Supplementary Information

Typing of acute leukemia by intelligent optical time-stretch imaging flow cytometry on a chip

Yueyun Weng,†[ab] Hui Shen,†[c] Liye Mei,†[a] Li Liu,[c] Yifan Yao,[a] Rubing Li,[a] Shubin Wei,[a] Ruopeng Yan,[a] Xiaolan Ruan,[d] Du Wang,[a] Yongchang Wei,[e] Yunjie Deng,[f] Yuqi Zhou,[f] Tinghui Xiao,[f] Keisuke Goda,[afg] Sheng Liu,[ab] Fuling Zhou*[c] and Cheng Lei*[af]

[a]*The Institute of Technological Sciences, Wuhan University, Wuhan, China.*
[b]*The Key Laboratory of Transients in Hydraulic Machinery of Ministry of Education, School of Power and Mechanical Engineering, Wuhan University, Wuhan, China.*
[c]*Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan, China.*
[d]*Department of Hematology, Renmin Hospital of Wuhan University, Wuhan, China.*
[e]*Department of Radiation & Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan, China.*
[f]*Department of Chemistry, University of Tokyo, Tokyo, Japan.*
[g]*Department of bioengineering, University of California, Los Angeles, UAS.*
[†]*These authors contributed equally to this work.*
[*]*Corresponding authors: zhoufuling@whu.edu.cn and leicheng@whu.edu.cn*

Supplementary Material List

- Supplementary Material and Methods
- Supplementary Table 1 – 3
- Supplementary Figures 1 – 10

Supplementary Material and Methods

The effectiveness of flow focusing in microchannel

The microfluidic chip is of great significance in OTS imaging flow cytometry, which can stably focus the cells on a smaller observation window for detection. The laminar flow is formed by simultaneously injecting two sheath flows and one sample flow to balance the inertial force and viscous drag force. So, the cells in the channel are focused on a line on the focal plane to form stable, high-speed, and linear single-cell flow. To verify the effectiveness of the channel we used in the experiment, we evaluate the performance of cell focusing using both simulation and experiments.

First, we employ COMSOL Multiphysics 5.6 to make a numerical study to investigate the hydrodynamic-focusing process. The geometric model of the channel is shown in Fig. S3(a), where the width, height, and length of the channel are 80 μm, 35 μm, and 1 mm respectively. The flow condition is set as laminar flow, and the static pressure of the outlet is set to 0. As shown in Fig. S3(b), the distribution of velocity in the flow field is calculated when the flow rate of sheath flow and core flow is set at 50 μL/min and 40 μL/min, respectively, which are the same as the settings in the AL typing experiments. Moreover, to visualize the flow field, we inject red ink into the core flow, and the simulation and experimental results are shown in Fig. S3(c) and Fig. S3(d), respectively. The experimental and convection-diffusion simulation results are consistent, indicating that the core flow is sandwiched by the sheath flow, and the microchannel can achieve good flow focusing on the x-y plane.

To further demonstrate the effectiveness of flow focusing, we also make a particle tracing investigation. As shown in Fig. S4(a), particles with three diameters, *i.e.* 10 μm, 15 μm, and 20 μm, are released at the inlet of core flow to simulate the movement of cells with different sizes in the microchannel. To simplify the model, we ignore the interactions between particles and only consider the drag fore and the wall-induced lift force. It can be seen from Fig. S4(b) that particles are gradually focused in the y-axis direction due to the viscous drag force which is dominant and located in the center of the channel. The biggest distance between the central positions of particles does not exceed 10 μm. In addition, as shown in Fig. S4(c), particles are gradually focused in the z-axis direction and also located in the center of the channel due to the effect of wall-induced lift force. The biggest distance between the central positions of particles does not exceed 5 μm. Therefore, the simulation results of particle tracking demonstrate

the effectiveness of flow focusing in the z-axis direction.

Additionally, we also experimentally show the focusing performance using cells with various sizes. As shown in Fig. S5. In the experiment, we capture 15 images consecutively each time. It should be noted that continuous capture means that the oscilloscope is triggered to sample 15 times continuously. Fig. S5(a), S5 (b), and S5 (c) represent the experimental results using three different microfluidic channels with the same dimensions as described above, while I and II represent two different experiments of each channel. In each panel, I shows consecutively captured images of single cells with different sizes, and II shows consecutively captured images of single cells, small particles, small cell debris, or cell clusters with different sizes. Since we cut the two edges of the Gaussian pulses in the experiment, there are two black areas on the two sides of the images. The experimental results shown in Fig. S5 indicate that the single cells, small particles, small cell debris, and cell clusters are all focused near the center of the channel. The fluctuation of the central positions of single cells, small particles, small cell debris, and cell clusters in the y-axis direction does not exceed 10 μm. And the contrast and brightness of the image of each particle are almost identical. Therefore, the experimental results in Fig. S5 demonstrate the effectiveness of flow focusing of the microfluidic channel in cell measurement.

