# Supplementary Materials

*Molecular Biology of the Cell*

Malkusch *et al*.

Supplementary information:

# Receptor tyrosine kinase MET ligand-interaction classified via machine learning from single-particle tracking data

Sebastian Malkusch[1,*], Johanna V. Rahm[2], Marina S. Dietz[2], Mike Heilemann[2], Jean-Baptiste Sibarita[3] and Jörn Lötsch[1,4]

[1] Institute of Clinical Pharmacology, Goethe University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany

[2] Institute of Physical and Theoretical Chemistry, Goethe University, Frankfurt am Main, Germany

[3] University Bordeaux, CNRS, Interdisciplinary Institute for Neuroscience, IINS, UMR 5297, F-33000 Bordeaux, France

[4] Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Theodor-Stern-Kai 7, 60596 Frankfurt am Main, Germany

* correspondence to: Sebastian Malkusch, Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany, e-mail: malkusch@med.uni-frankfurt.de; Phone: +49-69-6301-87818, Fax: +49-69-6301-7637

Keywords: single-particle tracking, hidden Markov modeling, data analysis, machine-learning, interpretable artificial intelligence, receptor tyrosine kinase, MET, internalin B

# Appendix

## Appendix A: Diffusion model

The concentration change over time is described by Fick's second law of diffusion:

$$\frac{\partial n}{\partial t} = D\Delta n \qquad \text{Equation 1}$$

With $n(r,t)$being the particle concentration at a distinct location $r$ and a distinct time point $t$. The parameter $D$ ist termed diffusion coefficient. $\Delta$ is the Laplace operator. When analyzing a single particle instead of a large ensemble, equation 1 rewrites to:

$$\frac{\partial p}{\partial t} = D\Delta p \qquad \text{Equation 2}$$

With $p(r,t \mid D)$ being the localization probability of the particle at a distinct location $r$ and time point $t$. As membrane bound receptors show a two-dimensional lateral diffusion with $r = \left(\widehat{e_x}x, \widehat{e_y}y\right)^T$the solution of equation 2 calculates to:

$$p(x,y,t \mid D) = \frac{1}{4\pi Dt} e^{-\left(\frac{x^2+y^2}{4Dt}\right)} \qquad \text{Equation 3}$$

The dimensionality of the diffusion model can be reduced to a single dimension by addressing only the Euclidean point-to-point distance $r = \sqrt{x^2+y^2}$. For this purpose equation 3 is transformed into polar coordinates and integrated over the line element $ds_\varphi = rd\varphi$:

$$p(r,t \mid D) = \frac{1}{4\pi Dt}\int_0^{2\pi} r\, e^{-\left(\frac{r^2}{4Dt}\right)} d\varphi = \frac{2r}{4Dt} e^{-\left(\frac{r^2}{4Dt}\right)} \qquad \text{Equation 4}$$

The expectation value of equation 4 calculates to:

$$< r^2 >= \int_0^\infty r^2\, p(r,t \mid D) dr = 4Dt$$ Equation 5

Inserting the expectation value of equation 5 into equation 4 results in the final diffusion model:

$$p(r,t \mid < r^2 >) = \frac{2r}{< r^2 >} e^{-(\frac{r^2}{<r^2>})}$$ Equation 6

A mixture model of multiple diffusive states is given by the superposition of the single state diffusion models:

$$P(r,t \mid \omega, < r^2 >) = \sum_{i=1}^{N} \omega_i \; P(r,t \mid < r^2 >_i)$$ Equation 7

# Appendix B: Error estimation

The main sources of error during data acquisition are first, errors during particle localization procedure caused by the Poisson statistics of the detected photon distribution, background noise (static error) and signal pixelation and second, particle movement during image acquisition time (dynamic error) (Savin and Doyle, 2005).

## Localization error

The localization error $\varepsilon$ describes how accurate the position $(\hat{\mu}_x, \hat{\mu}_y)$ of a fixed single emitter can be determined by fitting the two-dimensional projection of its diffracted photon density distribution to a model point spread function:

$$p(x,y \mid \mu_x, \mu_y, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2}\right)}$$ Equation 8

With $\mu_x$and $\mu_y$being the estimated probe's positions and $\sigma_x$and $\sigma_y$being the estimated sample standard deviations of the respective dimensions. The standard errors of the mean for each dimension $\varepsilon_x$and $\varepsilon_y$ are well approximated by:

$\varepsilon_{dim} \approx \frac{\sigma_{dim}}{\sqrt{N}}$ with $dim = x, y$ Equation 9

Here, $N$describes the sample size, by which means the detected number of photons. As the accurate determination of the static error in single molecule localization microscopy has been a research target for several years, it is a well solved problem. Therefore, $\varepsilon$ can either be estimated from the experimental conditions (photon distribution, signal-to-noise ratio and pixelation) (Mortensen *et al.*, 2010) or from the distribution of point-to-point distances measured between multiple localizations of the same static fluorescent probe (Endesfelder *et al.*, 2014).

## Static error

The static error estimates the deviation between the true ($\hat{r}$) and the determined ($r$) Euclidean point-to-point distance of two molecules. With the localization error $\varepsilon$at hand, the spatial probability density distribution of a single fluorescent probe's determined position in one dimension around its true position $\hat{\mu}_x$ is given by $p(x \mid \hat{\mu}_x, \varepsilon)$:

$$p(x \mid \hat{\mu}_x, \varepsilon) = \frac{1}{\sqrt{2\pi\varepsilon^2}} e^{-\frac{(x-\hat{\mu}_x)^2}{2\varepsilon^2}}$$ Equation 10

Assuming that the localization error is equal in both lateral dimensions ($\varepsilon = \varepsilon_x = \varepsilon_y$) the expected measured Euclidean point-to-point distance between two fluorescent probes that are separated by a true distance of $\hat{r} = \sqrt{(\hat{\mu}_{x1} - \hat{\mu}_{x2})^2 + \left(\hat{\mu}_{y1} - \hat{\mu}_{y2}\right)^2}$ calculates to:

$$< r^2 >= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$$ Equation 11

$$r^2\, p(x_1 |\, \hat{\mu}_{x1}, \varepsilon)\, p(x_2 | \hat{\mu}_{x2}, \varepsilon)\, p(y_1 | \hat{\mu}_{y1}, \varepsilon)\, p(y_2 | \hat{\mu}_{y2}, \varepsilon)\, dx_1 dx_2 dy_1 dy_2$$

