

# 1 Supplementary Information Content

2	<b>1 Derivation of the distribution for the observed distance between two fluorophores</b>	<b>2</b>
3	<b>2 Estimating the proximity probability of a pair</b>	<b>3</b>
4	<b>3 Data-driven selection of a threshold parameter for <math>d_{\text{obs}}</math></b>	<b>4</b>
5	<b>4 Estimating the true number of molecular couplings and the number of pairings from transient proximity</b>	<b>5</b>
6	<b>5 Pipeline to estimate the number of molecule couplings in an SMLM image</b>	<b>7</b>
7	<b>6 Generating simulated SMLM datasets</b>	<b>8</b>
8	A Generating two-channel signals . . . . .	8
9	B Stochastic simulation . . . . .	8
10	C An experimental distribution of localization precisions . . . . .	9
11	<b>7 Parameters used for various analyses</b>	<b>9</b>
12	<b>8 Comparing GMO's performance against an alternative method</b>	<b>11</b>
13	A Colocalization by the Minimal Pairwise Distance Method . . . . .	11
14	B Parameters used in GMO and MinDist . . . . .	11
15	C Results from comparing GMO to MinDist . . . . .	12
16	<b>9 Experimental materials and methods</b>	<b>14</b>
17	A Cells and DNA constructs . . . . .	14
18	B Confocal imaging . . . . .	15
19	C SMLM imaging . . . . .	16
20	<b>10 Analysis of experimental SMLM data</b>	<b>18</b>
21	A Preprocessing experimental SMLM images . . . . .	18
22	B Analysis of experimental SMLM dataset . . . . .	20
23	C Comments on choices made in the analysis . . . . .	20
24	<b>11 Comments on analysis of experimental images</b>	<b>21</b>
25	A Halo-TM-SNAP . . . . .	21
26	B Halo- $\beta$ 2AR/SNAP-CaaX . . . . .	21
27	C The estimated number of couplings scales with the square of detection rate . . . . .	22
28	<b>12 Validation at chemical equilibrium</b>	<b>23</b>
29	<b>13 The algorithm can estimate the level of coupling accurately across different density ratios</b>	<b>24</b>
30	<b>14 Inferring rate constants from kinetics data</b>	<b>25</b>
31	A Analytical solution . . . . .	25
32	B Inferring rate constants . . . . .	25

## Supplementary Information

### Supplementary Note 1: Derivation of the distribution for the observed distance between two fluorophores

Here we derive the distribution of the observed distance between two fluorophores in 2D SMLM. We show that the normalized observed distance squared and the normalized true distance squared follow a non-central chi-square relationship:

$$\frac{d_{\text{obs}}^2}{\sigma_A^2 + \sigma_B^2} \sim \chi_2'^2 \left( \frac{d_{\text{true}}^2}{\sigma_A^2 + \sigma_B^2} \right). \quad (\text{S1})$$

Consider two spectrally distinct fluorophores  $A$  and  $B$ , whose true positions are  $(A_x, A_y)$  and  $(B_x, B_y)$  respectively. Assume their observed positions  $O_A = (O_{Ax}, O_{Ay})$  and  $O_B = (O_{Bx}, O_{By})$  are drawn from Gaussian distributions centered at their true positions. For simplicity, we assume that the localization uncertainties are radially symmetric. Let the localization precision values for  $A$  and  $B$  be respectively  $\sigma_A$  and  $\sigma_B$  (1).

From  $O_A \sim \mathcal{N}((A_x, A_y), \sigma_A^2)$  and  $O_B \sim \mathcal{N}((B_x, B_y), \sigma_B^2)$ , we rescale to obtain

$$\frac{(O_{Ax} - O_{Bx}) - (A_x - B_x)}{\sqrt{\sigma_A^2 + \sigma_B^2}} \sim \mathcal{N}(0, 1) \quad \text{and} \quad \frac{(O_{Ay} - O_{By}) - (A_y - B_y)}{\sqrt{\sigma_A^2 + \sigma_B^2}} \sim \mathcal{N}(0, 1). \quad (\text{S2})$$

Converting to chi-square distributions, we get

$$\frac{(O_{Ax} - O_{Bx})^2}{\sigma_A^2 + \sigma_B^2} \sim \chi_1'^2 \left( \frac{(A_x - B_x)^2}{\sigma_A^2 + \sigma_B^2} \right) \quad \text{and} \quad \frac{(O_{Ay} - O_{By})^2}{\sigma_A^2 + \sigma_B^2} \sim \chi_1'^2 \left( \frac{(A_y - B_y)^2}{\sigma_A^2 + \sigma_B^2} \right), \quad (\text{S3})$$

where  $\chi_k'^2(\lambda)$  is the non-central chi-square distribution with degree of freedom  $k$  and noncentrality parameter  $\lambda$ . Note that the noncentrality parameter  $\lambda_{\sigma_A, \sigma_B, d_{\text{true}}}$  depends on  $\sigma_A$ ,  $\sigma_B$  as well as  $d_{\text{true}}$ . Then the square of the observed distance, normalized by  $\sigma_A^2 + \sigma_B^2$ , follows a non-central chi-square distribution with degree of freedom  $k = 2$  (2):

$$\frac{d_{\text{obs}}^2}{\sigma_A^2 + \sigma_B^2} = \frac{(O_{Ax} - O_{Bx})^2}{\sigma_A^2 + \sigma_B^2} + \frac{(O_{Ay} - O_{By})^2}{\sigma_A^2 + \sigma_B^2} \sim \chi_2'^2 \left( \frac{(A_x - B_x)^2}{\sigma_A^2 + \sigma_B^2} + \frac{(A_y - B_y)^2}{\sigma_A^2 + \sigma_B^2} \right) = \chi_2'^2 \left( \frac{d_{\text{true}}^2}{\sigma_A^2 + \sigma_B^2} \right). \quad (\text{S4})$$

Fig. S1(a-b) show examples of the distributions of the normalized observed distance.

From Fig. S1, we see that as localization precision  $\sigma_i$  increases, the mean of  $d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2}$  decreases. As  $d_{\text{true}}$  increases, the mean of  $d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2}$  also increases, and the distribution starts to resemble a Gaussian. For any fixed  $\sigma_A$  and  $\sigma_B$ , the monotonic dependence of  $\mathbb{E}[d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2}]$  on the normalized true distance (Fig. S1(c)) can be obtained using the exact formula for the mean of a non-central chi distribution (3). For completeness: the mean of  $\chi_k'(\lambda)$  is given by

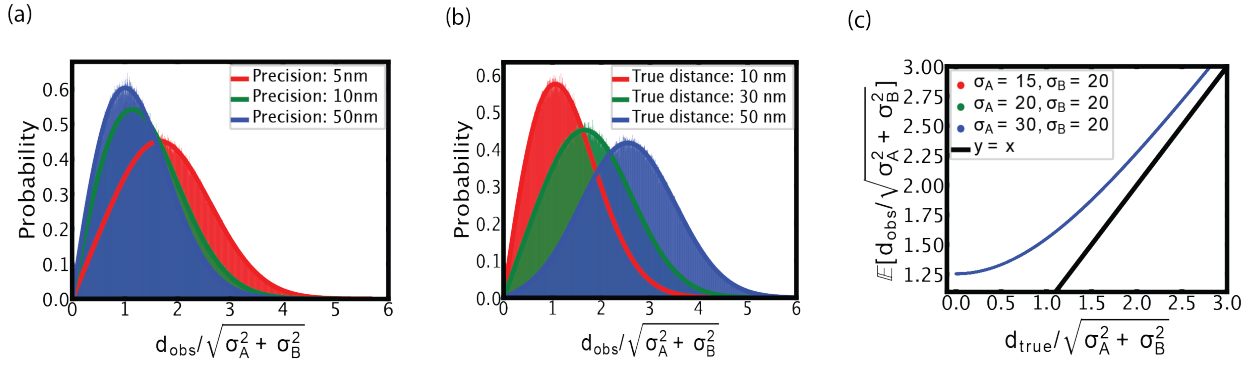
$$\mathbb{E}\chi_k'(\lambda) = \sqrt{2} \exp\left(-\frac{\lambda}{2}\right) \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \text{hyp1f1}\left(\frac{k+1}{2}, \frac{k}{2}, \frac{\lambda}{2}\right)$$

where  $\Gamma(z)$  is the gamma function,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt,$$

and hyp1f1 is the confluent hypergeometric function 1F1,

$$\text{hyp1f1}(a, b, x) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)/\Gamma(a)}{\Gamma(b+j)/\Gamma(b)j!} x^j.$$



**Figure S1.** Theoretical distributions and histograms ( $10^6$  sampled points) of  $d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2}$  simulated using Eq. (S4) for (a)  $d_{\text{true}} = 10$  nm and localization precisions  $\sigma_A = \sigma_B = 5, 10, 50$  nm, as well as (b)  $\sigma_A = \sigma_B = 15$  nm and  $d_{\text{true}} = 10, 30, 50$  nm. The simulated histograms align with the theoretical non-central chi-distributions characterized by a degree of freedom  $k = 2$ . (c) The mean of the random variable  $d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2}$  is independent of  $\sigma_A$  and  $\sigma_B$ .

## Supplementary Note 2: Estimating the proximity probability of a pair

In this section, we describe a procedure to estimate the proximity probability  $P_{\text{prox}}$  in Eq. (2) of the main text. As in the previous section, suppose emitters  $A$  and  $B$  are located at  $(A_x, A_y)$  and  $(B_x, B_y)$  respectively, with observed positions at  $O_A, O_B$ , and localization precisions  $\sigma_A, \sigma_B$ .

From Eq. (S2) above, let

$$Z_x = \frac{(O_{Ax} - O_{Bx}) - (A_x - B_x)}{\sqrt{\sigma_A^2 + \sigma_B^2}} \quad \text{and} \quad Z_y = \frac{(O_{Ay} - O_{By}) - (A_y - B_y)}{\sqrt{\sigma_A^2 + \sigma_B^2}},$$

so that  $Z_x, Z_y \sim \mathcal{N}(0, 1)$ . Then the proximity probability  $P_{\text{prox}}$  is given by

$$\begin{aligned} & \mathbb{P}(d_{\text{lower}} \leq d_{\text{true}} \leq d_{\text{upper}} | d_{\text{obs}}, \sigma_A, \sigma_B) \\ &= \mathbb{P}(d_{\text{lower}}^2 \leq |A_x - B_x|^2 + |A_y - B_y|^2 \leq d_{\text{upper}}^2 | d_{\text{obs}}, \sigma_A, \sigma_B) \\ &= \mathbb{P}(d_{\text{lower}}^2 \leq (O_{Ax} - O_{Bx} - \sqrt{\sigma_A^2 + \sigma_B^2} Z_x)^2 + (O_{Ay} - O_{By} - \sqrt{\sigma_A^2 + \sigma_B^2} Z_y)^2 \leq d_{\text{upper}}^2). \end{aligned} \quad (\text{S5})$$

The random variables in Eq. (S5) are  $Z_x$  and  $Z_y$ . We can estimate  $P_{\text{prox}}$  by generating many  $Z_x, Z_y$ , and counting how frequently the event in Eq. (S5) occurs. This process, whose pseudocode can be found in Algorithm 1, is computationally inexpensive, because it involves low-dimensional sampling. Typically  $N = 10^5$  trials is sufficient to obtain a reliable estimate of the proximity probability.

---

### Algorithm 1 Monte Carlo procedure to estimate the proximity probability of two emitters

---

**Input:**  $\sigma_A, \sigma_B, O_A, O_B, d_{\text{lower}}, d_{\text{upper}}$ , number of MC trials  $N$

**Output:** Proximity probability  $P_{\text{prox}}$

$m \leftarrow 0$

**for**  $i \leftarrow 1$  **to**  $N$  **do**

  Sample  $Z_x \sim \mathcal{N}(0, 1), Z_y \sim \mathcal{N}(0, 1)$

**if**  $d_{\text{lower}}^2 \leq (O_{Ax} - O_{Bx} - \sqrt{\sigma_A^2 + \sigma_B^2} Z_x)^2 + (O_{Ay} - O_{By} - \sqrt{\sigma_A^2 + \sigma_B^2} Z_y)^2 \leq d_{\text{upper}}^2$  **then**  
    $m \leftarrow m + 1$

**end if**

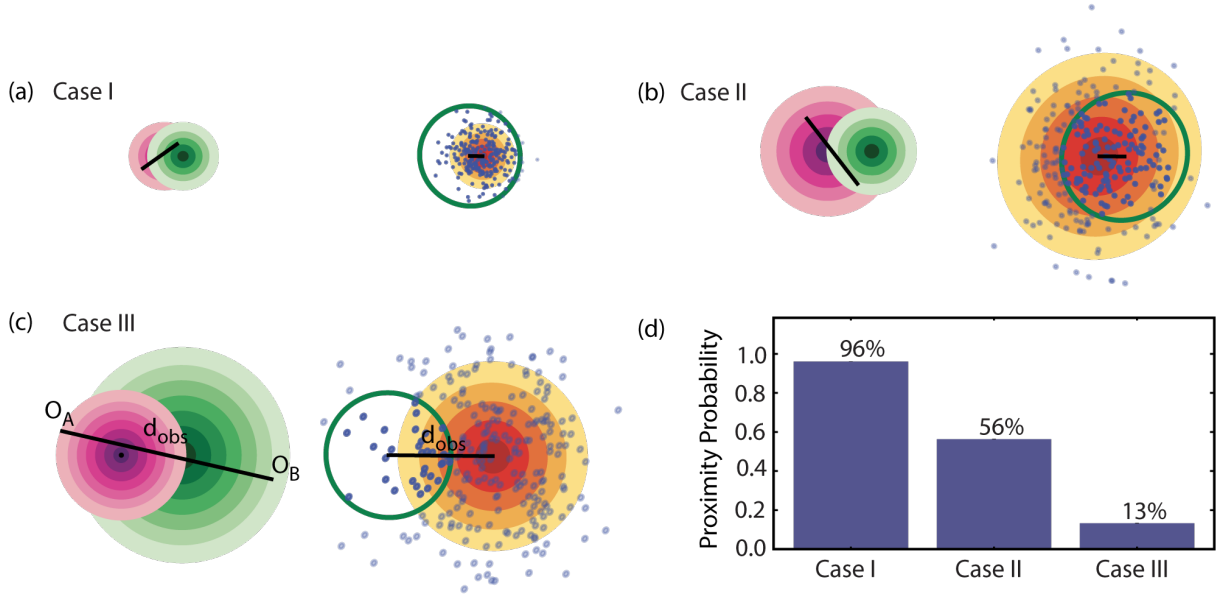
**end for**

**return**  $P_{\text{prox}} \approx \frac{m}{N}$

---

Fig. S2 shows the effect of observation parameters on estimating  $P_{\text{prox}}$ . In (a-c), the left panels show the true positions of the two emitters, with the black bar indicating  $d_{\text{obs}}$ ; the right panels show the Gaussians  $\mathcal{N}(O_A - O_B, \sigma_A^2 + \sigma_B^2)$ , and green rings marking the annulus  $0 = d_{\text{lower}} \leq r \leq d_{\text{upper}}$ . The proximity probability is maximized when most of the mass of  $\mathcal{N}(O_A - O_B, \sigma_A^2 + \sigma_B^2)$  lies within the annulus (Fig. S2(a,d)). The examples showcased in Fig. S2 are computed as follows, with all units in nm:

$$\begin{aligned}\mathbb{P}(0 \leq d_{\text{true}} \leq 20 | d_{\text{obs}} = 5, \sigma_A = \sigma_B = 5) &\approx 96\%, \\ \mathbb{P}(0 \leq d_{\text{true}} \leq 25 | d_{\text{obs}} = 10, \sigma_A = 15, \sigma_B = 10) &\approx 56\%, \\ \mathbb{P}(0 \leq d_{\text{true}} \leq 25 | d_{\text{obs}} = 50, \sigma_A = 15, \sigma_B = 25) &\approx 13\%.\end{aligned}$$



**Figure S2.** Monte Carlo procedure for estimating the proximity probability  $P_{\text{prox}}$  in three scenarios (not drawn to scale). (a-c) Pink (respectively green) rings represent the distribution for  $O_A \sim \mathcal{N}((A_x, A_y), \sigma_A^2)$  (respectively  $O_B \sim \mathcal{N}((B_x, B_y), \sigma_B^2)$ ); black bars mark observed distances  $d_{\text{obs}}$ . Yellow rings represent  $\mathcal{N}(O_A - O_B, \sigma_A^2 + \sigma_B^2)$ , from which points were drawn. The fraction of points that land inside the green ring estimates  $P_{\text{prox}}$ . (d) The approximated  $P_{\text{prox}}$  is largest when the localization uncertainties are small and the observed displacement lies in the annulus.