Details about the architecture of the CNN

Details about the architecture of the CNN shown in Fig. 1 are shown in Table S1, where the input size of the CNN is 128×368×368. Virtual adversarial training (VAT) in the CNN is a semi-supervised learning regularization method and an effective data augmentation technique that can enhance the local smoothness of the model given the conditional entropy. We employ the advanced semi-supervised regularization method based on VAT to avoid overfitting and improve the generalization ability, aiming at balancing the sample diversity and correcting the unevenness of cell grouping in the training. Specifically, we add some noise perturbations in the cell images and use the VAT algorithm to find the adversarial perturbation and transform the inputs. Then we update the weight of the model to minimize the Kullback-Leibler (KL) divergence, making the model more robust. The formula of VAT is as follows:

$$L_{VAT} = D[p(y \mid x), p(y|x + r_{adv}, \theta)] \quad (1)$$

$$r_{adv} = arg\max_{r} \{D[p(y \mid x), p(y \mid x + r, \theta]; \|r\|_2 \leq \varepsilon\} \quad (2)$$

where r_{adv} represents the added adversarial noise to the input cell image groups, and *r* represents a perturbation to the input data, which obeys a normal distribution with a mean of 0 and a variance of 1. ε represents the tolerance value, which is 0.5 in this work. $\|\cdot\|$ represents the L2 norm, *x* represents the input cell image groups, and *y* represents the output. $p(y \mid x)$ represents the conditional probability given the input *x*. *D* represents the non-negative measurement of the output after adding noise to the clean cell image groups. KL is mainly used to measure the relative entropy between features, also known as information divergence, which is an asymmetric criterion used to measure the differences in probability distributions between the perturbed category and the originally predicted category. The formula is shown in Equation (3):

$$D_{KL}(p||q) = \sum_{i=1}^{N} [p(x_i)\log p(x_i) - p(x_i)\log q(x_i)] \quad (3)$$

where *N* represents the number of the input cell image groups, $p(x_i)$ represents the actual prediction probability of sample *i*, and $q(x_i)$ represents the prediction probability after adding perturbations. Compared with the cross-entropy loss, VAT does not use real category information, but only uses the model output to generate perturbations, and then uses metric methods to measure the output probability before and after the perturbations, and finally performs optimization. The processing speed of the CNN network is 125 image groups per second, and the time required to train the network is about 4 hours. The training is stopped when validation accuracy does not decrease for 20 consecutive epochs, and we save the model weights. All the experiments are conducted in Pytorch under an Ubuntu server with an Intel Xeon(R) CPU E5-2680 v4 @2.40 GHz, 256 GB of RAM, and two NVIDIA 3090 GPUs with 24-GB memory.

Cell images screening

The number of all acquired images is greater than the number of valid single-cell images, and the proportion of invalid images in all images varies greatly depending on the sample. Therefore, invalid images need to be screened out from all acquired images.

In order to analyze the morphology of single cell, biophysical phenotype of cell is usually defined, such as the cell area, radius, perimeter, major axis, minor axis, and major/minor axis ratio, image mean, standard deviation, entropy, contrast, RMS brightness, third moment, energy gradient and other features. Each feature is defined in Table S3 in the supporting information. Based on the above features, we combine them in pairs, hoping to effectively distinguish valid images from invalid images in each sample. In order to determine the better combination of parameters, we first manually separate the valid images and invalid images of a representative sample based on their morphological features and select valid images from the sample as the benchmark for comparison. Then, we calculate the distributions based on the different combinations of features of acquired all images and screened images manually, shown in (i), (ii) in Fig. S8. As shown in Fig. S8, when the combination of entropy and RMS brightness is applied, it is more effective to distinguish valid images from invalid ones. However, even with this feature combination method, the boundaries (red dashed rectangular box) cannot be accurately divided, resulting in the loss of valid images or the screened images containing a certain number of invalid images. In this work, a rectangular gating is used because it is the most straightforward way to draw a straight lines by selecting the critical points on each biophysical phenotype feature axis. When multiple features are applied, the accuracy of the distinction can be improved, but due to the diversity of samples, the value of the feature is determined by each sample itself and cannot be unified, which leads to a large amount of calculation and complexity of images screening.

Therefore, in order to achieve high-precision screening and simple operation, we build a valid image screening CNN network based on the mainstream lightweight MobileNetV3 network, shown in Fig. S9. The MobileNetV3 contains four components. For the first component, it is a convolution layer with kernel size 3×3, followed by a batch normalized function and a h-swish activation function. The next part includes 11 bottleneck blocks and a convolution layer. The previous three bottleneck blocks employ kernel size 3×3, and the remaining bottleneck blocks use kernel size 5×5. Each bottleneck block comprises an expansion layer and depthwise (DW) separable

convolution. The expansion layer increases the dimensions for nonlinearity to get a more accurate outcome. DW separable convolution obtains features from the former layer efficiently, consisting of the DW convolution layer and pointwise convolution layer. For the last five bottleneck blocks, it introduces Squeeze-And-Excite (SE) architecture, which enables the network to automatically select potential feature maps. At the end of each bottleneck block, it adopts a shortcut layer with the kernel size 1×1 when the stride is equal to 1. After the 11 bottleneck training, it applies a convolution layer with kernel size 1×1, followed by a h-swish activation function. For the third part, the model uses an average pooling layer with kernel size 7×7 and a fully connected layer with 1,280 neurons. Finally, we use the fully connected layer as the classification layer. In the training stage, we use cross-entropy as the loss function, an early stopping method to prevent overfitting and save the well-trained model. In the work, to match the network input, the cellular images acquired with our OTS imaging flow cytometer are resized to 256 × 256 before being sent to the network. The processing speed of the valid image screening CNN network is 650 images per second, and the time required to train the network is about 3 hours. The hardware architecture and programming software used are the same as the CNN network in Fig. 1. In order to accurately establish the cell images screening network, we manually screened 26,431 images, which are divided into 3 categories as the training dataset. One category is 5,992 single-cell images as valid images, one is a group of invalid images composed of cell clusters, large cell debris, and large flocs, and the other is another group of invalid images composed of small particles, small cell debris, small flocs.

