density ($10^{-19}$ g/nm$^2$) | 16.65    | 13.62    | 23.70  | 15.99    | 15.32  | 14.30  |
|                                  | 3.05      | 2.42     | 4.33   | 1.79     | 1.73   | 1.39   |
| Surface area ($10^{12}$ nm$^2$)  | 172.16    | 114.37   | 376.22 | 71.57    | 31.36  | 17.96  |
|                                  | 109.01    | 100.90   | 225.56 | 31.83    | 13.90  | 6.93   |
| SAV ($10^2$/nm)                  | 18.71     | 17.37    | 21.46  | 13.87    | 10.17  | 8.59   |
|                                  | 3.53      | 4.92     | 4.19   | 2.08     | 1.34   | 1.04   |
| SDM ($10^{22}$ nm$^2$/g)         | 37.43     | 34.75    | 42.92  | 27.75    | 20.35  | 17.19  |
|                                  | 7.06      | 9.84     | 8.38   | 4.16     | 2.68   | 2.08   |
| PAV ($10^{-4}$/nm)               | 31.05     | 37.77    | 21.80  | 31.74    | 33.07  | 35.29  |
|                                  | 5.84      | 6.17     | 4.06   | 4.64     | 4.01   | 3.40   |
| Sphericity ($10^{-7}$)           | 6.42      | 8.45     | 4.50   | 11.26    | 16.13  | 21.39  |
|                                  | 1.95      | 3.16     | 1.44   | 15.59    | 3.77   | 4.12   |
| Phase variance ($10^3$ nm$^2$)   | 20.58     | 13.22    | 41.38  | 18.85    | 16.06  | 12.34  |
|                                  | 7.46      | 5.69     | 12.84  | 5.31     | 5.34   | 3.29   |
| Phase kurtosis ($10^{-6}$/nm$^4$) | 15.02    | 29.29    | 4.85   | 11.41    | 10.24  | 11.79  |
|                                  | 10.42     | 18.86    | 3.10   | 17.57    | 11.09  | 6.50   |
| Phase skewness ($10^{-5}$/nm$^3$) | -21.57    | -26.75   | -15.04 | -21.86   | -25.29 | -28.89 |
|                                  | 23.91     | 21.40    | 8.60   | 21.33    | 16.67  | 14.19  |
| Energy ($10^7$ nm$^2$)           | 34.97     | 19.30    | 94.82  | 18.72    | 10.64  | 6.61   |
|                                  | 21.54     | 15.01    | 52.08  | 7.40     | 4.60   | 2.38   |

In each table cell, the first and second values correspond to the average and to the standard deviation values, respectively.
* denotes $P$ values $<0.05$, ** denotes $P$ values $<0.005$, *** denotes $P$ values $<0.0005$, – denotes $P$ values $>0.05$.

surface area, dry mass, dry mass average density, phase volume, phase surface area, phase surface area to volume ratio, phase surface area to dry mass ratio, projected area to volume ratio, sphericity, phase variance, kurtosis, skewness, and energy.

The corresponding average and standard deviation values for each parameter for all cell lines are summarized in Table 1. As can be seen from this table, 13 out of 15 OPD-based parameters were statistically significant. Figures 4–6 show the histograms of each of 12 parameters for which the two groups of cells
compared were statistically significant (excluding the projected surface area parameter, to spare space, that was also statistically significant with $P$-values of <0.0005). Each subfigure shows the histograms of a pair of cell lines, where each cell line is shown in a different color (red or blue). All these parameters are statistically significant based on $P$-values of <0.05, <0.005, or <0.0005. These results demonstrate the parameters' ability to statistically discriminate between each cell line in each pair, even if the initial OPD map of the cells look similar.

The 13 successful parameters were used as an input for a PCA/SVM analysis (see machine learning details in the Method section) to extract the best combination of these parameters, which is useful for classification between each cell line pair. The ROC curves obtained by using PCA followed by an SVM classification for each pair of cells are presented in Figure 7. The best results of the classification were obtained for six-first principle components and for linear SVM kernel for all cell line pairs. As seen in this figure, the area under the curve (AUC) in all three classification tasks is high and around 0.9.