### 71 **Supplementary Note 3: Data-driven selection of a threshold parameter for** 72 $d_{\text{obs}}$

73 Intuitively, if  $A$  and  $B$  are extremely far away, i.e.,  $d_{\text{obs}} \gg d_{\text{true}}$ , we would not expect them to pair up. By considering  
74 only emitters that are reasonably close by, we cut down the computational time for Graph Matching Optimization  
75 (GMO). In this section, we quantify this threshold, and provide a heuristic for how to select the threshold parameter  
76 for  $d_{\text{obs}}$  based on the statistics of the localizations in a SMLM image.

77 In principle, GMO selects a matching from all possible pairs. However, using all possible pairs to construct the  
78 bipartite graph would be computationally expensive, and with diminishing return for the performance of GMO. For  
79 example, with  $N_A$  molecules  $A$  and  $N_B$  molecules  $B$ , there can be up to  $N_A N_B / 2$  possible pairs. Selecting the  
80 maximum weight matching from the input bipartite graph, as implemented in NetworkX (4; 5), is computationally  
81 expensive, with time complexity that scales as  $O(n^3)$ , where  $n$  denotes the number of nodes. Thus, when building  
82 the bipartite graph for GMO, we only consider pairs that are within a threshold distance away from each other. We do  
83 this by cutting out the thin tail of the distribution in Eq. (S4) and considering only pairs where  $d_{\text{obs}} / \sqrt{\sigma_A^2 + \sigma_B^2} \leq C$   
84 for some thresholding parameter  $C$ .

85 This truncation of the tail has minimal effect on the distribution. For example, for the distributions in Fig. S1(a),  
86 choosing  $C = 4$  captures the bulk of the distributions:

$$\begin{aligned}\mathbb{P}(d_{\text{obs}} / \sqrt{\sigma_A^2 + \sigma_B^2} \leq 4 | d_{\text{true}} = 10, \sigma_A = \sigma_B = 5) &= 99.13\%, \\ \mathbb{P}(d_{\text{obs}} / \sqrt{\sigma_A^2 + \sigma_B^2} \leq 4 | d_{\text{true}} = 10, \sigma_A = \sigma_B = 10) &= 99.87\%, \\ \mathbb{P}(d_{\text{obs}} / \sqrt{\sigma_A^2 + \sigma_B^2} \leq 4 | d_{\text{true}} = 10, \sigma_A = \sigma_B = 50) &= 99.96\%.\end{aligned}$$

87 Similarly, for the distributions in Fig. S1(b):

$$\mathbb{P}(d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2} \leq 4 | d_{\text{true}} = 10, \sigma_A = \sigma_B = 15) = 99.93\%,$$

$$\mathbb{P}(d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2} \leq 4 | d_{\text{true}} = 30, \sigma_A = \sigma_B = 15) = 99.13\%,$$

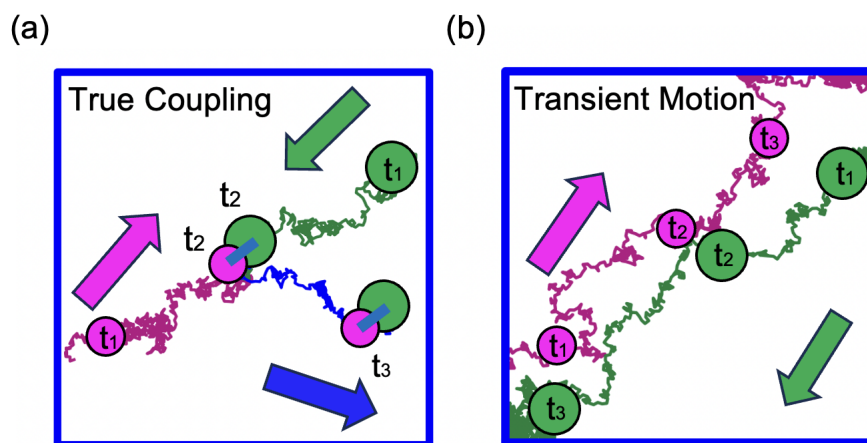
$$\mathbb{P}(d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2} \leq 4 | d_{\text{true}} = 50, \sigma_A = \sigma_B = 15) = 93.05\%.$$

88 Clearly, we cannot blindly choose  $C = 4$  for any SMLM image. To choose  $C$ , one must first estimate  $d_{\text{true}}$  for the  
89 two interacting molecule partners. Such an estimate may be obtained from fitting the distributions in Eq. (S1) from  
90 experiments such as iterative localization of the identical molecular complex (6; 2; 7). Next, one would plot the  
91 distribution of localization precision values of the SMLM image, and select a threshold  $\sigma_*$  such that it captures  
92 most of the precision values, e.g.,  $\sigma_i \leq \sigma_*$  captures the 90th percentile of all  $\sigma_A, \sigma_B$  from the dataset. One can  
93 then simulate the distribution of  $d_{\text{obs}}/\sqrt{2\sigma_*^2}$  according to Eq. (S4). From the resulting distribution similar to those in  
94 Fig. S1(a-b),  $C$  can be chosen such that  $\{d_{\text{obs}}/\sqrt{\sigma_A^2 + \sigma_B^2} \leq C\}$  captures the bulk of the distribution.

95 In principle, a larger value of  $C$  would capture more possible pairings. However, small increases in  $C$  might  
96 significantly increase the computational time due to the increased density of the graph. Choices must be made  
97 based on the dataset being analyzed, and the goal of the analysis.

## 98 Supplementary Note 4: Estimating the true number of molecular couplings 99 and the number of pairings from transient proximity

100 Because SMLM captures the observed positions of emitters at a fixed time, we cannot distinguish between molecular  
101 coupling and transient proximity. For example, if the SMLM image is a snapshot of Fig. S3 at  $t_2$ , we cannot  
102 distinguish between the two scenarios. We devised a process we call Iterative Monte Carlo Estimation of Molecular  
103 Couplings and Background Pairings (iMEC) as part of our colocalization algorithm, to estimate how many pairs are  
104 true couplings in a given SMLM dataset, and how many are due to transient proximity. More precisely, given  $N_A$   
105 and  $N_B$  emitters, of which there are  $N_{\text{pairs}}^*$  pairings identified by GMO, Algorithm 2 estimates how many of the  $N_{\text{pairs}}^*$   
106 pairings are transient, in the sense of Fig. S3(b). Thus, the number of true molecular couplings is what remains after  
107 subtracting the transient pairs.



**Figure S3.** Stochastic Brownian trajectories of two interacting partners. (a) Coupling occurs between two molecules at time  $t_2$ , after which they proceed together until  $t_3$  along the trajectory in blue. (b) Two molecules experience transient proximity during diffusion. A measurement only at time  $t_2$  cannot distinguish between the two scenarios.

108 We assume a system where  $A$  and  $B$  diffuse freely. If a molecule of  $A$  and a molecule of  $B$  are sufficiently close, they  
109 are capable of binding (Fig. S3(a)). However, if the system does not pass the activation energy barrier for the binding  
110 reaction to occur, the two molecules may simply diffuse away at a later time (Fig. S3(b)). A comprehensive stochastic  
111 model elucidating the pivotal role of colocalization in determining the probability of protein-protein interactions is  
112 detailed in the works (8; 9) by Batada et al..

113 In this work, we take a slightly different approach. In iMEC (Algorithm 2), we successively estimate the number  
 114 of pairings among molecules that we assume cannot interact, with a smaller number of molecules at the start of  
 115 each iteration. The inputs to Algorithm 2 are the number of signals corresponding to  $A$  ( $N_A$ ), the number of signals  
 116 corresponding to  $B$  ( $N_B$ ), the number of pairs identified by GMO ( $N_{\text{pairs}}^*$ ), and the area of the region being analyzed  
 117 ( $S$ ). In addition we specify the number of iterations per trial ( $T$ ), and the number of trials ( $M$ ).

118 In any given trial, for the first iteration, we assume all  $N_A$  and  $N_B$  molecules are non-interacting ( $N_{\text{coupled}}^0 = 0$ ), and  
 119 estimate the number of pairs  $N_{\text{bg}}^1$ . All pairs identified at this step are considered background pairs since we assumed  
 120 non-interacting particles. Then we set  $N_{\text{coupled}}^1 = N_{\text{pairs}}^* - N_{\text{bg}}^1$ . For the second iteration, since we assume there are  
 121 at least  $N_{\text{coupled}}^1$  coupled pairs, we repeat the process by assuming that the remaining  $N_A - N_{\text{coupled}}^1$  molecules of  
 122  $A$  and  $N_B - N_{\text{coupled}}^1$  molecules of  $B$  are non-interacting. After each iteration, we update the estimated number of  
 123 background pairs  $N_{\text{bg}}^i$ , and thus obtain an estimated number of true couplings  $N_{\text{coupled}}^i = N_{\text{pairs}}^* - N_{\text{bg}}^i$ . We stop the  
 124 trial after  $T$  iterations, and obtain estimates of  $N_{\text{bg}}^T$  and  $N_{\text{coupled}}^T$ . The output of Algorithm 2 is the average over  $M$   
 125 trials:  $N_{\text{coupled}}^* \approx \langle N_{\text{coupled}}^T \rangle$ .

126 An example of  $\langle N_{\text{bg}}^t \rangle$  and  $\langle N_{\text{coupled}}^t \rangle$  is shown in Fig. ?? of the main text. In the main text, we suppress the notation  
 127 for multiple MC trials for simplicity, i.e.,  $N_{\text{bg}}^T$  in the main text is in actuality  $\langle N_{\text{bg}}^T \rangle$ , and likewise for the number of  
 128 molecular couplings.

129 In principle, it is possible that  $N_{\text{bg}}^i > N_{\text{pairs}}^*$  at the end of an iteration, so  $N_{\text{coupled}}^i < 0$  by definition. This reflects  
 130 the situation where there are more background pairs expected than those identified by GMO. In such a case, the  
 131 next iteration starts with  $N_A - N_{\text{coupled}}^i$  and  $N_B - N_{\text{coupled}}^i$  molecules of  $A$  and  $B$  respectively. Note that here we  
 132 instantiate more molecules than in the dataset. If this occurs at the last iteration, we set  $N_{\text{coupled}}^T = 0$ . However, this  
 133 is unlikely to occur in a dataset for which molecular couplings are expected.

---

**Algorithm 2** Iterative Monte Carlo procedure to estimate the number of molecular couplings.

---

**Input:**  $N_A, N_B, N_{\text{pairs}}^*$ , area  $S$ , number of iterations  $T$  per trial, number of MC trials  $M$ .

**Output:**  $\langle N_{\text{bg}}^T \rangle, \langle N_{\text{coupled}}^T \rangle$

$S_{\text{bg}} \leftarrow 0$

$S_{\text{coupled}} \leftarrow 0$

**for**  $m \leftarrow 1$  **to**  $M$  **do**

$N_{\text{coupled}}^0 \leftarrow 0$

**for**  $i \leftarrow 1$  **to**  $T$  **do**

        Simulate  $N_A - N_{\text{coupled}}^{i-1}$  and  $N_B - N_{\text{coupled}}^{i-1}$  particles over area  $S$  uniformly random.

        Use pairing algorithm to find the most probable graph matching to estimate  $N_{\text{bg}}^i$ .

