

High-throughput label-free cell classification using machine learning

Claire L. Chen, Ata Mahjoubfar, Li-Chia Tai, Ian K. Blaby, Allen Huang, Kayvan R. Niazi, and Bahram Jalali
(Dated: November 4, 2015)

SUPPLEMENTARY METHODS

I. PCA

¶ PCA To speed up the computation for high dimensionality analysis, we employ PCA. We have performed PCA on the 16 dimensional data set acquired by our time stretch QPI system. This maps the data from the physical feature space to the PCA space. Fig. S1 shows the percent of the variance in data explained by each component (lower chart). The key observation is that most of the variance can be accounted for by the first two principle components. The upper portion of the plot shows the AUC for binary classification using each of the principle components. Interestingly, the first component with the highest explained variance is not necessarily the most important component for classification. This is not surprising since PCA components are optimized to the variance of data but not to the data labels. Therefore, apriori intuition about the physical significance of the features in the case here, is superior to PCA in eliminating dimensions that do not provide high value in classification. Nevertheless, in situations where physical intuition about meaning of features does not exist, PCA is a valuable tool. A second observation is that principle components that have low explained variance do have some value in classification. By revealing the structure in data that best explains the variance, PCA achieves data compression via dimensionality reduction.

¶ As shown in Fig. ??a and Fig. ??b, many of the 16 parameters are correlated and not all measured features in the data set produced by the time stretch quantitative phase imaging have the same amount of information. That result suggests that it may be possible to reduce the 16 dimensional data set to a smaller set of uncorrelated orthogonal dimensions without significantly compromising the classification accuracy. In that spirit, we have used principle component analysis (PCA) for dimensionality reduction. The PCA algorithm finds an alternative lower dimension space such that variance of data projected onto this subspace is maximized along subspace dimensions. By revealing the structure in data that best explains the variance, PCA achieves data compression via dimensionality reduction.

A. Regularization

¶ Regularization ¶ Cross validation

B. System performance and resolvable points

Lateral resolution of time stretch camera is decided by the limiting factor among Abbe diffraction limit of the objective lens, spectral resolvability of the diffraction grating pairs, spectral resolution in amplified dispersive Fourier transform, the photodetector rise-time and bandwidth, and the sampling rate of the back-end digitizer. Details of the limiting factors of lateral resolution and evaluation of these factors for our TS-QPI system can be found in Table S1. Field of view (FOV) is the area covered by the interrogation rainbow when the rainbow pulses hit the imaging plane. The rainbow pulses width is decided by the optical bandwidth selected from the laser source, $\Delta\lambda$, the magnification factor of the objective lens as well as the focal length of the other lens and parabolic mirrors, and the dimensions and blaze angles of the diffraction gratings.

The resolution of phase measurement along axial direction is determined by the effective number of bits (ENOB) of the digitizer and affected by the noise of laser source. Since pulse to pulse intensity and phase fluctuations are small, noise from laser source is not the limiting factor in our phase measurements. Supposing the ENOB of the digitizer is N , the minimum detectable optical path length difference, ΔL can be estimated as

$$\frac{1}{2} \sin\left(\frac{4\pi\Delta L}{\lambda + \Delta\lambda/2}\right) = 2^{-N} \quad (6)$$

where λ is the central wavelength of light, and $\Delta\lambda$ is the optical bandwidth. In our system, ENOB of the analog-to-digital converter is 5. Thus, OPD resolution along the axial direction is about 8.0 nm, corresponding to refractive index difference down to the order of 0.001 for cellular level measurements.

C. Microfluidic channel design and fabrication

The Polydimethylsiloxane (PDMS) microfluidic channel is custom-designed so that it could fit into the reflective optics design. Cells are hydrodynamically focused [1, 2] at the center of the channel flowing at a velocity of 1.3 m/s. The microfluidic device consists of a hydrodynamic focusing region and an imaging region targeted by the interrogation rainbow flashes in TS-QPI system. At the hydrodynamic focusing region, the sheath pressure focused the sample at the center of the channel by narrowing its flow width from 200 μm to about 40 μm with a sheath to sample volume ratio of 3:1. The dimension of the channel was chosen as 200 μm (width) \times 25 μm

