

# Label-free identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning

Jonghee Yoon<sup>1,2,+</sup>, YoungJu Jo<sup>1,+</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>  
and YongKeun Park<sup>1,2,4,\*</sup>

<sup>1</sup>*Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea*

<sup>2</sup>*KAIST Institute Health Science and Technology, Daejeon 34141, South Korea*

<sup>3</sup>*Department of Biological Sciences, KAIST, Daejeon 34141, South Korea*

<sup>4</sup>*Tomocube, Inc., Daejeon 34051, South Korea*

*+These authors contributed equally to this work*

## Correspondence:

### YongKeun Park, PhD

KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea

Phone: +82-42-350-2514; Fax: +82-42-350-2510

Email: [yk.park@kaist.ac.kr](mailto:yk.park@kaist.ac.kr)

**Identification of lymphocyte cell types is crucial for understanding their pathophysiological roles in human diseases. Current methods for discriminating lymphocyte cell types primarily relies 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 label-free identification of non-activated lymphocyte subtypes using refractive index tomography. From the measurements of three-dimensional refractive index maps of individual lymphocytes, the morphological and biochemical properties of the lymphocytes are quantitatively retrieved. Machine learning methods establish an optimized classification model using the retrieved quantitative characteristics of the lymphocytes to identify lymphocyte subtypes at the individual cell level. We show that our approach enables label-free identification of three lymphocyte cell types (B, CD4+ T, and CD8+ T lymphocytes) with high specificity and sensitivity. The present method will be a versatile tool for investigating the pathophysiological roles of lymphocytes in various diseases including cancers, autoimmune diseases, and virus infections.**

Lymphocytes consisting of various cell types including B, helper (CD4+) T, cytotoxic (CD8+) T, and regulatory T lymphocytes, 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 discriminate lymphocyte cell types. Because different kinds of lymphocytes have very similar cellular morphology such as a large nucleus with small cytosolic regions and round shapes, conventional optical methods such as bright-field microscopy or phase contrast microscopy have limited ability in classifying lymphocyte cell types<sup>10</sup>. To overcome this, a specific surface membrane proteins, known as surface markers, are recognized and tagged with magnetic beads or fluorescence molecules via antigen-antibody binding. And then, specific types of lymphocytes are identified and separated by magnetic forces or fluorescence signals<sup>11</sup>. Targeting surface markers is a precise and efficient manner to determine the cell types; however, labelling methods have potential risks of altering cellular functions by modifying membrane protein structures. Moreover, labelling methods have limitations such as they cannot simultaneously identify multiple cell types due to the limited numbers of distinguishable labelling agents<sup>12</sup>.

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 lymphocytes. Mass spectroscopy measures cellular biochemical properties which enable the profiling of lymphocyte proteins as well as the identification of lymphocyte subtypes. However, it has a limitation in live-cell analysis due to the homogenization process of the cells. Raman spectroscopy measures molecular vibrations and 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 (several seconds) to measure 2-D Raman signals, which limits its broad use in clinics.

Here, we present a label-free 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-dimension (3-D) refractive index (RI) tomogram of a sample which quantitatively provides morphological and biochemical information<sup>13,14</sup>. Previously, 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>34,35</sup>, phytoplankton<sup>36</sup>, and hair<sup>37</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 lymphocytes and macrophages could be discriminated based on quantitative morphological and biochemical properties; however, we were unable to identify lymphocyte cell types due to their indistinguishable cellular morphology and biochemical characteristics.

In the present study, to find morphological and biochemical differences in indistinguishable lymphocyte cell types, we used machine learning techniques. Machine learning methods construct classification models by combining multiple features in a data-driven manner, not a hypothesis-driven manner. This approach is especially powerful for high-dimensional data that are extremely difficult to manually process by a human due to the complicity and large size<sup>38</sup>. In biomedical research fields, machine learning methods have been widely used to solve complex biological problems: identification of bacterial species<sup>39</sup>, discrimination of WBC types<sup>40,41</sup>, and investigation of pathophysiologic conditions<sup>42-44</sup>. However, there has been no applications that use both 3-D RI tomography and machine learning for the purpose of biomedical studies. Here, we exploit machine learning techniques to establish classification models using the quantitative morphological and biochemical information of lymphocytes which is retrieved from the 3-D RI tomograms of individual cells and show that established classification models enable the identification of three lymphocyte subtypes (B, CD4+ T, and CD8+ T lymphocytes). The present approach is a versatile tool for determining lymphocyte cell types in a label-free manner.

## Results

The overall procedure for label-free identification of lymphocytes is summarised in Fig. 1. The present method involves three steps: (i) measurement of the 3-D RI tomograms of individual lymphocytes (Fig. 1a), (ii) establishment of a statistical classification model using the quantitative biochemical and morphological features of the lymphocytes obtained from the 3-D RI tomograms (Fig. 1b), and (iii) identification of the lymphocyte types using the established classifier at a single cell level (Fig. 1c).

