

 Open access • Posted Content • DOI:10.1101/107805

## **Label-free identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning** — [Source link](#)

Jonghee Yoon, YoungJu Jo, Min-Hyeok Kim, Kyoohyun Kim ...+3 more authors

**Institutions:** KAIST

**Published on:** 11 Feb 2017 - bioRxiv (Cold Spring Harbor Labs Journals)

Related papers:

- [Discrimination of immune cell activation using Raman micro-spectroscopy in an in-vitro & ex-vivo model.](#)
- [Purification of Human T and B Lymphocytes](#)
- [Assessing the molecular fingerprint of T lymphocyte activation](#)
- [Raman detection and identification of normal and leukemic hematopoietic cells.](#)
- [Markers for predicting the response of lymphocytes to a tumor and use thereof](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/label-free-identification-of-non-activated-lymphocytes-using-4x8n271acl>

# SCIENTIFIC REPORTS



OPEN

## Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning

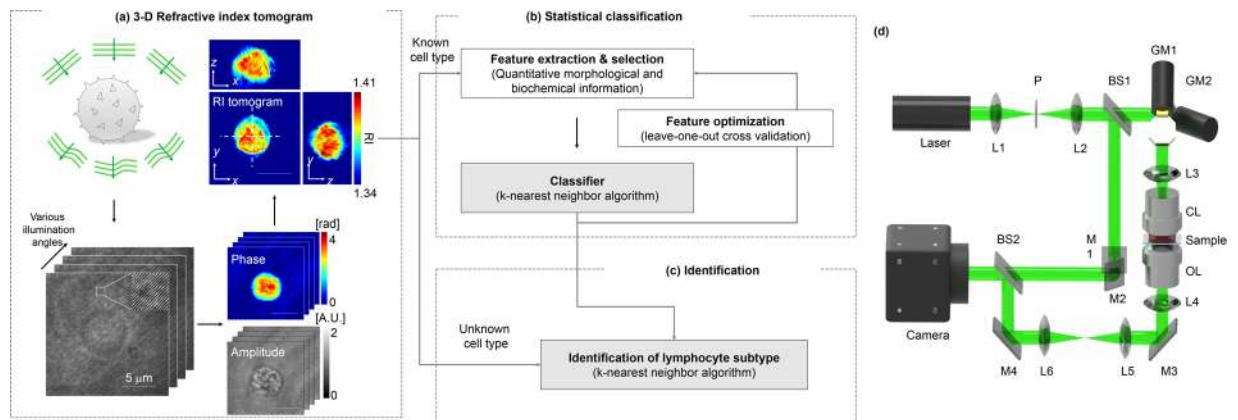
Jonghee Yoon<sup>1,2,5</sup>, YoungJu Jo<sup>1,2</sup>, Min-hyeok Kim<sup>3</sup>, Kyoohyun Kim<sup>1,2</sup>, SangYun Lee<sup>1,2</sup>, Suk-Jo Kang<sup>3</sup> & YongKeun Park<sup>1,2,4</sup>

Identification of lymphocyte cell types are crucial for understanding their pathophysiological roles in human diseases. Current methods for discriminating lymphocyte cell types primarily rely on labelling techniques with magnetic beads or fluorescence agents, which take time and have costs for sample preparation and may also have a potential risk of altering cellular functions. Here, we present the identification of non-activated lymphocyte cell types at the single-cell level using refractive index (RI) tomography and machine learning. From the measurements of three-dimensional RI maps of individual lymphocytes, the morphological and biochemical properties of the cells are quantitatively retrieved. To construct cell type classification models, various statistical classification algorithms are compared, and the  $k$ -NN ( $k = 4$ ) algorithm was selected. The algorithm combines multiple quantitative characteristics of the lymphocyte to construct the cell type classifiers. After optimizing the feature sets via cross-validation, the trained classifiers enable identification of three lymphocyte cell types (B, CD4+ T, and CD8+ T cells) with high sensitivity and specificity. The present method, which combines RI tomography and machine learning for the first time to our knowledge, could be a versatile tool for investigating the pathophysiological roles of lymphocytes in various diseases including cancers, autoimmune diseases, and virus infections.

Lymphocytes consist of various cell types including B, helper (CD4+) T, cytotoxic (CD8+) T, and regulatory T cells, and play crucial roles in the adaptive immune system<sup>1</sup>. Each lymphocyte cell type has different functions: B lymphocytes produce antibodies, and T lymphocytes recognize a specific antigen and execute effector functions. The lymphocyte population and function are tightly regulated to defend the host against harmful invaders or abnormal conditions<sup>1,2</sup>. Disturbances in lymphocyte function and regulation are related to various diseases including cancers<sup>3-5</sup>, autoimmune diseases<sup>6,7</sup>, and virus infections<sup>8,9</sup>.

To understand the roles of different types of lymphocytes, several methods based on labelling techniques have been developed to identify and separate lymphocyte cell types. Because different types of non-activated lymphocytes have very similar cellular morphology such as a large nucleus with small cytosolic regions and round shapes, it is impossible to discriminate lymphocyte cell types with conventional optical methods such as bright-field microscopy or phase contrast microscopy<sup>10</sup>. To overcome this difficulty, specific surface membrane proteins, known as surface markers, are recognized and tagged with magnetic beads or fluorescence molecules via antigen-antibody binding. Then each type of lymphocytes can be distinguished and separated by magnetic forces or fluorescence signals<sup>11</sup>. Targeting surface markers is a precise and efficient approach to determine the cell types; however, labelling methods have potential risks of altering cellular functions by modifying membrane protein structures. In addition, labelling methods have limitations in the number of cell types that can be identified simultaneously due to the limited multiplexing capability of the labelling agents<sup>12</sup>.

<sup>1</sup>Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, 34141, Republic of Korea. <sup>2</sup>KAIST Institute Health Science and Technology, Daejeon, 34141, Republic of Korea. <sup>3</sup>Department of Biological Sciences, KAIST, Daejeon, 34141, Republic of Korea. <sup>4</sup>Tomocube, Inc., Daejeon, 34051, Republic of Korea. <sup>5</sup>Present address: Department of Physics, University of Cambridge, Cambridge, CB3 0HE, UK. Jonghee Yoon and YoungJu Jo contributed equally to this work. Correspondence and requests for materials should be addressed to Y.P. (email: [yk.park@kaist.ac.kr](mailto:yk.park@kaist.ac.kr))



**Figure 1.** Schematic diagrams of the label-free identification of individual lymphocytes using optical diffraction tomography and machine learning. **(a)** Procedures for label-free measurement of 3-D RI tomograms of lymphocytes. Multiple holograms of a lymphocyte are measured by changing the angle of illumination. Optical field information with various incident angles was retrieved from the measured holograms, and then 3-D RI tomograms were reconstructed using an optical diffraction tomography algorithm. Scale bar, 5  $\mu\text{m}$ . **(b)** Training a statistical cell type classifier for identifying lymphocyte cell types using the  $k$ -NN algorithm. The multiple quantitative morphological and biochemical features of individual lymphocytes were combined to recognize and exploit the cell-type-specific fingerprints via supervise learning. **(c)** Identifying the cell types of newly observed individual lymphocytes with the established classifier. **(d)** Schematic of the experimental setup. A Mach-Zehnder interferometric microscope equipped with a 2-D scanning galvanomirror (GM) was used for measuring the holograms of individual lymphocytes. L1–6, lenses; P, pinhole; BS1–2, beam splitters; M1–4, mirrors; CL, condenser lens; OL, objective lens.