Table 2 summarizes the results obtained using the SVM learning combined with PCA for each pair of cells, with the
AUC and the matching working points (sensitivity/specificity) chosen on the ROC curves. Note that for the definition of sensitivity and specificity, the Hs 895.Sk, SW 480, and WM 115 cell lines were defined as "negatives," and the Hs 895.T, SW 620, and WM 266-4 cell lines were defined as "positives."

These results demonstrate an automatic algorithm with an ability to classify cells in different cancer stages using the OPD-map-based parameters. We achieved high classification rates and AUC values for unattached cells, whereby no morphological differences can be observed in the phase maps by the naked eye (see Fig. 3). The AUC values correspond with the separation between the groups presented in the histograms (Figs. 4–6), whereby better classification with higher percentages corresponds with lower $P$-values between the groups in the histograms. Thus, for the WM pair (Fig. 5), we can observe better separation than the Hs pair (Fig. 4) and the SW pair (Fig. 6).

Significantly, since the best PCA/SVM parameters were achieved for all cell line pair classification tasks (six-best principle components and linear SVM kernel were finally chosen...
for all cases), it can be assumed that a global machine for cancer staging can be developed.

**DISCUSSION AND CONCLUSIONS**

Early detection of cancer can prevent tedious and painful treatment, may prevent recurrence, and improve survival. To diagnose cancer, pathologists typically use fresh samples that are frozen and sectioned, or fixed samples that are dehydrated, embedded in paraffin, sectioned and stained with dyes and/or antibodies to specific tumor antigens. Our work describes the first steps in the development of optical signatures that can distinguish normal cells from cancer in situ and cancer in situ from metastatic forms without fixation, sectioning or any type of labeling. Thus, the scores can be performed rapidly and automatically and are not only based on simple qualitative cell parameters based on 2D imaging (such as cell size and general shape), but rather use the cellular optical thickness dimension as well. This yields new parameters based on the cell topological map that have not been available to the pathologists before, with which automatic cell identification can be performed.

To address the clinical need for cancer diagnosis, we identified OPD-based diagnostic signatures for live and unattached cells in different stages of oncogenesis. We proposed using spatial morphological and texture parameters, which are based on the cell OPD maps, to compare tumor-derived cancer cell lines of different cancer levels and cell lines derived from a healthy tissue of the same individuals. We showed the

---

**Figure 6.** Histograms of the parameters based the OPD maps for colorectal adenocarcinoma colon cells SW 480 (red) versus metastatic from lymph node of colorectal adenocarcinoma cells SW 620 (blue): (a) mean, (b) median, (c) phase volume, (d) phase variance, (e) dry mass, (f) dry mass average density, (g) surface area, (h) phase surface area to volume ratio, (i) phase surface area to dry mass ratio, (j) projected area to volume ratio, (k) energy, and (l) sphericity. *** denotes P-values <0.0005. [Color figure can be viewed at wileyonlinelibrary.com]

**Figure 7.** ROC curves of true-positive rate (sensitivity) versus false-positive rate (1 – specificity), as obtained using SVM learning combined with PCA for the: Hs cell lines (blue curve), WM cell lines (red curve), and SW cell lines (green curve). The resulting AUC is written after the cell line pair name. The small circles denote the working points of the classification tasks. [Color figure can be viewed at wileyonlinelibrary.com]

