Supporting information

Analysis of the Confidence of the Bead Separation Results	
Receiver Operator Characteristic Method	
Figure S-1	(a) The terminology in ROC analysis; (b) The ROC curve and the definition of fractional separation.
Figure S-2	a) The histogram of percentage size distribution of 356 4T1 cells with an average diameter of 17.63 µm and a standard deviation of 3.04 µm; b) The histogram of percentage size distribution of 302 bone marrow cells with an average diameter of 8.08 µm and a standard deviation of 1.39 µm.

Analysis of the Confidence of the Bead Separation Results

In this section, we attempt to determine the minimum number of 1.97 μm or 4.84 μm beads required to obtain statistically meaningful results, as listed in Table 1. Given the assumption that the small and large beads segregate independently, we can consider the binomial probability, p_{big}, that a big bead goes to the upper channel. From the preliminary data a good lower bound of this probability is 0.9. The same is true for p_{small}, the probability that a small bead goes to the lower channel. This lower bound of $p > 0.9$ holds for all speeds. Now the question is how many beads we must observe in order to reduce the margin of error below 4%? This is given by solving the following inequality:

$$1.96\sqrt{p(1-p)/n} < 4\% \ .$$

The solution yields $n > 216$ beads. Using the more accurate Wilson Confidence Interval Estimate yields $n > 219$ beads. So provided that the true probability of segregation is larger than 0.9, then any measured fraction will be within 4% of the truth (95% confidence) if we conservatively measure 220 or more events.

In the experiment of separating the binary mixture of beads, we have continuously detected 323 beads (288 4.84 μm beads and 35 1.97 μm beads) in the upper branch and 321 beads (314 1.97 μm beads and 7 4.84 μm beads) in the lower branch at a sorting rate of 0.778 beads/sec. Similarly, the numbers of 4.84 μm beads and 1.97 μm beads recorded in the upper branch at 0.485 beads/sec were 257 and 7, compared with 3 4.84 μm beads and 306 1.97 μm beads in the lower branch. Since the number of beads interrogated is beyond 220 for both the upper and

lower branches in each experiment, we believe that the analysis based on the recorded data is

statistically meaningful.

Receiver Operator Characteristic Method

The receiver operator characteristic (ROC) is a powerful approach in statistical analysis, and

the resulted fractional separation provides an ideal parameter to quantitatively characterize the

separation performance. The detailed description of ROC can be widely accessed in literature.

In brief, for two overlapped distributions, a selection of a separation threshold could define a pair

of parameters, the sensitivity and specificity corresponding to the choice of the threshold. As

shown in Fig. S-1a, the threshold divides the two distributions into four regimes: True Positive

(TP: the section of the right curve (positive curve) above the threshold), False Positive (FP: the

section of the left curve (negative curve) above the threshold), False Negative (FN: the section of

the right curve below the threshold), True Negative (TN: the section of left curve below the

threshold). In the case of measured cell size distributions, these values are the cell counts in each

regime. Based on these values, a pair of parameters, sensitivity and specificity, can be defined

as: $Sensitivity = \dfrac{TP}{FN + TP}$, and $Specificity = \dfrac{TN}{FP + TN}$. For each choice of the threshold, the

pair of parameters (sensitivity, 1-specificity) define one point on the parametric ROC curve (Fig.

S-1b). If two populations are identical and fully overlap with each other, then the ROC curve is

the diagonal line in the ROC plane, $Sensitivity + Specificity = 1$. On the other hand, if two