Figure 1a shows the procedures for reconstructing the 3-D RI tomograms of the lymphocytes. To reconstruct the 3-D RI tomograms, multiple 2-D holograms of a cell are measured at various angles of illuminations using an interferometric microscope (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 with a microscope. At the camera plane, the sample beam interferes with the other arm and generates 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 (See Methods). Then, a 3-D RI tomogram of a lymphocyte is reconstructed using the retrieved multiple optical amplitude and phase information via an optical diffraction tomography algorithm<sup>14,45</sup>.

To test whether 3-D RI tomograms of lymphocytes enable the identification of their cell types, three types of lymphocytes (B cell, CD4<sup>+</sup> T, and CD8<sup>+</sup> T lymphocytes) were obtained from mice peripheral blood through specific surface marker staining and flow cytometry (See Methods) prior to measuring on the 3-D RI tomograms. Figure 2 shows representative 3-D RI tomograms of each lymphocyte type. Cross-sectional slices of the representative 3-D RI tomograms of B cell, CD4<sup>+</sup> T cell, and CD8<sup>+</sup> T cell in the  $x$ - $y$ ,  $y$ - $z$ , and  $x$ - $z$  planes are also shown in Figs. 2a-c, respectively. The measured RI distribution clearly represents the cellular boundaries and internal organelles such as the nuclear membrane and nucleoli. The B cell shows a well-defined nucleus and nucleoli with RI values ranging from 1.34 to 1.41, while the RI values of the cytosolic areas of CD4<sup>+</sup> and CD8<sup>+</sup> T cells are higher than that of the B cell. 3-D RI tomograms of each lymphocyte type in Figs. 2a-c are 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 (Figs. 2d-f and Supplementary Videos 1-3). These 3-D rendered images show their round cellular boundaries and intracellular components. Even though each lymphocyte type looks like it has a slightly different RI distribution, they are basically indistinguishable due to their identical cellular size and morphology.

To investigate the differences among the lymphocyte cell types, a statistical analysis was conducted of the quantitative characteristics of the three lymphocyte subtypes. Quantitative morphological (cellular surface area, cellular volume, and sphericity) and biochemical (cellular dry density and dry mass) parameters were calculated from the 3-D RI distribution of the lymphocytes (See Methods). Figures 2d-f show the quantitative characteristics of the B, CD4<sup>+</sup> T, and CD8<sup>+</sup> lymphocytes, respectively. The cellular surface area and volume of lymphocytes were simply calculated from the voxel information of the 3-D RI tomograms. And then, the sphericity, a dimensionless parameter that indicates the roundness of the cellular morphology, was obtained by the ratio of the calculated surface area and volume. The biochemical information was retrieved from the RI distribution of the lymphocytes because the RI values are linearly proportional to the local concentration of non-aqueous molecules (mostly protein).

Figure 3 shows scatter plots of the quantitative morphological and biochemical information of each lymphocyte type and the results of the statistical analyses. The cellular surface areas of the B cell ( $n = 149$ ), CD4+ ( $n = 95$ ), and CD8 T lymphocytes ( $n = 111$ ) 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 cell subtypes ( $p$ -value  $< 0.001$ ), while there was no statistical difference between the CD4+ and CD8+ T cells. In addition, the result of quantifying the cellular volume of the lymphocytes also shows a similar tendency with the result of the cellular surface area analysis. The cellular volume of the B lymphocytes ( $133.43 \pm 26.47 \text{ fL}$ ) was significantly small ( $p$ -value  $< 0.001$ ) compared to that of the CD4+ ( $155.73 \pm 35.14 \text{ fL}$ ) and CD8+ ( $152.77 \pm 26.52 \text{ fL}$ ) lymphocytes (Fig. 3b). However, the CD4+ and CD8+ T cells had a similar cellular volume. The sphericities were  $0.86 \pm 0.06$ ,  $0.84 \pm 0.06$ , and  $0.86 \pm 0.05$  for the B cells, CD4+ and CD8+ T cells, respectively (Fig. 3c). Although the sphericity of the CD4+ T cells was statistically smaller than that of the B cells ( $p$ -value  $< 0.01$ ) and CD8+ T cells ( $p$ -value  $< 0.05$ ), all the lymphocyte subtypes had a high sphericity which means the lymphocytes have round shapes.

Next, we compared the biochemical properties of the lymphocytes. The cellular dry mass density of the B cells, CD4+ and CD8+ T cells was  $15.43 \pm 1.88$ ,  $14.81 \pm 2.54$ , and  $16.66 \pm 1.88 \text{ g/dL}$ , respectively (Fig. 3d). The CD8+ T cells had a significantly increased cellular dry mass density compared to that of the others ( $p$ -value  $< 0.001$ ). The total cellular dry mass, calculated by integrating the dry mass density over the cell volume, was  $20.28 \pm 2.97$ ,  $22.65 \pm 4.49$ , and  $25.19 \pm 3.51 \text{ pg}$  for the B cells, CD4+ and CD8+ T cells, respectively (Fig. 3e). There were significant differences in the total cellular dry mass among the cell types ( $p$ -values  $< 0.001$ ). The B cells had a smaller cellular mass compared to the T cell subsets. Moreover, the CD8+ T cells were statistically heavier than that of the CD4+ T cells. These results reveal that the biochemical characteristics of each lymphocyte type have statistical differences. However, even though the quantitative morphological and biochemical analyses show statistical differences among the lymphocyte population, a single lymphocyte cannot be identified by its cell type using a single retrieved quantitative characteristic because of cell-to-cell variations.