$N_{\text{coupled}}^i \leftarrow N_{\text{pairs}}^* - N_{\text{bg}}^i$

**end for**

$S_{\text{bg}} \leftarrow S_{\text{bg}} + N_{\text{bg}}^T$

$S_{\text{coupled}} \leftarrow S_{\text{coupled}} + N_{\text{coupled}}^T$

**end for**

$\langle N_{\text{bg}}^T \rangle \leftarrow S_{\text{bg}}/M$

$\langle N_{\text{coupled}}^T \rangle \leftarrow S_{\text{coupled}}/M$

**if**  $\langle N_{\text{coupled}}^T \rangle < 0$  **then**

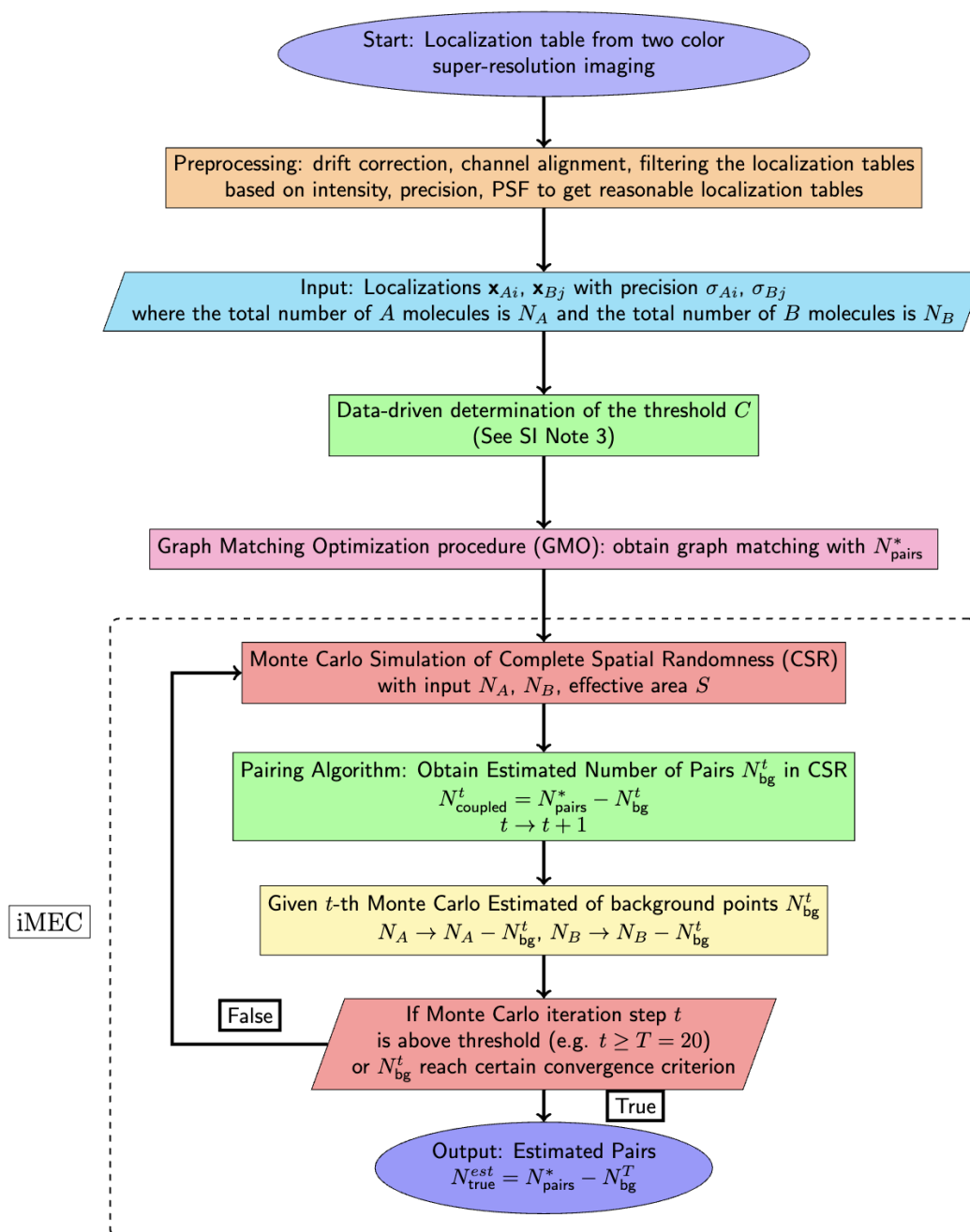
    Set  $\langle N_{\text{coupled}}^T \rangle = 0$

**end if**

**return**  $\langle N_{\text{bg}}^T \rangle, \langle N_{\text{coupled}}^T \rangle$

---

134 **Supplementary Note 5: Pipeline to estimate the number of molecule**  
 135 **couplings in an SMLM image**



**Figure S4.** Full pipeline to detect molecular pairings and count the number of true couplings in a given SMLM localization table. The core components include a graph matching optimization procedure (GMO) and an iterative Monte Carlo estimation of molecular couplings (iMEC).

## 136 Supplementary Note 6: Generating simulated SMLM datasets

137 Many of our results were based on simulated SMLM datasets. To generate a simulated dataset, we instantiated  
138 the positions of molecules  $A$ ,  $B$ , and  $AB$  either from a uniform distribution, or from a stochastic reaction-diffusion  
139 simulation.

### 140 A. Generating two-channel signals

141 To generate the localizations in the two channels from the positions of  $A$ ,  $B$ , and  $AB$ , we kept the localizations  
142 for  $A$  and  $B$ , and split each  $AB$  into two localizations. For each  $AB$  located at  $(x, y)$ , we sampled from the  
143 uniform distribution a stochastic distance  $d_{\text{obs}} \sim \text{Unif}(0, 10 \text{ nm})$ , and an angle  $\theta \sim \text{Unif}(0, 2\pi)$ . We placed signals  
144 corresponding to  $A$  and  $B$  at

$$\left(x + \frac{d_{\text{obs}}}{2} \cos \theta, y + \frac{d_{\text{obs}}}{2} \sin \theta\right) \quad \text{and} \quad \left(x - \frac{d_{\text{obs}}}{2} \cos \theta, y - \frac{d_{\text{obs}}}{2} \sin \theta\right)$$

145 respectively. For the localization uncertainty of each signal, we either set it as a constant (e.g., for Fig. 4 of the main  
146 text; see Tables S3), or drew from the distribution in Fig. S6.

### 147 B. Stochastic simulation

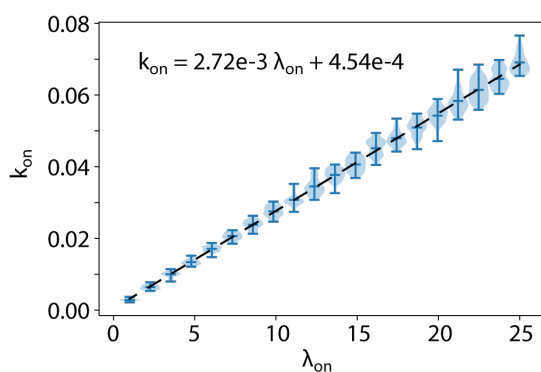
148 The stochastic simulations reported in Fig. 6 of the main text were generated using a particle-based  
149 reaction-diffusion model, implemented using the package ReaDDY, version 2.0.12 (10).

150 We set the simulation box to be  $\sqrt{\text{Area}} \times \sqrt{\text{Area}} \times 1 \text{ nm}$ , with periodic boundaries. This effectively simulates the  
151 system as if it is 2-dimensional for  $\text{Area} \gg 1$ . Diffusion is modeled by  $\frac{d\vec{x}_i(t)}{dt} = \sqrt{2D_i}\vec{\xi}_i(t)$ , where  $\vec{x}_i(t)$  is the position  
152 of particle  $i$ , with diffusion coefficient  $D_i$ , and  $\{\vec{\xi}_i\}_i$  are independent Gaussian random variables.

153 Each specified reaction occurs with probability  $p(\lambda; dt) = 1 - e^{-\lambda dt}$ , where  $\lambda$  plays a similar role to a rate constant.  
154 For a reaction involving a single reactant, e.g.,  $AB \rightarrow A + B$ , the propensity constant is given by the macroscopic  
155 rate constant:  $\lambda_{\text{off}} = k_{\text{off}}$ . For binary reactions, e.g.,  $A + B \rightarrow AB$ , a reaction would occur with probability  $p(\lambda; dt)$   
156 only if the reacting molecules,  $A$  and  $B$  in this case, are no more than  $d_{\text{educt}}$  apart, where the educt distance  $d_{\text{educt}}$   
157 is a chosen parameter. For this simulation, we chose  $d_{\text{educt}} = 30 \text{ nm}$ . Note that for binary reactions,  $k_{\text{on}} \neq \lambda_{\text{on}}$  in  
158 general. For  $d_{\text{educt}} = 30 \text{ nm}$ , the relationship for  $\lambda_{\text{on}} \leq 25 \text{ s}^{-1}$  was numerically determined to be

$$k_{\text{on}} = 2.72 \times 10^{-3} \lambda_{\text{on}} + 4.54 \times 10^{-4},$$

159 as shown in Fig. S5. For  $\lambda_{\text{on}} = 20 \text{ s}^{-1}$ , the corresponding rate constant is  $k_{\text{on}} = 0.0549 \mu\text{m}^2 \text{molecule}^{-1} \text{s}^{-1}$ .



**Figure S5.** Conversion between the rate constant  $k_{\text{on}}$  and the propensity constant  $\lambda_{\text{on}}$  for  $A + B \rightarrow AB$ , at  $d_{\text{educt}} = 30 \text{ nm}$ .

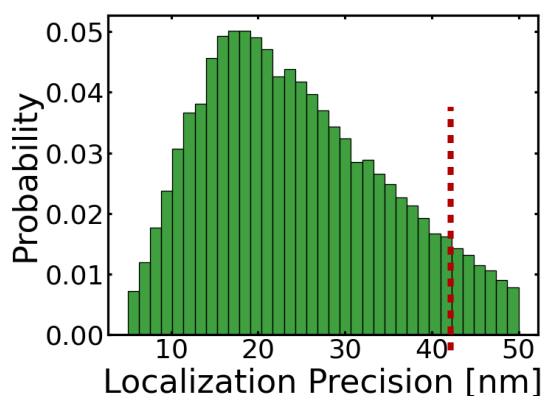
160 We simulated the system with initially only  $A$  and  $B$ , but no  $AB$ . At the densities stated in Fig. 6(c) (pink  $30 A/\mu\text{m}^2$ ,  
161  $20 B/\mu\text{m}^2$ ; blue  $20 A/\mu\text{m}^2$ ,  $20 B/\mu\text{m}^2$ ; yellow  $10 A/\mu\text{m}^2$ ,  $20 B/\mu\text{m}^2$ ), the particles were uniformly distributed in  
162 the simulation box. We ran the simulation, and extracted the positions of  $A$ ,  $B$ , and  $AB$  at various time points. The  
163 parameters for the stochastic simulations are presented in Table S1.

**Table S1.** Stochastic Simulation Parameters in Fig. 6

Timestep $dt$	$0.5 \times 10^{-3}$ s
Total Simulation Time	3 s
Area	$100 \mu\text{m}^2$
$\lambda_{\text{on}}$	$20 \text{ s}^{-1}$
$\lambda_{\text{off}}$	$1 \text{ s}^{-1}$
$d_{\text{educt}}$	30 nm
Diffusion coefficient $D_A$	$0.3 \mu\text{m}^2/\text{s}$
Diffusion coefficient $D_B$	$0.3 \mu\text{m}^2/\text{s}$
Diffusion coefficient $D_{AB}$	$0.01 \mu\text{m}^2/\text{s}$
Densities of $(A, B)$	(30,20), (20,20) or (10,20) molecules/ $\mu\text{m}^2$

### 164 C. An experimental distribution of localization precisions

165 The average experimental precision in our datasets typically ranged from 15 to 30 nm. In several of our simulated  
 166 SMLM datasets, we used localization precision values drawn from the experimental distribution in Fig. S6, with  
 167  $\sigma_i \leq 40$  nm to model localizations with moderate to high signal-to-noise ratios. We found that the localization precision  
 168 distributions between different cells (under the same experimental conditions) were remarkably similar. Fig. S6 is  
 169 representative of the typical distributions observed; experimental protocol can be found in Supplementary Note 9.



**Figure S6.** An experimental localization precision distribution with the peak at approximately 20 nm. The cutoff localization precision used in all simulations was 40 nm.

### 170 Supplementary Note 7: Parameters used for various analyses

171 For Figures 3-6 in the main text, to produce the results, we ran GMO and iMEC. This section consists of tables of  
 172 parameters used in the analysis.

**Table S2.** Simulation Parameters for Fig. 3(b-c)

Area $S$	$100 \mu\text{m}^2$
Number of AB	1545 or 1100
Number of background A $N_A$	455 or 900
Number of background B $N_B$	455 or 900
True distance $d_{\text{true}}$	Sampled from Unif(0, 10 nm)
Localization Precision $\sigma_A, \sigma_B$	Sampled from Fig. S6
$d_{\text{lower}}$	0 nm
$d_{\text{upper}}$	20 nm
Search distance factor threshold $C$ from SN 3	4
Number of sampling points to estimate probability $N$ from Algorithm 1	$10^5$
Number of iterations $T$ to estimate $N_{\text{bg}}^*$ from Algorithm 2	15
Number of MC trials $M$ in each estimation of $N_{\text{bg}}^*$ in Algorithm 2	10

**Table S3.** Simulation Parameters in Fig. 4

	Fig. 4(a)	Fig. 4(b)
Density $\rho$	$\rho = 5, 10$ or $20$ (see plot)	Varying (see $x$ -axis)
Area $S$	$S = 100 \mu\text{m}^2$	$100 \mu\text{m}^2$
Number of AB	$\rho S/2$	$\rho S/2$
Number of background A	$\rho S/2$	$\rho S/2$
Number of background B	$\rho S/2$	$\rho S/2$
True distance $d_{\text{true}}$	Sampled from Unif(0,10 nm)	Sampled from Unif(0,10 nm)
Localization Precision $\sigma_A, \sigma_B$	Varying (see $x$ -axis)	$\sigma_A = \sigma_B = 10, 20$ or $30$ nm (see plot)
$d_{\text{lower}}$	0 nm	0 nm
$d_{\text{upper}}$	20 nm	20 nm
Search distance factor threshold $C$ from SN 3	4	4
Number of sampling points $N$ to estimate probability from Algorithm 1	$10^5$	$10^5$
Number of iterations $T$ to estimate $N_{\text{bg}}^*$ from Algorithm 2	25	25
Number of MC trials $M$ in each estimation of $N_{\text{bg}}^i$ in Algorithm 2	10	10

**Table S4.** Simulation Parameters in Fig. 5(d)

Percentage of couplings $c$	$c = 0\%, 6\%$ or $21\%$
Density $\rho_A$	$\rho_A = 5 \mu\text{m}^{-2}$
Density $\rho_B$	$\rho_B = 5, 10, 15 \mu\text{m}^{-2}$
Area $S$	$S = 100 \mu\text{m}^2$
Number of AB	$\rho_A S c$
Number of background A	$\rho_A S (1 - c)$
Number of background B	$\rho_B S (1 - c)$
$K_{\text{eq}}$	$\min(\rho_A, \rho_B) c / (1 - c)^2 \rho_A \rho_B$
True distance $d_{\text{true}}$	Sampled from Unif(0,10 nm)
Localization Precision $\sigma_A, \sigma_B$	Sampled from Fig. S6
$d_{\text{lower}}$	0 nm
$d_{\text{upper}}$	20 nm
Search distance factor threshold $C$ from SN 3	4
Number of sampling points to estimate probability $N$ from Algorithm 1	$10^5$
Number of iterations $T$ to estimate $N_{\text{bg}}^*$ from Algorithm 2	25
Number of MC trials $M$ in each estimation of $N_{\text{bg}}^i$ in Algorithm 2	10

**Table S5.** Simulation Parameters in Fig. 6(c-d)

Density $\rho$	$\rho_A = 30, 20, 10, \rho_B = 20 \mu\text{m}^{-2}$
Area $S$	$100 \mu\text{m}^2$
Number of AB at time $t$	From stochastic simulation
Number of background A at time $t$	From stochastic simulation
Number of background B at time $t$	From stochastic simulation
True distance $d_{\text{true}}$	Sampled from Unif(0,10 nm)
Localization Precision $\sigma_A, \sigma_B$	Sampled from Fig. S6
$d_{\text{lower}}$	0 nm
$d_{\text{upper}}$	20 nm
Search distance factor threshold $C$ from SN 3	4
Number of sampling points to estimate probability $N$ from Algorithm 1	$10^5$
Number of iterations $T$ to estimate $N_{\text{bg}}^*$ from Algorithm 2	25
Number of MC trials $M$ in each estimation of $N_{\text{bg}}^i$ in Algorithm 2	10

## 173 **Supplementary Note 8: Comparing GMO's performance against an** 174 **alternative method**

175 Most colocalization methods focus on evaluating spatial correlations between the two channels, and do not directly  
176 address colocalization between two specific signals (11; 12; 13; 14; 15). In order to compare our method (GMO)  
177 against an alternative, we designed an algorithm based on the ideas in (16). We called this the *Minimal Pairwise*  
178 *Distance Method* (MinDist), as it aims to minimize distances between pairs.

