Supplementary Information

Intelligent Optofluidic Analysis for Ultrafast Single Bacterium Profiling of Cellulose Production and Morphology

Jiaqing Yu[a], Guoyun Sun[a], Nicholas Weikang Lin[a], Sundaravadanam Vishnu Vadanan[b], Sierin Lim[b] and Chia-Hung Chen[c*]

a. Department of Biomedical Engineering, National University of Singapore, Singapore 117575
b. School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore 637457
c. Department of Biomedical Engineering, City University of Hong Kong 83 Tat Chee Avenue, Kowloon Tong, Hong Kong

Content:

Supporting information 1: Comparison of bacterial cellulose assays

	Direct detection without labeling	Long-term culture	High throughput
Flow cytometry	×	×	✓✓
Microwell	✓	✓	✓
Microfluidics	✓	✓	✓✓

Table S1: Comparison of assays to measure bacterial cellulose production. Flow cytometry is a high-throughput tool to screen bacteria in a continuous-flow setting. However, it is not easy to incubate bacteria for days to generate sufficient bacterial cellulose (BC) with internal structures for screening. Microwell technology is used to culture bacteria to produce sufficient BC for analysis but has limitations in high-throughput measurement. Here, the microfluidic approach was demonstrated to incubate single bacteria in agarose particles for days to produce BC for high-throughput measurement.

Supporting information 2: Bacteria encapsulation and proliferation

To encapsulate single bacteria in droplets, a pinched microchannel was developed to regulate bacterial distribution along a microchannel [R. Ramji et al, Biomicrofluidics, 2014, 8, 034104]. The average droplet diameter generated was approximately 30 μm, and the bacterial concentration was adjusted to 5×10^7 bacteria/mL. If the encapsulation process followed the Poisson distribution, λ would be approximately 0.5, and the encapsulation rate would follow the dotted line, as shown in **Fig. S1a**. The experimental encapsulation rate of the improved microfluidic structure is shown with the bars. The single-bacterium encapsulation rate was significantly enhanced, while the number of empty droplets or droplets containing multiple cells decreased. **Fig. S1b** shows the proliferation of encapsulated bacteria in the agarose hydrogel particle during cell incubation. The number of bacteria in one hydrogel particle continuously increased from 1 to approximately 20 in 3 days (72 hours).