To identify the lymphocyte types at an individual cell level, machine learning methods were exploited to construct classification models using multiple quantitative characteristics of the lymphocytes. To demonstrate a proof of concept, a training-and-identification method, called statistical classification or supervised machine learning, was used. Seventy percent of the total lymphocytes were randomly chosen and systematically analysed to extract the unique characteristics for each lymphocyte type (Fig. 1b), and then, the cell types of the remaining lymphocytes were identified with the extracted features (Fig. 1c). For the statistical model, a  $k$ -NN algorithm was used<sup>46,47</sup>, which has been widely used for classification purposes in various research fields. The  $k$ -NN algorithm consists of a supervised machine learning which establishes a classifier based on quantitative characteristics of the training lymphocytes from known cell types. In this study, a four nearest neighbour ( $k$ -NN,  $k = 4$ ) algorithm was used to establish the classification models.

To exploit the structural and biochemical information of the intracellular organelles of the lymphocyte as features of the  $k$ -NN ( $k = 4$ ) algorithm, the quantitative characteristics of the lymphocytes were calculated at various threshold RI values by increasing the threshold from 1.34 to 1.378 with an increment of 0.002. We found that the 3-D RI distribution, which had higher RI values than the threshold RI, tends to reveal information about the intracellular components as the threshold RI values increase (Supplementary Fig. 1). All combinations of structural and biochemical information obtained at a single RI threshold or two different RI thresholds are tested to establish an optimized classifier with cross-validation, and then, a

classifier, which shows the best performance, is selected. Then, the remaining lymphocytes are identified using the established classifier, and the identification accuracy (sensitivity and specificity) is measured.

Figure 4 and Table 1 show the overall cross-validation accuracy, sensitivity (true positive results overall positive inputs), and specificity (true negative results overall negative inputs) of the training and test results. We performed statistical classification on three different combinations of lymphocyte types: (i) B and T lymphocytes, (ii) T lymphocyte subsets (CD4+ and CD8+), and (iii) all three types of lymphocytes. First, the T cell subsets were considered as one T cell type, and the classification model is optimized to have the best overall cross-validation accuracy. Features used for establishing the classifier were the surface area, sphericity, and dry mass (RI threshold: 1.342) and all the quantitative information (RI threshold: 1.368). The overall accuracy of the optimized classifier for identifying the B and T cells was 93.15%, and the test accuracy was 89.81% (Fig. 4a). Second, the CD4+ and CD8+ T cells were analysed. To optimize the statistical classification model, the surface area and sphericity (RI threshold: 1.342 and 1.362) of the CD4+ and CD8+ T cells, respectively, were used as features. The classification accuracy was 87.41% and 84.38% for the training and test, respectively (Fig. 4b). Lastly, the statistical classification was performed to identify the three types of lymphocytes. The classification model was optimized with the surface area, dry mass density (RI threshold: 1.340), and sphericity, dry mass density, dry mass (RI threshold: 1.370) as features. The accuracy of the training and test were 80.65% and 75.93%, respectively. These results indicate that machine learning enables the identification of lymphocyte cell types with an accuracy of over 75%.

## Discussion

We demonstrated label-free identification of lymphocyte cell types at a single cell level using ODT and statistical classification. ODT provides quantitative morphological and biochemical information on the lymphocytes by measuring their 3-D RI distribution. We found that there were significant differences in the quantitative characteristics among the lymphocyte cell population; however, individual lymphocytes in different lymphocyte cell types are indistinguishable due to cell-to-cell variations. To overcome this limitation, the  $k$ -NN ( $k = 4$ ) algorithm was exploited as a statistical classification method to establish classification models using the quantitative characteristics of the lymphocytes. The optimized classification models can discriminate B and T cells with high accuracy. In addition, the T cell subsets were identified using the  $k$ -NN ( $k = 4$ ) algorithm with an overall accuracy of over 80%. Moreover, the three types of lymphocytes were identified using the classification model with an overall accuracy of over 75%.