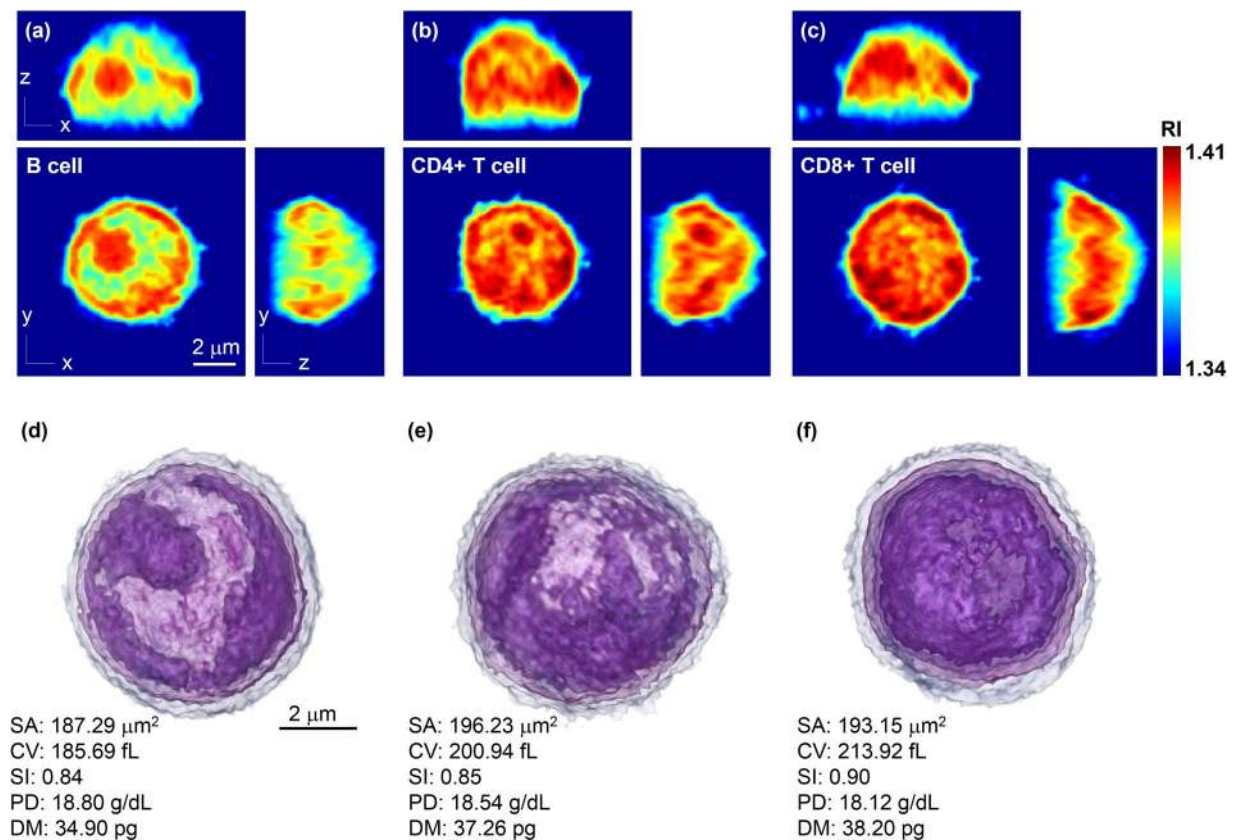
Label-free approaches such as mass spectroscopy<sup>12</sup> and Raman spectroscopy<sup>10</sup> have also been introduced to overcome the limitations of labelling methods because these spectroscopic methods exploit intrinsic biochemical properties of cells. Mass spectroscopy measures cellular biochemical properties which enable the profiling of lymphocyte proteins as well as the identification of lymphocyte cell types. However, it has a limitation in live-cell analysis due to the homogenization process of the cells. Raman spectroscopy measures molecular vibrations and also characterizes biochemical properties of a sample. Raman spectroscopy permits label-free live-cell analysis of lymphocytes with high accuracy; however, it requires a bulky optical system and long acquisition time (typically several seconds per cell) which limits its practical use.

Here, we present a method to identify lymphocyte cell types by exploiting optical diffraction tomography (ODT) and machine learning. ODT is a label-free imaging technique that measures a three-dimensional (3-D) refractive index (RI) tomogram of a sample which provides quantitative morphological and biochemical information<sup>13,14</sup>. ODT has been widely used to study various biological samples including red blood cells<sup>15–22</sup>, white blood cells (WBC)<sup>23,24</sup>, hepatocytes<sup>25</sup>, cancer cells<sup>16,26–32</sup>, neurons<sup>32,33</sup>, bacteria<sup>34–36</sup>, phytoplankton<sup>37</sup>, and hair<sup>38</sup>. In our previous study, we reported that ODT enables the quantitative analysis of WBCs including lymphocytes and macrophages<sup>23</sup>; we demonstrated that the two WBC subtypes could be discriminated using ODT. However, we were unable to simply identify lymphocyte cell types due to their nearly indistinguishable cellular morphology and biochemical characteristics.

In the present study, we use machine learning techniques to systematically interrogate the subtle differences between the lymphocyte cell types. Since RI is an intrinsic property of each biochemical component, the measured 3-D RI tomograms should encode the cell-type-specific fingerprints. However, it is difficult to manually discover such fingerprint information due to the complexity of 3-D tomograms. To solve this difficulty, statistical classification methods construct classification models by combining multiple features in a data-driven manner, instead of conventional hypothesis-driven investigations. This approach is especially powerful for high-dimensional data that are extremely difficult to be manually processed by humans due to the complexity and large size<sup>39</sup>, and thus machine learning techniques have been widely used to solve complex biological problems: identification of bacterial species<sup>40,41</sup>, discrimination of WBC subtypes<sup>42,43</sup>, investigation of pathophysiological conditions<sup>44–46</sup>, and classification of kinetic cell states<sup>47</sup>. Here we combine 3-D RI tomography and machine learning for the first time; we exploit statistical classification techniques to establish the cell type classifiers using the quantitative morphological and biochemical information extracted from the 3-D RI tomograms of individual lymphocytes. The trained classifiers enable identification of three lymphocyte cell types (B, CD4+ T, and CD8+ T cells) with high sensitivity and specificity.