### 179 **A. Colocalization by the Minimal Pairwise Distance Method**

180 MinDist sorts all pairwise distances  $d_{\text{obs}}$  (within a threshold  $d_{\text{thres}}$ ), and selects colocalized pairs starting with the  
181 smallest distances. After each pair has been selected, their corresponding signals are no longer available for further  
182 colocalization, so we delete their pairwise distances to any other localizations from the list. The pseudo-code is  
183 provided in Algorithm 3.

---

#### **Algorithm 3** Minimal Pairwise Distance Method (MinDist)

---

**Input:**  $d_{\text{thres}}$ , list of observed positions  $\{O_{A_1}, \dots, O_{A_N}\}, \{O_{B_1}, \dots, O_{B_M}\}$

**Output:** list of pairs  $C = \{(i, j)\}$  of putative colocalizations

```
 $I \leftarrow []$   
 $D \leftarrow []$   
 $C \leftarrow []$   
for  $n \leftarrow 1$  to  $N$ ,  $m \leftarrow 1$  to  $M$  do  
  if  $d_{nm} = \text{dist}(O_{A_n}, O_{B_m}) < d_{\text{thres}}$  then  
    Add  $d_{nm}$  to  $D$   
    Add  $(n, m)$  to  $I$   
  end if  
  Sort  $I$  and  $D$  by ascending order of  $D$   
  while  $I$  is not empty do  
    Add the first entry  $(n, m) \in I$  to  $C$   
    Delete all entries  $(n, j)$  from  $I$  and the corresponding entries  $d_{n,j}$  from  $D$   
    Delete all entries  $(i, m)$  from  $I$  and the corresponding entries  $d_{i,m}$  from  $D$   
  end while  
end for  
return  $C$ 
```

---

184 Although MinDist is very intuitive, it has two obvious drawbacks. 1) It does not take into account localization precision  
185 values, and 2) it is unclear how  $d_{\text{thres}}$  should be chosen. The purpose of  $d_{\text{thres}}$  is to speed up computation. As  
186  $d_{\text{thres}} \rightarrow \infty$ , all possible pairs are being considered. For finite  $d_{\text{thres}}$ , as it increases, more putative pairs are being  
187 considered, leading to increases in both true and false positives.

### 188 **B. Parameters used in GMO and MinDist**

189 We compared GMO and MinDist in two scenarios: short interaction distance ( $15 \leq d_{\text{true}} \leq 25$  nm) and long interaction  
190 distance ( $55 \leq d_{\text{true}} \leq 65$  nm). Parameters used in each scenarios are presented in Table S6.

191 As noted earlier, a larger value of  $d_{\text{thres}}$  in MinDist in general increases both the number of true positives (TP) and  
192 false positives (FP). To ensure a fair comparison, we set  $d_{\text{thres}}$  so that FP for GMO is nearly identical to that of MinDist  
193 (see right-panels of Figs. S7(a) and S8(a)).

**Table S6.** Parameters used in the comparison of GMO to MinDist

	Small interaction distance	Large interaction distance
Density $\rho$	Varying (see $x$ -axis of Fig. S7)	Varying (see $x$ -axis of Fig. S7)
Area $S$	$S = 100$	100
Number of AB	$\rho S$	$\rho S$
Number of background A	0	0
Number of background B	0	0
True distance $d_{\text{true}}$	Sampled from Unif(15, 25 nm)	Sampled from Unif(55, 65 nm)
Localization Precision $\sigma_A, \sigma_B$	Sampled from Fig. S6	Sampled from Fig. S6
$d_{\text{lower}}$	10 nm	45 nm
$d_{\text{upper}}$	30 nm	75 nm
Search distance factor threshold $C$ from SN 3	4	4
Number of sampling points to estimate probability	$10^5$	$10^5$
$d_{\text{thres}}$ in minimal pairwise distance method	85 nm	90 nm

### 194 C. Results from comparing GMO to MinDist

195 We evaluated the performance of the two methods by comparing the number of true positives per micron (TP), the  
 196 number of false positives per micron (FP), the TP ratio of GMO to MinDist, recall, and precision, where

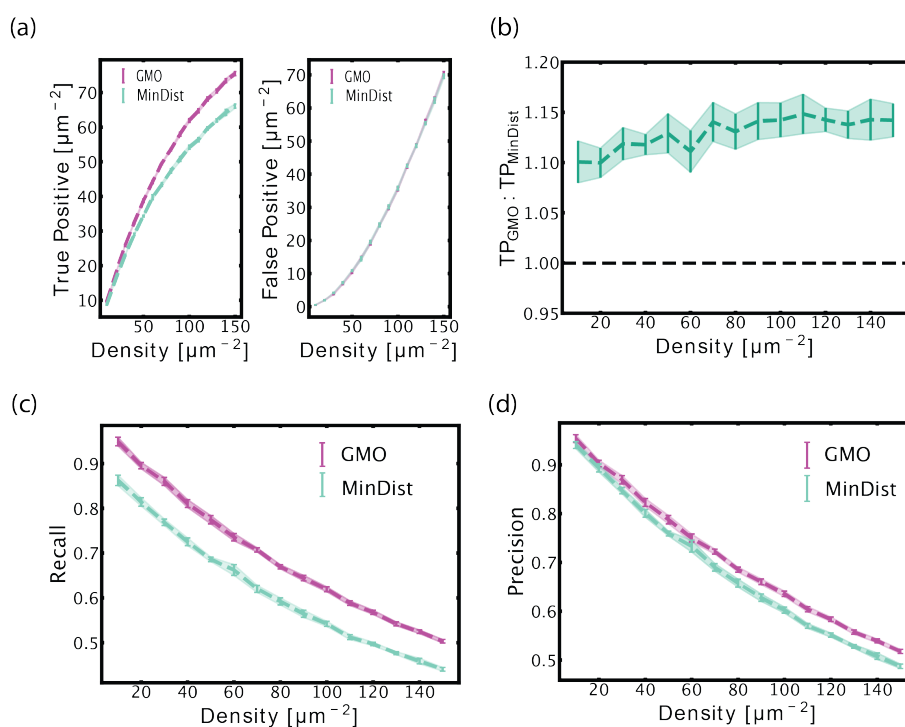
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

197 In all these metrics, GMO outperformed MinDist.

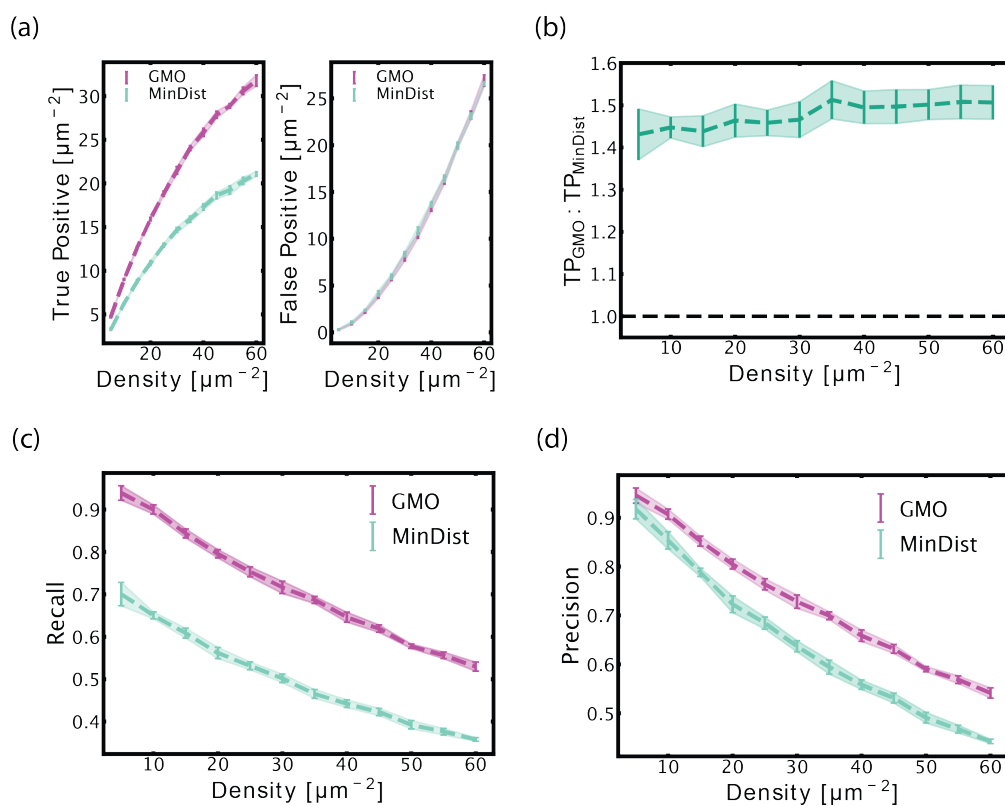
198 Consider first the case of short interaction distance, where we sampled  $d_{\text{true}}$  from [15, 25] nm uniformly. Fig. S7(a)  
 199 plots TP and FP for GMO and MinDist. We chose  $d_{\text{thres}}$  to minimize the difference in FP. GMO clearly outperformed  
 200 MinDist by at least 10%, as demonstrated by the ratio of TPs in Fig. S7(b). Moreover, this performance improved  
 201 slightly as the density of localization signals increased. Fig. S7(c-d) show recall and precision of the two algorithms.  
 202 GMO's recall was constantly  $\sim 10\%$  better than MinDist's, while precision was marginally better.

203 We also considered the scenario where the interaction distance was large. For example, (13) reported an interaction  
 204 distance of  $\sim 60$  nm between PSD95 and Homer, which are involved in the functioning of synapses. For long  
 205 interaction distance, we uniformly sampled  $d_{\text{true}}$  from [55, 65] nm. Fig. S8(a) shows TP and FP for both GMO  
 206 and MinDist, where again,  $d_{\text{thres}}$  for MinDist was chosen to minimize the difference in FP's. Here GMO performed  
 207 significantly better than MinDist at all densities, by  $\sim 40\text{-}50\%$  (Fig. S8(b)). In terms of recall, Fig. S8(c) shows  
 208 that GMO outperformed MinDist by  $\sim 25\%$  across all densities, and GMO's precision was increasingly better than  
 209 MinDist's at higher density (Fig. S8(d)).

210 At all interaction distances, GMO constantly outperformed MinDist both in the number of true positives discovered  
 211 and recall, across all densities of localizations. The outperformance was more significant for longer interaction  
 212 distances. This is unsurprising as MinDist prioritizes *shorter* pairwise distances for its colocalizations.



**Figure S7.** Performance of GMO versus MinDist at short interaction distance. (a) True Positives (TP) and False Positives (FP) discovered at different densities. (b) The ratio of TPs from the two methods, where  $\text{TP}_{\text{GMO}}:\text{TP}_{\text{MinDist}} > 1$  means GMO outperforms MinDist at identifying TPs. In terms of (c) recall and (d) precision, GMO outperformed MinDist. Error bars indicate  $\pm 1$  standard deviation over 10 simulated datasets.



**Figure S8.** Performance of GMO versus MinDist at long interaction distance. (a) True Positives (TP) and False Positives (FP) discovered at different densities. (b) The ratio of TPs from the two methods, where  $\text{TP}_{\text{GMO}}:\text{TP}_{\text{MinDist}} > 1$  means GMO does better picking up TPs. In terms of (c) recall and (d) precision, GMO outperformed MinDist. Error bars indicate  $\pm 1$  standard deviation over 10 simulated datasets.

## 213 **Supplementary Note 9: Experimental materials and methods**

### 214 **A. Cells and DNA constructs**

215 **A.1. Cell lines.** Two cell lines were used: Human Embryonic Kidney 293FT (HEK293FT) cells and HEK293 cells with  
216 a CRISPR/Cas9 knockout of the G-protein Gs (HEK293 ΔGs). HEK293 ΔGs cells were received as a gift from Dr.  
217 Asuka Inoue at Tohoku University (17; 18; 19).

218 **A.2. Cell culture and transfection.** Cells were cultured in Dulbecco's Modification of Eagle's Medium (DMEM) with 4.5  
219 g/L glucose, L-glutamine & sodium pyruvate (Corning CAT #10-013-CV). DMEM (500 mL) was supplemented with 50  
220 mL of Fetal Bovine Serum (FBS) (VWR CAT #89510-186) and 5 mL of 100x Penicillin-Streptomycin solution (Corning  
221 CAT #30-002-CI). Media was sterilized by filtration with a 0.2 μm PES bottle filter (CAT #10040-436). Culture media  
222 for live-cell confocal imaging was prepared the same way except that Phenol-Red-free FluoroBrite DMEM (Fisher  
223 Scientific CAT #A1896701) was used.

224 Cells were cultured in 25 cm<sup>2</sup>, 75 cm<sup>2</sup>, and 175 cm<sup>2</sup> treated culture flasks (VWR #10062-872, GenClone #25209,  
225 Corning #353136, VWR #10062-864) using standard sterile techniques. Passaging was done using a solution of  
226 0.05% Trypsin + 0.53 mM EDTA in Phosphate-Buffered Saline (PBS) without calcium and magnesium (Corning CAT  
227 #21-040-CV). A 10x stock of this buffer was made by diluting 2.5% trypsin (ThermoFisher #15090046) and powdered  
228 EDTA (CalbioChem CAT #4050) to 0.5% trypsin + 5.3 mM EDTA, pH = 7.4. Stock solution was frozen in 1 mL aliquots  
229 at -20°C in 15 mL conical centrifuge tubes. Before use, the 1 mL aliquot was thawed by adding 9 mL of PBS directly  
230 to the tube.

231 Confluency of 80% was targeted for optimal transfection. Cells were seeded either in 6-well or 12-well plates for  
232 transfection. Catalog numbers for 6, 12, 24, 48, and 96 well plates used at various points in these experiments  
233 are VWR CAT #10062-892, 10062-894, 10062-896, 10062-898, 10062-900. Cells were transfected using either  
234 Lipofectomene 3000 (ThermoFisher CAT #L3000008) or Polyplus JetOptimus (Polyplus-Sartorius CAT #101000006)  
235 according to manufacturer's protocols. DNA amounts were scaled according to the surface area of the well plate  
236 used. For example, the following amounts of DNA were used for a 6-well plate or a 35 mm diameter dish:

- 237 • Halo-TM-SNAP: 1 μg
- 238 • Halo-β2AR: 1 μg
- 239 • GαS-SNAP: 1 μg
- 240 • SNAP-CaaX: 2.5 μg
- 241 • Gβ: 0.5 μg
- 242 • Gγ: 0.2 μg

243 **A.3. Plasmids.** The following plasmids were used. Full circular sequence maps for all plasmids are provided as  
244 downloadable files.