The identification results show that the classification models more precisely discriminate between B and T lymphocytes rather than the T cell subsets, which implies that differences in cellular morphology and biochemical properties between B and T cells are more distinct than those between the CD4+ and CD8+ T cells. The results are consistent with previous knowledge of the lymphocyte-differentiation pathway<sup>48</sup>. B and T lymphocytes originate from hematopoietic stem cells and then mature in different organs. Thus, lymphocytes have identical cellular phenotypes such as one large nucleus and spherical shapes; however, B and T lymphocytes have entirely different cellular functions. Even though our method established the classifiers by optimizing features without biological relevance, the machine learning algorithm found distinct differences in the morphological and biochemical properties among the lymphocyte subtypes.

The present method combines 3-D RI tomography and machine learning, which provides several advantages. First, the present method enables label-free identification of lymphocyte subtypes which cannot be achieved by optical microscopic techniques without using fluorescent methods because of similar phenotypes in lymphocyte subtypes. ODT measures the 3-D RI distribution of the lymphocytes, and machine learning methods find significant differences among the lymphocyte subsets to identify their cell types. Second, the present method has a simple and cost-effective optical setup compared to flow cytometry 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 reduces the required time for measuring a 3-D RI tomogram by exploiting a digital micromirror device<sup>49</sup>. Thus, the present method 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 WBCs, 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 classify various cell types.

There are several points to be improved in future works. The present study proves a proof of concept; however, the overall accuracies for identifying each lymphocyte type should be enhanced. Several classification algorithms including two  $k$ -NN algorithms ( $k = 4$  and  $k = 6$ ), linear discrimination, quadratic discrimination, naïve Bayes, and decision tree were tested, and the  $k$ -NN ( $k = 4$ ) algorithm showed the best performance for identifying the lymphocyte types (Supplement Fig. 2). However, these classifiers are at the basic level of machine learning methods. Recently, deep learning methods have been introduced and widely used in various research fields including image recognition<sup>50</sup>, speech recognition<sup>51</sup>, and biomedical research<sup>52-54</sup>. Thus, the identification accuracy could be more improved by exploiting ODT and deep learning methods. In addition, the speed of measuring the 3-D RI tomograms could be improved. For single lymphocyte imaging, we found and measured lymphocytes sparsely placed on a cover glass which limits the speed of the tomogram measurements. We expect

that microfluidic approaches might be a solution to increase the speed of the 3-D RI tomograms of lymphocytes and become a practical method for discriminating cell types.

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 samples, and machine learning enables the classification of cell types using the measured quantitative information. The present method can be further used to study immunology, cancer, and neuroscience.

## Methods

### Mice

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

### Flow cytometry and 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) was used for dead cell exclusion. Sorting was performed on an Aria II or III system (BD Biosciences) using an 85- $\mu\text{m}$  nozzle or Astrios system (Beckman Coulter) using a 70- $\mu\text{m}$  nozzle. Antibodies for flow cytometry were purchased from BD Biosciences, eBioscience, Biolegend. 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).

### Measurement of 3-D refractive index tomograms

To reconstruct the 3-D RI tomograms of lymphocytes, a Mach-Zehnder interferometric microscope was used<sup>55</sup> (Fig. 2). A laser beam from a diode-pumped solid-state laser ( $\lambda = 532$  nm, 100 mW, Shanghai Dream Laser Co., Shanghai, 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, Newton, NJ, USA), and the other is used as a reference beam. The sample is placed between a condenser lens (UPLSAPO Water 60 $\times$ , the numerical aperture (NA) = 1.2, Olympus, Japan) and an objective lens (PLAPON Oil 100 $\times$ , NA = 1.4, Olympus, Japan). 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., San Diego, CA, USA). For reconstructing a 3-D RI tomogram, a total of 300 holograms of a sample are measured by changing the angle of illuminations which takes less than 1 sec. Then, the optical field information (amplitude and phase) of the measured holograms are retrieved using a field retrieval algorithm based on Fourier transformation<sup>56,57</sup>. From the retrieved multiple amplitude and phase information, a 3-D RI tomogram is reconstructed using an optical diffraction tomography algorithm. 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>58</sup>. Details on reconstructing 3-D RI tomograms can be found elsewhere<sup>15,59</sup>.

### Quantitative characterization of the structural and biochemical information of the lymphocytes

Quantitative structural and biochemical information of the lymphocytes were calculated from the measured 3-D RI tomograms. To calculate the cellular volume  $V$  and surface area  $S$  of a lymphocyte, the voxels of a 3-D RI tomogram of a lymphocyte, which had higher RI values compared to the threshold RI value, were selected. From the selected voxels, the cellular volume and surface area were calculated corresponding to the numbers of voxels and cellular boundaries, respectively. Sphericity, which is a dimensionless parameter and indicates the roundness of a lymphocyte, was obtained from the measured cellular volume and surface area as follows:  $Sphericity = \pi^{1/3} \cdot (6V)^{2/3} / S$ . The biochemical information

(dry mass density and cellular dry mass) were obtained from the RI values due to the linear relation between an RI value and the local concentration of non-aqueous molecules (i.e., proteins, lipids, and nucleic acids inside cells). RI values were converted to the concentration of non-aqueous molecules (mostly proteins) with the following relationship:  $n = n_0 + \alpha C$ , where  $n$  is an RI value of a voxel;  $n_0$  is an RI value of the background, and  $\alpha$  is the refractive index increment (RII) of proteins. Because 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 dry mass density over the cellular volume. Details on calculating the quantitative information of a sample can be found elsewhere<sup>23,25</sup>.