The problem can be simplified by rotating the coordinate system in a way that $\hat{\mu}_{x1} = 0$, $\hat{\mu}_{x2} = \hat{r}$,$\hat{\mu}_{y1} = 0$ and $\hat{\mu}_{y2} = 0$ and by adding $0 = \hat{r}^2 + 2x_2\hat{r} - \hat{r}^2 - 2x_2\hat{r}$:

$$< r^2 >= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$$ Equation 12

$$(x_1^2 - 2x_1x_2 + (x_2 - \hat{r})^2 + y_1^2 - 2y_1y_2 + y_2^2 + 2x_2\hat{r} - \hat{r}^2)$$

$$\frac{1}{\sqrt{2\pi\varepsilon^2}}e^{-\frac{x_1^2}{2\varepsilon^2}} \quad \frac{1}{\sqrt{2\pi\varepsilon^2}}e^{-\frac{(x_2-\hat{r})^2}{2\varepsilon^2}} \quad \frac{1}{\sqrt{2\pi\varepsilon^2}}e^{-\frac{y_1^2}{2\varepsilon^2}} \quad \frac{1}{\sqrt{2\pi\varepsilon^2}}e^{-\frac{y_2^2}{2\varepsilon^2}} \quad dx_1dx_2dy_1dy_2$$

Finally, equation 11 can be simplified to:

$$< r^2 >= \hat{r}^2 + 4\varepsilon^2$$ Equation 13

Thus, the expected measured mean squared displacement systematically overestimates the true mean squared displacement (Savin and Doyle, 2005).

## Dynamic error

The dynamic error is caused by the fact that the position of the fluorescent probe is determined from the superposition of all its photons detected during the camera exposure time. As a result, the measured particle position at a distinct time point ($t$) is not an accurate snapshot but an average of all positions taken by the particle during acquisition time ($\tau$).

$\underline{\mu}_{dim} = \frac{1}{\tau}\int_0^{\tau}\mu_{dim}\,(t-\xi)d\xi$ with $dim\ =\ x,y$ Equation 14

This means that dynamics within a time period $t \leq \tau$cannot be resolved (Savin and Doyle, 2005).

## Model correction

Based on previous work from Savin and Doyle (Savin and Doyle, 2005), the diffusion coefficient of a single molecule can be calculated from the measured apparent mean squared displacement, if the static error is known.

$< r^2 >= 4D\left(t - \frac{\tau}{3}\right) + 4\varepsilon^2$ for $t \geq \tau$ Equation 15

# Appendix C: The chemical master equation

The HMM postulates a reaction network consisting of first order reactions at discrete time steps. To translate this reaction network into a continuous time space, the state transition probabilities of the HMM need to be transferred into first order reaction rate constants. A first order reaction follows a mono-exponential decay function

$$f_{decay}(t) = Ae^{-kt}, \qquad \text{Equation 16}$$

with $k\left[\frac{1}{s}\right]$ defining the reaction rate constant. The probability that a single molecule will transit within the time span of a discrete time step of the HMM ($\Delta t$) is given by the integral of the normalized decay function form $t = 0$to $t = \Delta t$.

$$p = \int_0^{\Delta t} k\, e^{-kt} dt \qquad \text{Equation 17}$$

Solving the integral of equation 16 results in a conversion formula:

$$k = \frac{-ln(1-p)}{\Delta t} \qquad \text{Equation 18}$$

A closed bio-chemical reaction system that is set up from first-order chemical equations can thus be described by the chemical master equation:

$$\frac{dx_i}{dt} = \sum_{j \neq i} k\,(j \to i)\left[x_j\right] - k(i \to j)[x_i] \qquad \text{Equation 19}$$

Here, $k(i \to j)[x_i]$ describes the reactions that reduce the concentration $[x_i]$ and $k(j \to i)[x_j]$ describes reactions that increase the concentration $[x_i]$, while they reduce the concentration $[x_j]$.

# Appendix D: Confusion matrix

The performance of the classification algorithms is measured by several metrics that are originated in the four entries of the confusion matrix: True positive (TP), false positive (FP), true negative (TN), false negative (FN). The metrics are composed of these as follows:

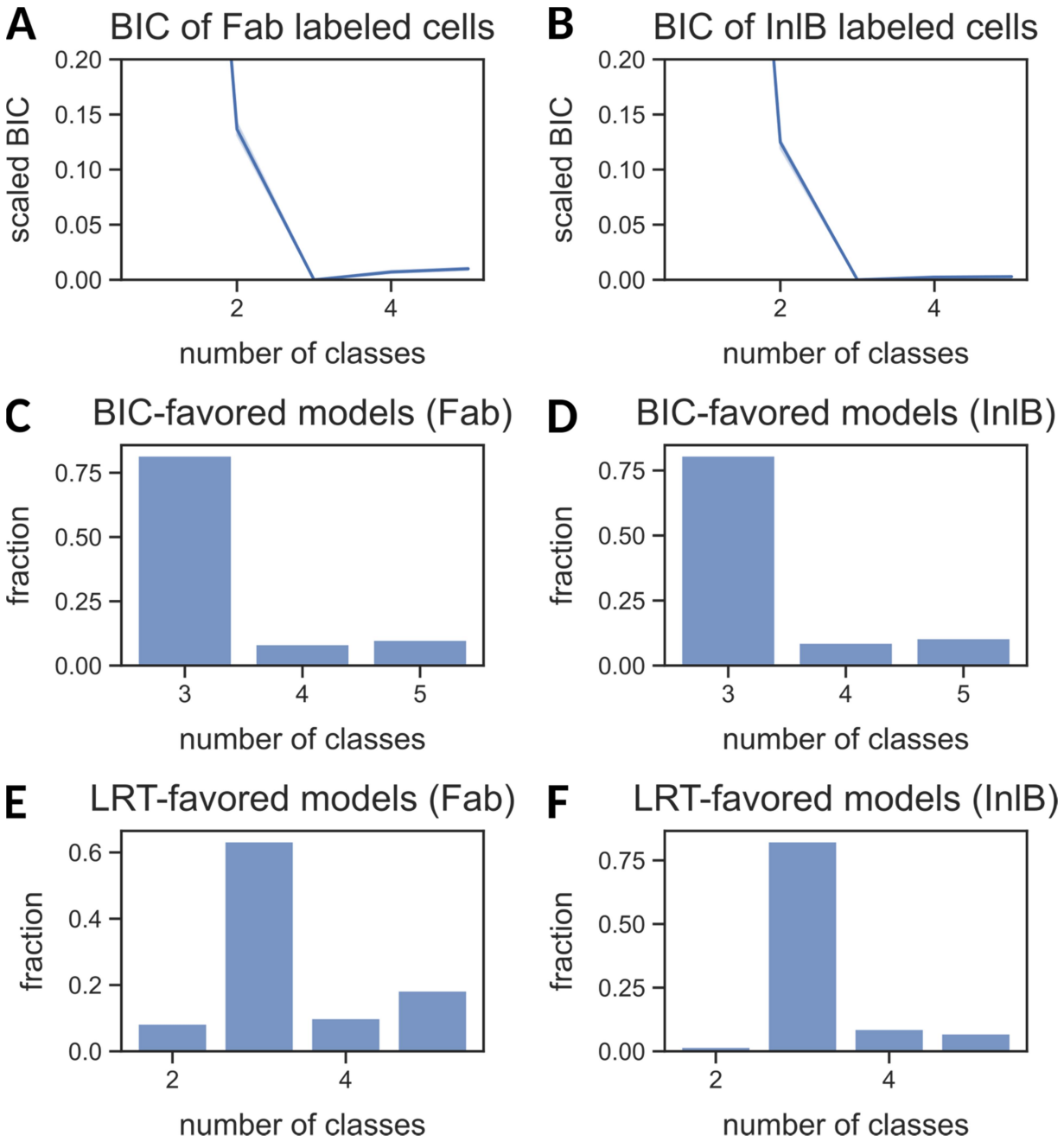
$$balanced\ accuracy = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right)$$ Equation 20

$$F1\ score = \frac{2TP}{2TP+FP+FN}$$ Equation 21

$$precision = \frac{TP}{TP+FP}$$ Equation 22

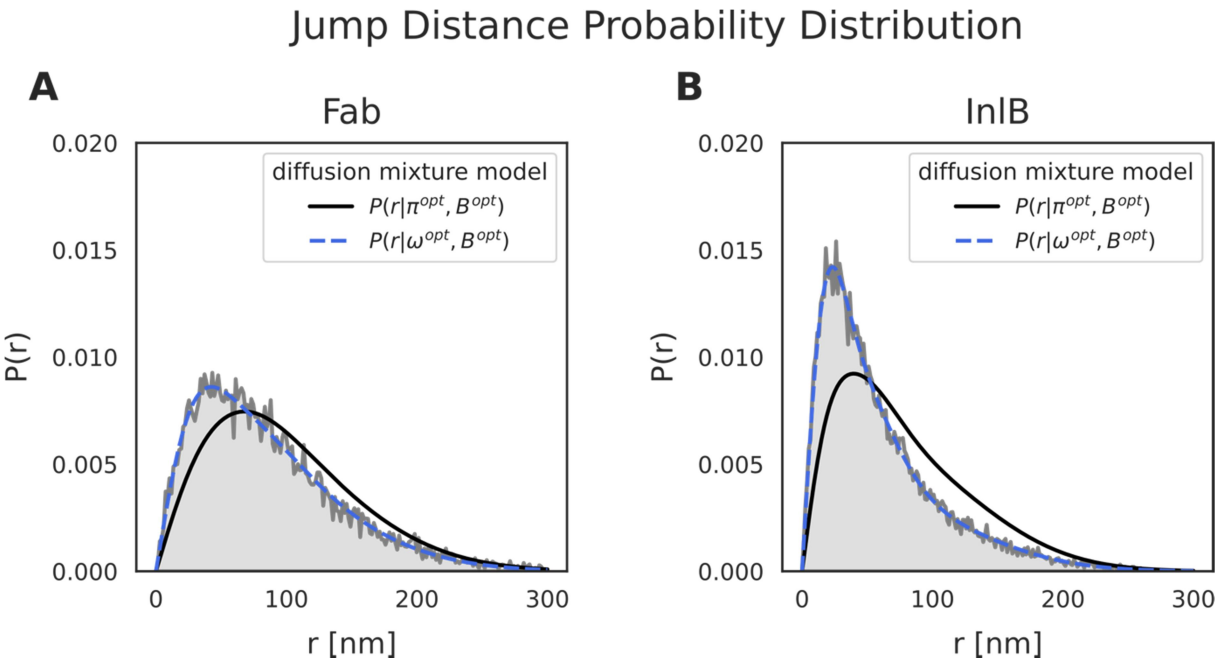
$$recall = \frac{TP}{FP+FN}$$ Equation 23