In order to verify the effectiveness of the image screening network, two samples are selected from the ALL, AML patients and healthy donors respectively. One is that there are fewer invalid images with lower complexity, and the other is that there are more invalid images with higher complexity. First, we use the trained images screening model to select valid images from 6 samples. Correspondingly, we manually select valid images from 6 samples as the benchmark for comparison. Secondly, we also calculate the distributions based on the combination of entropy and RMS brightness of acquired all images, valid images screened by CNN and screened images manually,

shown in (i), (ii), (iii) in Fig. S10. As shown in Fig. S10, by dividing the boundary, the image is divided into three categories, one is the valid images in a red dashed rectangular box, one is the combination of small particles, small cell debris, and small flocs in the upper left corner, and the other is the combination of cell clusters, large cell debris and large flocs in the lower right corner. It can be seen that the divided boundary between the valid image and the invalid image in the lower complexity sample of Fig. S10(a), Fig. S10(c), and Fig. S10(e) is clear, with fewer valid images lost or the invalid images included, and the screening accuracy is 73.03%, 83.25%, 88.13%, respectively. It is difficult to accurately divide the divided boundary between valid images and invalid images in highly complex samples of Fig. S10(b), Fig. S10(d), and Fig. S10(f), the screening accuracy drops sharply, and the screening accuracy rates are 35.51%, 46.65%, 69.62%, respectively. For comparison, we calculated the accuracy of the valid images screened using the CNN model, which is 91.28%, 93.29%, and 98.31%, respectively for lower complexity samples and 87.28%, 95.51%, 97.52%, respectively for higher complexity samples, with an average accuracy of 94.52%. The above results show that the CNN model we built has high accuracy and wide adaptability, which meets the needs of the clinical diagnosis of leukemia. Therefore, we apply this model to screen the valid images from the 30 samples.

Supplementary Table 1 – 3

Table S1. Details about the architecture of the CNN

Layers	**kernel size**	**Stride**	**Padding**	**Input channel**	**Feature map size**	**Activation function**
Conv1	7×7	2	3	128	64×184×184	ReLU
CBAM1	-	-	-	64	64×184×184	Sigmoid
Maxpooling	3×3	2	1	64	64×92×92	-
Resnet Block1	3×3	1, 1 1, 1	0	64, 64 64, 64	64×92×92	ReLU
Resnet Block2	3×3	2, 1 1, 1	0	64, 128 128, 128	128×46×46	ReLU
Resnet Block3	3×3	2, 1 1, 1	0	128, 256 256, 256	256×23×23	ReLU
Resnet Block4	3×3	2, 1 1, 1	0	256, 512 512, 512	512×12×12	ReLU
CBAM2	-	-	-	512	512×12×12	Sigmoid
Avgpooling	12×12	2	-	512	512×1×1	-
FC	-	-	-	512	3	Softmax

Table S2. Data distribution and diagnosis results of 12 clinical samples

Sample	Clinical Diagnosis	OTS Diagnosis	Amount of Valid Images	Number of Groups
1	ALL	ALL	7,322	57
2	ALL	ALL	8,434	65
3	AML	AML	12,134	94
4	AML	AML	13,583	106
5	AML	AML	1,520	11
6	AML	AML	5,574	43
7	AML	AML	8,585	67
8	AML	AML	5,896	46
9	AML	AML	3,597	28
10	Healthy	Healthy	17,012	132
11	Healthy	Healthy	5,476	42
12	Healthy	Healthy	6,973	54

Table S3. Definitions of features of the biophysical phenotype of cell

Feature	Symbol	Unit	Description or Equation
Cell Spatial Coordinate	x,y	-	The coordinate position of the cell corresponding to the pixel in the image
Major Axis	l	μm	The length of the major axis of the ellipse contour
Minor Axis	w	μm	The length of the minor axis of the ellipse contour
Major/minor Axis Ratio	k	-	The ratio of the major axis and the minor axis
Radius	r	um	The mean value of the distance from the centroid of the ellipse contour to each point on the contour
Perimeter	p	μm	Cell contour length
Area	A	μm^2	The area within the cell contour
Image Entropy	E	-	$\sum_{x=1}^{N}\sum_{y=1}^{N} I(x,y) log I(x,y)$
Image Mean	μ	-	$\frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N} I(x,y)$
Image Variance	δ		$\frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N} (I(x,y)-\mu)^2$
Image Standard Deviation	σ		$\sqrt{\frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N} (I(x,y)-\mu)^2}$
Image Third Moment	δ'		$\sqrt[3]{\frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N} (I(x,y)-\mu)}$
Image Contrast	I_C		$\sum_{x=1}^{N}\sum_{y=1}^{N} (i-j)^2 I(x,y)$
Image RMS Brightness	I_{RMS}		$\sqrt{\frac{\sum_{x=1}^{N}\sum_{y=1}^{N} I(x,y)^2}{N^2}}$
Image Energy Gradient	E_g		$\sum_{x=1}^{N-1}\sum_{y=1}^{N-1} ((I(x+1,y)-I(x,y))^2 + (I(x,y+1)-I($
Image Brenner Gradient	B_g		$\sum_{x=1}^{N-2}\sum_{y=1}^{N} (I(x+2,y)-I(x,y))^2$