### **Image processing and statistical analysis**

Image processing was performed with Matlab\_R2014b and ImageJ. Statistical analysis was done by the GraphPad Prism software. The RI isosurfaces were rendered by commercial software (TomoStudio, Tomocube Inc., Korea).

### **ACKNOWLEDGEMENTS**

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

### **AUTHOR CONTRIBUTIONS**

Y.P. conceived the idea and directed the work. Y.P. and S.K. analysed the data. K.K. designed optics system. M.K. isolated lymphocytes from mice peripheral blood. J.Y. performed the experiments. J.Y. and Y.J. established statistical classification model. All authors wrote the manuscript.

### **COMPETING FINANCIAL INTERESTS**

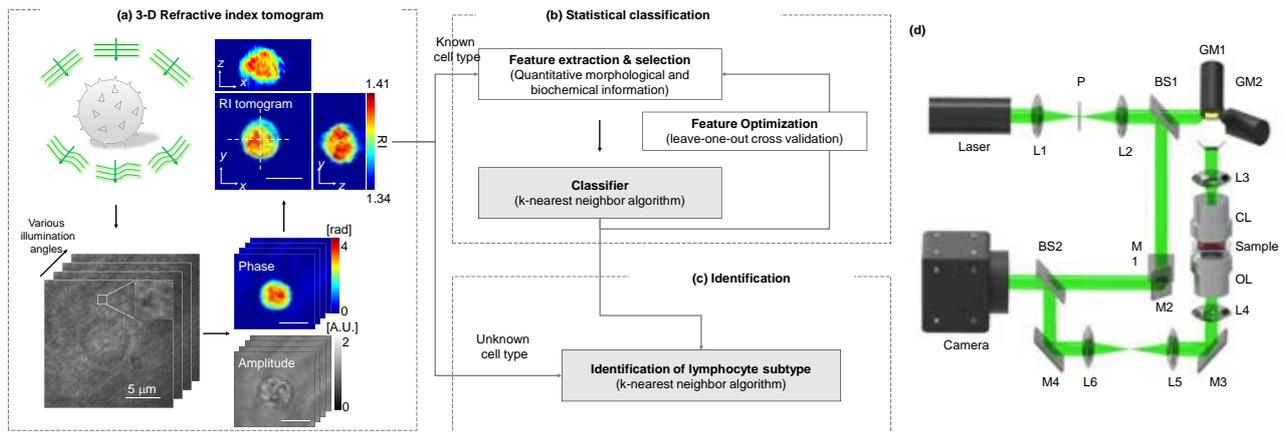
Prof. Park has financial interests in Tomocube Inc., a company that commercializes optical diffraction tomography and quantitative phase imaging instruments and is one of the sponsors of the work.