## Results

The overall procedures for identification of non-activated lymphocytes are summarized in Fig. 1. The present approach involves three steps: (i) measurement of the 3-D RI tomograms of individual lymphocytes (Fig. 1a), (ii) construction of the statistical cell type classifiers using the quantitative biochemical and morphological features



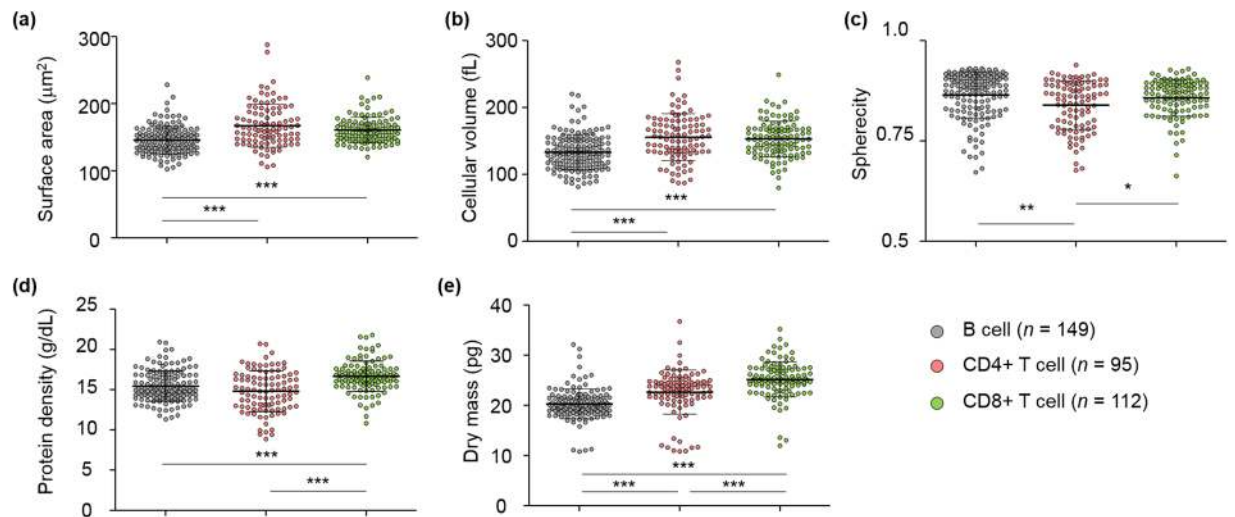
**Figure 2.** Representative 3-D RI tomograms of each lymphocyte cell type and the 3-D rendered images with quantitative characterization. Cross-sectional slices of a RI tomogram of (a) B cell, (b) CD4+ T cell, and (c) CD8+ T cell. Scale bar, 2  $\mu\text{m}$ . (d–f) 3-D rendered tomograms and quantitative characterization of the morphological and biochemical features of (a–c). Scale bar, 2  $\mu\text{m}$ . SA, surface area; CV, cellular volume; SI, sphericity; PD, protein density; DM, dry mass.

extracted from the tomograms (Fig. 1b), and (iii) identification of the new individual lymphocytes using the established classifiers (Fig. 1c).

Figure 1a shows the procedures for reconstructing the RI tomograms of the individual lymphocytes. To reconstruct a 3-D RI tomograms, multiple 2-D holograms of a cell are measured at various angles of illuminations using an interferometric microscope<sup>48</sup> (Fig. 1d). A coherent laser beam is split into two arms by a beam splitter. One arm passes through a sample, and then the diffracted light from the sample is projected onto a camera plane through a microscope. At the camera plane, the sample beam interferes with the other arm to generate a spatially modulated hologram. The angle of the beam impinging onto the sample is controlled by a dual-axis galvanomirror. From the measured holograms, complex optical fields consisting of both the amplitude and quantitative phase images are retrieved using a field retrieval algorithm<sup>49,50</sup>. Then, a 3-D RI tomogram of a lymphocyte is reconstructed using the multiple optical amplitude and phase information via an optical diffraction tomography algorithm<sup>14,51</sup> (see *Methods*).

We obtained and sorted three lymphocyte cell types, B, CD4+ T, and CD8+ T cells, from mice peripheral blood (see *Methods*). The measured 3-D RI tomograms of the individual lymphocytes are shown in Fig. 2. The cross-sectional slices of the representative 3-D tomograms of the three cell types are shown in Fig. 2a–c, respectively. To facilitate visualization, the tomographic data was 3-D rendered using a customized transfer function in a commercialized software (TomoStudio<sup>TM</sup>, Tomocube Inc., Republic of Korea) to resemble haematoxylin and eosin staining (Fig. 2d–f and Supplementary Videos 1–3). Clearly, the measured RI distribution visualizes the cellular boundaries and internal organelles such as nuclear membranes and nucleoli. The B cell shows a well-defined nucleus and nucleoli with RI values ranging from 1.34 to 1.41. We also note that the RI values of the cytosolic regions of the CD4+ and CD8+ T cells are higher than that of the B cell. Despite of these slight differences in RI distribution, the cell-type-specific fingerprints for cell type identification could not be clearly defined through visual interrogation, mainly due to the cell-to-cell variations.

Next we extracted the quantitative characteristics of the individual lymphocytes from the 3-D RI tomograms as illustrated in Fig. 3 (n = 149, 95, and 112 for B, CD4+ T, and CD8+ T cells, respectively). The five quantitative morphological (surface area, volume, and sphericity) and biochemical (protein density and dry mass) parameters were calculated from the tomograms (see *Methods*). The cellular surface area and volume of the lymphocytes were simply calculated from the segmented (RI threshold = 1.340) voxel information of the 3-D tomograms, and then the sphericity, a dimensionless parameter that indicates the roundness of the cellular morphology, was obtained



**Figure 3.** Quantitative analysis of the morphological and biochemical parameters of the individual B, CD4+ T, and CD8+ T cells. (a–c) The scatter plots of the morphological features: (a) surface area, (b) cell volume, and (c) sphericity. (d,e) The scatter plots of the biochemical features: (d) protein density and (e) dry mass. Each symbol indicates a single-cell measurement. Note that a single 3-D RI tomogram simultaneously provides all five parameters. Horizontal black lines, mean values; vertical lines with intervals, standard deviation. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

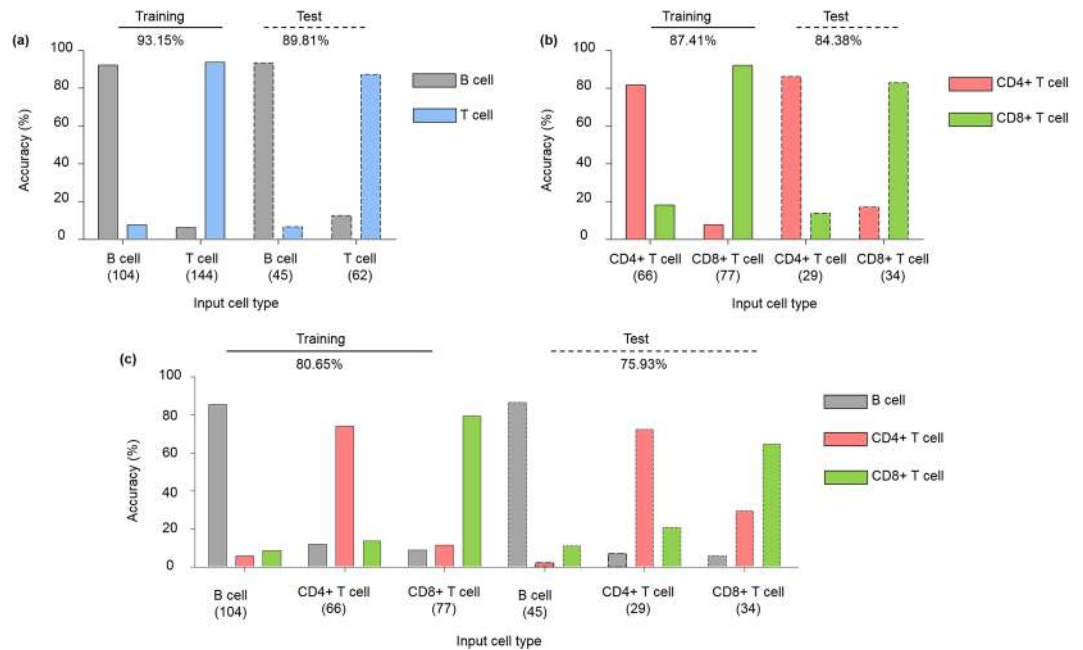
by the ratio of the surface area and volume. The cellular protein density and dry mass were retrieved using the RI values that are linearly proportional to the local concentration of non-aqueous molecules (mostly proteins).