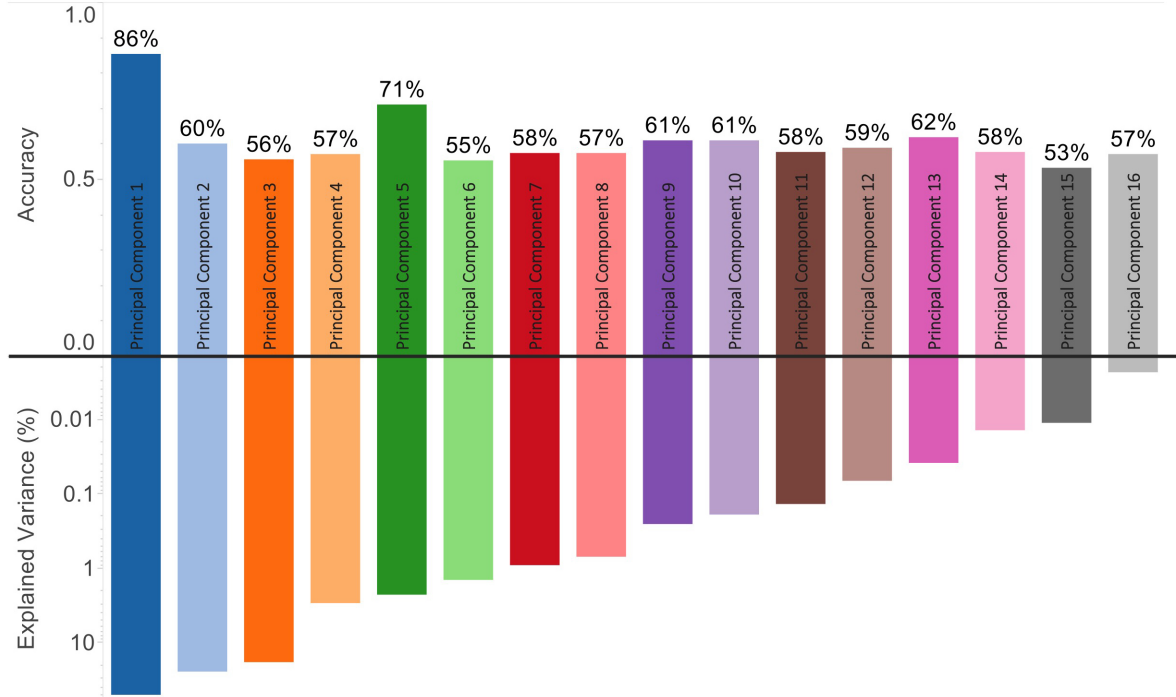


FIG. S1. Principal component analysis (PCA); Upper bar chart shows accuracy of classification by each individual principal component, and lower bar chart shows the percentage of the total variance explained by each principal component, accounting for the variability expressed in the data. As expected, principal components with larger variability do not necessarily give high accuracy in classification.

TABLE S1. Resolution Limiting Factors in TS-QPI

| System category | Component | Number of resolvable points | Lateral resolution |
|---------------------|-------------------------|--|--------------------|
| Free-space optics | Diffraction gratings | $N_{grating} = \frac{\Delta\lambda}{\delta\lambda_{grating}} = \Delta\lambda / (\lambda \cdot \frac{d}{m \cdot 2w_0}) \quad (1)$ <p>where $\Delta\lambda$ is the optical bandwidth, λ is the central wavelength, m is the order of diffraction, w_0 is the beam waist, and d is the groove spacing.</p> | 3.09 μm |
| | Lenses and mirrors | $N_{Abbe} = \frac{FOV}{\delta x_{diffraction}} = \frac{FOV}{(\frac{\lambda + \Delta\lambda/2}{2 \cdot NA})} \quad (2)$ <p>where FOV is field of view, NA is numerical aperture of the objective lens.</p> | 2.00 μm |
| Time Stretch | Group delay dispersion | $N_{DFT} = \frac{\Delta\lambda}{\delta\lambda} = \frac{\Delta\lambda}{\lambda \cdot \sqrt{\frac{2}{DL_f \cdot c}}} \quad (3)$ <p>where D is the group velocity dispersion, L_f is the dispersive fiber length.</p> | 0.73 μm |
| Electronic back-end | Photodetector bandwidth | $N_{PD} = \frac{\Delta t}{\delta t} = \frac{DL_f \Delta\lambda}{0.35/B} \quad (4)$ <p>where B is the bandwidth of the photodetector.</p> | 0.28 μm |
| | ADC sampling rate | $N_{ADC} = DL_f \Delta\lambda f_{ADC} \quad (5)$ <p>where f_{ADC} is the sampling rate of digitizer.</p> | 0.10 μm |

(height) so that the cells will be imaged within depth of view with a narrow lateral distribution. The size of the entire PDMS channel is optimized for fitting on a 2 inch diameter dielectric mirror with sufficient space at the edges to achieve strong bonding. The thickness of the channel top layer is optimized for stabilizing peek tubes performance reliability while accommodate the working distance of the objective lens.