- 245 • pHR-CMV-Tet02\_Twin-Strep: pHR-CMV-Tet02\_Twin-Strep was purchased from Addgene (plasmid #113833).
- 246 • Halo-TM-SNAP was received as a gift from Dr. Johannes Broichhagen at Leibniz-Forschungsinstitut for  
247 Molecular Pharmacology (20). Amino acid sequence of Halo-TM-SNAP:

248 MVPSSDPLVTAASVLEFGLGISTMETDTLLLVLLLVPGSTGDYPYDVPDYAGAQPARGSEIG  
249 TGFPPDPHYVEVLGERMHYVDVGRDGTPLVFLHGNPTSSVWRNIIPHVAPTHRCIAPDLIGMG  
250 KSDKPDLGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPVERVKGIAFMFIRPI  
251 PTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDRE  
252 PLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAV  
253 DIGPGLNLLQEDNPDIGSEIARWLSTLEISGVDEQKLISEEDLNAVGGQDTQEVIIVPHSLPFKIV  
254 VVISAILALVVLTIISLIILIMLWQKKPRGAQPARSMDKDCMKRTTLDSPLGKLELSGCEQGLH  
255 EIIIFLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVVFQSESF  
256 TRQVLWKLKLVVVFGEVISYSHLAALAGNPAATAAVKTALSGNPVPIIPCHRVRVQGDLDVGGYE  
257 GGLAVKEWLLAHEGHRGKPGGLGRLEVLFGQVD

- 258 • Halo-β2AR was purchased from Addgene (plasmid #66994) and transformed into backbone vector #113833.  
259 Amino acid sequence of Halo-β2AR:

260 MATGSRTSLLLAFLGLCLPWLQESAFPTIPLSGSEIGTGFPPDPHYVEVLGERMHYVDVGRDGT  
261 TPVFLHGNPTSSYLWRNIIPHVAPSHRCIAPDLIGMGKSDKPDLGYFFDDHVRVYLDVDFIEALGL

262 EEVVLVIHDWGSALGFHWAKRNPVERVKGIACMEFIRPIPTWDEWPEFARETFQAFRTADVGRELI  
263 IDQNAFIEGALPMGVVRPLTEVEMDHYREPFLKPVDPREPLWRFPNELPIAGEPANIVALVEAYMN  
264 WLHQSPVPKLLFWGTGPGVLIPPAEAARLAESLPNCKTVDIGPGLFLLQEDNPDIGSEIARWLP  
265 LAGSGGGSMGQPGNGSAFLAPNRSHAPDHDVTQQRDEVVVVGMGIVMSLIVLAIVFGNVLVIT  
266 AIAKFERLQTVTNYFITSLACADLRVMGLAVVPFGAAHILMKMWTFGNFWCEFWTSIDVLCVTASI  
267 ETLCVIAVDRYFAITSPFKYQSLLTKNKARVILMVVIVSGLTSFLPIQMHWRATHQEAINCYA  
268 NETCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRVFQEAQRQLQKIDKSEGRFHVQNLQSVEQD  
269 GRTGHGLRRSSKFLCKEHLKALKTGIMGFTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIGY  
270 VNSGFNPLIYCRSPDFRIAFQELLCLRRSSLKAYGNGYSSNGNTGEQSGYHVEQEKENKLLCEDL  
271 PGTEDFVGHQGTVPSDNIDSQGRNCSTNDSLL

- 272 • G $\alpha$ S-SNAP was designed in our lab and synthesized by TWIST Biosciences. The gene was sub-cloned into  
273 backbone vector #113833 using Gibson Assembly. SNAP-tag placement was chosen according to functional  
274 studies with a BRET-based sensor (21). Amino acid sequence of G $\alpha$ S-SNAP:

275 MGCLGNSKTEDQRNEEKAQREANKKIEKQLKDKQVYRATHRLLLLGAGESGKSTIVKQMRILHV  
276 NGFNEGGEEDPQAARSNSDGEKATKVDIKNNLKEAIIETIVAAMSNLVPVELANPENQFRVDY  
277 ILSVMNSGGGGTRSTGMDKDCMKRRTTLDSPGLKLELSGCEQGLHRIIFLGKGTSAADAVEVPAP  
278 AAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLKVVKFGVEISY  
279 SHLAALAGNPAATAAVKTALSGNPVPIIPCHRVVQGDLDVGGYEGGLAVKEWLLAHEGHRLGKP  
280 GLGGTATGGGGSVPDFDFPPEFYEHAKALWEDEGV RACYERSNEYQLIDCAQYFLDKIDVIKQAD  
281 YVPSDQLLRCRVLTSGLFETKQVQKVNFMFDVGGQRDERRKWIQCFNDVTAIIFVVAASSYN  
282 MVIREDNQTNRLQEALNLFKSIWNNRWRRTISVILFLNKQDLAELVLAGKSKIEDYFPEFARYT  
283 TPEDATPEPGEDPRVTRAKYFIRDEFRLISTASGDGRHYCYPHFTCAVDTENIRRVFNDCRDI IQ  
284 RMHLRQYELL

- 285 • G $\beta$  was purchased from Addgene (plasmid #133856). It was sub-cloned into backbone vector #113833 without  
286 the fluorescent tag. Amino acid sequence of G $\beta$ :

287 MSELQQLRQEAQLKNQIRDARKACADATLSQITNNIDPVGRIQMRTRRTLRLGHLAKIYAMHWGT  
288 DSRLLSVASQDGKLIWDSYTTNKVHAIPLRSSWVMTCAAPSGNYVACGGLDNICSIIYNLKTRE  
289 GNVRSRELAGHTGYLSCCRFLDNQIVTSSGDTTCALWDIETGQTTTFTGHTGDVMSLSLAPD  
290 TRLFVSGACDASAKLWDVREGMCRQFTTGHESDINAICFFPNGNAFATGSDDATCRLFDLRADQE  
291 LMTYSHDNIICGITSVSFSKSGRLLLAGYDDFNCNVWDALKADRAGVLAGHDNRVSLGVTDDGM  
292 AVATGSWDSFLKIWN

- 293 • G $\gamma$  was purchased from Addgene (plasmid #133857). It was sub-cloned into backbone vector #113833 without  
294 the fluorescent tag. Amino acid sequence of G $\gamma$ :

295 MASNNTASIAQARKLVEQLKMEANIDRIKVSAAAADLMAYCEAHAKEDPLLTPVPAENPFREKK  
296 FFCAIL

- 297 • SNAP-CaaX: pSNAP-CaaX was purchased from Addgene (plasmid #101133). Amino acid sequence of  
298 SNAP-CaaX:

299 MELHRGGGRDIKLTMDKDCMKRRTTLDSPGLKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAA  
300 VLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLKVVKFGVEISYQQ  
301 LAALAGNPAATAAVKTALSGNPVPIIPCHRVS SSGAVGGYEGGLAVKEWLLAHEGHRLGKPG  
302 GPAGCMSCKCVLS

## 303 B. Confocal imaging

304 **B.1. Fluorescent labeling of cells for confocal imaging.** We used coated glass-bottom dishes directly from the  
305 manufacturer with no additional cleaning. A list of dishes used: Matek CAT #P35G-1.5-14-C, VWR CAT #75856-742,  
306 Cellvis CAT #d35-14-1.5-n. Cells were seeded onto glass-bottom dishes and labeled with Janelia Fluor (JF)  
307 dyes. The dyes used were JF479-HaloTag, JF552-HaloTag, JF552-cpSNAP-tag, JF525-HaloTag, JF549i-SNAP-tag,  
308 JF549i-HaloTag, and JFX646-SNAP-tag. Identities of dyes used in specific experiments can be found in figure  
309 captions. Long-term, all dye stocks were stored at 500  $\mu$ M at -20°C, and at 4°C in a dark box for short-term storage.

310 For cell labeling, a dye solution in cell media (DMEM + 10% FBS + PS) was prepared at a final concentration of 2  
311  $\mu$ M. Solution volume was sufficient to fully cover the bottom surface of the dish (e.g., 500  $\mu$ L for a 35 mm dish). For

312 the labeling reaction, the dye solution was warmed to 37°C and cells were incubated with this solution at 37°C for  
313 15-20 minutes, then washed. The protocol for one wash cycle was as follows:

- 314 • Dye/DMEM solution was aspirated.
- 315 • PBS at 37°C was added gently and swirled. Enough solution was added to cover the bottom of the dish  
316 completely.
- 317 • PBS was aspirated.
- 318 • Steps 2 and 3 were repeated once.
- 319 • Fresh cell media (DMEM + 10% FBS + PS) was added and cells were placed in the incubator.

320 After a 15-minute incubation, the entire wash cycle was repeated but in the final wash step, phenol-red-free  
321 FluoroBrite DMEM was used. Next, confocal imaging was performed.

322 We found that without an incubation step between wash cycles, background dye remained inside the cell. Additional  
323 incubation time allowed this non-specifically bound dye to diffuse out of the cell, which improved imaging contrast.

324 **B.2. Confocal image acquisition.** Confocal image acquisition was performed at the Harvard Center for Biological  
325 Imaging (HCBI) using a Zeiss LSM 880 confocal microscope with a Zeiss Plan-Apochromat 63x / 1.40 NA  
326 oil DIC M27 objective (Zeiss CAT #420782-9900-000). Immersol 518F imaging oil (Microscope World CAT  
327 #444960-0000-000) was used. Images were acquired at a pixel depth of 16 bits. Images were acquired in frame  
328 scanning mode with line averaging set to 2 per pixel. Pixel dwell time was 12.20  $\mu$ s for Halo-TM-SNAP with a pixel  
329 size of 98 nm, and 16.38  $\mu$ s for Halo- $\beta$ 2AR/SNAP-CaaX with a pixel size of 88 nm. Green fluorescence was excited  
330 with a 488 nm multi-line argon laser. Red fluorescence was excited with a DPSS 561 nm laser. Far-red fluorescence  
331 was excited with a HeNe 633 nm laser. The lasers were directed to the objective via a 458/561 multi-beamsplitter  
332 for green and red excitation and a 514/561/633 multi beamsplitter for far red excitation. The following laser powers  
333 and gain settings were used to acquire images presented in figures:

334 Halo- $\beta$ 2AR/SNAP-CaaX:

335 561: laser power 0.056 mW ( $\sim$ 44 kW/cm<sup>2</sup>), gain 750, detection window 570-651 nm

336 633: laser power 0.062 mW ( $\sim$ 39 kW/cm<sup>2</sup>), gain 800, detection window 638-755 nm

337

338 Halo-TM-SNAP:

339 488: laser power 0.165 mW ( $\sim$ 173 kW/cm<sup>2</sup>), gain 811.4, detection window 490-570 nm

340 561: laser power 0.282 mW ( $\sim$ 224 kW/cm<sup>2</sup>), gain 650, detection window 566-679 nm

341

342 The pinhole was set to 1 AiryUnit. The spectral point detector used was a 32 channel GaAsP PMT for  
343 green and red emission plus the IR enhanced PMT 2 for far red emission.  
344

345 **B.3. Confocal image processing.** Confocal images were adjusted for brightness and contrast using Fiji software. All  
346 original images are available upon request.

## 347 C. SMLM imaging

348 **C.1. Preparation of mounting surfaces.** We mounted coverslips on plain pre-cleaned glass slides, 25 x 75 mm, 1 mm  
349 thick (VWR CAT #48300-026). Glass slides were cleaned with methanol (Sigma Aldrich CAT #34860-1L-R) right  
350 before mounting to remove any dust. The mounting surface was covered with methanol using a dropper pipette and  
351 wiped once in one direction with a KimWipe. Any remaining methanol quickly evaporated, and the slides were ready  
352 for mounting. We used 22 x 22 mm No. 1.5 glass coverslips (VWR CAT #48366-227). Coverslips were placed on a  
353 Wash-N-Dry coverslip rack (Sigma Aldrich CAT #Z688568) to expose both sides of the surface. The rack was placed  
354 in a glass beaker and submerged in 1 M HCl. The beaker was covered with 2 layers of aluminum foil and heated to  
355 60°C for 4-6 hours. Coverslips were then washed twice by immersing the entire coverslip rack into fresh DI water  
356 each time. Finally, coverslips were washed one additional time using 100% ethanol. The coverslips were stored in  
357 ethanol for up to 1 hour. Next, coverslips were removed from 100% ethanol and allowed to dry inside of a sterile TC  
358 hood. One coverslip was placed in each well of a 6 well tissue culture plate and 2 mL of DMEM + 10% FBS + PS  
359 were added.

360 **C.2. Fluorescent labeling.** Cells seeded in a multi-well plate were labeled with photo-activatable (PA) Janelia Fluor  
361 (JF) dyes. We used PA-JF549-HaloTag and PA-JF646-SNAP-tag dyes. Identities of dyes used in specific

362 experiments can be found in figure captions. All dyes were stored at 500  $\mu$ M at -20°C for long-term storage. In  
363 contrast to confocal dyes, SMLM dyes were not stored at 4°C.

364 For cell labeling, a dye solution in cell media (DMEM + 10% FBS + PS) was prepared at a final concentration of 2  
365  $\mu$ M. Solution volume was sufficient to fully cover the bottom surface of the dish (e.g., 500  $\mu$ L for a 35 mm dish). For  
366 the labeling reaction, the dye solution was warmed to 37°C and cells were incubated with this solution at 37°C for  
367 15-20 minutes, then washed. The protocol for one wash cycle was as follows:

- 368 • Dye/DMEM solution was aspirated.
- 369 • PBS at 37°C was added gently and swirled. Enough solution was added to cover the bottom of the dish  
370 completely.
- 371 • PBS was aspirated.
- 372 • Steps 2 and 3 were repeated once.
- 373 • Fresh cell media (DMEM + 10% FBS + PS) was added and cells were placed in the incubator.

374 After a wash cycle cells were allowed to incubate at 37°C for 15 minutes. We repeated this wash cycle with the  
375 15-minute incubation step 6 total times. Trypsin + EDTA (prepared as described in “B. Cell Culture and Transfection”)  
376 was added to detach cells. Next, cells were seeded at low confluency (~25%) onto the HCl-washed coverslips (from  
377 “C.1. Preparation of Mounting Surfaces”).

378 **C.3. Fixation buffer.** Fixation buffer was prepared using 4% w/v granular paraformaldehyde (PFA) (Electron  
379 Microscopy Sciences CAT #19208) in PBS (~50-500 mL total volume). PFA/PBS mixture was heated to 50-60°C  
380 using a heating plate while stirring with a stir bar. Initially, the solution was cloudy as not all PFA dissolved. Next, 1  
381 M NaOH in Milli-Q water was added drop-wise until the PFA/PBS solution turned clear. The solution was stirred at  
382 50-60°C until all PFA dissolved, which typically took 5-10 minutes. The pH was adjusted to 7.4 and the solution was  
383 frozen at -20°C in 10 mL aliquots in 15 mL conical centrifuge tubes. When needed, aliquots were thawed in water at  
384 room temperature and sonicated until completely clear. Once thawed, aliquots were not refrozen.