Supplementary Figures 1 – 10

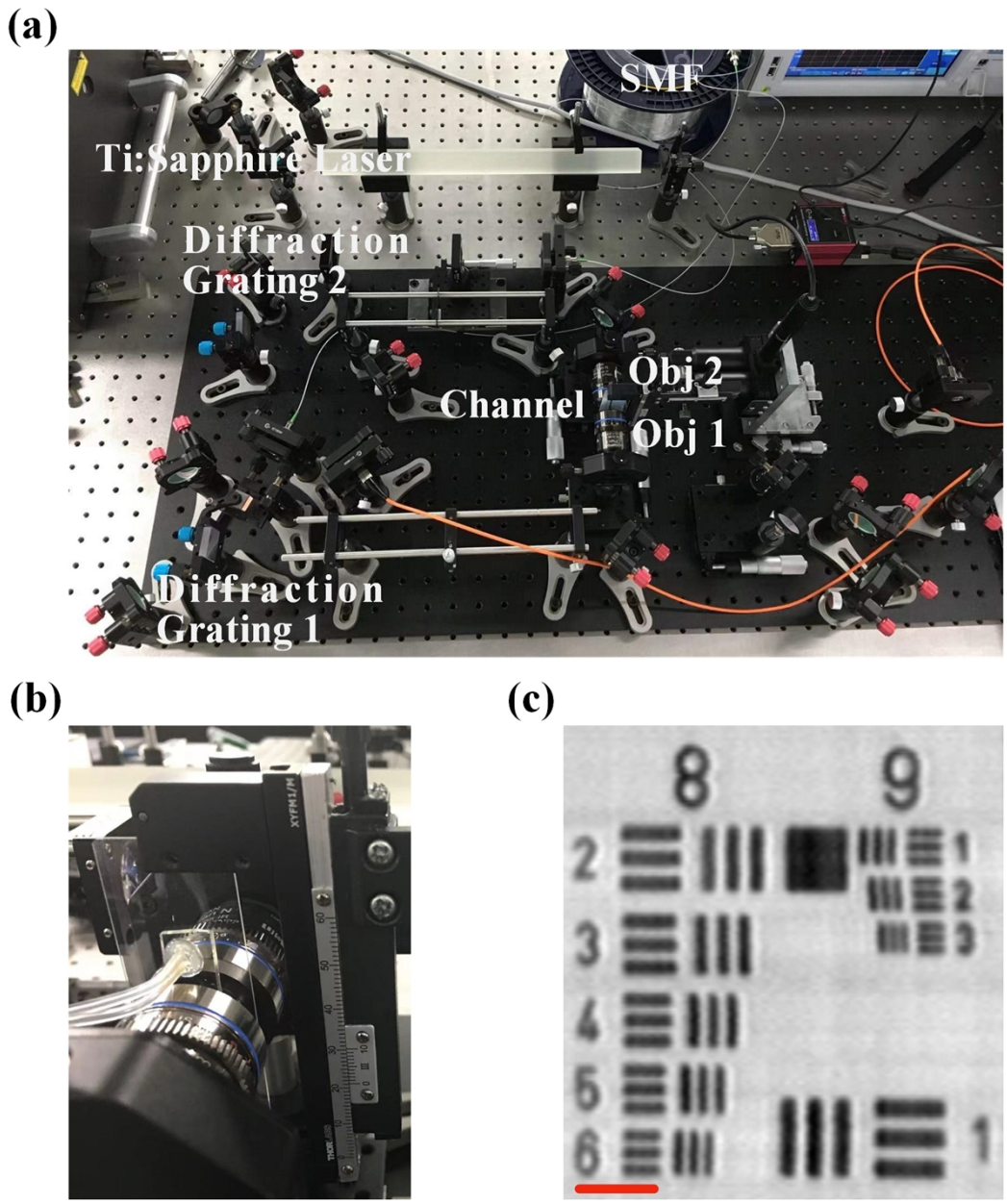

Figure S1. Setup and performance of the OTS imaging flow cytometry. (a) The picture of the OTS imaging flow cytometry. (b) The picture of the position of the microfluidic chip in OTS imaging flow cytometry. (c) Imaging the USAF-1951 resolution chart by OTS imaging flow cytometry with a spatial resolution better than 780 nm. Scale bar: 10 μm.

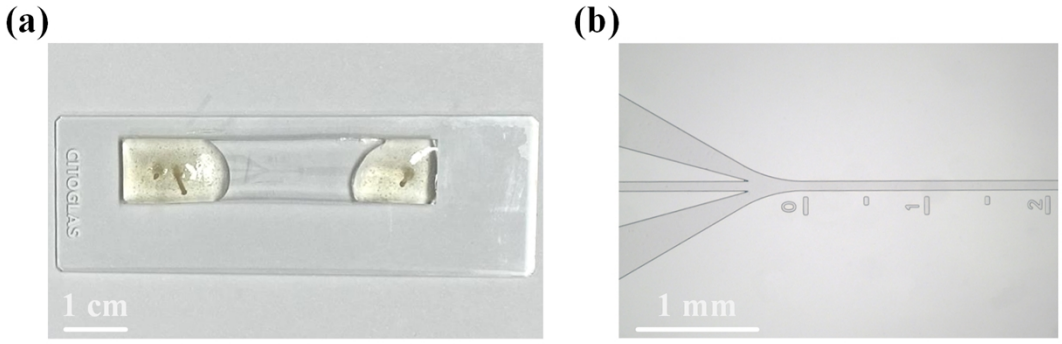

Figure S2. Pictures of the microfluidic chip and channel. (a) The picture of the microfluidic chip. (b) Microscopic picture of microchannel.

