

Supplementary material

S1. Size of the circular patterns of p-selectin

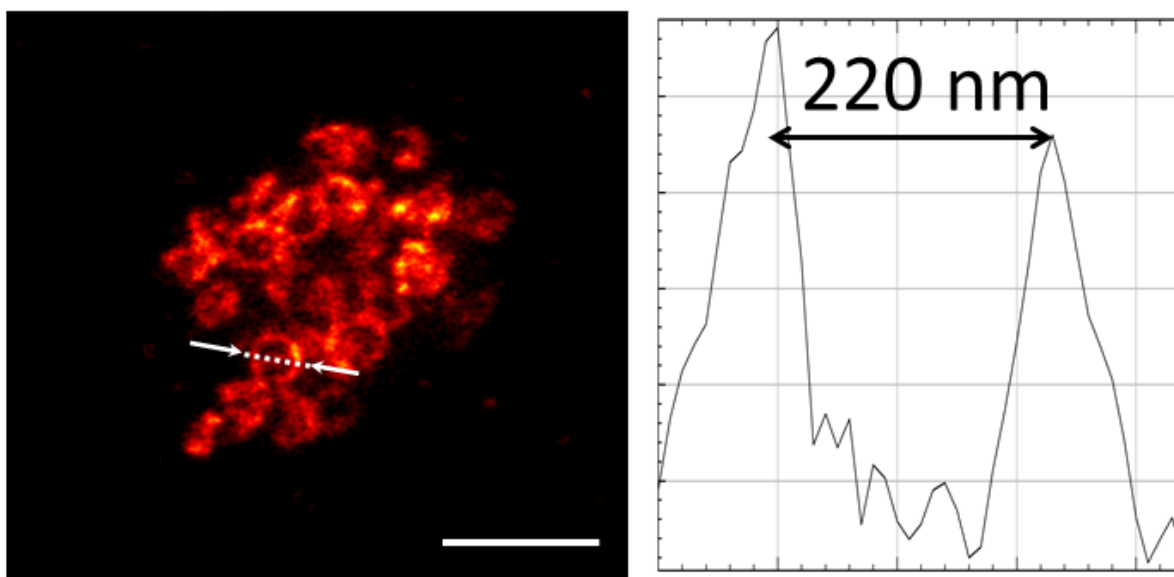


Figure S1 Representative raw STED image of a platelet stained for P-selectin, co-cultured with MDA-MB231 tumor cells. The circular P-selectin pattern can be clearly seen with circles of diameters of about 200-300 nm, as indicated by the line profile.

S2. Dictionary learning

The dictionary learning algorithm (implemented in Python using the Scikit package^{S1}) uses a computer to construct a dictionary from a training set of images (Figure 4A). The dictionary consists of a set of image patches, in our case 30x30 pixel image patches corresponding to 300x300 nm regions. The patches are built in such a way that every training image can be well described (according to conditions described in^{S2}) by a linear combination of the elements in the dictionary using least number of elements (sparsity)^{S2}.

To use such trained dictionary for classification requires training the dictionary on a large number of images that are representative for the images that will be classified later on. This number should be as large as possible, and we used 20000 training images, which is a feasible number to handle with the computing power of a standard PC. Given that we did not have

access to a sufficient number of experimental platelet images (less than 1000 in total, or approximately 100 per category) to train such a dictionary, and the difficulties for computers to detect and classify images based on circular patterns, we applied a modified strategy. Rather than experimental STED images, the dictionary was trained on 20000 computer simulated training images ($4\mu\text{m}\times 4\mu\text{m}$ corresponding to 400×400 pixels), mimicking platelets with a random distribution of P-selectin. These images were generated using Matlab2013b and contained cluster-like structures of dots (convolved with a Gaussian point spread function), varying in size between 20 and 40 nm, randomly and uniformly distributed (between 5 and 500 dots per image) within an elliptic area with minor and major axis randomly distributed between 1-4 μm in size and with different noise levels. The brightness of the dots was randomly distributed (such that signal-to-noise ratio took on values between 2 and 20) to resemble images of P-selectin in platelets displaying no circular patterns (see Fig. 2A).

Training on such images yields a dictionary that efficiently (sparsely) describes images containing no circular patterns. At the same time, images containing clear circular structures would not be as efficiently described by the same dictionary. The experimental images and the extent to which they display circular structures could then be classified by how well (or bad) the dictionary can describe the images.

S3. Radial distribution

The analysis for radial distribution was done in MATLAB2013b by first calculating the center of mass (CoM) of a given platelet, based on the intensity in every pixel in the image and given by the formula

$$(i_{\text{CoM}}, j_{\text{CoM}}) = \frac{1}{\sum_{i,j} \text{Inten}(i,j)} \sum_{i,j} i \cdot \text{Inten}(i,j) \quad \text{S1}$$

where $(i_{\text{CoM}}, j_{\text{CoM}})$ are the coordinates for the CoM within the image and $\text{Inten}(i,j)$ is the pixel intensity at pixel (i,j) in the image. From the CoM point, straight lines were drawn outwards, towards the periphery of the platelet (Figure 5A). The intensity in every pixel on the line was stored in an array, thus creating an intensity trace $I(r)$ for each line, where r is the distance from the CoM of the platelet (Figure 5A). Furthermore, the intensity traces were normalized such that the sum over all elements in $I(r)$ equals one, i.e. $\sum_r I(r) = 1$. For every