To investigate the differences among the lymphocyte cell types, a statistical analysis was conducted on the quantitative parameters of the three lymphocyte cell types. First we studied the inter-type differences in the quantitative morphological features. The cellular surface areas of the B, CD4+ T, and CD8+ T lymphocytes were  $145.87 \pm 20.25$ ,  $167.23 \pm 32.08$ , and  $160.80 \pm 19.12 \mu\text{m}^2$ , respectively (Fig. 3a). The B cells had significantly smaller cellular surface areas compared to the T cells ( $P < 0.001$ ), while there was no significant difference between the CD4+ and CD8+ T cells. The cellular volumes of the lymphocytes also show a similar tendency with the result of the cellular surface area analysis. The cellular volumes of the B cells ( $133.43 \pm 26.47 \text{ fL}$ ) were significantly smaller ( $P < 0.001$ ) compared to those of the CD4+ T ( $155.73 \pm 35.14 \text{ fL}$ ) and CD8+ T ( $152.77 \pm 26.52 \text{ fL}$ ) cells (Fig. 3b), while the CD4+ and CD8+ T cells had similar cellular volumes. The sphericities were  $0.86 \pm 0.06$ ,  $0.84 \pm 0.06$ , and  $0.86 \pm 0.05$  for the B, CD4+ T, and CD8+ T cells, respectively (Fig. 3c). The sphericities of the CD4+ T cells were statistically smaller than those of the B cells ( $P < 0.01$ ) and CD8+ T cells ( $P < 0.05$ ). Note that all the lymphocyte cell types had high sphericity values which suggest the round shapes of the lymphocytes.

We also compared the biochemical properties of the lymphocyte cell types. The cellular protein densities of the B, CD4+ T, and CD8+ T cells were  $15.43 \pm 1.88$ ,  $14.81 \pm 2.54$ , and  $16.66 \pm 1.88 \text{ g/dL}$ , respectively (Fig. 3d). The CD8+ T cells had significantly higher cellular protein densities compared to those of the others ( $P < 0.001$ ). The cellular dry mass was  $20.28 \pm 2.97$ ,  $22.65 \pm 4.49$ , and  $25.19 \pm 3.51 \text{ pg}$  for the B, CD4+ T, and CD8+ T cells, respectively, and showed significant differences ( $P < 0.001$ ) between all the cell types (Fig. 3e). The B cells had a smaller cellular dry mass compared to the T cells. Moreover, the CD8+ T cells were statistically heavier than the CD4+ T cells. To summarize, we observed several statistical differences in both morphological and biochemical features at the *population* level. However, we failed to manually establish cell type classifiers based on these parameters at the *single-cell* level, due to the high dimensionality of the feature space and the cell-to-cell variations.

To achieve accurate identification of individual lymphocytes, we employed a machine learning approach to combine and exploit multiple features encoded in the 3-D RI tomograms. First we enlarged the feature space by extracting the quantitative parameters calculated at 20 different threshold RI values (from 1.340 to 1.378 with an increment of 0.002) to reveal the information specific to intracellular components in addition to the overall morphology (Supplementary Fig. 1). Then we systematically investigated the 100-dimensional feature space (5 parameters per threshold value), which is impractical to be manually explored, using statistical classification models. We employed the well-known  $k$ -nearest neighbours ( $k$ -NN) algorithm<sup>52, 53</sup> (see *Methods*) with  $k = 4$  after comparing several models (Supplementary Fig. 2).

The statistical classification, or supervised machine learning, was performed through the training and test stages as explained earlier (Fig. 1b,c). We randomly split each lymphocyte subtype data into 70% and 30% of training ( $n = 104$ , 66, and 77 for B, CD4+ T, and CD8+ T cells, respectively) and test sets ( $n = 45$ , 29, and 35 for B, CD4+ T, and CD8+ T cells, respectively), respectively. First we constructed the cell type classifiers by combining subsets of the features extracted from the training data set. We exhaustively searched the optimal combinations of the features via cross-validation (see *Methods*) and the classifier with the best training accuracy was selected. The established classifiers exploit the cell-type-specific fingerprints recognized from the high-dimensional feature space. To test if these fingerprints are general to new lymphocytes, we identified the individual lymphocytes categorized as the test data. The test accuracy and its sub-parameters, called sensitivity (true positive results over all



**Figure 4.** Identification performance of the optimized lymphocyte cell type classifiers. The performance of cell type identification was illustrated for (a) binary classification of B and T cells, (b) binary classification of CD4+ and CD8+ T cells, and (c) multiclass classification of all three lymphocyte cell types, for both training and test sets. Note the small difference between the training and test cases, suggesting nice generalization of the trained classifiers. The numbers below the name of each cell type indicate the number of cells used.

Training				Test			
Input cell type	Output cell type (number of cell)	Sensitivity (%)		Input cell type	Output cell type (number of cell)	Sensitivity (%)	
		B cell	T cell			B cell	T cell
B cell	B cell	96	8	B cell	B cell	42	3
	T cell	9	135		T cell	8	55
Specificity (%)		93.75	92.31	Specificity (%)		81.30	93.33
				93.15 (overall)			
				89.81 (overall)			

**Table 1.** Detailed performance of the B and T lymphocyte cell type classifiers. The identification accuracy and its sub-parameters, sensitivity and specificity, were also presented.

Training				Test			
Input cell type	Output cell type (number of cell)	Sensitivity (%)		Input cell type	Output cell type (number of cell)	Sensitivity (%)	
		CD4+ T cell	CD8+ T cell			CD4+ T cell	CD8+ T cell
CD4+ T cell	CD4+ T cell	54	12	CD4+ T cell	CD4+ T cell	25	4
	CD8+ T cell	6	71		CD8+ T cell	6	29
Specificity (%)		92.21	81.82	Specificity (%)		82.86	86.21
				87.41 (overall)			
				84.38 (overall)			

**Table 2.** Detailed performance of the CD4+ and CD8+ T lymphocyte cell type classifiers. The identification accuracy and its sub-parameters, sensitivity and specificity, were also presented.

positive inputs) and specificity (true negative results over all negative inputs), were calculated by comparing the machine-predicted and true cell types.