385 **C.4. Fixation protocol.** Media was aspirated from dishes with fluorescently labeled cells (from “C.2. Fluorescent  
386 Labeling”) and replaced with fixation buffer. Cells were then kept in fixation buffer at room temperature for 15  
387 minutes. During this process the dishes were swirled gently every 5 minutes. After 15 minutes of fixation, cells were  
388 washed with room-temperature PBS three times and stored at 4°C if not imaged immediately.

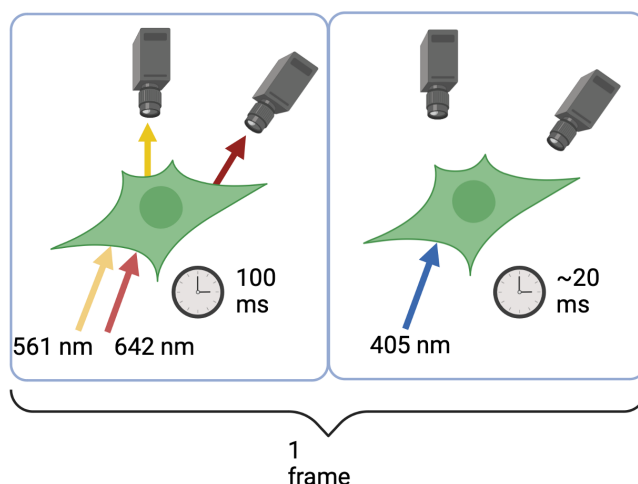
389 Immediately before imaging, PBS was gently aspirated. 0.1  $\mu$ m TetraSpek Microspheres (Thermo-Fisher CAT  
390 #T7279) were added to the glass coverslip containing cells to act as fiducial markers. The concentration from  
391 the manufacturer was reported to be  $\sim 1.8 \times 10^{11}$  particles/mL. We diluted Microspheres by a factor of 200 in  
392 Milli-Q water and sonicated before each use. This dilution factor led to the desired surface density of Microspheres  
393 and was determined empirically by serial dilution and imaging. We aimed for 5-10 Microspheres per field of view. 10  
394  $\mu$ L of diluted Microspheres were added to each coverslip. Diluted Microspheres were stored at 4°C.

395 **C.5. Mounting coverslips with cells for SMLM imaging.** To mount coverslips, a few  $\mu$ L of Fluoromount-G mounting  
396 media (ThermoFisher CAT #00-4958-02) were added to the cleaned microscope slide (from “C.1. Preparation of  
397 Mounting Surfaces”). Fluoromount-G is extremely viscous. To ensure there were no bubbles, 20  $\mu$ L of Fluoromount-G  
398 were drawn up in a pipette and a small portion was expelled. Next, a drop of Fluoromount-G was added to  
399 the microscope slide. A fresh pipette tip was then used to remove any air bubbles within the deposited drop of  
400 Fluoromount-G. Coverslips with cells facing the mounting media were mounted on top of Fluoromount-G and allowed  
401 to set for 10 minutes. After 10 minutes, the edges of the coverslips were sealed with clear nail polish and allowed to  
402 dry for  $\sim$ 10 minutes. The samples were then moved to a light-proof slide storage box. Samples were stored at 4°C.

403 **C.6. SMLM image acquisition.** SMLM image acquisition was performed on a Zeiss Elyra microscope at the HCBI with  
404 a Zeiss Plan-Apochromat 63x / 1.40 NA oil DIC M27 objective. Immersol 518F was used.

405 PA-JF549-HaloTag was excited with a DPSS 561 nm laser. 561 nm laser power measured after the objective without  
406 oil was 30.1 mW. With the objective removed, the power at the back focal plane was 40.3 mW. In parallel with 561  
407 nm excitation, PA-JF646-SNAP-tag was excited with a 642 nm HeNe laser. HeNe laser power measured after the  
408 objective without oil was 30.2 mW. With the objective removed, the power at the back focal plane was 44.0 mW.  
409 Under these illumination conditions, the on-time of a single dye molecule was between 1 and 3 frames (100-300  
410 ms). The sample was also illuminated with a 405 nm photo-activation laser between 561 nm and 641 nm excitation  
411 and collection. A schematic of a single frame of imaging is provided in Fig. S9. 405 nm laser power was increased

412 over the duration of the experiment – as localizations became less frequent – from 0.031 mW to 0.626 mW measured  
413 after the objective without oil, corresponding to 0.052 mW and 1.04 mW respectively in the back focal plane.



**Figure S9.** A single frame of SMLM illumination. Each frame has an exposure time of 100 ms for the 561 and 642 nm lasers. For approximately 20 ms during frame transfer and camera data processing, 405 nm laser light stochastically activates more fluorophores. 405 nm light is not present during camera acquisition.

414 The sample was initially illuminated with 561 nm and 642 nm laser lines (without 405 nm illumination) to photo-bleach  
415 prematurely photo-activated fluorophores. After 50-100 frames, the 405 laser was turned on.

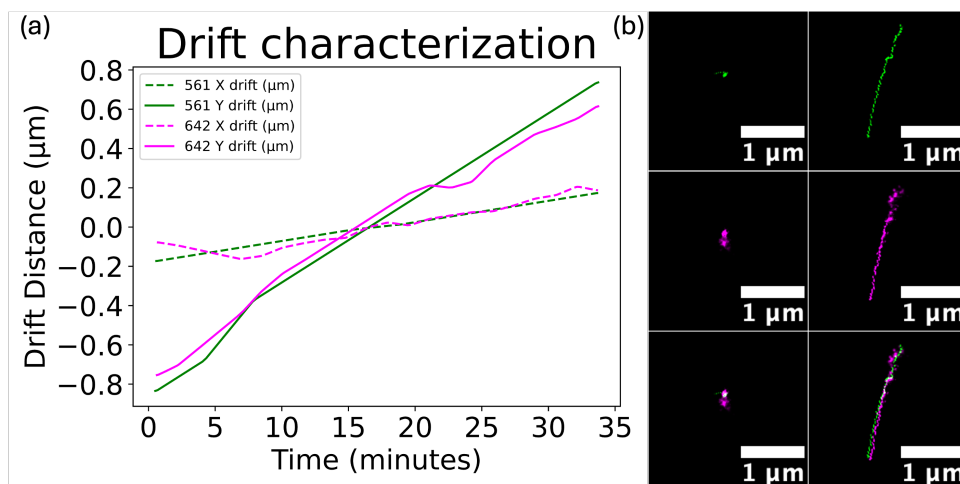
416 Emission from both dyes was collected on two spatially registered cameras. The main beamsplitter for both channels  
417 was a quad 405/488/561/642 mirror. The emission was directed to two cameras through a DuoLink unit using a  
418 band-pass 490-560 + long-pass 640 dichroic camera splitter that reflected red emission (560-640) to camera 1 and  
419 far red emission (640+) to camera 2. The cameras used were Dual PCO.edge 4.2 CLHS sCMOS camera. The frame  
420 exposure time was set to 100 ms at 16 bits. The frame size was 49x49  $\mu\text{m}$ .

## 421 **Supplementary Note 10: Analysis of experimental SMLM data**

### 422 **A. Preprocessing experimental SMLM images**

423 **A.1. Generating localizations from SMLM images.** SMLM images were processed using Zeiss Zen software to identify  
424 localizations. To select localizations, peak mask size was set to 9 pixels and fit a Gaussian mask over potential  
425 localizations. Peak intensity-to-noise ratio was set to 6, which selected among potential localizations and kept  
426 localizations if the standard deviation of its intensity against the rest of the image was above 6. Overlapping  
427 localizations (as defined by the mask of an event) were discarded. Localizations were filtered by localization precision  
428 (7 to 40 nm), number of photons (700-8000 photons) and experimental PSF size (90-180 nm). Localization precision  
429 was estimated by Zen using a Monte-Carlo Simulation of the experimental PSF including noise. These localization  
430 filters were selected based on respective localization histograms. They primarily excluded out-of-focus fluorophores  
431 and any dim fluorescent background. Spatially correlated localizations were grouped to aggregate emission of  
432 a single fluorophore that persisted over multiple camera frames. Localizations were then corrected for drift as  
433 described below.

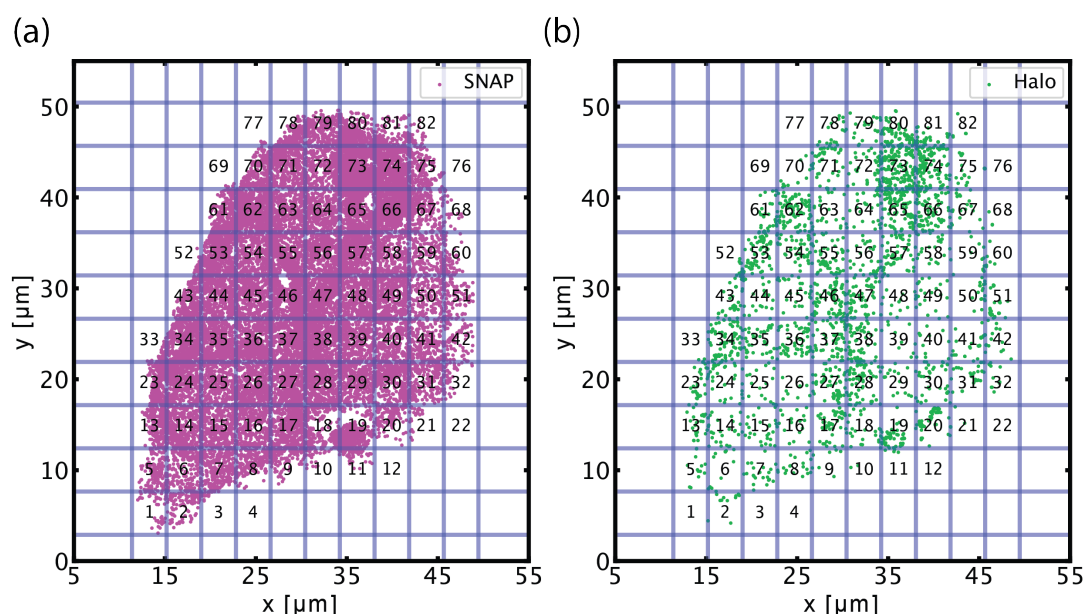
434 **A.2. Drift correction.** Typical SMLM data acquisition lasted from 20 to 40 minutes per image. To characterize drift,  
435 we imaged 100 nm TetraSpek Microspheres mounted between a glass slide and a coverslip as described above.  
436 We acquired 10,120 frames over approximately 33 minutes in two colors and observed drift on the order of several  
437 hundred nanometers to microns (Fig. S10(a)). Drift was corrected using Zen's model-based drift correction algorithm.  
438 Zoomed-in images of a bead before and after drift correction are shown in Fig. S10(b). After drift correction was  
439 applied, the full width at half maximum of the cluster of localizations was 18 nm for the green channel and 29 nm for  
440 the magenta channel.



**Figure S10.** Model based drift correction algorithm was applied to the imaging signal derived from a 100 nm TetrakSpek fluorescent Microsphere. (a) Drift in the  $x$  and  $y$  direction for both 561 and 642 nm channels. (b) Drift correction as a function of frame number was applied in each channel. Before and after images of the two channels show the drift in the two channels as well as the efficacy of the correction.

441 **A.3. Removing areas of high density and dividing into subregions.** Next, we generated a scatter plot based on the  
 442 positions of drift-corrected localization signals in each cell. In order to remove extraneous localizations outside of  
 443 the cell, we drew a boundary made of line segments to approximate the shape of the cell. The resulting polygon  
 444 defined  $\Omega$ , the effective region of analysis. Any localizations outside of  $\Omega$  were discarded from further analysis. Next,  
 445 we eliminated localization signals that were part of tight clusters, using the clustering algorithm DBSCAN (22), with  
 446 parameters  $\text{eps} = 75 \text{ nm}$  and  $\text{min\_samples} = 10$ .

447 Then we divided  $\Omega$  into smaller subregions, by overlaying  $\Omega$  with  $d \times d$  equally sized boxes, where by default  $d = 10$ .  
 448 See Fig. S11 for an example. Each subregion  $R_{ij}$  was defined as where  $\Omega$  and one of these boxes overlap, with  
 449 effective area  $S_{ij}$ . Any subregion, where in at least one of two channels the density was  $> 35 \mu\text{m}^{-2}$ , was again  
 450 discarded.



**Figure S11.** Divide into subregions by overlaying with  $d \times d$  boxes ( $d = 10$ ). (a) SNAP localizations, and (b) Halo localizations in a cell expressing Halo-TM-SNAP.

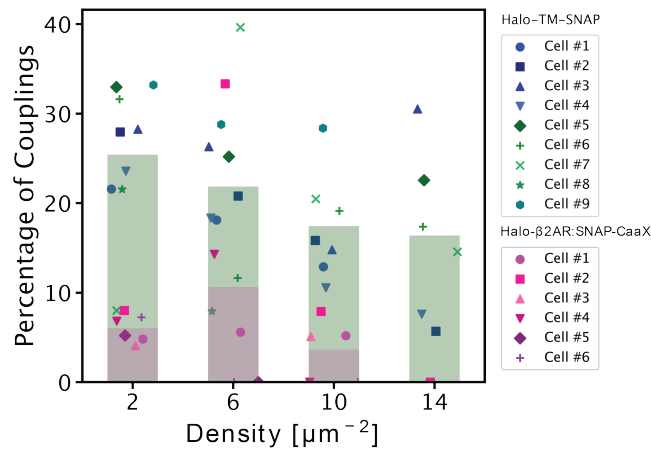
451 **B. Analysis of experimental SMLM dataset**

452 On each subregion  $R_{ij}$ , we applied our pipeline (Fig. S4) to the  $N_{A,ij} + N_{B,ij}$  localizations to estimate the number  
 453 of coupled pairs  $N_{\text{coupled},ij}$  in  $R_{ij}$ . Note that in principle, iMEC can estimate more background pairs than the number  
 454 of pairs in a subregion (i.e.,  $R_{ij}$  has less couplings than uniform random distributions of signals), in which case  
 455  $N_{\text{coupled},ij} = 0$ . We calculated the % of coupling as

$$q_{ij} = \frac{N_{\text{coupled},ij}}{\min(N_{A,ij}, N_{B,ij})}$$

456 In Fig. 5 of the main text we reported the average  $\langle q_{ij} \rangle$  for each cell.

457 The average over entire cells supported the same qualitative conclusion as from considering subregions with  
 458 similar densities, that the percentage of couplings in cells expressing Halo-TM-SNAP was significantly higher than  
 459 cells expressing Halo- $\beta$ 2AR/SNAP-CaaX. In Fig. S12, we plotted  $\langle q_{ij} \rangle$ , averaged over subregions with similar  
 460 densities. More specifically, these were subregions whose densities  $\min\{\rho_{A,ij}, \rho_{B,ij}\}$  lie in  $[0,4]$ ,  $[4,8]$ ,  $[8, 12]$ ,  
 461 or  $[12, 16] \mu\text{m}^{-2}$  respectively. The percentage  $\langle q_{ij} \rangle$  displayed larger variability between different density ranges, but  
 462 overall, the percentage of coupling is significantly higher for cells expressing Halo-TM-SNAP than those expressing  
 463 Halo- $\beta$ 2AR/SNAP-CaaX.