---

**Table 2.** Machine learning classification results between the groups (PCA/SVM analysis). AUC = Area under curve.

|            | SENSITIVITY (%) | SPECIFICITY (%) | AUC     |
|------------|-----------------|-----------------|---------|
| Hs 895.Sk  | 81              | 83              | 0.878   |
| Hs 895.T   | 93              | 99              | 0.985   |
| WM 115     | 82              | 81              | 0.897   |
| WM 266-4   |                 |                 |         |
| SW 480     |                 |                 |         |
| SW 620     |                 |                 |         |
feasibility to distinguish between different cell conditions in a label free manner, and we have quantified differences among microscopically similar looking cells with statistical significance. We demonstrated that diagnostic optical signatures can be derived from comparisons between single cells, normal fibroblasts derived from living biopsy tissue compared to fibroblasts isolated from melanoma or from comparison of fibroblasts derived from melanoma to cells derived from a melanoma lymph metastasis. Thus, we showed that these quantitative phase based parameters are able to distinguish cancerous cells from healthy cells and metastatic cancer cells from primary cancer cells with high accuracy.

The optical signatures that can be built using the OPD parameters are advantageous in that tumor cells can be imaged without time consuming labeling, sectioning, sequencing or other methods in use today to identify tumor cells by genomic changes.

Specifically, to compare healthy and cancerous cells, normal human skin fibroblasts isolated from a 48 year old Caucasian female (Hs 895.5k) were compared to fibroblasts isolated from a melanoma tumor from the skin of the same individual (Hs.895.T), with classification results of 81% sensitivity and 83% specificity. To compare primary cancer cells and metastatic cancer cells, human fibroblasts isolated from a melanoma tumor in situ from a 58-year-old female (WM 115) were compared to fibroblasts isolated from a lymph metastasis in the same individual (WM 266-4), with classification results of 93% sensitivity and 99% specificity. Additionally, cells from colorectal adenocarcinoma in situ (SW 480) were compared to cells from colorectal adenocarcinoma lymph metastasis from the same individual (SW 620), with classification results of 82% sensitivity and 81% specificity.

Tumor microenvironment may differ between melanoma in situ and distant sites, contributing to reorganization of cellular morphology and interaction with quantitative phase. Indeed, metastatic forms of different cancers share morphological similarities leading to the idea that many cancers evolve into a more common metastatic state that is stratified by heterogeneity in tumor microenvironment, which differs depending on the site of metastasis, interaction with the immune system, and chemotherapeutic resistance. In Figures 4–6, which could be a result from differences in cell cycle length and nonsynchronization across the population; low P-values were still obtained between the groups, so that these groups are statistically different. No statistical significance was observed for kurtosis and skewness, which might be explained by the fact that the experiments were done on floating, uniformly spherical cells, so the measures of peaks, flatness, or symmetry were less applicable.

It can be seen in Figures 4–6 that the parameter values decreased with cancer progression (healthy > cancer > metastatic) except for the projected area to volume ratio and sphericity that increased with cancer progression, since the mathematical relation for these parameters is opposite. In any case, as can be seen from these results, a similar trend of progression is retained in the average parameters values. We suppose that the observed trends will be kept for other similar cell line pairs. However, since our experiments are based on three cell line pairs, further research is needed to prove this hypothesis. The greater differences in the calculated parameters between the WM cell line cells than between the SW cell lines might be explained by the fact that that the WM fibroblasts isolated from in situ melanoma compared to the WM fibroblasts isolated from metastatic melanoma are likely to be more similar (both flat well-spread) than the SW epithelial cells isolated from colorectal adenocarcinoma, which are more cuboidal in situ, compared to SW metastatic cells, which are more spread and more motile.

The challenge with label-based flow cytometry for cancer diagnosis is the number of cells that are needed, which in turn is dependent on the brightness of the marker. In general, 1,000–1,000,000 cells are needed from a dispersed tumor biopsy depending on the proportion of cancer cells to non-cancer cells in the sample and tumor heterogeneity. There is no universal marker that can be used to detect all cancers. Flow cytometry is used to detect leukemia, other bone marrow-associated cancers and to determine success of stem cell transplantsations, but certainly cannot be used to detect circulating tumor cells where greater than or equal to five metastatic breast cancer cells are found in 7.5 mL peripheral blood. In contrast, our label-free method can distinguish between normal and cancerous cells and primary cancer from metastatic cancer cells, with as few as tens of cells, without using fixation, embedding and tumor markers, but still has to face with the fast acquisition and analysis challenges.