Figure 4 and Tables 1–3 illustrate the identification performance for both training and test stages. We performed statistical classification on three different combinations of the lymphocyte cell types: (i) binary classification of B and T lymphocytes, (ii) binary classification of the two T lymphocyte types (CD4+ and CD8+), and (iii) multiclass classification of all three types of lymphocytes. First, the two T cell types were considered as one class to train a binary classifier of B and T cells (Fig. 4a). The accuracy of the optimized classifier was 93.15% and 89.81%

Training					Test						
		Output cell type (number of cell)			Sensitivity (%)			Output cell type (number of cell)			Sensitivity (%)
		B cell	CD4+ T cell	CD8+ T cell				B cell	CD4+ T cell	CD8+ T cell	
Input cell type	B cell	89	6	9	85.58	Input cell type	B cell	39	1	5	86.67
	CD4+ T cell	8	49	9	74.24		CD4+ T cell	2	21	6	72.41
	CD8+ T cell	7	9	62	79.49		CD8+ T cell	2	10	22	64.71
Specificity (%)		89.58	91.76	89.41	80.65 (overall)	Specificity (%)		93.65	86.08	89.41	75.93 (overall)

**Table 3.** Detailed performance of the three lymphocyte cell types classifiers. The identification accuracy and its sub-parameters, sensitivity and specificity, were also presented.

for the training and test cases, respectively (selected features: surface area (RI threshold = 1.342, 1.368), volume (1.368), sphericity (1.342, 1.368), protein density (1.368), and dry mass (1.342, 1.368)). Second, the CD4+ and CD8+ T cells were statistically classified also in a binary fashion (Fig. 4b). The accuracy was 87.41% and 84.38% for the training and test sets, respectively (selected features: surface area (1.342, 1.362) and sphericity (1.324, 1.362)). Lastly, the multiclass cell type classifier of the three lymphocyte cell types was constructed. The identification accuracy for the training and test were 80.65% and 75.93%, respectively (selected features: surface area (1.340), sphericity (1.370), protein density (1.370), and dry mass (1.370)). The small differences in accuracy (i.e. negligible overfitting) suggest that the trained cell type classifiers make use of the general characteristics of each lymphocyte cell type; thus the trained classification models would accurately identify the newly observed individual lymphocytes.

## Discussion

We demonstrated the identification of non-activated lymphocyte cell types at the single-cell level using ODT and machine learning. ODT provides quantitative morphological and biochemical information on the individual lymphocytes by measuring their 3-D RI distribution. We extracted the quantitative parameters from the 3-D tomograms and observed significant differences between the cell types at the population level. However, we failed to construct accurate cell type classifiers at the single-cell level, mainly due to the high dimensionality of the feature space and the cell-to-cell variations. To overcome this limitation, the  $k$ -NN ( $k = 4$ ) algorithm was employed as a statistical classification method to systematically extract and exploit the cell-type-specific fingerprints encoded in the 3-D RI tomograms. The optimized cell type classifiers can discriminate B and T cells with high accuracy of approximately 90%. The CD4+ and CD8+ T cells could be distinguished as well with an overall accuracy of over 80%. In addition, the simultaneous multiclass identification of the three lymphocyte cell types presented an overall accuracy of over 75%.

The identification performance shows that the classification models discriminate B and T lymphocytes more precisely compared to the T cell subtype identification (CD4+ and CD8+ T cells); these results imply that the differences in cellular morphology and biochemical properties between the B and T cells are more distinct than those between the CD4+ and CD8+ T cells. This observation is consistent with the previous knowledge of the lymphocyte-differentiation pathway<sup>54</sup>. The B and T lymphocytes originate from hematopoietic stem cells and then mature in different organs. Thus, the lymphocytes have similar cellular phenotypes such as one large nucleus and spherical shapes; however, the B and T cells have entirely different cellular functions. Even though our method established the cell type classifiers by optimizing the features based on the statistical performance instead of biological relevance, the machine learning algorithm automatically recognizes and exploits the distinct differences in the morphological and biochemical properties between the lymphocyte cell types.

The present approach combines 3-D RI tomography and statistical classification, for the first time to our knowledge, which provides several advantages. First, the present method enables the identification of lymphocyte cell types exploiting intrinsic optical properties of cells, which cannot be achieved by conventional optical microscopic techniques without using fluorescent methods. As ODT measures intrinsic 3-D RI distribution of individual lymphocytes, it provides consistent and highly reproducible results. In contrast, labeling methods such as fluorescence microscopy techniques can provide molecular-specific localization information. However, their signals are generally qualitative, and may vary significantly depending on experimental protocols, skills, and equipment. This variability of labeling methods may decrease overall identification accuracy. Alternatively, existing bright-field or phase contrast microscopy can be used to measure 2-D morphological features, such as projected area, aspect ratio, and nucleus size. However, these qualitative image data obtained with these conventional bright-field or phase contrast microscopy only provide limited information. We found that statistical classification performed by morphological information only (surface area, cellular volume, and sphericity) lowers overall accuracy (Supplementary Fig. 3). This result clearly indicates the advantages of using the 3-D RI tomogram. Second, the present method uses a simple and cost-effective optical setup compared to fluorescence-activated cell sorting or other label-free techniques such as Raman spectroscopy. Recently, a 3-D holographic microscope has become commercially available, which simplifies the optical system and achieves over 100 tomograms per second by exploiting a digital micromirror device to reduce the required time for measuring a 3-D RI tomogram<sup>55, 56</sup>. Thus, the present approach can be easily transferred to basic research facilities and clinics. Lastly, there is no limitation in applying the present method to discriminate other types of cells including RBCs, cancer cells, neurons,

and glial cells. Because ODT has been widely used to measure various biological samples, the present approach can be readily used to identify various cell types.

One of the limitations of the current study is that antibody-labelled lymphocytes were used to demonstrate the proof of principle. Because labeling and sorting procedures are inevitable to confirm the presence of antibodies in lymphocytes, and the use of antibody labeling is a standard technique to assure the subtypes of antibody-specific classification. Nonetheless, this approach still uses the intrinsic imaging contrast – 3-D RI tomography, and the labeling agents used to specify antibodies have negligible effects to the measured 3-D RI signal. CD4 antibody bound per cell value is about 48,000 and the total number of protein per cell of a eukaryotic cell is in the order of billions<sup>57, 58</sup>. Thus, antibody staining accounts for less than 0.0005% of total protein numbers, which can be ignored in the measured 3-D RI tomogram because the RI signals attribute to total protein distributions.

From the algorithmic point of view, there are several points to be improved for practical use of the proposed technique. As described earlier, we compared several statistical classification algorithms including the  $k$ -NN ( $k = 4$  and  $k = 6$ ), linear discrimination analysis, quadratic discrimination analysis, naïve Bayes, and decision tree (Supplementary Fig. 2), and selected the  $k$ -NN ( $k = 4$ ) as the classification model. While we tested several machine learning algorithms exploiting quantitative features with different thresholds widely used in ODT-based studies, employing advanced features extraction methods and statistical classification models could enhance the overall identification performance. Unfortunately, designing powerful features for 3-D biological microscopy data, especially for 3-D RI tomograms, has been largely unexplored and is beyond the scope of this proof-of-concept study. The methods developed in different disciplines could be translated to facilitate this direction: scale-invariant feature transform for X-ray computerized tomography<sup>59</sup> or histogram-based features for lidar-based point clouds<sup>60</sup>. However, these ‘shallow’ descriptors require laborious optimization procedures specific to samples and imaging setups employed.