# Supplementary Figures

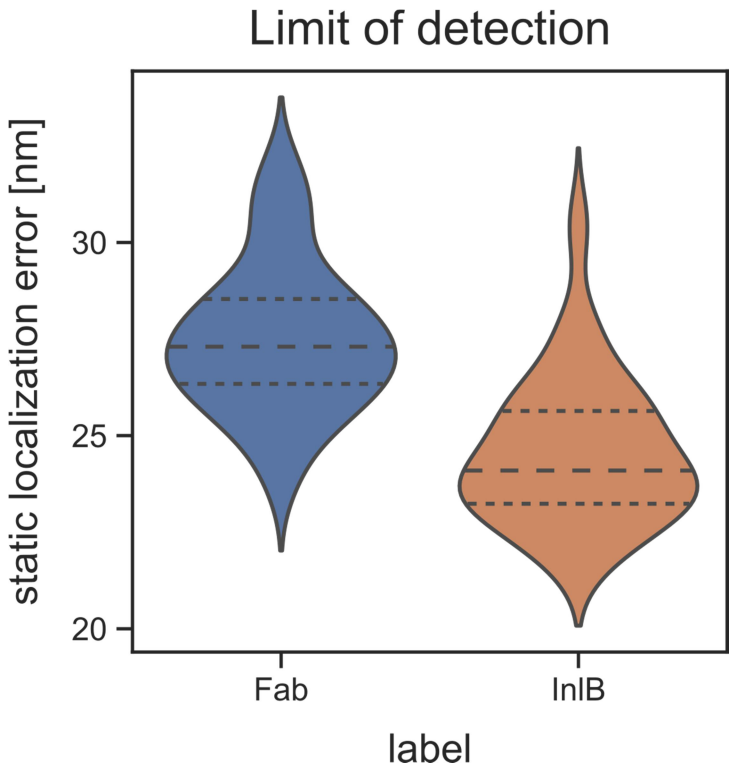



Supplementary Figure 1: Model Selection. For each measurement of the dataset, 5 hidden Markov models (HMMs) were trained with different complexity (number of classes n=1 to 5). For

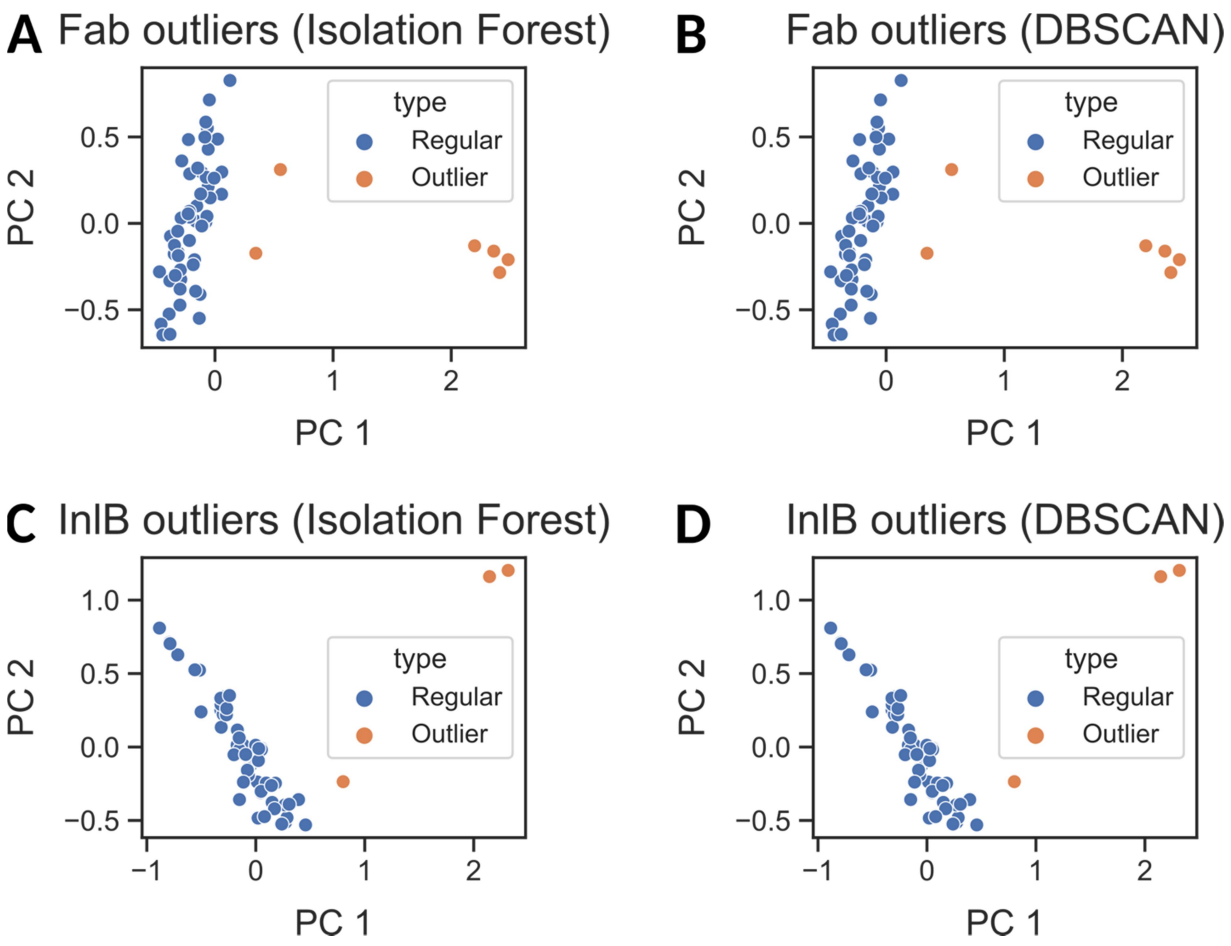
each model the Bayesian information criterion (BIC) was calculated. The value range of all calculated BIC values was scaled to an interval of [0; 1]. The resulting distribution and it's 95% confidence interval is shown for untreated cells (A) and InlB treated cells (B). C and D: The HMM with a minimal BIC value is identified per cell and the resulting distribution of optimal HMMs is depicted as a bar chart for untreated cells (C) and InlB treated cells (D). Consequently, the most frequently populated model is used for further analyses. e and f: As an alternative approach, the optimal HMM is identified using likelihood ratio tests. Here the optimal model is defined as the simplest models, for which a comparison with the next most complex model using the likelihood ratio test does not lead to a significant improvement (the significance threshold was set to $\alpha = 0.05$). Again, the optimal HMMs per cell are identified and their distribution is depicted as a bar chart for untreated cells (E) and InlB treated cells (F). BIC-based HMM analysis was performed using the ermine package.