## Reference

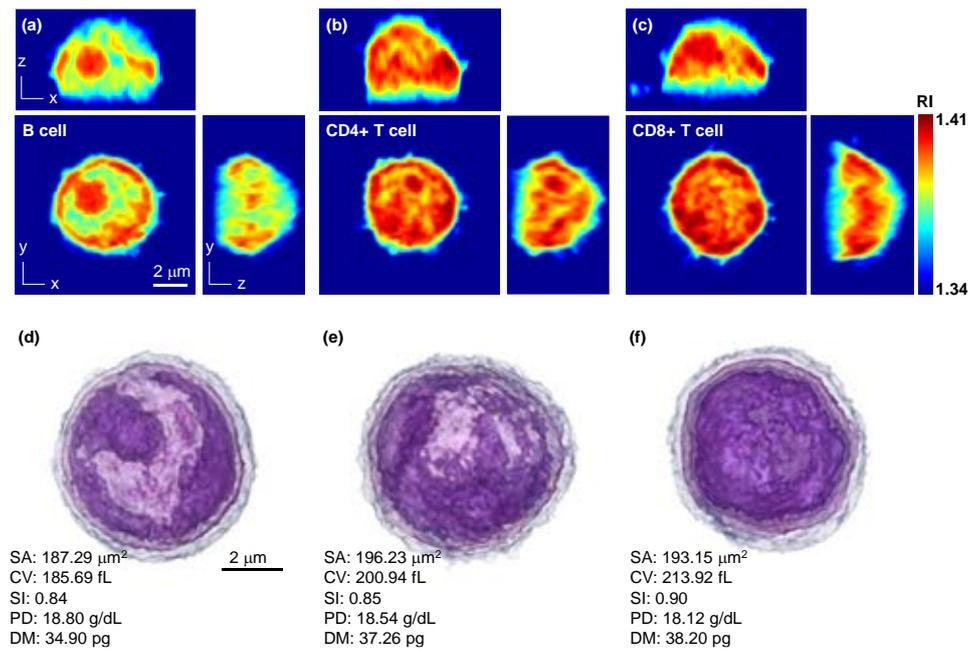
- 1 Radtke, F., Wilson, A., Mancini, S. J. C. & MacDonald, H. R. Notch regulation of lymphocyte development and function. *Nat Immunol* **5**, 247-253 (2004).
- 2 Moser, B. & Loetscher, P. Lymphocyte traffic control by chemokines. *Nat Immunol* **2**, 123-128 (2001).
- 3 Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511, doi:10.1038/35000501 (2000).
- 4 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).
- 5 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).
- 6 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).
- 7 von Boehmer, H. & Melchers, F. Checkpoints in lymphocyte development and autoimmune disease. *Nat Immunol* **11**, 14-20 (2010).
- 8 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).
- 9 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).
- 10 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).
- 11 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).
- 12 Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology* **27**, 378-386 (2009).
- 13 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).
- 14 Kim, K. *et al.* Optical diffraction tomography techniques for the study of cell pathophysiology. *Journal of Biomedical Photonics & Engineering* **2**, 020201 (2016).
- 15 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. *Journal of Biomedical Optics* **19**, doi:10.1117/1.Jbo.19.1.011005 (2014).
- 16 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). *Optics Express* **23**, 18996-18996, doi:10.1364/Oe.23.018996 (2015).
- 17 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).
- 18 Park, H. *et al.* Characterizations of individual mouse red blood cells parasitized by Babesia microti using 3-D holographic microscopy. *Scientific Reports* **5**, doi:10.1038/Srep10827 (2015).
- 19 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).
- 20 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).
- 21 Lee, S., Park, H., Jang, S. & Yongkeun, P. (Am Soc Hematology, 2016).
- 22 Park, H. *et al.* Measuring cell surface area and deformability of individual human red blood cells over blood storage using quantitative phase imaging. *Scientific Reports* **6** (2016).
- 23 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. *Scientific reports* **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. *Optics Express* **17**, 266-277, doi: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. *Journal of Biomedical Optics* **19**, doi: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**, e49502 (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  
(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. *Optics Letters* **39**, 2210-2213, doi:10.1364/Ol.39.002210 (2014).
- 32 Cotte, Y. *et al.* Marker-free phase nanoscopy. *Nature Photonics* **7**, 113-117 (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. *bioRxiv*, doi:10.1101/080937 (2016).
- 34 Bennet, M., Gur, D., Yoon, J., Park, Y. & Faivre, D. A Bacteria-Based Remotely Tunable Photonic Device.  
*Advanced Optical Materials* (2016).
- 35 Kim, T. *et al.* White-light diffraction tomography of unlabelled live cells. *Nature Photonics* **8**, 256-263 (2014).
- 36 Lee, S. *et al.* High-Resolution 3-D Refractive Index Tomography and 2-D Synthetic Aperture Imaging of Live  
Phytoplankton. *Journal of the Optical Society of Korea* **18**, 691-697 (2014).
- 37 Lee, S. *et al.* Measurements of morphology and refractive indexes on human downy hairs using three-  
dimensional quantitative phase imaging. *Journal of biomedical optics* **20**, 111207-111207 (2015).
- 38 Kotsiantis, S. B., Zaharakis, I. & Pintelas, P. (2007).
- 39 Jo, Y. *et al.* Label-free identification of individual bacteria using Fourier transform light scattering. *Optics  
Express* **23**, 15792-15805, doi:10.1364/Oe.23.015792 (2015).
- 40 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).
- 41 Chen, C. L. *et al.* Deep Learning in Label-free Cell Classification. *Scientific Reports* **6** (2016).
- 42 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 (2012).
- 43 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 (2014).
- 44 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 (2012).
- 45 Wolf, E. Three-dimensional structure determination of semi-transparent objects from holographic data. *Optics  
Communications* **1**, 153-156 (1969).
- 46 Miska, E. A. MicroRNA expression profiles classify human cancers. *Cytom Part B-Clin Cy* **72b**, 126-126 (2007).
- 47 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 (2006).
- 48 Ogawa, M. Differentiation and Proliferation of Hematopoietic Stem-Cells. *Blood* **81**, 2844-2853 (1993).
- 49 Shin, S. *et al.* in *SPIE BiOS*. 971814-971814-971818 (International Society for Optics and Photonics).
- 50 Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning Hierarchical Features for Scene Labeling. *Ieee T  
Pattern Anal* **35**, 1915-1929, doi:10.1109/Tpami.2012.231 (2013).
- 51 Hinton, G. *et al.* Deep Neural Networks for Acoustic Modeling in Speech Recognition. *Ieee Signal Proc Mag*  
**29**, 82-97, doi:10.1109/Msp.2012.2205597 (2012).
- 52 Ma, J. S., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep Neural Nets as a Method for Quantitative  
Structure-Activity Relationships. *J Chem Inf Model* **55**, 263-274, doi:10.1021/ci5000747n (2015).
- 53 Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*  
**500**, 168-+, doi:10.1038/nature12346 (2013).
- 54 Leung, M. K. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. Deep learning of the tissue-regulated splicing code.  
*Bioinformatics* **30**, 121-129, doi:10.1093/bioinformatics/btu277 (2014).
- 55 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**, 011005-  
011012 (2014).
- 56 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).
- 57 Debnath, S. K. & Park, Y. Real-time quantitative phase imaging with a spatial phase-shifting algorithm. *Optics  
Letters* **36**, 4677-4679 (2011).
- 58 Lim, J. *et al.* Comparative study of iterative reconstruction algorithms for missing cone problems in optical  
diffraction tomography. *Optics Express* **23**, 16933-16948, doi:10.1364/Oe.23.016933 (2015).
- 59 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. *Scientific Reports* **6** (2016).