We expect that ‘deep’ learning, a state-of-the-art machine learning technique based on multilayered neural networks, could be a powerful and generic feature extraction strategy for 3-D biological microscopy. Recently, our group successfully combined 2-D holographic microscopy and deep learning for the label-free screening of multiple pathogens<sup>61</sup>. Upon extension to 3-D ODT, the remarkable flexibility and learning abilities of deep neural networks would let us fully exploit the complex information encoded in the 3-D RI tomograms and dramatically enhance the identification performance. An important step in this direction is to combine the proposed method with high-throughput imaging technologies<sup>62</sup> to obtain a large size of training data sufficient for deep learning.

In summary, we envision that ODT combined with machine learning will be a useful tool in biomedical research. ODT quantitatively provides the morphological and biochemical characteristics of the samples, and then machine learning enables the label-free identification of cell types using the measured quantitative information. The present method can be widely used in the study of immunology, cancer biology, and neuroscience.

## Methods

**Mice.** C57BL/6J mice (gender and age-matched, 6–8 weeks) were purchased from Daehan Biolink (Republic of Korea). Animal care and experimental procedures were performed under approval of the Institutional Animal Care and Use Committee of KAIST (KA2010-21, KA2014-01 and KA2015-03). All the experiments in this study were carried out in accordance with the approved guidelines.

**Flow cytometry for lymphocyte sorting.** White blood cells were isolated from the blood harvested from the heart of mice. Erythrocytes were removed by ACK lysis. Cells were blocked with anti-CD16/32 and then stained for surface molecules. DAPI (4,6-diamidino-2-phenylindole; Roche, Switzerland) was used for dead cell exclusion. Sorting was performed on an Aria II or III system (BD Biosciences, CA) using an 85- $\mu$ m nozzle or Astrios system (Beckman Coulter, CA) using a 70- $\mu$ m nozzle. Antibodies for flow cytometry were purchased from BD Biosciences, eBioscience (CA), Biolegend (CA). The antibodies used were CD3 $\epsilon$  (clone 17A2), CD4 (GK1.5), CD8 $\alpha$  (53-6.7), CD19 (1D3), CD45R (B220, RA3-6B2), NK1.1 (PK136).

**3-D refractive index tomography.** To reconstruct the 3-D RI tomograms of lymphocytes, a Mach-Zehnder interferometric microscope was used<sup>15</sup> (Fig. 1d). A laser beam from a diode-pumped solid-state laser ( $\lambda = 532$  nm, 100 mW, Shanghai Dream Laser Co., China) is split into two arms using a beam splitter. One arm illuminates a sample with various illumination angles ranging from  $-60^\circ$  to  $60^\circ$  in air at the sample plane with respect to the optic axis, which is systematically controlled with a dual-axis galvanomirror (GVS012, Thorlabs, NJ), and the other is used as a reference beam. The sample is placed between a condenser lens (UPLSAPO Water 60 $\times$ , numerical aperture (NA) = 1.2, Olympus, Japan) and an objective lens (PLAPON Oil 100 $\times$ , NA = 1.4, Olympus, Japan). For a single cell level analysis, we locate a single cell in the field of view using a manual translation stage. The diffracted light from the sample is then collected by the objective lens and projected onto the camera plane. At the camera plane, the sample beam interferes with the reference beam, generating spatially modulated holograms, which are then captured by a CMOS camera (1024 PCI, Photron USA Inc., CA). For reconstructing a 3-D RI tomogram, a total of 300 holograms of a sample are measured with a frame rate of 1000 Hz by changing the angle of illuminations which takes 0.3 s. Then, the optical field information (amplitude and phase) of the measured holograms are retrieved using a field retrieval algorithm based on Fourier transform<sup>49, 50</sup>. From the retrieved multiple amplitude and phase information, a 3-D RI tomogram is reconstructed using an optical diffraction tomography algorithm<sup>14, 51</sup>. An iterative regularization algorithm with a non-negativity constraint was used to fill the missing cone information which results from the limited NA of the condenser and objective lenses<sup>63</sup>. Details on reconstructing 3-D RI tomograms can be found elsewhere<sup>15, 64</sup>. The experimental resolution of our setup estimated by imaging a micro-bead was 373 nm and 496 nm for lateral and axial directions, respectively, which is consistent with the theoretical values<sup>65</sup>.



**Image processing and statistical analysis.** Image processing was performed with Matlab (R2014b; MathWorks Inc., MA) and ImageJ (the National Institutes of Health, MD). The RI isosurfaces were rendered with commercial software (TomoStudio™, Tomocube Inc., Republic of Korea). Statistical analysis was done with GraphPad Prism (GraphPad Software Inc., CA). *P*-values were calculated by Student's *t*-test.

**Calculation of the quantitative structural and biochemical characteristics.** The quantitative structural and biochemical parameters of the individual lymphocytes were calculated from the measured 3-D RI tomograms. To calculate the cellular surface area *S* and volume *V*, the voxels with the RI values higher than the threshold RI value were selected for segmentation from a 3-D RI tomogram of a lymphocyte. The surface area and volume were calculated from the number of voxels at the boundary and inner region of the segmented region, respectively. The sphericity, which is a dimensionless parameter that indicates the roundness of a lymphocyte, was obtained from the calculated surface area and volume as follows:  $Sphericity = \pi^{1/3} \cdot (6V)^{2/3} / S$ . The biochemical characteristics (protein density and dry mass) were obtained from the RI values due to the well-characterized linear relation between the RI value and the local concentration of non-aqueous molecules (i.e., proteins, lipids, and nucleic acids inside cells; mostly proteins). RI values were converted to the protein density *C* with the following relation:  $n = n_0 + \alpha C$ , where *n* and *n*<sub>0</sub> are the RI values of a voxel and the medium, respectively, and  $\alpha$  is the refractive index increment (RII). Because it is known that most proteins have similar RII values, we used a RII value of 0.2 mL/g in this study. The total dry mass of a lymphocyte was calculated by simply integrating the protein density over the cellular volume. Details on calculating the quantitative information from 3-D RI tomograms can be found elsewhere<sup>23, 25</sup>.

**Machine learning.** We investigated the 100-dimensional feature space as described in the main text. We selected the *k*-NN (*k* = 4) as the classification model after comparing several algorithms (Supplementary Fig. 2). The *k*-NN algorithm predicts the class of a newly observed data by choosing the most frequent class labels of *k* nearest neighbour data points in the feature space. We standardized all features prior training and test because *k*-NN is sensitive to pre-processing. Since there exists substantial redundancy between the features and it is desirable to choose minimal number of features to reduce overfitting, it was crucial to select the optimal feature set. We exhaustively searched all combinations of the morphological and biochemical features obtained at a single or two different RI threshold values. The feature set with the highest cross-validation accuracy was selected. The optimized classifier was tested using the data that was not utilized for training.