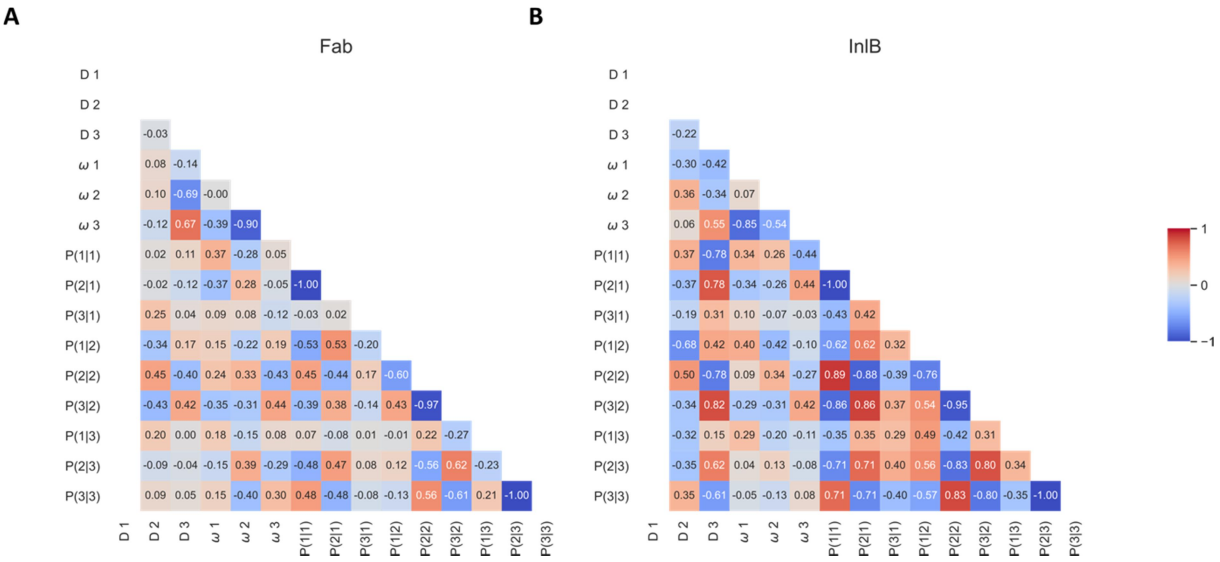

Supplementary Figure 2: Modeling the jump-distance probability density function: Overlay of the jump-distance probability density function (filled grey area) and two three state diffusion mixture models: For the first model (black solid line) the HMM parameter $\pi^{opt}$ was used as state occupancy and the HMM parameter $B^{opt}$ was used as mobility ($P(r|\pi^{opt}, B^{opt})$). For the second model (blue dashed line) the state occupancy $\omega^{opt}$ was learned by expectation maximization while keeping the mobility HMM parameter $B^{opt}$fix ($P(r|\omega^{opt}, B^{opt})$) . The figures show the results for a representative Fab labeled cell (A) and a representative InlB labeled cell (B).



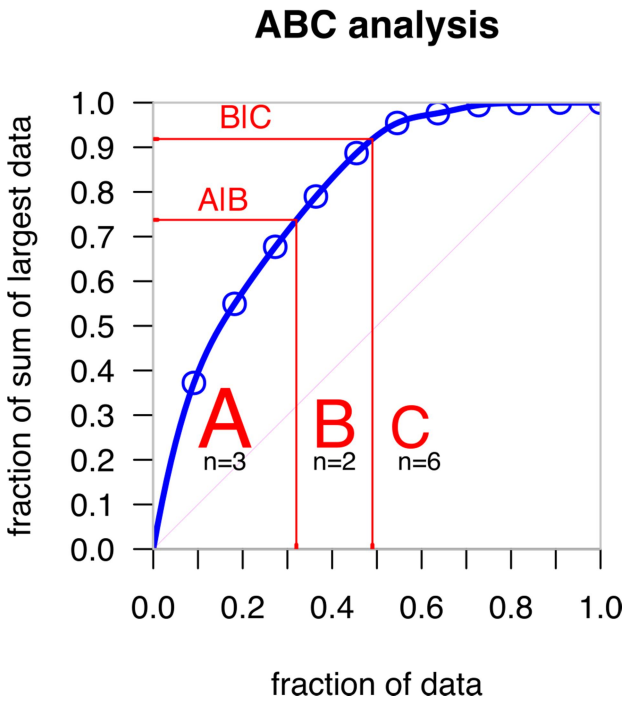

Supplementary Figure 3: Limit of detection (LOD). The static localization error of single molecule detection is determined in the form of the theoretically achievable localization precision. The localization precision distribution is visualized separately for untreated (blue, Fab) and InlB treated cells (orange, InlB). The distribution quartiles are highlighted as dashed lines. The average localization precision measured for untreated cells is $\sigma_{Fab} = 27.52 \pm 1.92\,nm$ and $\sigma_{InlB} = 24.58 \pm 1.91\,nm$ for InlB treated cells. Substituting the localization precision for an immobile particle ($D = 0\ \frac{\mu m^2}{s}$) into equation 14 results in the LOD for particle movement. It is given by the smallest measurable MSD which is $LOD_{Fab} = 3043.57 \pm 431.57\,nm^2$ for untreated cells and $LOD_{InlB} = 2430.6 \pm 392.73\,nm^2$ for InlB treated cells. Values are given as mean ± standard deviation.