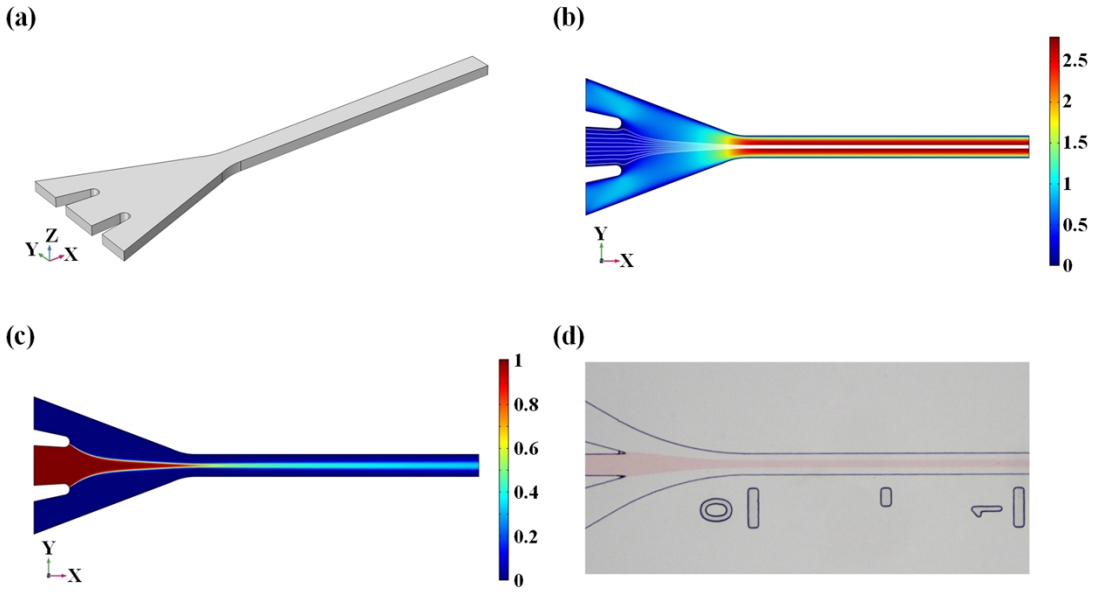

Fig. R3. The simulation and experimental results for the visualization of the flow field. (a) The geometrical model for numerical simulation. (b) The velocity profile along the microchannel in the simulation, where the streamlines are denoted by white lines and the color bar indicates the velocity. (c) The concentration profile along the microchannel in the simulation and the color bar indicates normalized concentration. (d) The concentration profile along the microchannel in the experiment.

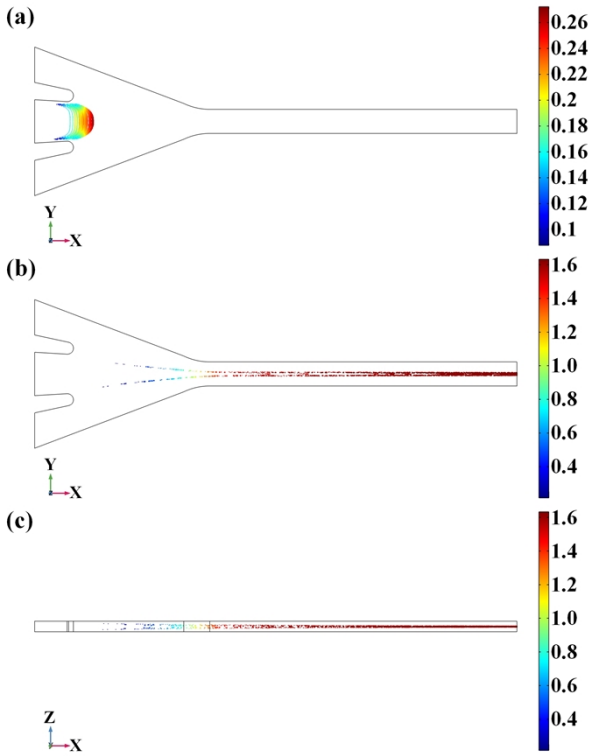

Fig. R4. The simulation results for the particle tracing within the microchannel. (a) and (b) Particle trajectories in the x - y plane after 0.8 ms and 2.3 ms of particle release. (c) Particle trajectories in the x - z plane after 2.3 ms of particle release.