Currently, IPM of unattached cells in flow can be performed faster than video processing rate on a regular computer, including 2D phase unwrapping and much faster if using the graphics processing unit (GPU) of the computer. Our method can be integrated into clinical imaging system using compact modules that can be connected in the output of a conventional microscope. By combining IPM...
with real-time fast processing algorithms (56,77,78) and automatic cell detection and machine learning algorithms for evaluating cell condition as presented in our work, quantitative phase microscopy has potential to become a powerful clinical screening tool for cancer diagnosis and might allow pathologists to inexpensively grade circulating tumor cells in liquid biopsies in real time. Thus, together with the analysis tools presented here for cancer monitoring, we believe that in the future IPM can be useful for detecting and monitoring cancer from body fluids in flow cytometry, for routine cell analysis or for the investigation and detection of pathological conditions in a semi-automated way.

In this study, we used state-of-the-art machine-learning algorithms based on SVM to demonstrate the feasibility of our approach for automatic classification of cancer cells based on IPM data. However, note that there is a great variety of other machine learning algorithms that could be implemented. One of the possible extensions of this work is to apply neural network related algorithms for deep learning, which require a greater number of samples.

To summarize, the proposed quantitative imaging technique shows preliminary clinical potential for automatic cell flow cytometry. This technique does not require cell staining as conventional flow cytometry, and since it uses the quantitative phase profile it has access to parameters which are not available using bright-field microscopy or other nonquantitative phase imaging techniques. Still, future studies are needed to confirm if the unique quantitative phase signatures developed here can distinguish normal from circulating cancer cells isolated from liquid biopsies, and determine the diagnostic and prognostic value of quantitative phase signatures in determining tumor grade and metastatic potential.

Acknowledgments
This study was supported by Horizon 2020 European Research Council (ERC) grant no 678316. We thank Ruth Gottlieb for significant biological support and advising, Dr. Ksavery Kalinowski for useful technical discussions, and Dr. Itay Barnea for useful biological discussions.

Literature Cited
1. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, Radich J, Anderson G, Hartwell L. The case for early detection. Nat Rev Cancer 2003;3:243–252.
2. Gupta GP, Massagué J. Cancer metastasis: Building a framework. Cell 2006;127:679–695.
3. Hamilton W, Walter FM, Rubin G, Neal RD. Improving early diagnosis of symptomatic cancer. Nat Rev Clin Oncol 2016;13:740–748.
4. Crowley E, D’Nicolantonio E, Lospakia P, Bardelli A. Liquid biopsy: Monitoring cancer-gene expression in the blood. Nat Rev Clin Oncol 2013;10:472–484.
5. Bast RC, Jr, Hensley M, Mills GB. Quality in cancer diagnosis. CA Cancer J Clin 2010;60:679–695.
6. Elmoori I, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, Radich J, Anderson G, Hartwell L. The case for early detection. Nat Rev Cancer 2003;3:243–252.
7. Raab SS, Grzybicki DM. Quality in cancer diagnosis. CA Cancer J Clin 2010;60:139–165.
8. Paternini-Bréchet P, Benali NL. Circulating tumor cells (CTC) detection: Clinical impact and future directions. Cancer Lett 2007;253:180–204.
9. Yu M, Stott S, Toner M, Maheswaran S, Haber DA. Circulating tumor cells: Approaches to isolation and characterization. J Cell Biol 2011;192:373–382.
10. Alba-Panadero C, Pantel K. Circulating tumor cells: Liquid biopsy of cancer. Clin Chem 2013;59:110–118.
11. Steeg PS. Tumor metastasis: Mechanistic insights and clinical challenges. Nat Med 2006;12:895–904.
subcellular topographic imaging technique using full-field quantitative phase microscopy coupled with interference reflection microscopy. Biomed Opt Express 2012;3:2175.