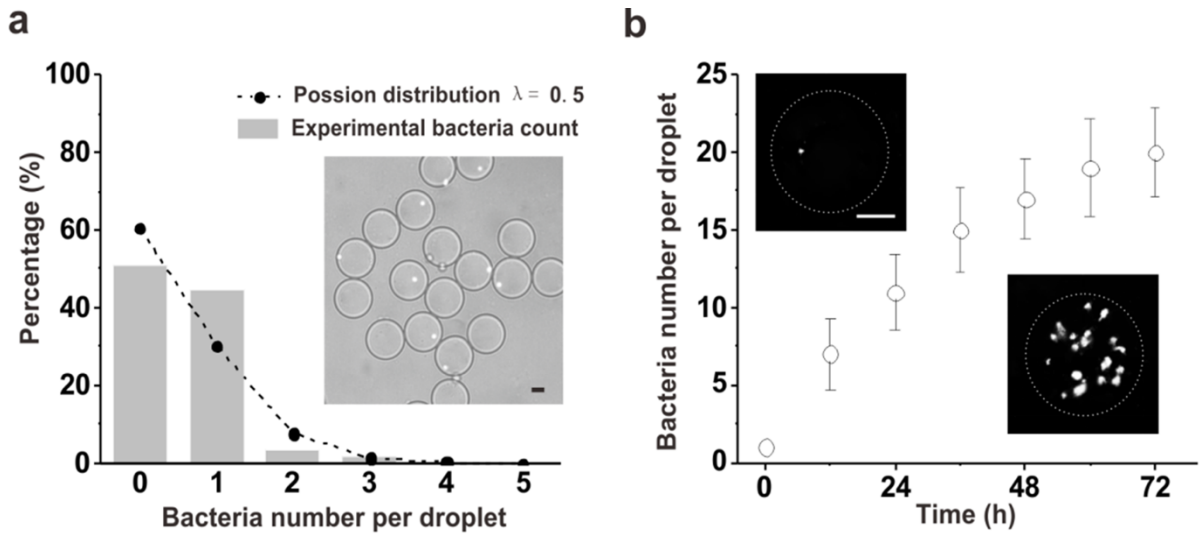

Figure S1: Encapsulation and proliferation of bacteria in agarose hydrogel particles: (a) Comparison of the experimental count and the Poisson distribution ($\lambda = 0.5$) for the encapsulation rates of 0 to 5 bacteria per droplet; (b) bacteria number per droplet counted over 72 h showing the proliferation of bacteria under the agitated culture condition; scale bar: 10 μm.

Reference:

Single cell kinase signaling asay using pinched flow coupled droplet microfluidics. R. Ramji, M. Wang, A. A. S. Bhagat, D. S. W. Tan, N. V. Thakor, C. T. Lim and C. H. Chen, *Biomicrofluidics*, 2014, 8, 034104.

Supporting information 3: Applications based on bacterial cellulose types

Based on the morphology of BC, we manually classified the secreted BC into five categories: I. small amount of cellulose; II. medium amount of BC with speckles; III. large amount of BC with speckles; IV. medium amount of BC agglomerates; and V. large amount of BC agglomerates. Based on previous research [A. Basu et al., *Scientific Reports*, 2018, 8, 5780, J. Wang et al., *Carbohydrate Polymers*, 2019, 210, 63], BC with different morphological properties will display different features and thus have varied applications. BC with a dense and compact morphology (category IV and V in our study) presents high mechanical strength and could be used to fabricate synthetic supporting materials and artificial blood vessels. On the other hand, BC with loose, porous and large surface areas (category II and III in our study) will have such advantages as high water-holding capacity and high transparency and are suitable for applications in wound dressings, membranes and scaffolds for tissue regeneration. In this study, the presented platform was used to rapidly identify the five defined classes of BC, which would be helpful to determine the bacteria that could produce suitable BC for desired applications. A description of the functions of the five defined categories is attached below in **Table S2** (definitions of the five BC types and their applications).

Table S2 Five categories classified based on BC morphology in agarose hydrogel particles

Category	I	II	III	IV	V
Morphology type	Small amount of BC	Medium amount of BC speckles	Large amount of BC speckles	Medium amount of BC agglomerates	Large amount of BC agglomerates
Schematic					
Feature	Very little BC in particle	Loose, porous, large surface area; high water-holding capacity, high transparency		Compact and smooth surface; high mechanical strength	
Applications	N/A	Wound dressings, membranes and scaffolds for tissue regeneration		Synthetic supporting materials, artificial blood vessels	
Bacteria strain	With low BC productivity	*K. rhaeticus* *G. xylinus*		*G. hansenii* *G. xylinus*	

References:

A. Basu, S. V. Vadanan et al., "A Novel Platform for Evaluating the Environmental Impacts on Bacterial Cellulose Production", *Scientific Reports*, 2018, 8, 5780.

J. Wang, J. Tavakoli et al., "Bacterial cellulose production, properties and applications with different culture methods – a review.", *Carbohydrate Polymers*, 2019, 210, 63.

Supporting information 4: Workflow to prepare the convolutional neural network database

To prepare the data for convolutional neural network (CNN) analysis, a schematic to introduce data preparation is shown in **Figure S2**. In the first step, the five categories were manually labeled based on morphologies of bacterial cellulose (BC) in an agarose hydrogel particle, which are illustrated on page 3 as follows: I. small amount of BC; II. medium amount of BC speckles; III. large amount of BC speckles; IV. medium amount of BC agglomerates; and V. large amount of BC agglomerates.

The second step was to construct a database of five categories. The scattering images and their corresponding brightfield images were recorded together in the static state, as shown in Figure 3. The classes of cellulose were labeled manually based on brightfield images. For each category, 500 images were recorded to form an image database with 2,500 images in total.

The third step was to establish a specialized CNN. The CNN was a classifier with an input of a monochromatic 128×128 pixel scattering image and an output of five defined categories. CNN can conduct feature learning of a scattering image through a series of convolutional layers with rectified linear units (ReLU) and pooling processes. A total of 2,500 monochromatic 128×128 pixel scattering images (500 for each category) with BC morphologies were chosen randomly for algorithm training. The training process was accomplished when accuracy could not be further improved by the algorithm.