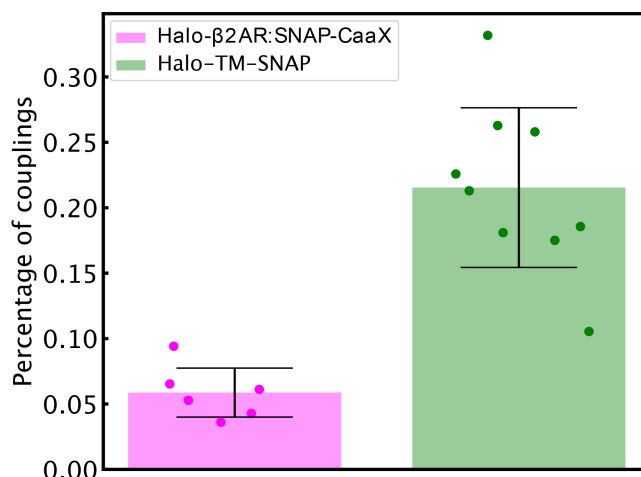


**Figure S12.** Percentage of couplings in subregions of certain densities (each marker represents a single cell). Across all densities, more couplings were identified in cells expressing Halo-TM-SNAP than those expressing Halo- $\beta$ 2AR/SNAP-CaaX.

464 **C. Comments on choices made in the analysis**

465 **On dividing into subregions.** Different local regions of the same cell spanned a range of densities. We opted to run  
 466 the analysis with  $d = 10$  as well as  $d = 1$ . With  $d = 10$  (Fig. 5 of the main text), the % of couplings for both cells  
 467 lines were nearly identical to when  $d = 1$  (Fig. S13), with 21% couplings identified for Halo-TM-SNAP and 6% for  
 468 Halo- $\beta$ 2AR/SNAP-CaaX.

469 **On removing clusters and subregions with high densities.** Clusters could have arisen for various reasons, including  
 470 overcounting or biological factors. Biological factors include cellular recycling into vesicles and endosomal pockets  
 471 that may trap non-specifically bound dye. These clusters exhibited significantly elevated local densities, typically  
 472 exceeding  $500 \mu\text{m}^2$ , which is beyond the range in which our algorithm performs well. Thus, clusters were removed  
 473 by DBSCAN prior to analysis. Similarly, subregions with densities greater than  $35 \mu\text{m}^{-2}$  were excluded from analysis  
 474 to maintain a certain level of performance.



**Figure S13.** Percentage of couplings without dividing the cell into subregions (6 cells used for Halo-β2AR/SNAP-CaaX, and 9 cells used for Halo-TM-SNAP). Error bars indicate  $\pm 1$  standard deviation.

## Supplementary Note 11: Comments on analysis of experimental images

In an ideal scenario, we would expect the percentage of couplings for Halo-TM-SNAP to be 100%, and 0% for Halo-β2AR/SNAP-CaaX. The experimental data deviated from this scenario, with only  $\sim 20\%$  of couplings identified for cells expressing Halo-TM-SNAP (Figs. 5(d), S12, S13), and in certain subregions,  $>0\%$  for Halo-β2AR/SNAP-CaaX (Fig. S12). In this section we discuss potential explanations for these observations.

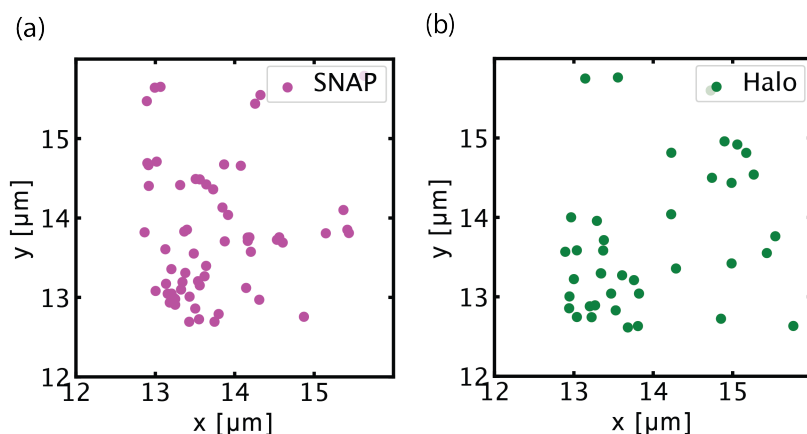
### A. Halo-TM-SNAP

The protein Halo-TM-SNAP is a transmembrane protein recombinantly fused to a HaloTag and a SNAP-tag (Fig. 5(a, right) of main text). Thus, ideally, we should have observed 100% couplings, however, not all tags were observed. The missing localizations could be ascribed to several causes. First, the labeling efficiency and the detection rate were not perfect. Second, certain localizations were excluded in the preprocessing step, based on localization precision, point spread function half-width, and what were believed to be instances of overcounting. Third, some dye molecules were activated prior to imaging, which was started after  $\sim 100$  frames.

In a positive control such as Halo-TM-SNAP, missing localizations affect the percentage of couplings identified. Consider for example, if only the HaloTag failed to be identified in a protein, the SNAP-tag that was detected might be paired with the HaloTag of a different protein by our algorithm. This situation may become too complex to parse in terms of the weighted bipartite graph in GMO. In Section C below, we showed that if molecules are detected at rate  $1 - r$  and the true number of couplings is  $N$ , then the expected number of couplings identified by our algorithm would be  $N(1 - r)^2$ . For example, a detection rate of  $r = 55\%$  would lead to a percentage of couplings  $\sim 20\%$ .

### B. Halo-β2AR/SNAP-CaaX

Halo-β2AR and SNAP-CaaX are not expected to interact. Our pipeline identified a positive percentage of couplings for certain subregions of some cells, even though the expected percentage of couplings is 0%. In these subregions, we noticed that the two proteins were not uniformly distributed in space, but there were clusters. However, these clusters were not eliminated by DBSCAN (Fig. S14), which we suspect was due to our choice of parameters ( $\text{eps}=75$  nm and  $\text{min\_samples}=10$ ). We also observed that when these clusters were present, localizations in both spectral channel were also dense in the same subregion, as demonstrated by Fig. S14.

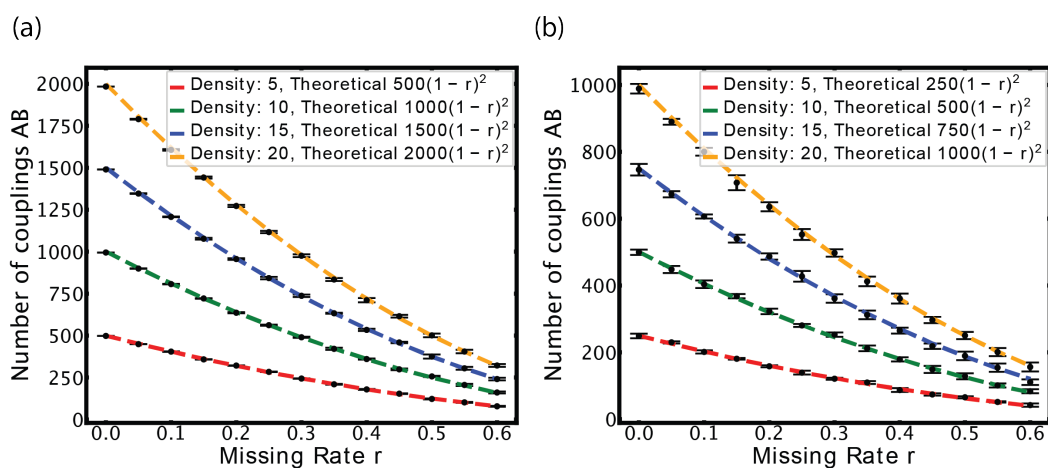


**Figure S14.** A subregion in a cell expressing Halo- $\beta$ 2AR/SNAP-CaaX. DBSCAN was not able to eliminate some clusters. The two channels also appeared to be correlated in terms of density in this subregion.

### 500 C. The estimated number of couplings scales with the square of detection rate

501 Here we considered the effect of missing detection on the percentage of couplings. In a system with  $N_{\text{coupled}}$   
 502 couplings, if localizations of  $A$  and  $B$  were dropped at rate  $r$ , i.e., a detection rate of  $1 - r$ , then in theory, the  
 503 detected number of couplings is expected to be  $N_{\text{coupled}}(1 - r)^2$ , which was the conclusion from our numerical  
 504 experiments.

505 We conducted two numerical experiments at varying densities. In the first, we modelled a positive control system  
 506 like Halo-TM-SNAP ( $N_A = N_B = N_{\text{coupled}}$ ); in the second, we assumed equilibrium between the interacting partner  
 507 and the complex ( $N_A = N_B = 2N_{\text{coupled}}$ ). In each experiment, we simulated SMLM datasets, then randomly deleted  
 508 localizations of  $A$  and  $B$  at rate  $r$ , and ran our algorithm on the resulting datasets. The parameters used can be  
 509 found in Table S7. Fig. S15 shows that in both numerical experiments, the numbers of couplings detected by our  
 510 pipeline matched the theoretical prediction  $N_{\text{coupled}}(1 - r)^2$  closely. Note that in these numerical experiments,  $N_A$   
 511 and  $N_B$  were kept constant, so the number of couplings was proportional to the percentage of couplings.



**Figure S15.** The number of couplings identified scales as  $(1 - r)^2$  where  $1 - r$  is the detection rate. (a) Positive control, where all molecules are pairs. (b) Equilibrium, where half the population are paired. Dashed Line: Theoretical Expression; Dots: Algorithm-Predicted Values. Error bars indicate  $\pm 3$  standard deviations across 10 trials of iMEC.

**Table S7.** Parameters used to generate Fig. S15

Density $\rho_A = \rho_B = \rho$	See Fig. S15
Area $S$	$S = 100 \mu\text{m}^2$
Number of AB	$\rho S$ or $\rho S/2$
Number of background A	0 or $\rho S/2$
Number of background B	0 or $\rho S/2$
True distance $d_{\text{true}}$	Sampled from Unif(0, 10 nm)
Localization Precision	Sampled from Fig. S6
$d_{\text{lower}}$	0 nm
$d_{\text{upper}}$	20 nm
Search distance factor threshold $C$	4
Number of sampling points to estimate probability	$10^5$
Number of Monte Carlo steps to estimate $N_{\text{bg}}^*$	15
Number of Monte Carlo trails in each estimation of $N_{\text{bg}}^*$	10

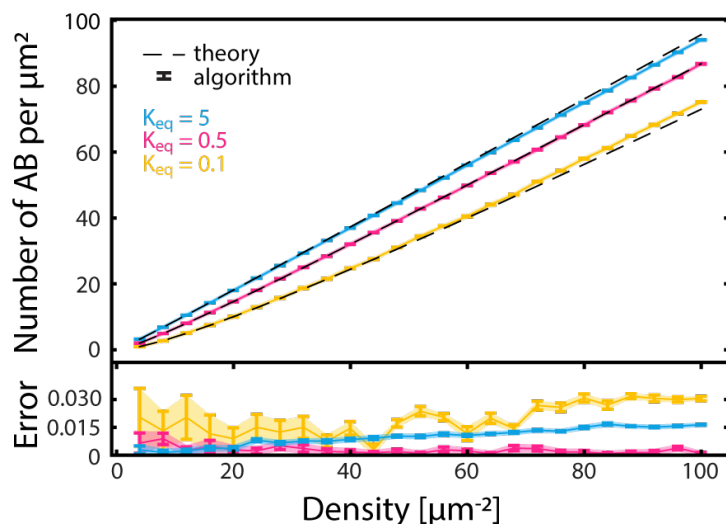
## 512 Supplementary Note 12: Validation at chemical equilibrium

513 To see whether the algorithm could be used to infer the extent of couplings at the equilibrium, we applied it to the  
 514 system  $A + B \rightleftharpoons AB$  at equilibrium with equilibrium constant  $K_{\text{eq}}$ . The percentage of couplings is proportional to  
 515 the concentration of  $AB$ , whose analytical expression can be derived:

$$[AB] := \frac{N_{AB}}{S} = \rho + \frac{1}{2K_{\text{eq}}} - \sqrt{\frac{1}{4K_{\text{eq}}^2} + \frac{\rho}{K_{\text{eq}}}}, \quad (\text{S6})$$

516 where the density  $\rho$  refers to the total amount over an area  $S$ , i.e.,  $\rho = (N_A + N_{AB})/S$  and  $\rho = (N_B + N_{AB})/S$ .  
 517 We generated simulated dataset for  $K_{\text{eq}} = 0.1, 0.5, 1$  and  $\rho$  up to 100 molecules  $\mu\text{m}^{-2}$ . All parameters used can be  
 518 found in Table S8.

519 As shown in Fig. S16, error remained below 3.5% across all densities and equilibrium constants. The error exhibited  
 520 a slight tendency to grow as density increased. Note that when the number of interacting partners  $N_{AB}$  was small  
 521 (e.g., for  $K_{\text{eq}} = 0.1$  and low density), the large error was due to a small denominator in the definition of error.



**Figure S16.** (Top) The density of  $AB$  estimated by the algorithm (markers) and the theoretical density (dashed lines) at different  $K_{\text{eq}}$  and densities. (Bottom) Error rate. Error bars indicate  $\pm 1$  standard deviation across 10 trials of iMEC.