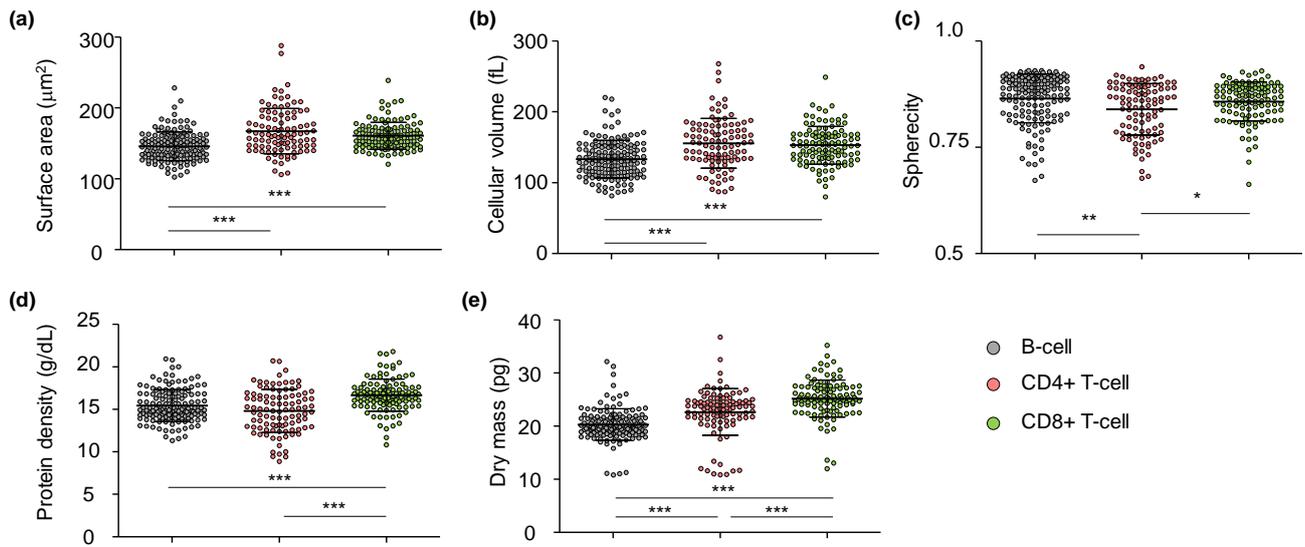
## Figure legends



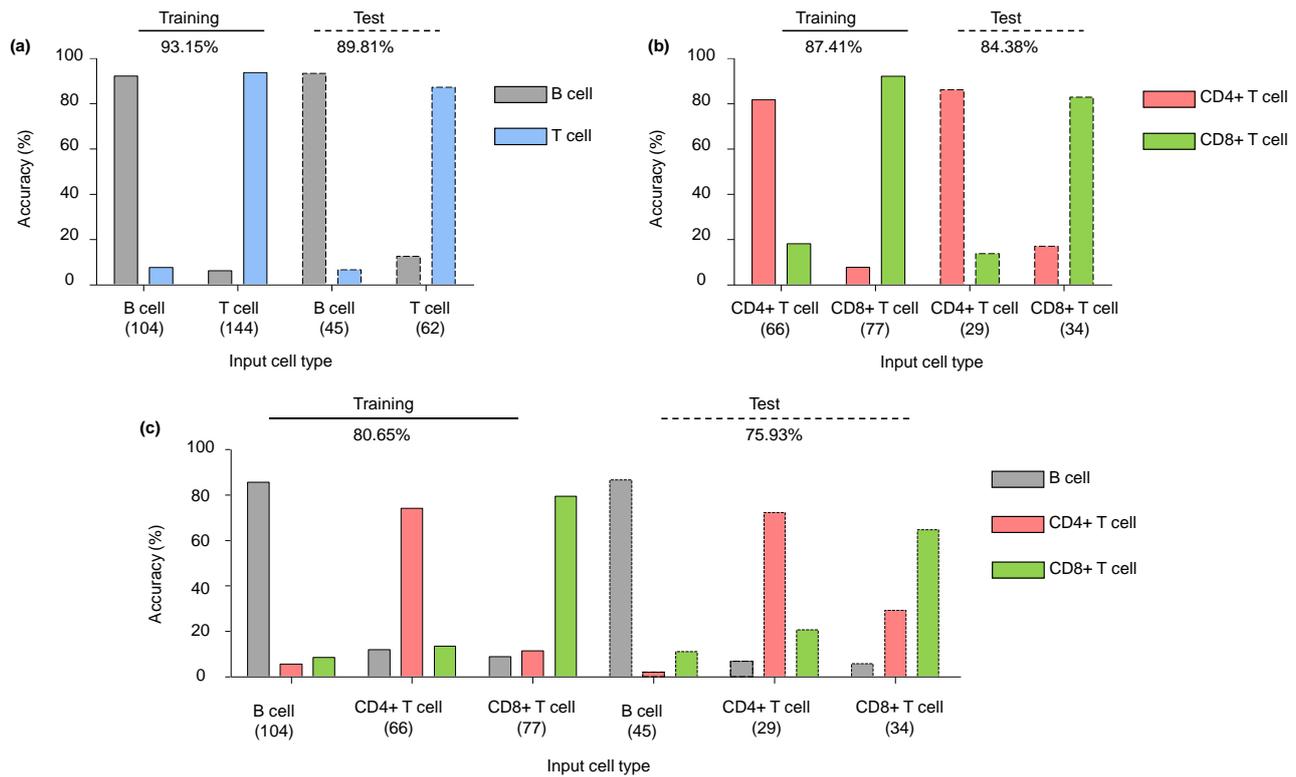
**Figure 1 | Schematic diagrams of the label-free identification of lymphocyte cell types using optical diffraction tomography and machine learning algorithm.** (a) Procedures for measuring label-free 3-D tomograms of lymphocytes. Multiple holograms of lymphocytes are measured by changing the angle of illumination. Optical field information of the lymphocytes was retrieved from the measured holograms, and then, 3-D refractive index tomograms were reconstructed using the retrieved optical field information. Scale bar is 5 μm. (b) Establishing a statistical classifier for identifying lymphocyte cell types using *k*-nearest neighbour algorithm. Quantitative morphological and biochemical features of lymphocytes known their cell types (supervise learning) were used to generate the classifier. (c) Determining the types of unidentified cells 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 lymphocytes. BS1–2, beam splitters; L1–6, lens; SF, spatial filtering; CL, condenser lens; OL, objective lens; M1–4, mirrors.



**Figure 2 | Representative 3-D RI tomograms of each lymphocyte 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 is 2 μm. (d-f) 3-D rendered RI tomograms and quantitative characterization of the morphological and biochemical features of (a-c), respectively. Scale bar is 2 μm. SA, surface area; CV, cellular volume; SI, sphericity; PD, protein density; DM, dry mass.



**Figure 3 | Quantitative analysis of the morphological and biochemical features of the B cell, CD4+ T cell, and CD8+ T cell.