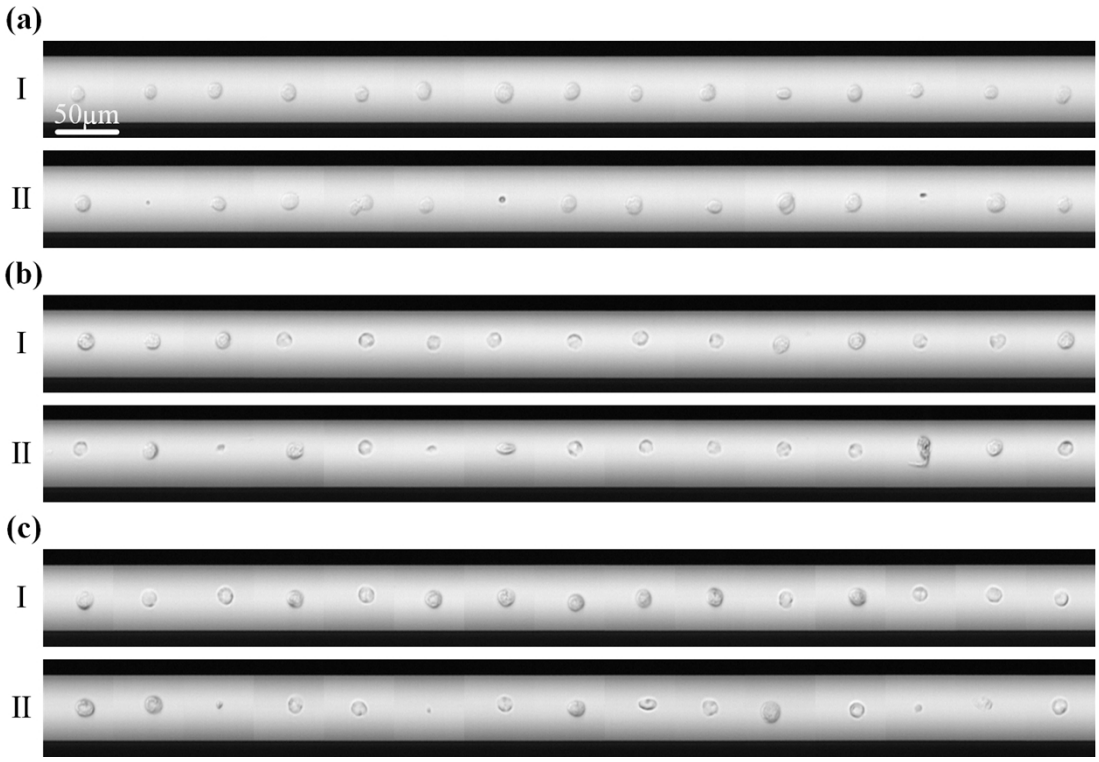

Figure S5. Consecutively captured images of cells in different microchannels.

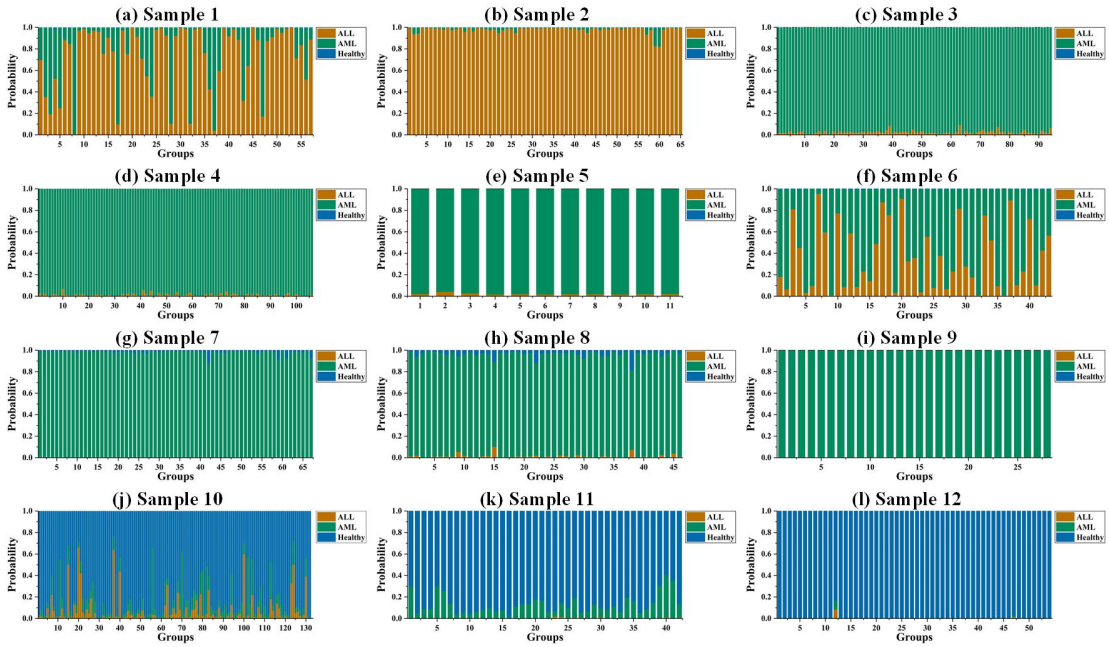

Figure S6. The probability of the typing of each group and the numbers of the groups in 12 samples. (a), (b) The number of groups belonging to the ALL is the most. (c) – (i) The number of groups belonging to the AML is the most. (j) – (l) The number of groups belonging to the healthy donors is the most.