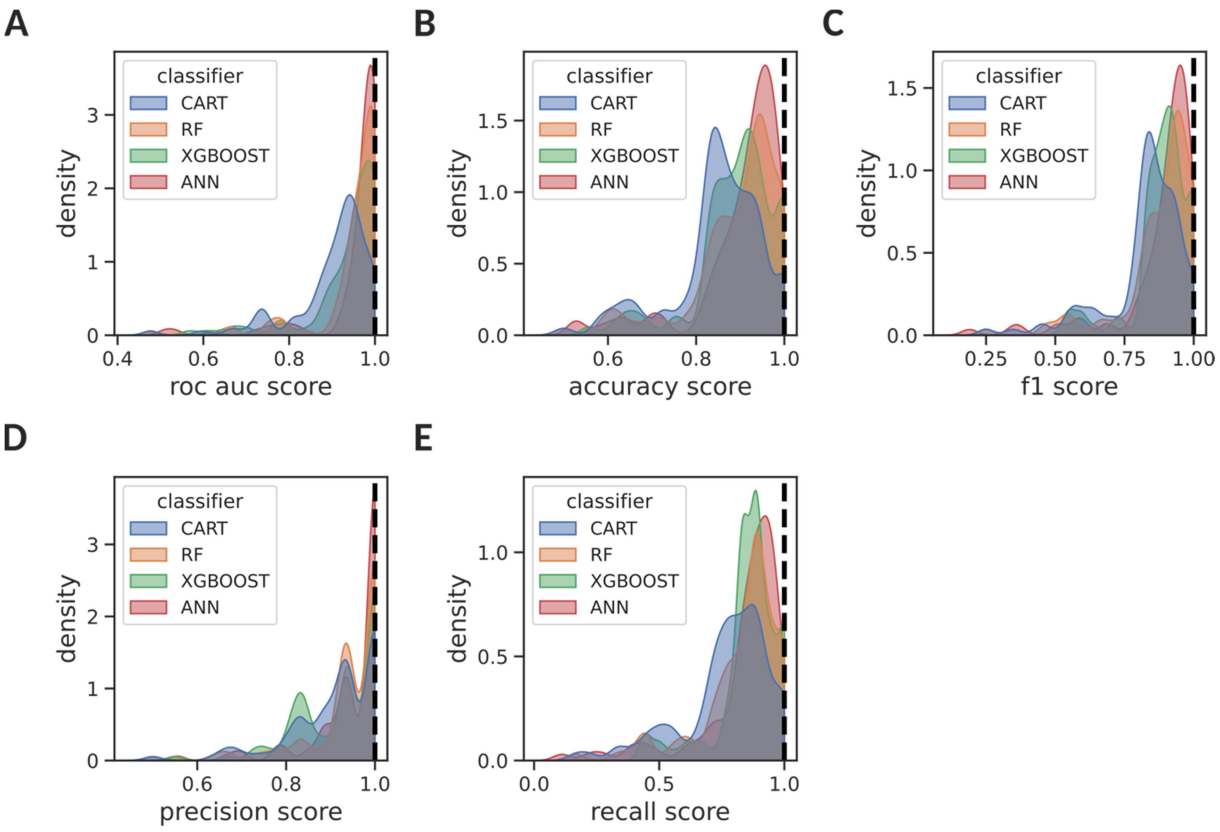

Supplementary Figure 4: Outlier Detection. Outliers in the hidden Markov model (HMM) parameter list were identified using either an isolation forest or by performing density-based spatial clustering of applications with noise (DBSCAN). Prior to DBSCAN analysis the values of each feature were scaled to a range of [0; 1]. Both groups (untreated and InlB treated cells) were analyzed separately. For the purpose of visualization the 11-dimensional space of the HMM-parameter list was projected onto two dimensions, by applying a principal component analysis (PCA) and plotting the first two principal components (PC1 and PC2). Outliers are highlighted in orange. A: Untreated outliers as identified by an isolation forest. B: Untreated outliers as identified by DBSCAN. C: InlB treated outliers as identified by an isolation forest. D: InlB treated outliers as identified by DBSCAN. Machine-learning based detection of outliers was performed using the scikit-learn package.




Supplementary Figure 5: Pairwise hidden Markov model (HMM) parameter correlation. Spearman's rank correlation matrix of HMM parameter distribution for untreated (A) and InlB treated (B) cells. Parameter pairs with a Spearman's rank correlation coefficient $> 0.9$ are defined as highly correlated. Consequently, filtering P(1|1), P(2|2) and P(3|3) from the HMM-parameter space results in a parameter set without highly correlated parameter pairs. Correlation analysis was performed using the pandas package.




Supplementary Figure 6: ABC analysis by item categorization of hidden Markov model (HMM) parameters. The importance of HMM-parameters to the classification task of identifying untreated and InlB treated cells was characterized by calculating the mean absolute SHAP values for the decisions made by a random forest classifier. Item categorization was performed by ABC analysis. The ABC plot (blue line) shows the cumulative distribution function of the parameter importance as characterized by the mean absolute SHAP values with the limits between sets A, B and C indicated as red lines. The analysis shows that 27% of the HMM-parameters account for 68% of the SHAP-based feature importance. Consequently, the three HMM-parameters $(D_2, \omega_3, P(2|3))$ that belonged to ABC set "A" were considered as most relevant to the classification task (see table 3). Item categorization was performed using the R package ABCanalysis.




Supplementary Figure 7: Negative control of classifier performance measure trained on a reduced feature space based on validation metrics. The performance of different classifiers on classifying single cells as either untreated or InlB treated was measured. The analysis was performed in four steps: First a subset of randomly chosen HMM parameters that matches the dimensionality of parameters classified as highly important by the ABC-analysis are selected. Second, the reduced parameter list is split into a training and a validation data set (⅔ to ⅓). Third, the training data set was permuted. Third, an artificial neural network (ANN), a classification and regression tree (CART), a random forest classifier (RF) and an eXtreme Gradient Boosting algorithm (XGBoost) were trained on the training data set to predict whether cells are Fab or InlB labeled from the variables learned from the HMM. To rule out a random result, the experiment was repeated 100 times with different HMM parameter subsets. Fourth, the classifier performance was validated on the validation data set by means of different metrics: (A) The area under the receiver operating characteristic curve (ROC-AUC), (B) the balanced accuracy, (C) the f1-score, (D) the precision and (E) the recall. The dashed lines represent the mean performance of the random forest classifier

trained on the HMM-parameter subset that only consists of the “important few” parameters categorized as “A” by a previously performed ABC-analysis. Classifier training was performed using the scikit-learn, keras and xgboost packages.