** (a) Surface area, (b) cell volume, (c) sphericity, (d) protein density, and (e) dry mass. Each symbol indicates an individual cell measurement. Horizontal black lines indicate the mean values; vertical lines, indicating standard deviation. Statistical analysis was done with Student's *t*-test.



**Figure 4 | Identification of lymphocyte cell types using a  $k$ -nearest neighbour algorithm.** Identification results of (a) B and T cell types, (b) T cell subtypes (CD4+, CD8+), and (c) all three lymphocyte types. Numbers below each cell type indicate the number of cells used for supervised learning and identification.

## Training

## Test

		Output cell type (number of cell)		Sensitivity (%)
		B cell	T cell	
Input cell type	B cell	96	8	92.31
	T cell	9	135	93.75
Specificity (%)		92.31	92.31	93.15 (overall)

		Output cell type (number of cell)		Sensitivity (%)
		CD4 T cell	CD8 T cell	
Input cell type	CD4 T cell	54	12	81.82
	CD8 T cell	6	71	92.21
Specificity (%)		92.21	81.82	87.41(overall)

		Output cell type (number of cell)			Sensitivity (%)
		B cell	CD4 T cell	CD8 T cell	
Input cell type	B cell	89	6	9	85.58
	CD4 T cell	8	49	9	74.24
	CD8 T cell	7	9	62	79.49
Specificity (%)		89.58	91.76	89.41	80.65 (overall)

		Output cell type (number of cell)			Sensitivity (%)
		B cell	CD4 T cell	CD8 T cell	
Input cell type	B cell	39	1	5	86.67
	CD4 T cell	2	21	6	72.41
	CD8 T cell	2	10	22	64.71
Specificity (%)		93.65	86.08	89.41	75.93 (overall)

**Table 1** Results of the identification of lymphocyte cell types using a  $k$ -nearest neighbour ( $k = 4$ ) algorithm.