The fourth step was the continuous screening process. In our platform, the aqueous flowrate was 10 μL/min (no oil flow). Approximately 100 particles passed through the sensing area of the CCD camera per second. The camera capture speed was 200 fps. The synchronization between flowrate and camera capture speed could be improved by further characterization, but although some particle images were missed due to asynchrony, the number of images collected was already sufficient for model build-up. In this study, with the capability of high-throughput screening, statistical information was effectively obtained for analysis. The optofluidic image pattern of empty agarose hydrogel particles was collected and labeled in our database and would not affect our reading outcome.

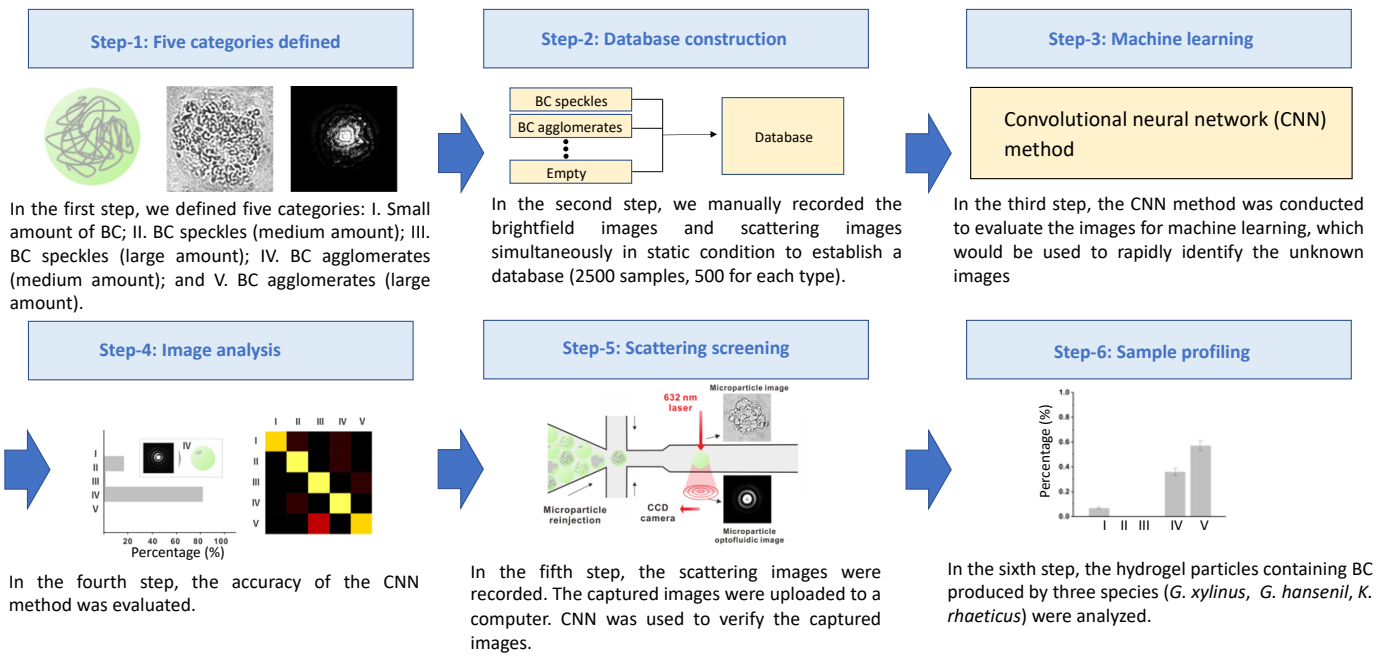

Figure S2: Workflow to prepare the database for CNN analysis of BC production in agarose hydrogel particles.

Supporting information 5: Convolutional neural network (CNN) training process

A convolutional neural network (CNN) training process of the optofluidic images of single hydrogel particles for three categories (high, medium and small amounts of cellulose in the agarose hydrogel particles) was developed. The CNN approach was applied to obtain the results shown Figure 4c. Five hundred images were prepared in each category. To train the CNN algorithm, 500 epochs were processed. Each epoch contained 11 iterations. The readouts were validated every 30 iterations. After 18 epochs (198 iterations), the accuracy of the CNN model approached 100% (**Figure S3**).