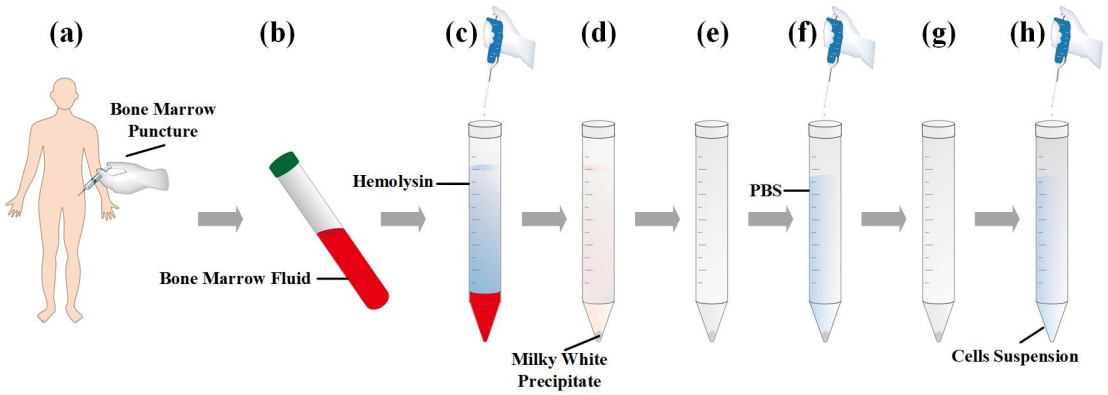

Figure S7. Detailed Procedure for Sample Preparation. (A) The 2-3ml of bone marrow fluid is extracted by bone marrow puncture. (B) The bone marrow fluid is dripped into a heparin anticoagulation tube. (C) The evenly shaken bone marrow effusion and hemolysin are poured into the centrifuge tube successively. (D) The sample is evenly mixed and centrifuged. (E) The supernatant is sucked and a milky white precipitate is left. (F) A PBS solution is added to the centrifuge tube, shaken evenly, and centrifuged. (G) The supernatant is sucked and a milky white precipitate is left. (H) The centrifuged sample is again added with PBS solution and shaken evenly.

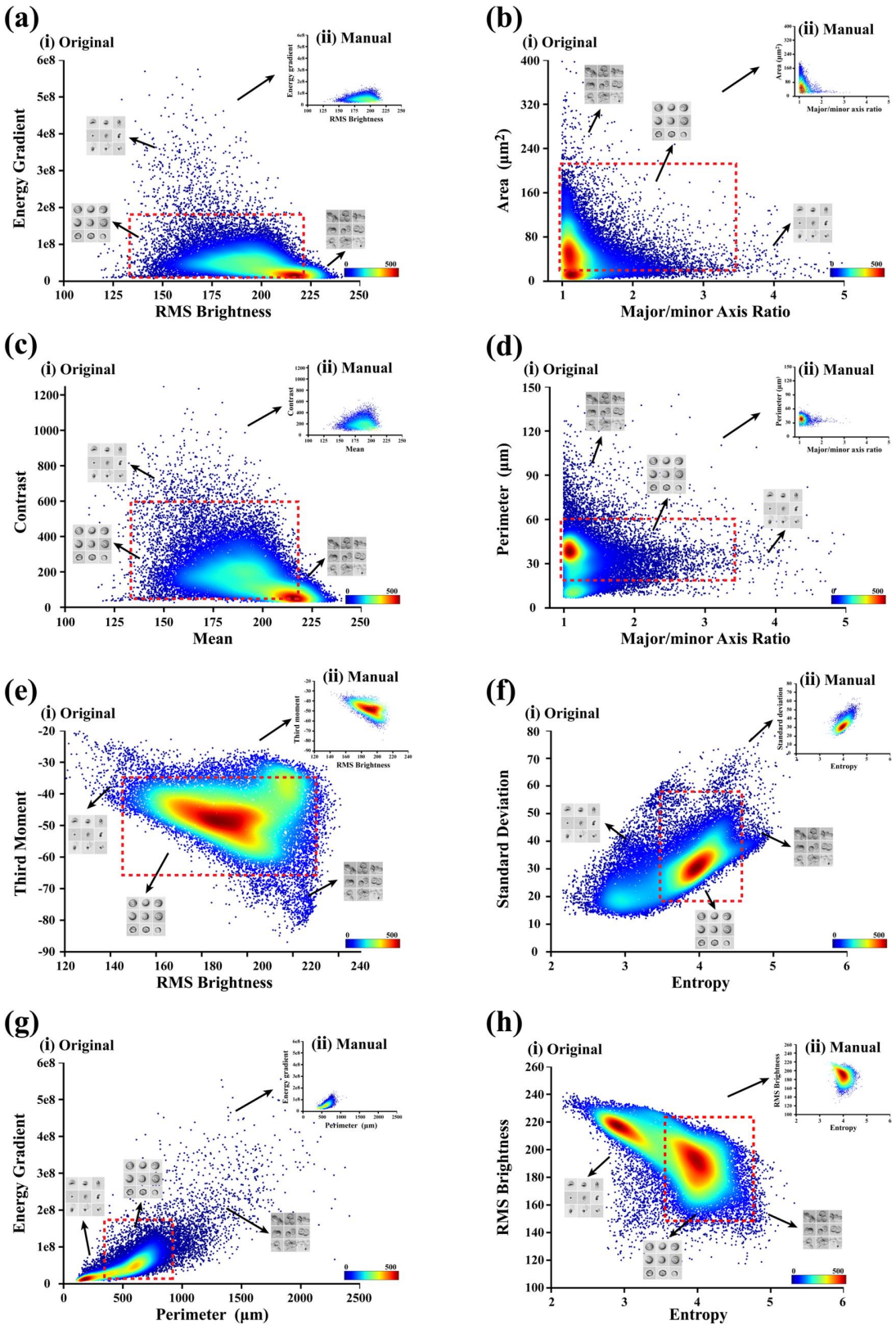

Figure S8. Comparison of image screening with different combinations of feature parameters. (a) The feature combination of RMS brightness and energy gradient. (b) The feature combination of major/minor axis ratio and area. (c) The feature combination of mean and contrast. (d) The feature combination of major/minor axis ratio and perimeter. (e) The

feature combination of RMS brightness and third moment. (f) The feature combination of entropy and standard deviation. (g) The feature combination of perimeter and energy gradient. (h) The feature combination of entropy and RMS brightness. (i) Distribution of all cell images under different feature combinations. (ii) Distribution of valid images by manual screening under different feature combinations.