The PDMS microfluidic channel (Fig. S2) is fabricated using standard soft lithography. The mask was designed in AutoCAD and printed with a resolution down to 1 μm . Then a 4-inch silicon wafer was spin-coated with 75 μm thickness of a negative photoresist (SU-8 from MicroChem) and was exposed under the mask using an aligner. After post-exposure baking, the wafer was developed at room temperature, rinsed with isopropyl alcohol (IPA), and placed in a petri dish. A PDMS mixture (Sylgard 184 Silicone Elastomer, Dow Corning) was poured onto the patterned wafer, degassed in a vacuum chamber for 30 min and cured at 80 °C for one hour. Once cured, the PDMS channel was cut out and peeled off from the master wafer. We used 1.25 μm diameter hollow needle to punch the inlet and outlet holes. The punched PDMS channel was then cleaned with nitrogen gun and magic tape (3M), treated with oxygen plasma (Enercon Dyne-A-Mite 3D Treater) for 2 min, and bonded to a 2-inch broadband dielectric mirror (Thorlabs BB2-E04) for obtaining high reflectance from channel substrate at near infrared spectral window. Finally microtubes (PE-50 tubing, .023x.038in) with steel catheter couplers (In-

stech, 22 ga x 15 mm) are connected to the inlet and outlet punctures.

D. Preparation of algae cell lines

Chlamydomonas reinhardtii strains used were *cw15* (*nit1 NIT2 mt⁺⁻*) and *sta6* (*cw15 nit1 NIT2 arg7-7 sta6-1::ARG7 mt⁺*), available as CC-4568, CC-4348 respectively from the Chlamydomonas resource center (CRC)[3].

Cells were grown in tris-acetate-phosphate (TAP) medium supplemented with arginine (100 $\mu\text{g mL}^{-1}$). Cultures were grown in Innova incubators (New Brunswick Scientific, Edison, NJ) at 24 °C, agitated at 180 rpm with continuous light (95 $\mu\text{mol m}^{-2} \text{s}^{-1}$, 6 cool white fluorescent bulbs at 4100 K and 3 warm white fluorescent bulbs at 3000 K per incubator). To induce lipid production, cells were cultured to mid-log phase in regular TAP prior to deprivation of N by transfer to ammonium-free (i.e. nitrogen-free) TAP medium, as described previously [4]. Briefly, cells subjected to nitrogen deprivation were grown to 4×10^6 cells mL^{-1} and collected by centrifugation at 1006 $\times g$ for 5 min at room temperature. The supernatant was discarded, and the cells were washed in nitrogen-free TAP. Cells were then resuspended in nitrogen-free TAP to a final cell count of 2×10^6 cells mL^{-1} . Cell densities were determined using a hemocytometer.

-
- [1] James B Knight, Ashvin Vishwanath, James P Brody, and Robert H Austin, "Hydrodynamic focusing on a silicon chip: mixing nanoliters in microseconds," *Physical Review Letters* **80**, 3863 (1998).
 - [2] Gwo-Bin Lee, Chih-Chang Chang, Sung-Bin Huang, and Ruey-Jen Yang, "The hydrodynamic focusing effect inside rectangular microchannels," *Journal of Micromechanics and Microengineering* **16**, 1024 (2006).
 - [3] Matt Laudon, "Chlamydomonas resource center, university of minnesota," <http://chlamycollection.org/>

(Data of access:20/07/2015), online.

- [4] Ian K Blaby, Anne G Glaesener, Tabea Mettler, Sorel T Fitz-Gibbon, Sean D Gallaher, Bensheng Liu, Nanette R Boyle, Janette Kropat, Mark Stitt, Shannon Johnson, *et al.*, "Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a *chlamydomonas reinhardtii* starchless mutant," *The Plant Cell Online* **25**, 4305–4323 (2013).

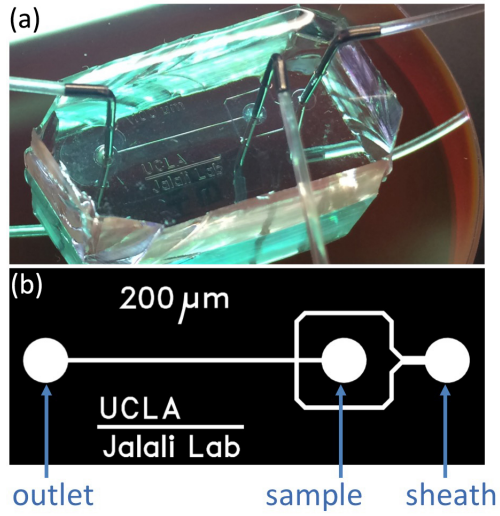


FIG. S2. PDMS microfluidic channel mounted on a highly reflective surface with near-infrared dielectric coating; The microfluidic device consists of a hydrodynamic focusing region and an imaging region targeted by the interrogation rainbow flashes in TS-QPI system. (a) Sample solution with suspended cells is fed into the channel through the sample inlet, and deionized water as the sheath flow is sent through sheath inlet. At the hydrodynamic focusing region, the sheath pressure focused the sample at the center of the channel by narrowing its flow width from $200\text{ }\mu\text{m}$ to about $40\text{ }\mu\text{m}$ with a sheath to sample volume ratio of 3:1. (b) The pattern of the mask used to imprint microfluidic channel design on silicon wafer with photoresist. The circles are inlet and outlet reservoirs.