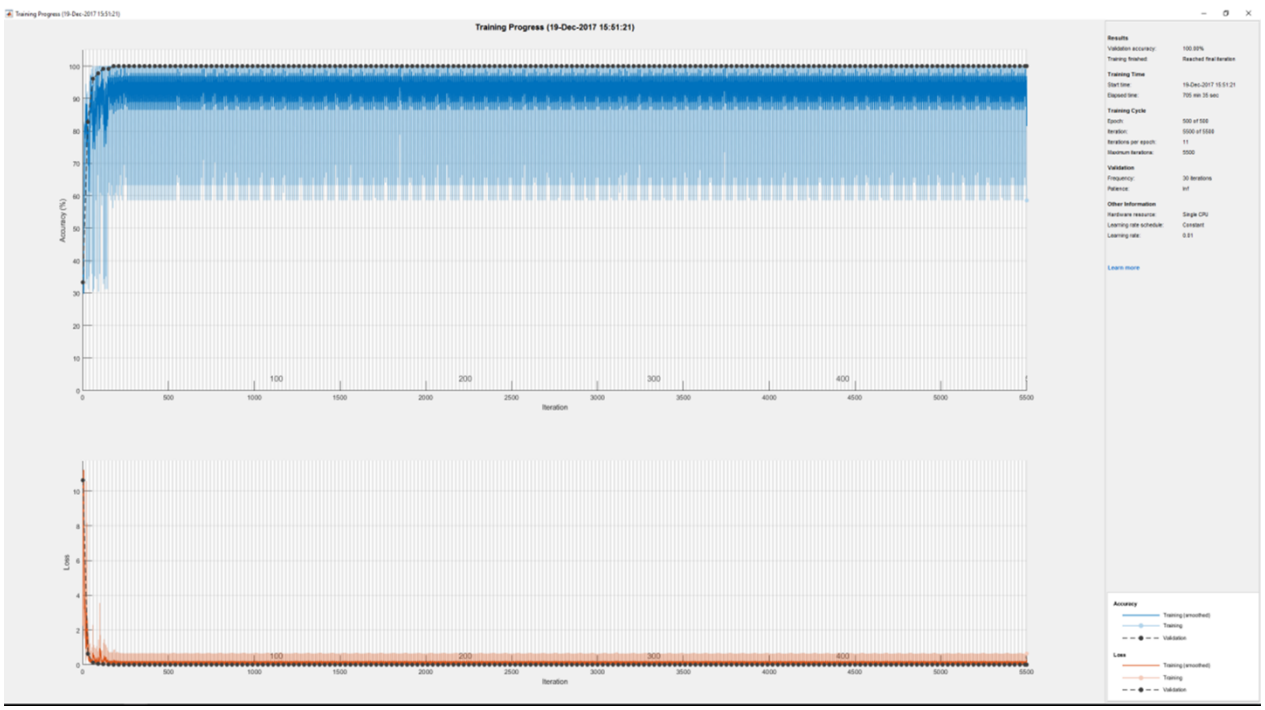

Figure S3: Convolutional neural network (CNN) training process of the optofluidic images of single hydrogel particle analysis.

Supporting information 6: Convolutional neural network (CNN) code for data production

Five steps were used to construct our convolutional neural network (CNN) algorithm: 1. Load the database, 2. Count the images for each label, 3. Set the number of each class for training, 4. Define the CNN architecture, and 5. Train the CNN. The complete CNN code with all parameters is provided here for data reproduction.