platelet, 72 such intensity traces were constructed, i.e. the intensity traces for each individual platelet were registered along 72 lines, drawn from the CoM with an angle of 5° between them. For each intensity trace, we calculated the first and second order moment $m1$ and $m2$. $m1$ corresponds to the average distance from the center of mass in the direction of the particular intensity trace under consideration, and $m2$ is the variance of the distance of the same intensity trace. $m1$ and $m2$ are given by

$$m1_j = \sum_i r_i I_j(r_i) \quad S2$$

$$m2_j = \sum_i (r_i - m1_j)^2 I_j(r_i) \quad S3$$

Here r_i is the pixel i at distance r from the CoM of the platelet and $I_j(r_i)$ is the intensity trace taken along the j :th line (Figure 5A). In this way a total of 72 different values of $m1$ and $m2$ were calculated for each platelet.

S4. Platelet categorization, combining SSIM and radial distribution features

The probability for a test platelet to belong to an activation condition is given by the conditional probability that given certain set of parameters (in our case; SSIM, first order moment $m1$ and second order moment $m2$ of the radial distribution), what is the probability for a test platelet to belong to an activation condition/category $c_i = \{\text{resting, ADP, thrombin, TXA2, 184A1, MCF10A, EFO21, MDA-MB231, MCF7}\}$. With the parameters for a test platelet labeled $s = \{SSIM, m1, m2\}$, then the probability for this test platelet to belong to category c_i , given the set s is given by Bayes' rule as

$$P(c_i|s) = \frac{P(s|c_i)P(c_i)}{P(s)} \quad S4$$

Here, $P(s|c_i)$ is the conditional probability for the test platelet to have the parameters s given the category c_i , $P(c_i)$ is the unconditional probability that the test platelet belongs to category c_i and $P(s)$ is the unconditional probability for the test platelet to have the parameters s .

To estimate the probabilities on the right-hand side in equation S4 we constructed histograms for the parameters $m1, m2, SSIM$, including all platelets for each particular activation condition (Figure 5C), as well as histograms for $m1, m2, SSIM$ including all platelets and all activation conditions (Figure S2). These histograms were normalized such that the area

underneath them equaled unity, so that they represented probability densities for the corresponding parameters.

The conditional probability $P(s|c_i)$ is calculated as the product of the probabilities of $m1, m2, and SSIM$ for each category c_i , obtained from the histograms for each separate category (Figure 5C).

It is important to note that both $m1$ and $m2$ are sets of 72 values for each platelet (i.e. one value for each of the 72 intensity traces in every individual platelet). Therefore, the probability for a set of values in the parameters $m1$ and $m2$, as well as $SSIM$, is given by

$$P(s|c_i) = \sum_{j=1}^{72} \sum_{k=1}^{72} P(m1_j|c_i)P(m2_k|c_i)P(SSIM|c_i) \quad S5$$

The unconditional probability $P(c_j)$ can be calculated as the empirical uniform probability given by the number of platelets in category c_j divided by the total number of platelets. The second unconditional probability $P(s)$ can also be estimated by equation S5, but now using the probabilities for a certain set of $m1, m2, and SSIM$ for all platelets (from all categories), as plotted in (Figure S2). Each test platelet was then categorized into the category that gave the highest probability, given the parameters $m1, m2, and SSIM$ for that particular test platelet.

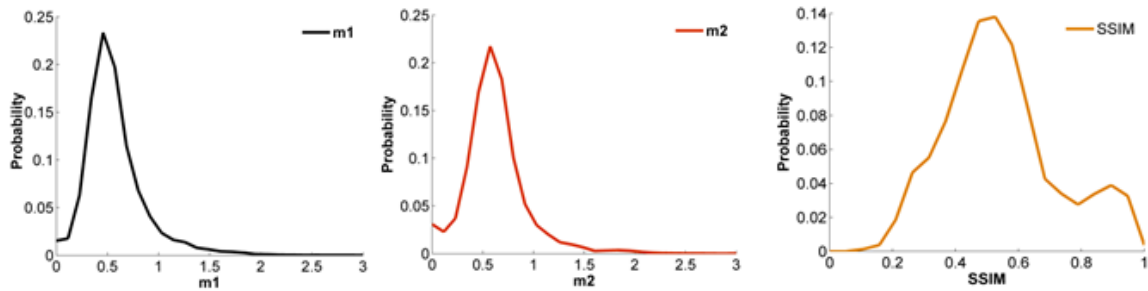


Figure S2 Histograms constructed of all values of $m1$ (Left), $m2$ (Middle) and $SSIM$ (Right) values, taken together for all the different activation conditions. All histograms were normalized such that the area underneath them equals unity and can therefore be regarded as probability densities.

References:

- (S1) Pedregosa F. , Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 **2011** 2825-2830
- (S2) Mairal J, Bach F, Ponce J, Sapiro G. Online Learning for Matrix Factorization and Sparse Coding. *Journal of Machine Learning Research* 11 **2010** 19-60