# Supplementary Tables

Supplementary Table 1: Quartiles of parameters associated with membrane bound MET dynamics. The parameters were machine-learned by 108 three-state hidden Markov models (HMMs) and jump-distance mixture models in two steps. First, an individual HMM was trained on each single-cell trajectory dataset resulting in a list of optimized HMM-parameter sets. Second, The state occupancy for each cell was determined by optimizing a jump-distance mixture model on the trajectory datasets. For simplicity reasons, the resulting list of optimized model parameters $(\omega^{opt}, A^{opt}, D^{opt})$ is termed optimized HMM-parameters. Differences between the HMM-parameter distributions of untreated and InlB treated cells were tested by the use of a two-sided Mann-Whitney U test against the 0-Hypothesis $H_0: Fab = InlB$ with an alternative hypothesis $H_a: Fab \neq InlB$. The p-values are corrected for multiple testing using the method of Bonferroni. The number of tests performed is 14. Testing was performed using the scipy package.

| HMM parameter | Fab quartiles | InlB quartiles | Test statistic | p value (Ha: Fab ≠ InlB) |
|---|---|---|---|---|
| $\omega_1$ | 0.132; 0.148; 0.161 | 0.205; 0.224; 0.261 | 86 | $4.96 \cdot 10^{-16}$ |
| $\omega_2$ | 0.332; 0.366; 0.400 | 0.416; 0.423; 0.445 | 396 | $9.69 \cdot 10^{-10}$ |
| $\omega_3$ | 0.446; 0.479; 0.521 | 0.317; 0.345; 0.369 | 2825 | $6.45 \cdot 10^{-16}$ |
| $D_1$ | 0.000; 0.000; 0.000 | 0.000; 0.000; 0.000 | - | - |
| $D_2$ | 0.061; 0.067; 0.073 | 0.039; 0.041; 0.043 | 2916 | $4.73 \cdot 10^{-18}$ |

| | | | | |
|---|---|---|---|---|
| $D_3$ | 0.241; 0.248; 0.259 | 0.207; 0.218; 0.227 | 2769 | $1.14 \cdot 10^{-14}$ |
| $P(1\|1)$ | 0.948; 0.954; 0.962 | 0.935; 0.946; 0.952 | 2143 | $4.64 \cdot 10^{-4}$ |
| $P(2\|1)$ | 0.038; 0.046; 0.052 | 0.048; 0.054; 0.064 | 780 | $4.40 \cdot 10^{-4}$ |
| $P(3\|1)$ | 0.000; 0.000; 0.000 | 0.000; 0.000; 0.000 | 1195 | $1.49 \cdot 10^{-0}$ |
| $P(1\|2)$ | 0.017; 0.019; 0.021 | 0.028; 0.033; 0.037 | 128 | $4.35 \cdot 10^{-15}$ |
| $P(2\|2)$ | 0.932; 0.942; 0.952 | 0.895; 0.911; 0.926 | 2526 | $7.57 \cdot 10^{-10}$ |
| $P(3\|2)$ | 0.030; 0.040; 0.048 | 0.045; 0.056; 0.068 | 635 | $6.06 \cdot 10^{-6}$ |
| $P(1\|3)$ | 0.000; 0.000; 0.000 | 0.000; 0.000; 0.000 | 989 | $5.59 \cdot 10^{-2}$ |
| $P(2\|3)$ | 0.030; 0.035; 0.047 | 0.073; 0.089; 0.106 | 95 | $7.94 \cdot 10^{-16}$ |
| $P(3\|3)$ | 0.953; 0.965; 0.970 | 0.894; 0.911; 0.926 | 2821 | $7.94 \cdot 10^{-16}$ |

Supplementary Table 2: Comparison of the state occupancy machine-learned by the hidden Markov model (HMM, $\omega^{opt}$) based analysis of trajectory datasets with the equilibrium state occupancy generated by the simulation of stochastic Petri nets (PN, $\omega^{sim}$). The analysis was performed in three steps: First, HMM-parameters were optimized. Second, the state occupancy was determined by using a diffusion mixture model. Third, for each cell a stochastic Petri net (PN) was designed and parameterized with the optimized transition probabilities learned by the HMM. For this reason reaction rate constants ($k^{sim}$) were calculated from the HMM-learned transition probability matrix ($A^{opt}$). The time evolution of small ensembles of 1000 molecules, initially all fast diffusing, was simulated with the stochastic PNs until equilibrium was reached. For each simulation, the average equilibrium population ratio was calculated from the last 1000 time steps. For each experiment the occupancy distributions are compared state-wise using two-sided and one-sided Mann-Whitney U tests against a 0-Hypothesis $H_0: S_1 = S_2$. The alternative hypothesis of the two-sided Mann-Whitney U test is given by $H_a: S_1 \neq S_2$ while the alternative hypotheses of the one-sided Mann-Whitney U tests are given either by $H_a: S_1 < S_2$ or $H_a: S_1 > S_2$. For each condition $\omega_{ligand}^{experiment}$ three tests were performed and the p-values are corrected for multiple testing using the method of Bonferroni. Testing was performed using the scipy package.