1. Load the database

```
digitDatasetPath = fullfile(matlabroot,'bin','database');
digitData = imageDatastore(digitDatasetPath,...
    'IncludeSubfolders',true,'LabelSource','foldernames');
```

2. Count the images for each label

(note: confirm all classes have 500 images, not a necessary step)

```
labelCount = countEachLabel(digitData)
```

3. Set the number of each class for training

(note: 80% from database)

```
trainNumFiles = 400;
[trainDigitData,valDigitData] = splitEachLabel(digitData,trainNumFiles,'randomize');
```

4. Define the CNN architecture

```
layers = [
    imageInputLayer([128 128 1])    (note: 128*128 is the size of image)
convolution2dLayer(3,16,'Padding',1)
(note: filter size is 3-by-3; the number of filters is 16;)
batchNormalizationLayer
(note: this layer is used to normalize the activations and gradients propagating through a network)
    reluLayer          (note: this is a nonlinear activation function)
maxPooling2dLayer(2,'Stride',2)
```

convolution2dLayer(3,32,'Padding',1)

(note: reduces the spatial size of the feature map and removes redundant spatial information)

batchNormalizationLayer

reluLayer

maxPooling2dLayer(2,'Stride',2)

convolution2dLayer(3,64,'Padding',1)

batchNormalizationLayer

reluLayer

fullyConnectedLayer(7)

(note: this layer is used to combine the features for image classification)

softmaxLayer

classificationLayer];

5. Train the CNN

options = trainingOptions('sgdm',...

'MaxEpochs',500, ... (note: set training iteration)

'ValidationData',valDigitData,...

'ValidationFrequency',30,... (note: set validation frequency to 30)

'Verbose',false,...

'ValidationPatience',30,... %could be inf

'Plots','training-progress');

net = trainNetwork(trainDigitData,layers,options);

Supporting information 7: Regression suggested to improve accuracy

In this study, we did not use a regression strategy. The accuracy for identifying the class of BC was ~80%. It is expected that by including regression strategies, accuracy could be improved for a wide range of different applications (such as immune cell and cancer cell identification). L2 regularization (ridge regression) has been suggested to improve the accuracy of identifications. This strategy adds dropout layers and batch normalization layers and can be used to optimize our current model. To create a less complex (parsimonious) model, L2 regularization could be conducted to address the over-fitting issue by adding squared magnitude. In addition, more labeled images could be added to enlarge our training dataset to further improve the accuracy of identification.

Supporting information 8: Scanning electron microscopy (SEM) images

Scanning electron microscopy (SEM) images of the cellulose structures were obtained using an FEI Quanta 200 scanning electron microscope (Holland Philips Co.). The single-cell morphology and BC produced by *Gluconacetobacter xylinus, Gluconacetobacter hansenii* and *Komagataeibacter rhaeticus* in bulk conditions are shown in **Figure S4**.

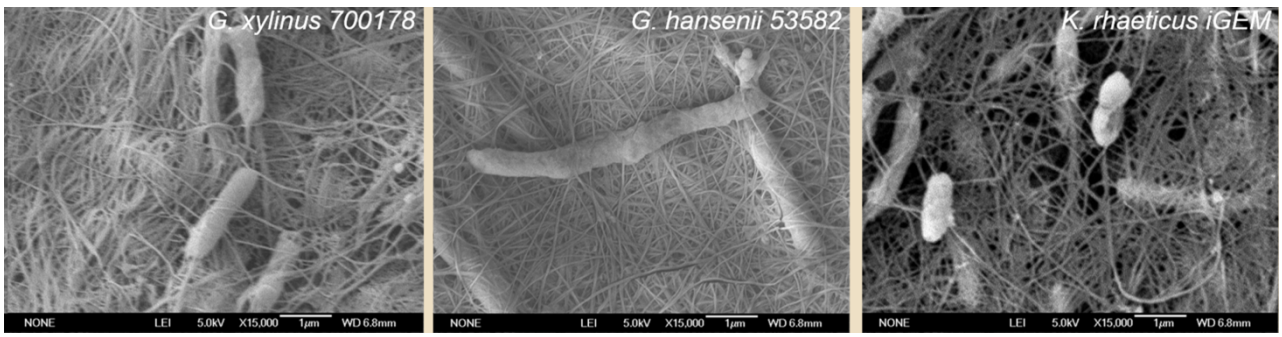

Figure S4: Scanning electron microscopy (SEM) photos of *Gluconacetobacter xylinus* (700178)*, Gluconacetobacter hansenii* (53582) and *Komagataeibacter rhaeticus* (iGEM) and the cellulose they produced after 7 days of static culture.