## References

- Radtke, F., Wilson, A., Mancini, S. J. C. & MacDonald, H. R. Notch regulation of lymphocyte development and function. *Nat Immunol* **5**, 247–253 (2004).
- Moser, B. & Loetscher, P. Lymphocyte traffic control by chemokines. *Nat Immunol* **2**, 123–128 (2001).
- Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511, doi:10.1038/35000501 (2000).
- Rose, M. G. & Berliner, N. T-cell large granular lymphocyte leukemia and related disorders. *Oncologist* **9**, 247–258, doi:10.1634/theoncologist.9-3-247 (2004).
- de Visser, K. E., Eichten, A. & Coussens, L. M. Paradoxical roles of the immune system during cancer development. *Nat Rev Cancer* **6**, 24–37, doi:10.1038/nrc1782 (2006).
- Ueda, H. *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423**, 506–511, doi:10.1038/nature01621 (2003).
- von Boehmer, H. & Melchers, F. Checkpoints in lymphocyte development and autoimmune disease. *Nat Immunol* **11**, 14–20 (2010).
- Cabrera, R. *et al.* An immunomodulatory role for CD4(+)CD25(+) regulatory T lymphocytes in hepatitis C virus infection. *Hepatology* **40**, 1062–1071, doi:10.1002/hep.20454 (2004).
- Saez-Cirion, A. *et al.* HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection *ex vivo* and peculiar cytotoxic T lymphocyte activation phenotype. *P Natl Acad Sci USA* **104**, 6776–6781, doi:10.1073/pnas.0611244104 (2007).
- Hobro, A. J., Kumagai, Y., Akira, S. & Smith, N. I. Raman spectroscopy as a tool for label-free lymphocyte cell line discrimination. *Analyst* **141**, 3756–3764 (2016).
- Fischer, K. *et al.* Isolation and characterization of human antigen-specific TCR alpha beta(+) CD4(-)CD8(-) double-negative regulatory T cells. *Blood* **105**, 2828–2835 (2005).
- Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378–386 (2009).
- Kim, D. *et al.* Refractive index as an intrinsic imaging contrast for 3-D label-free live cell imaging. *bioRxiv*, doi:10.1101/106328 (2017).
- Kim, K. *et al.* Optical diffraction tomography techniques for the study of cell pathophysiology. *J Biomed Photonics Eng* **2**, 020201 (2016).
- Kim, K. *et al.* High-resolution three-dimensional imaging of red blood cells parasitized by *Plasmodium falciparum* and *in situ* hemozoin crystals using optical diffraction tomography. *J Biomed Opt* **19**, doi:10.1117/1.Jbo.19.1.011005 (2014).
- Kim, Y. *et al.* Common-path diffraction optical tomography for investigation of three-dimensional structures and dynamics of biological cells (vol 22, pg 10398, 2014). *Opt Express* **23**, 18996–18996, doi:10.1364/Oe.23.018996 (2015).
- Park, H. *et al.* Three-dimensional refractive index tomograms and deformability of individual human red blood cells from cord blood of newborn infants and maternal blood. *J Biomed Opt* **20**, 111208, doi:10.1117/1.jbo.20.11.111208 (2015).
- Park, H. *et al.* Characterizations of individual mouse red blood cells parasitized by *Babesia microti* using 3-D holographic microscopy. *Sci Rep* **5**, doi:10.1038/Srep10827 (2015).
- Lee, S. Y., Park, H. J., Best-Popescu, C., Jang, S. & Park, Y. K. The Effects of Ethanol on the Morphological and Biochemical Properties of Individual Human Red Blood Cells. *PLoS one* **10**, e0145327 (2015).
- Hur, J., Kim, K., Lee, S., Park, H. & Park, Y. Melittin-induced alterations in morphology and deformability of human red blood cells using quantitative phase imaging techniques. *bioRxiv* 091991 (2016).
- Lee, S., Park, H., Jang, S. & Yongkeun, P. Refractive index tomograms and dynamic membrane fluctuations of red blood cells from patients with diabetes mellitus. *Blood* **128**(22), 4813–4813 (2016).
- Park, H. *et al.* Measuring cell surface area and deformability of individual human red blood cells over blood storage using quantitative phase imaging. *Sci Rep* **6** (2016).
- Yoon, J. *et al.* Label-free characterization of white blood cells by measuring 3D refractive index maps. *Biomed Opt Express* **6**, 3865–3875 (2015).