43. Watanabe E, Hoshiba T, Javid B. High-precision microscopic phase imaging without phase unwrapping for cancer cell identification. Biomed Opt Express 2016;7:922–931.

44. Lannin TB, Thege FI, Kirby B. Comparison and optimization of machine learning methods for automated classification of circulating tumor cells. Cytometry Part A 2016;89:922–931.

45. Javid B, Moon J, Yeom S, Carapezza E. Three-dimensional imaging and recognition of microorganisms using single-exposure on-line (SEOL) digital holography. Opt Express 2005;13:4492–4506.

46. Feizi A, Zhang Y, Greenbaum A, Guziak A, Luong M, Chan RYL, Berg B, Ortkan H, Luo W, Wu M, et al. Rapid, portable and cost-effective yeast cell viability and concentration analysis using lensfree on-chip microscopy and machine learning. Lab Chip 2016;16:4350–4358.

47. Park HS, Rinehart MT, Walter KA, J, Chi I TA, Wax A. Automated detection of F. pallidum using machine learning algorithms with quantitative phase images of unstained cells. PLoS ONE 2016;11:e0163045.

48. Asano M, Yokita A, Suzuki H. Wide spectrum of antitumor activity of a neutralizing monoclonal antibody to human vascular endothelial growth factor. Jpn J Cancer Res 1999;90:93–100.

49. Nakamura T, Sato K, Hamada H. Effective gene transfer to human melanomas via integrin-targeted adenoviral vectors. Hum Gene Ther 2002;13:613–626.

50. Bishut H, Gabai H, Girshovitz P, Shaked NT. Optical-mechanical signatures of cancer cells based on fluctuation profiles measured by interferometry. J Biophotonics 2014;7:624–630.

51. Kito M, Aka Y, Ohishi N, Yagi K. Induction of apoptosis in cultured colon cancer cells by transfection with human interferon beta gene. Biochem Biophys Res Commun 1999;257:771–776.

52. Kang JC, Chen JS, Lee CH, Chung JJ, Shih YS. Intratumoral macrophage counts correlate with tumor progression in colorectal cancer. J Surg Oncol 2010;102:242–248.

53. Doge J, Freo M, Rodriguez MA. Expression and function of the nuclear factor of activated T cells in colon carcinoma cells. Involvement in the regulation of cyclooxynase-2. J Biol Chem 2005;280:8686–8693.

54. Kitazono M, Sumizawa T, Takeo Y, Chen ZS, Furukawa T, Nagayama S, Tani A, Takao S, Aikou T, Akiyama SI. Multidrug resistance and the lung resistance-associated protein in human colon carcinoma SW-620 cells. J Natl Cancer Inst 1999;91:1647–1653.

55. Kitazono M, Okumura H, Ikeda R, Sumizawa T, Furukawa T, Nagayama S, Sato K, Aikou T, Akiyama SI. Reversal of LRP-associated drug resistance in colon carcinoma SW-620 cells. Int J Cancer 2001;93:126–131.

56. Asano M, Yokita A, Suzuki H. Wide spectrum of antitumor activity of a neutralizing monoclonal antibody to human vascular endothelial growth factor. Jpn J Cancer Res 1999;90:93–100.

57. Kitazono M, Sumizawa T, Takeo Y, Chen ZS, Furukawa T, Nagayama S, Tani A, Takao S, Aikou T, Akiyama SI. Multidrug resistance and the lung resistance-associated protein in human colon carcinoma SW-620 cells. J Natl Cancer Inst 1999;91:1647–1653.

58. Wadell H. Volume, shape, and roundness of microorganism using quantitative phase imaging. PLoS ONE 2013;8:e55676.

59. Canham RB, Burton AC. Distribution of size and shape in populations of normal human red cells. Circ Res 1968;22:405–422.

60. Mir M, Wang Z, Tangella K, Popescu G. Diffraction phase cytometry: Blood on a CD-ROM. Opt Express 2009;17:2579–2585.

61. Matekura A, Strzelecki M. Texture Analysis Methods—a Review. Technical University of Lods 1998, COST B11—Technical Report, 1998. pp 873–887.