**Table S8.** Parameters used to generate Fig. S16

Density $\rho$	Varying $\rho$ (see $x$ -axis)
Area $S$	$S = 100 \mu\text{m}^2$
Number of AB	$N_{AB}$ (from Eq. S6)
Number of background A	$\rho S - N_{AB}$ from
Number of background B	$\rho S - N_{AB}$ from )
True distance $d_{\text{true}}$	Sampled from Unif(0, 10 nm)
Localization Precision $\sigma_A, \sigma_B$	Sampled from Fig. S6
$d_{\text{lower}}$	0 nm
$d_{\text{upper}}$	20 nm
Search distance factor threshold $C$ from SN 3	4
Number of sampling points to estimate probability $N$ from Algorithm 1	$10^5$
Number of iterations $T$ to estimate $N_{\text{bg}}^*$ from Algorithm 2	25
Number of MC trials $M$ in each estimation of $N_{\text{bg}}^z$ in Algorithm 2	10

## Supplementary Note 13: The algorithm can estimate the level of coupling accurately across different density ratios

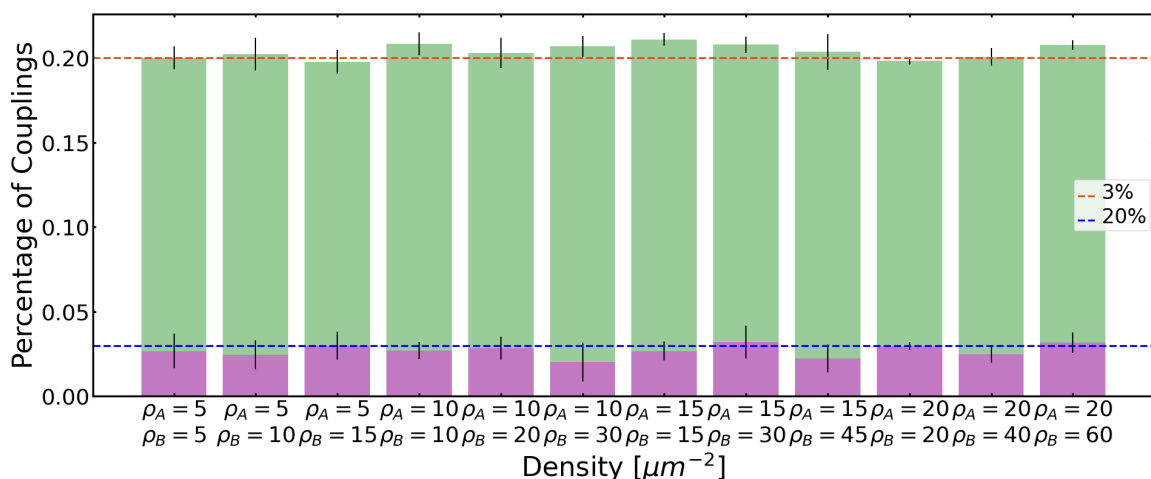
Colocalization analysis has been notably affected by density fluctuations. Measuring the degree of coupling becomes challenging when there are variations in the density ratio between  $A$  and  $B$  (23; 24; 15). A comprehensive assessment requires comparisons across diverse cells with varying densities of molecules  $A$  and  $B$ . Imagine a scenario where molecules of  $A$  (with density  $\rho_A$ ) colocalize with molecules of  $B$  (with density  $\rho_B$ ) at some percentage of colocalization. Now, if we increase the densities to  $2\rho_B$  and  $3\rho_B$  while maintaining the same percentage of couplings, an ideal colocalization algorithm should consistently yield the same level of couplings for these different density ratios. However, achieving this ideal outcome has proven challenging. Several approaches on incorporating densities to estimate the level of colocalization (using different definitions of colocalization) have been recently developed (14; 15).

Recall that we defined the percentage of couplings  $q$  as the number of coupling molecules divided by the minimum of the number of molecules of  $A$  and  $B$ . Here we would like to simulate a percentage of couplings  $\gamma$  to either be 3% (low binding) or 20% (high binding), and see whether our pipeline returns a percentage of coupling  $\langle q \rangle$  that is close to  $\gamma$  or not. In a region with area  $S$ ,  $N_A$  and  $N_B$  molecules of  $A$  and  $B$  respectively, i.e., at densities  $\rho_A = N_A/S$ ,  $\rho_B = N_B/S$ , one would expect the number of coupling to be  $N_{AB} = S\gamma \min\{\rho_A, \rho_B\}$ . Correspondingly, the number of uncoupled  $A$  would be  $N_A = (\rho_A - \gamma \min\{\rho_A, \rho_B\})S$ . A similar calculation for the number of coupled  $B$  gives  $N_B = (\rho_B - \gamma \min\{\rho_A, \rho_B\})S$ . Thus, for different combinations of  $\rho_A$  and  $\rho_B$ , we generated a dataset with  $N_A$ ,  $N_B$ , and  $N_{AB}$  molecules of  $A$ ,  $B$  and  $AB$  respectively over a box of area  $S$ . Then we applied our algorithm and calculated the percentage of coupling  $q(\rho_A, \rho_B)$ . All parameters used are shown in Table S9.

A good algorithm would return a percentage of coupling  $q(\rho_A, \rho_B)$  that is independent of  $\rho_A, \rho_B$ . Indeed, this is what we found; in Fig. S17 we plotted  $q(\rho_A, \rho_B)$  for different combinations of  $(\rho_A, \rho_B)$ , and found that  $q(\rho_A, \rho_B)$  is approximately  $\gamma$ , which was set at either 3% or 20%.

**Table S9.** Simulation Parameters in Fig. S17

Density $\rho_A, \rho_B$	Varying (see $x$ -axis)
Area $S$	$100 \mu\text{m}^2$
Percentage $\gamma$	$\gamma = 3\%$ or $20\%$
Number of AB	$S\gamma \min(\rho_A, \rho_B)$
Number of background A	$(\rho_A - \gamma \min(\rho_A, \rho_B))S$
Number of background B	$(\rho_B - \gamma \min(\rho_A, \rho_B))S$
True distance	Sampled from Unif(0,10 nm)
Localization Precision	Sampled from Fig. S6
$d_{\text{lower}}$	0 nm
$d_{\text{upper}}$	20 nm
Search distance factor threshold $C$	4
Number of sampling points to estimate probability	$10^5$
Number of Monte Carlo steps to estimate $N_{\text{bg}}^*$	15
Number of Monte Carlo trails in each estimation of $N_{\text{bg}}^z$	10



**Figure S17.** Estimated percentage of couplings  $q$  under different combinations of densities for  $A$  ( $\rho_A$ ) and  $B$  ( $\rho_B$ ), where the dataset generated were consistent with percentage of couplings  $\gamma = 3\%$  and  $20\%$  respectively. Error bars represent one standard deviation across 10 trials of iMEC.

## Supplementary Note 14: Inferring rate constants from kinetics data

We ask whether the algorithm can be used to infer kinetic parameters. We used the simulated non-equilibrium dataset in Fig. 6(d) of the main text, and the number of coupled molecules  $N_{\text{coupled}}$  from our pipeline (GMO+iMEC) to estimate the rate constants  $k_{\text{on}}$ ,  $k_{\text{off}}$  of the binding reactions  $A + B \rightleftharpoons AB$ .

### A. Analytical solution

In order to infer rate constants from our pipeline, we fitted the number of complexes  $AB$  to the solution of the differential equations modelling this system. Consider the reaction  $A + B \rightleftharpoons AB$  with rate constants  $k_{\text{on}}$  and  $k_{\text{off}}$  for the forward and backward reactions respectively. Under mass-action kinetics, with no  $AB$  initially, the concentrations evolve according to the system of differential equations

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -\frac{d[AB]}{dt} = -k_1[A][B] + k_2[AB],$$

with initial state  $[A](0) = A_0$ ,  $[B](0) = B_0$  and  $[AB](0) = 0$ . The solution is given by

$$[AB](t) = \frac{A_0 B_0 (e^{-2\sqrt{c_2} k_1 t} - 1)}{(c_1 - \sqrt{c_2}) e^{-2\sqrt{c_2} k_1 t} - c_1 + \sqrt{c_2}}, \quad (\text{S7})$$

$$[A](t) = A_0 - [AB](t),$$

$$[B](t) = B_0 - [AB](t),$$

where  $c_1 = (A_0 + B_0 + k_2/k_1)/2$  and  $c_2 = c_1^2 - A_0 B_0$ .

### B. Inferring rate constants

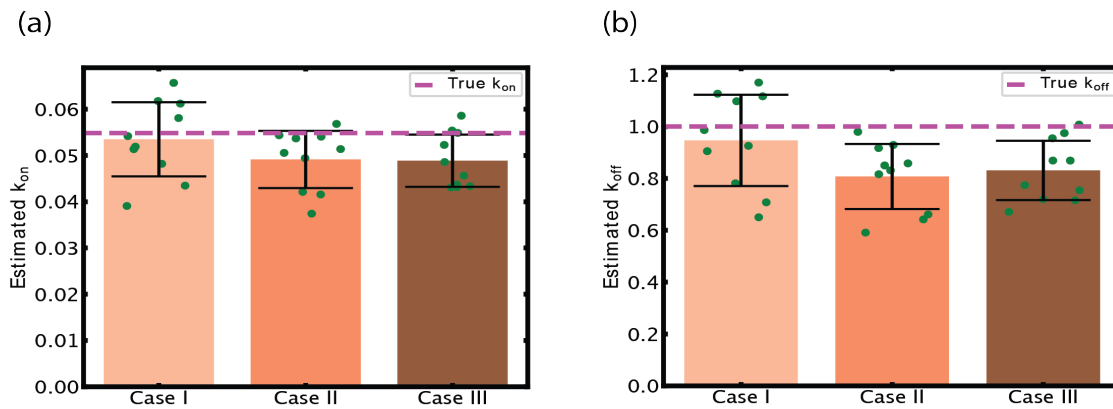
The rate constants  $k_{\text{on}}$  and  $k_{\text{off}}$  were obtained by fitting the density of  $AB$  identified ( $N_{\text{coupled}}/S$ ) at time  $t$ , where  $S$  is the area of the simulation box, to  $[AB](t)$  from Eq. (S7). Simulation and analysis parameters used can be found in Tables S1 and S5.

Compared to the true parameters  $k_{\text{on}} = 0.0549 \text{ molecules}^{-1} \mu\text{m}^2 \text{s}^{-1}$  (see Fig. S5 and Supplementary Note 6B) and  $k_{\text{off}} = 1 \text{ s}^{-1}$ , the rate constants from fitting were found to be

$$\hat{k}_{\text{on}} = \frac{0.0535 + 0.0492 + 0.0489}{3} = 0.0505 \text{ molecules}^{-1} \mu\text{m}^2 \text{s}^{-1},$$

$$\hat{k}_{\text{off}} = \frac{0.947 + 0.808 + 0.831}{3} = 0.862 \text{ s}^{-1},$$

562 where we averaged over the three densities (“cases”) in Fig. 6(d) of the main text. Fig. S18 shows the fitted rate  
 563 constants for each case separately. Note that estimation of the rate constants was more accurate at low density  
 564 (Case I); this is consistent with our observation that the performance of our algorithm degrades with increasing  
 565 density (Fig. 4(b) of the main text).



**Figure S18.** Inference of rate constants  $k_{on}$  and  $k_{off}$  from stochastic simulation, by fitting to Eq. (S7). (a) Estimated  $k_{on}$  and (b)  $k_{off}$  for Case I:  $d_A = 10, d_B = 20 \mu\text{m}^{-2}$ , Case II:  $d_A = 20, d_B = 20 \mu\text{m}^{-2}$  and Case III:  $d_A = 30, d_B = 20 \mu\text{m}^{-2}$ . Error bars indicate  $\pm 1$  standard deviation over 10 different trials.

## 566 Bibliography

- 567 1. D Sage, et al., Super-resolution fight club: assessment of 2d and 3d single-molecule localization microscopy software. *Nature Methods* **16**, 387–395 (2019).
- 568 2. LS Churchman, H Flyvbjerg, JA Spudich, A non-gaussian distribution quantifies distances measured with fluorescence  
 569 localization techniques. *Biophysical Journal* **90**, 668–671 (2006).
- 570 3. J Park Jr, Moments of the generalized rayleigh distribution. *Quarterly of Applied Mathematics* **19**, 45–49 (1961).
- 571 4. A Hagberg, P Swart, D S Chult, Exploring network structure, dynamics, and function using networkx, (Los Alamos National  
 572 Lab.(LANL), Los Alamos, NM (United States)), Technical report (2008).
- 573 5. Z Galil, Efficient algorithms for finding maximum matching in graphs. *ACM Computing Surveys (CSUR)* **18**, 23–38 (1986).
- 574 6. LS Churchman, Z Ökten, RS Rock, JF Dawson, JA Spudich, Single molecule high-resolution colocalization of Cy3 and  
 575 Cy5 attached to macromolecules measures intramolecular distances through time. *Proceedings of the National Academy of  
 576 Sciences* **102**, 1419–1423 (2005).
- 577 7. S Niekamp, et al., Nanometer-accuracy distance measurements between fluorophores at the single-molecule level.  
 578 *Proceedings of the National Academy of Sciences* **116**, 4275–4284 (2019).
- 579 8. NN Batada, LA Shepp, DO Siegmund, Stochastic model of protein–protein interaction: Why signaling proteins need to be  
 580 colocalized. *Proceedings of the National Academy of Sciences* **101**, 6445–6449 (2004).
- 581 9. NN Batada, LA Shepp, DO Siegmund, M Levitt, Spatial regulation and the rate of signal transduction activation. *PLoS  
 582 Computational Biology* **2**, e44 (2006).
- 583 10. M Hoffmann, C Fröhner, F Noé, Readdy 2: Fast and flexible software framework for interacting-particle reaction dynamics.  
 584 *PLoS computational biology* **15**, e1006830 (2019).
- 585 11. SV Costes, et al., Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical Journal*  
 586 **86**, 3993–4003 (2004).
- 587 12. S Malkusch, et al., Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochemistry  
 588 and Cell Biology* **137**, 1–10 (2012).
- 589 13. T Lagache, et al., Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nature  
 590 Communications* **9**, 698 (2018).
- 591 14. F Levet, et al., A tessellation-based colocalization analysis approach for single-molecule localization microscopy. *Nature  
 592 Communications* **10**, 2379 (2019).
- 593 15. AL Ejdrup, et al., A density-based enrichment measure for assessing colocalization in single-molecule localization  
 594 microscopy data. *Nature Communications* **13**, 4388 (2022).
- 595 16. W Trabesinger, et al., Statistical analysis of single-molecule colocalization assays. *Analytical Chemistry* **73**, 1100–1105  
 596 (2001).
- 597 17. A Inoue, et al., Illuminating G-protein-coupling selectivity of GPCRs. *Cell* **177**, 1933–1947 (2019).
- 598 18. JS Smith, et al., Noncanonical scaffolding of  $G_{\alpha i}$  and  $\beta$ -arrestin by G protein–coupled receptors. *Science* **371**, eaay1833  
 599 (2021).
- 600 19. Y Ono, et al., Generation of  $G_{\alpha i}$  knock-out HEK293 cells illuminates  $G_{\alpha i}$ -coupling diversity of GPCRs. *Communications  
 601 Biology* **6**, 112 (2023).
- 602 20. R Birke, et al., Sulfonated red and far-red rhodamines to visualize SNAP-and HALO-tagged cell surface proteins. *Organic &  
 603 Biomolecular Chemistry* **20**, 5967–5980 (2022).
- 604

- 605 21. H Schihada, R Shekhani, G Schulte, Quantitative assessment of constitutive g protein–coupled receptor activity with  
606 bret-based g protein biosensors. *Science Signaling* **14**, eabf1653 (2021).
- 607 22. M Ester, HP Kriegel, J Sander, X Xu, , et al., A density-based algorithm for discovering clusters in large spatial databases  
608 with noise in *kdd*. Vol. 96, pp. 226–231 (1996).
- 609 23. S Bolte, FP Cordelières, A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy*  
610 **224**, 213–232 (2006).
- 611 24. C Tameling, et al., Colocalization for super-resolution microscopy via optimal transport. *Nature Computational Science* **1**,  
612 199–211 (2021).