24. Kim, K., Yoon, J. & Park, Y. Simultaneous 3D visualization and position tracking of optically trapped particles using optical diffraction tomography. *Optica* **2**, 343–346 (2015).
25. Kim, K. *et al.* Three-dimensional label-free imaging and quantification of lipid droplets in live hepatocytes. *Sci Rep* **6**, 36815 (2016).
26. Kim, D. *et al.* Label-free high-resolution 3-D imaging of gold nanoparticles inside live cells using optical diffraction tomography. *bioRxiv* 097113 (2016).
27. Sung, Y. J. *et al.* Optical diffraction tomography for high resolution live cell imaging. *Opt Express* **17**, 266–277, doi:[10.1364/Oe.17.000266](https://doi.org/10.1364/Oe.17.000266) (2009).
28. Kus, A., Dudek, M., Kemper, B., Kujawinska, M. & Vollmer, A. Tomographic phase microscopy of living three-dimensional cell cultures. *J Biomed Opt* **19**, doi:[10.1117/1.Jbo.19.4.046009](https://doi.org/10.1117/1.Jbo.19.4.046009) (2014).
29. Sung, Y., Choi, W., Lue, N., Dasari, R. R. & Yaqoob, Z. Stain-Free Quantification of Chromosomes in Live Cells Using Regularized Tomographic Phase Microscopy. *Plos One* **7**, doi:[10.1371/journal.pone.0049502](https://doi.org/10.1371/journal.pone.0049502) (2012).
30. Su, J. W., Hsu, W. C., Chou, C. Y., Chang, C. H. & Sung, K. B. Digital holographic microtomography for high-resolution refractive index mapping of live cells. *J Biophotonics* **6**, 416–424, doi:[10.1002/jbio.201200022](https://doi.org/10.1002/jbio.201200022) (2013).
31. Hsu, W. C., Su, J. W., Tseng, T. Y. & Sung, K. B. Tomographic diffractive microscopy of living cells based on a common-path configuration. *Opt Lett* **39**, 2210–2213, doi:[10.1364/Ol.39.002210](https://doi.org/10.1364/Ol.39.002210) (2014).
32. Cotte, Y. *et al.* Marker-free phase nanoscopy. *Nat Photo* **7**, 418–418, doi:[10.1038/nphoton.2013.116](https://doi.org/10.1038/nphoton.2013.116) (2013).
33. Yang, S.-A., Yoon, J., Kim, K. & Park, Y. Measurements of morphological and biochemical alterations in individual neuron cells associated with early neurotoxic effects in Parkinson's disease using optical diffraction tomography. *Cytometry Part A*, doi:[10.1002/cyto.a.23110](https://doi.org/10.1002/cyto.a.23110) (2017).
34. Bennet, M., Gur, D., Yoon, J., Park, Y. & Faivre, D. A Bacteria-Based Remotely Tunable Photonic Device. *Adv Opt Mater* (2016).
35. Kim, T. *et al.* White-light diffraction tomography of unlabelled live cells. *Nat Photon* **8**, 256–263 (2014).
36. Kim, T. I. *et al.* Antibacterial Activities of Graphene Oxide–Molybdenum Disulfide Nanocomposite Films. *ACS Appl Mater Interfaces* **9**, 7908–7917 (2017).
37. Lee, S. *et al.* High-Resolution 3-D Refractive Index Tomography and 2-D Synthetic Aperture Imaging of Live Phytoplankton. *J Opt Soc Korea* **18**, 691–697 (2014).
38. Lee, S. *et al.* Measurements of morphology and refractive indexes on human downy hairs using three-dimensional quantitative phase imaging. *J Biomed Opt* **20**, 111207–111207 (2015).
39. Kotsiantis, S. B., Zaharakis, I. & Pintelas, P. Machine learning: a review of classification and combining techniques. *Artif Intell Rev* **26**, 159 (2007).
40. Jo, Y. *et al.* Label-free identification of individual bacteria using Fourier transform light scattering. *Opt Express* **23**, 15792–15805, doi:[10.1364/Oe.23.015792](https://doi.org/10.1364/Oe.23.015792) (2015).
41. Jo, Y. *et al.* Angle-resolved light scattering of individual rod-shaped bacteria based on Fourier transform light scattering. *Sci Rep* **4**, 5090, doi:[10.1038/srep05090](https://doi.org/10.1038/srep05090) (2014).
42. Vercruyse, D. *et al.* Three-part differential of unlabeled leukocytes with a compact lens-free imaging flow cytometer. *Lab on a Chip* **15**, 1123–1132 (2015).
43. Chen, C. L. *et al.* Deep Learning in Label-free Cell Classification. *Sci Rep* **6** (2016).
44. Ramos-Pollan, R. *et al.* Discovering Mammography-based Machine Learning Classifiers for Breast Cancer Diagnosis. *J Med Syst* **36**, 2259–2269, doi:[10.1007/s10916-011-9693-2](https://doi.org/10.1007/s10916-011-9693-2) (2012).
45. Ozolek, J. A. *et al.* Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning. *Med Image Anal* **18**, 772–780, doi:[10.1016/j.media.2014.04.004](https://doi.org/10.1016/j.media.2014.04.004) (2014).
46. Bergner, N. *et al.* Identification of primary tumors of brain metastases by Raman imaging and support vector machines. *Chemometr Intell Lab* **117**, 224–232, doi:[10.1016/j.chemolab.2012.02.008](https://doi.org/10.1016/j.chemolab.2012.02.008) (2012).
47. Hejna, M., Jorapur, A., Song, J. S. & Judson, R. L. High accuracy label-free classification of kinetic cell states from holographic cytometry. *bioRxiv* 127449 (2017).
48. Lee, K. *et al.* Quantitative phase imaging techniques for the study of cell pathophysiology: from principles to applications. *Sensors* **13**, 4170–4191 (2013).
49. Takeda, M., Ina, H. & Kobayashi, S. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J Opt Soc Am* **72**, 156–160 (1982).
50. Debnath, S. K. & Park, Y. Real-time quantitative phase imaging with a spatial phase-shifting algorithm. *Opt Lett* **36**, 4677–4679 (2011).
51. Wolf, E. Three-dimensional structure determination of semi-transparent objects from holographic data. *Opt Commun* **1**, 153–156 (1969).
52. Miska, E. A. MicroRNA expression profiles classify human cancers. *Cytometry B Clin Cytom* **72b**, 126–126 (2007).
53. Chen, X. W., Zhou, X. B. & Wong, S. T. C. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *Ieee T Bio-Med Eng* **53**, 762–766, doi:[10.1109/Tbme.2006.870201](https://doi.org/10.1109/Tbme.2006.870201) (2006).
54. Ogawa, M. Differentiation and Proliferation of Hematopoietic Stem-Cells. *Blood* **81**, 2844–2853 (1993).
55. Shin, S. *et al.* Optical diffraction tomography using a digital micromirror device for stable measurements of 4D refractive index tomography of cells. *SPIE BiOS*. 971814-971814-971818 (2016).
56. Shin, S., Kim, K., Yoon, J. & Park, Y. Active illumination using a digital micromirror device for quantitative phase imaging. *Opt Lett* **40**, 5407–5410 (2015).
57. Wang, L. *et al.* Discrepancy in measuring CD4 expression on T-lymphocytes using fluorescein conjugates in comparison with unimolar CD4-phycoerythrin conjugates. *Cytometry B Clin Cytom* **72**, 442–449 (2007).
58. Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **35**, 1050–1055 (2013).
59. Flitton, G. T., Breckon, T. P. & Bouallagu, N. M. Object recognition using 3d sift in complex ct volumes. *BMVC* (2010).
60. Behley, J., Steinhage, V. & Cremers, A. B. Performance of histogram descriptors for the classification of 3D laser range data in urban environments In *Robotics and Automation (ICRA)*, 4391–4398 (2012).
61. Jo, Y. *et al.* Holographic deep learning for rapid optical screening of anthrax spores. *bioRxiv*, 109108 (2017).
62. Sung, Y. *et al.* Three-dimensional holographic refractive-index measurement of continuously flowing cells in a microfluidic channel. *Phys Rev Appl* **1**, 014002 (2014).
63. Lim, J. *et al.* Comparative study of iterative reconstruction algorithms for missing cone problems in optical diffraction tomography. *Opt Express* **23**, 16933–16948, doi:[10.1364/Oe.23.016933](https://doi.org/10.1364/Oe.23.016933) (2015).
64. Kim, K., Choe, K., Park, I., Kim, P. & Park, Y. Holographic intravital microscopy for 2-D and 3-D imaging intact circulating blood cells in microcapillaries of live mice. *Sci Rep* **6** (2016).
65. Lauer, V. New approach to optical diffraction tomography yielding a vector equation of diffraction tomography and a novel tomographic microscope. *J Microsc* **205**, 165–176, doi:[10.1046/j.0022-2720.2001.00980.x](https://doi.org/10.1046/j.0022-2720.2001.00980.x) (2002).

## Acknowledgements

This work was supported by KAIST, Tomocube Inc., and the National Research Foundation of Korea (2015R1A3A2066550, 2014M3C1A3052567, 2014K1A3A1A09063027 to Y.P., 2012M3A9B4027955 to S.K.). Y.J. acknowledges support from KAIST Presidential Fellowship.

## Author Contributions

Y.P. conceived of the idea and directed the work. J.Y. performed the optical experiments and processed the tomographic data. Y.J. designed and implemented the classification models. J.Y. and Y.J. optimized the cell type classifiers. M.K. isolated and sorted the lymphocytes from mice peripheral blood. K.K. designed the optical system. J.Y., Y.J., M.K., S.K., and Y.P. analysed the data. All authors wrote and revised the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-06311-y](https://doi.org/10.1038/s41598-017-06311-y)

**Competing Interests:** Y.P. has financial interests in Tomocube Inc., a company that commercializes optical diffraction tomography and quantitative phase imaging instruments and is one of the sponsors of this study.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017