populations are perfectly separated then the ROC curve is the union of the segments (0, sensitivity) and (1-specificity, 1), where the values of sensitivity and 1-specificity vary monotonically from 0 to 1 as the threshold value scans through the two distributions. Any overlapped pair of distributions will produce an ROC curve lies between these two extremes (the upper left triangle of the ROC plane). The fractional separation (FS) of two distributions is proportional to the area under the ROC curve (AUC), which is normalized according to the following definition: $FS = 2 \times AUC - 1$. The fractional separation provides a single quantitative parameter characterizing to what extent two distributions overlap, which we find is an ideal parameter to determine the performance of size-based separation for cells of continuous size distribution.
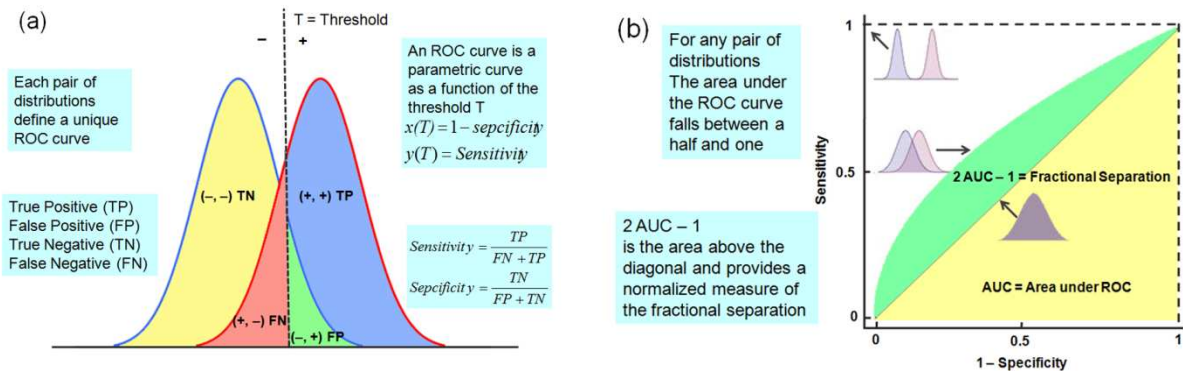

Figure S-1. (a) The terminology in ROC analysis; (b) The ROC curve and the definition of fractional separation.

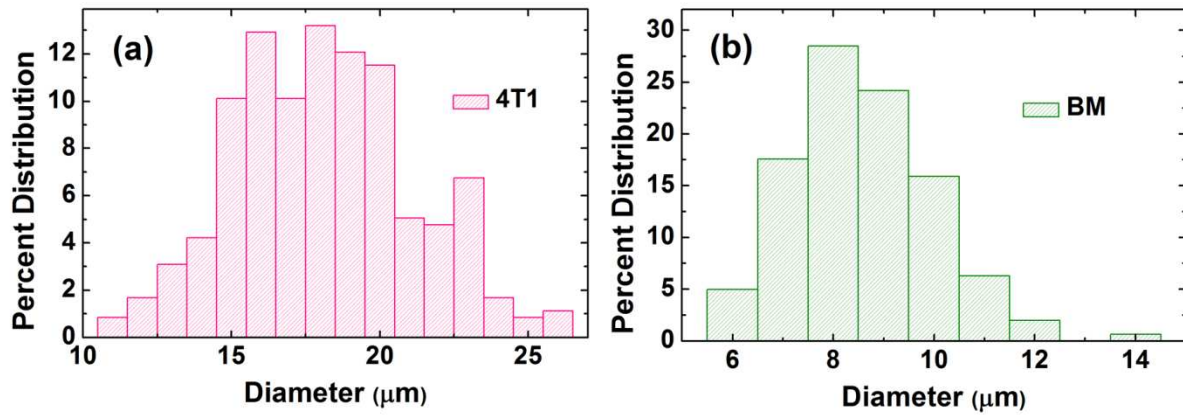**Size Distribution of the 4T1 and Murine Bone Marrow Cells**

Figure S-2. a) The histogram of percentage size distribution of 356 4T1 cells with an average diameter of 17.63 µm and a standard deviation of 3.04 µm; b) The histogram of percentage size distribution of 302 bone marrow cells with an average diameter of 8.08 µm and a standard deviation of 1.39 µm.