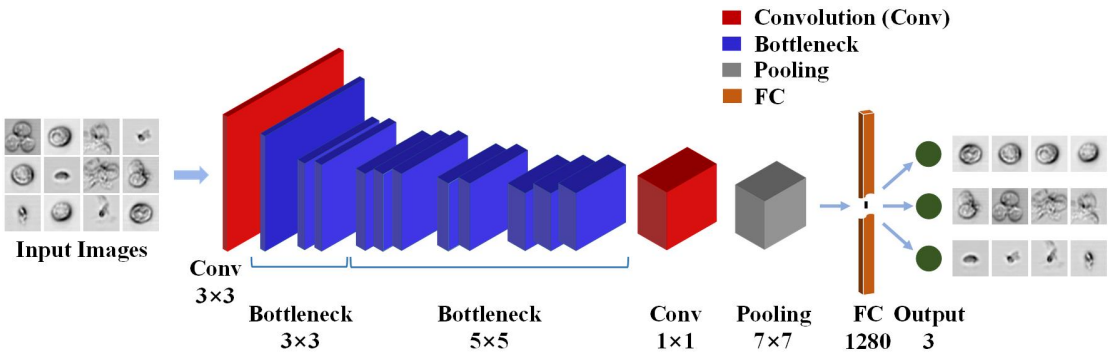

Figure S9. Images screening model based on the mainstream lightweight MobileNetV3 network.

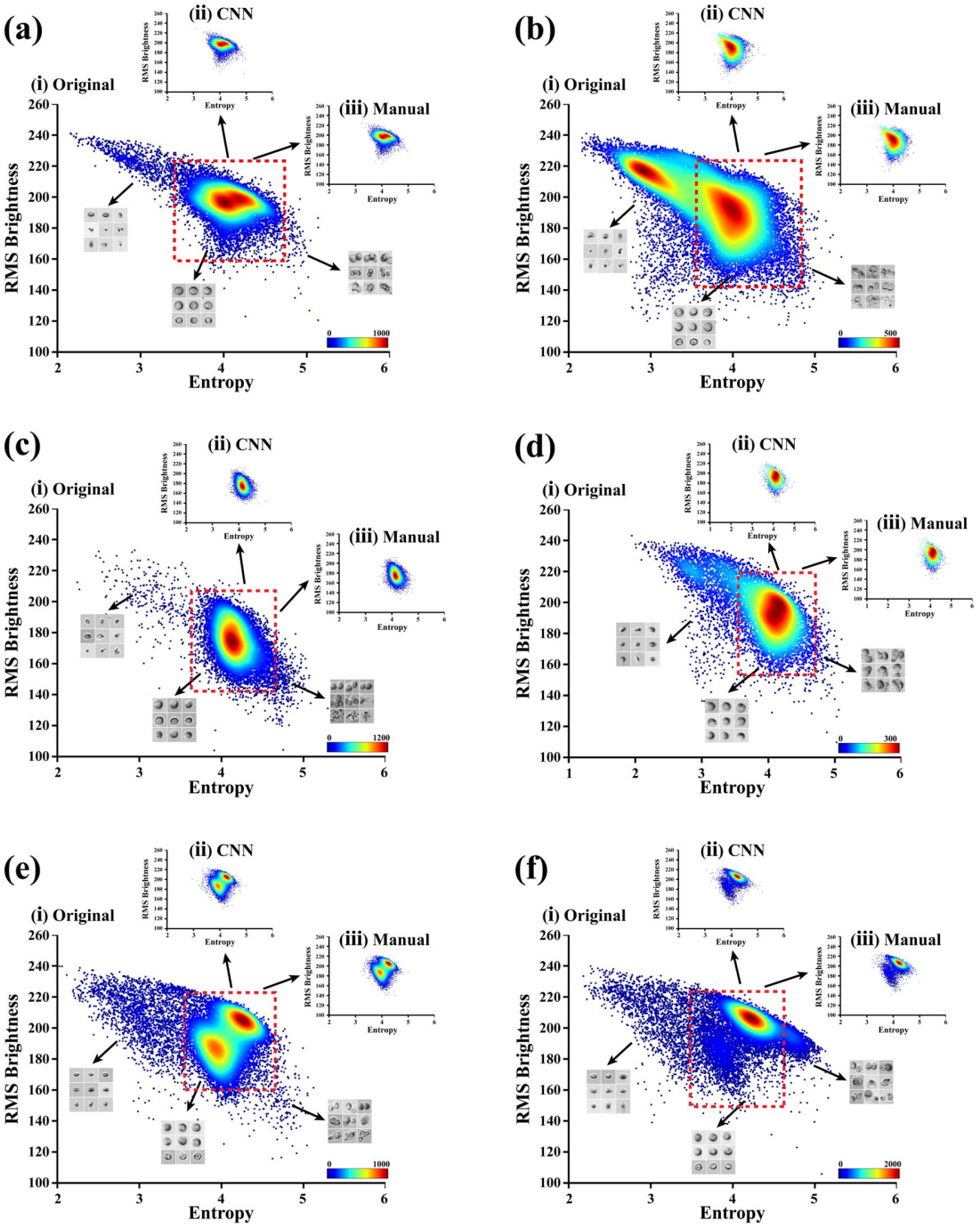

Figure S10. Comparison of biophysical phenotype feature, CNN, and manual cell image screening methods. (a), (c), (e) Images screening results in the lower complexity sample of ALL, AML, and healthy donors sample, respectively. (b), (d), (f) Images screening results in the high complexity sample of ALL, AML, and healthy donors sample, respectively. (i) Distribution of all images. (ii) Distribution of images screened by CNN from all images. (iii) Distribution of images screened by manual from all images.