| Experiment | State S1 | State S2 | Test statistic | p value (Ha: S1 ≠ S2) | p value (Ha: S1 < S2) | p value (Ha: S1 > S2) |
|---|---|---|---|---|---|---|
| $\omega_{Fab}^{opt}$ | immobile | slow | 0 | $1.01 \cdot 10^{-18}$ | $5.07 \cdot 10^{-19}$ | 3.00 |
| $\omega_{Fab}^{opt}$ | immobile | fast | 0 | $1.01 \cdot 10^{-18}$ | $5.07 \cdot 10^{-19}$ | 3.00 |
| $\omega_{Fab}^{opt}$ | slow | fast | 204 | $4.02 \cdot 10^{-14}$ | $2.01 \cdot 10^{-14}$ | 3.00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $\omega_{InlB}^{opt}$ | immobile | slow | 0 | $1.01 \cdot 10^{-18}$ | $5.07 \cdot 10^{-19}$ | 3.00 |
| $\omega_{InlB}^{opt}$ | immobile | fast | 88 | $1.18 \cdot 10^{-16}$ | $5.90 \cdot 10^{-17}$ | 3.00 |
| $\omega_{InlB}^{opt}$ | slow | fast | 2782 | $1.26 \cdot 10^{-15}$ | 3.00 | $6.32 \cdot 10^{-16}$ |
| $\omega_{Fab}^{sim}$ | immobile | slow | 0.00 | $1.01 \cdot 10^{-18}$ | $5.07 \cdot 10^{-19}$ | 3.00 |
| $\omega_{Fab}^{sim}$ | immobile | fast | 0.00 | $1.01 \cdot 10^{-18}$ | $5.07 \cdot 10^{-19}$ | 3.00 |
| $\omega_{Fab}^{sim}$ | slow | fast | 951 | $5.57 \cdot 10^{-3}$ | $2.79 \cdot 10^{-3}$ | 3.00 |
| $\omega_{InlB}^{sim}$ | immobile | slow | 0 | $1.01 \cdot 10^{-18}$ | $5.07 \cdot 10^{-19}$ | 3.00 |
| $\omega_{InlB}^{sim}$ | immobile | fast | 873 | $9.87 \cdot 10^{-4}$ | $4.93 \cdot 10^{-4}$ | 3.00 |
| $\omega_{InlB}^{sim}$ | slow | fast | 2916 | $1.01 \cdot 10^{-18}$ | 3.00 | $5.07 \cdot 10^{-19}$ |

Supplementary Table 3: Negative control of classifier performance measure based on validation metrics. The performance of different classifiers on classifying single cells as either untreated or InlB treated was measured. The analysis was performed in four steps: First the HMM parameter list $(\omega^{opt}, A^{opt}, D^{opt})$ was filtered for highly correlated parameters and split into a training and a validation data set (⅔ to ⅓). Second, the training data set was permuted. Third, an artificial neural network (ANN), a classification and regression tree (CART), a random forest classifier (RF) and an eXtreme Gradient Boosting algorithm (XGBoost) were trained on the training data set to predict whether cells are Fab or InlB labeled from the variables learned from the HMM. To rule out a random result, the experiment was repeated 100 times with different permutations of the training data set. Fourth, the classifier performance was validated on the validation data set by means of different metrics: The area under the receiver operating characteristic curve (ROC-AUC), the balanced accuracy, the f1-score, the precision and the recall. The values are given as mean ± standard deviation. As the training data is permuted, the classifiers are expected to perform random guesses. Classifier training was performed using the scikit-learn, keras and xgboost packages.

| Classifier | ROC-AUC | Accuracy | F1-score | Precision | Recall |
|---|---|---|---|---|---|
| ANN | 0.497 ± 0.208 | 0.489 ± 0.168 | 0.481 ± 0.179 | 0.487 ± 0.172 | 0.487 ± 0.206 |
| CART | 0.492 ± 0.150 | 0.502 ± 0.150 | 0.481 ± 0.146 | 0.526 ± 0.182 | 0.468 ± 0.167 |
| RF | 0.457 ± 0.197 | 0.473 ± 0.149 | 0.475 ± 0.143 | 0.482 ± 0.150 | 0.481 ± 0.160 |
| XGBoost | 0.501 ± 0.176 | 0.499 ± 0.135 | 0.498 ± 0.130 | 0.508 ± 0.142 | 0.497 ± 0.138 |

Supplementary Table 4: Positive control of classifier performance measure based on validation metrics. The performance of different classifiers on classifying single cells as either untreated or InlB treated was measured. The analysis was performed in four steps: First the HMM parameters that are characterized as highly important to the classification task by the previously performed ABC-analysis are selected. Second, the reduced parameter list is split into a training and a validation data set (⅔ to ⅓). Third, the training data set was permuted. Third, an artificial neural network (ANN), a classification and regression tree (CART), a random forest classifier (RF) and an eXtreme Gradient Boosting algorithm (XGBoost) were trained on the training data set to predict whether cells are Fab or InlB labeled from the variables learned from the HMM. To rule out a random result, the experiment was repeated 100 times with different seed-based classifier initializations. Fourth, the classifier performance was validated on the validation data set by means of different metrics: The area under the receiver operating characteristic curve (ROC-AUC), the balanced accuracy, the f1-score, the precision and the recall. The values are given as mean ± standard deviation. As the HMM-parameter subset used within the experiment only consists of the "important few" parameters categorized as "A" by a previously performed ABC-analysis, the classifiers are expected to perform similarly as if trained on the complete HMM-parameter set (table 2). Classifier training was performed using the scikit-learn, keras and xgboost packages.

| Classifier | rank | ROC-AUC | Accuracy | F1-score | Precision | Recall |
|---|---|---|---|---|---|---|
| ANN | 3 | 1.000 ± 0.000 | 0.992 ± 0.013 | 0.992 ± 0.013 | 1.000 ± 0.000 | 0.984 ± 0.025 |
| CART | 1 | 0.964 ± 0.028 | 0.920 ± 0.067 | 0.912 ± 0.075 | 0.975 ± 0.055 | 0.864 ± 0.114 |
| RF | 4 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.000 |
| XGBoost | 2 | 0.998 ± 0.007 | 0.978 ± 0.038 | 0.976 ± 0.040 | 0.992 ± 0.028 | 0.963 ± 0.062 |

# Literature

Endesfelder, U., Malkusch, S., Fricke, F., and Heilemann, M. (2014). A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. Histochem Cell Biol *141*, 629-638.

Mortensen, K.I., Churchman, L.S., Spudich, J.A., and Flyvbjerg, H. (2010). Optimized localization analysis for single-molecule tracking and super-resolution microscopy. Nat Methods *7*, 377-381.

Savin, T., and Doyle, P.S. (2005). Static and dynamic errors in particle tracking microrheology. Biophysical Journal *88*, 623-638.