62. Zhang J, Zhang XD, Ha SW. A novel approach using PCA and SVM for face detection. In: Fourth International Conference on Natural Computation IEEE, vol. 3; 2008. pp 29–33.

63. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines. UK: Cambridge University Press; 2000.

64. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2011;2:1–27.

65. Jaye DL, Bray RA, Gebel HM, Harris WA, Waller EK. Translational applications of flow cytometry in clinical practice. J Immunol 2012;188:4715–4719.

66. Budd GT, Cristofanilli M, Ellis MJ, Stopeck A, Borden E, Miller MC, Matera J, Rebollet M, Doyle GV, Terstappen LW, et al. Circulating tumor cells versus imaging predicting overall survival in metastatic breast cancer. Clin Cancer Res 2006;12:5403–6409.

67. Steeg PS. Targeting metastasis. Nat Rev Cancer 2016;16:201–218.

68. Girshovitz P, Shaked NT. Real-time quantitative phase reconstruction in off-axis digital holography using multiplexing. Opt Lett 2014;39:2262–2265.

69. Backoach O, Kariv S, Girshovitz P, Shaked NT. Generalized cell morphological parameters based on interferometric phase microscopy and their application to cell life cycle characterization. Biomed Opt Express 2012;3:1757–1775.

70. Rappaz B, Cano E, Colomb T, Kuhn J, Depeursinge C, Simanis V, Magistretti PJ, Manquet P. Noninvasive characterization of the fission yeast cell cycle by monitoring dry mass with digital holographic microscopy. J Biomed Opt 2009;14:54049.

71. Girshovitz P, Shaked NT. Generalized cell morphological parameters based on interferometric phase microscopy and their application to cell life cycle characterization. Biomed Opt Express 2012;3:1757–1775.

72. Mir M, Ding H, Wang Z, Reedy J, Tangella K, Popescu G. Blood screening using diffraction phase cytometry. J Biomed Opt 2010;15:27016.

73. Girshovitz P, Shaked NT. Generalized cell morphological parameters based on interferometric phase microscopy and their application to cell life cycle characterization. Biomed Opt Express 2012;3:1757–1775.

74. Porter PK. Allometry of mammalian cellular oxygen consumption. Cell Mol Life Sci 2001;58:815–822.

75. Asano M, Yukita A, Suzuki H. Wide spectrum of antitumor activity of a neutralizing monoclonal antibody to human vascular endothelial growth factor. Jpn J Cancer Res 1999;90:93–100.

76. Porte A, Aikou T, Akiyama SI. Reversal of LRP-associated drug resistance in colon carcinoma SW-620 cells. Int J Cancer 2001;93:126–131.

77. Taniguchi K, Hoshida S, Hoshida Y, Kitahara N, Okumura K, Nagayama S, Tani A, Aikou T. Akiyama SI. Multidrug resistance and the lung resistance-associated protein in human colon carcinoma SW-620 cells. J Natl Cancer Inst 1999;91:1647–1653.

78. Kitazono M, Okumura H, Ikeda R, Sumizawa T, Furukawa T, Nagayama S, Seto K, Aikou T, Akiyama SI. Reversal of LRP-associated drug resistance in colon carcinoma SW-620 cells. Int J Cancer 2001;93:126–131.

79. Girshovitz P, Shaked NT. Fast phase processing in off-axis holography using multiplexing with complex encoding and live-cell fluctuation map calculation in real-time. Opt Express 2015;23:8773–8787.

80. Pham HV, Bhaduri B, Tangella K, Best-Popescu C, Popescu G. Real time blood testing using quantitative phase imaging. PLoS ONE 2013;8:e55676.

81. Pham HV, Bhaduri B, Tangella K, Best-Popescu C, Popescu G. Real time blood testing using quantitative phase imaging. PLoS ONE 2013;8:e55676.