Published in final edited form as:

Methods Inf Med. 2012 ; 51(4): 332–340. doi:10.3414/ME11-02-0041.

Intelligent Data Analysis to Model and Understand Live Cell Time-Lapse Sequences

Alistair Paterson¹, Marjan Ashtari¹, David Ribé², Gudrun Stenbeck², and Allan Tucker^{1,*}

¹School of Information Systems Computing and Mathematics, Brunel University, West London, UK

²Centre for Cell and Chromosome Biology, Brunel University, West London, UK

Summary

Background—One important aspect of cellular function, which is at the basis of tissue homeostasis, is the delivery of proteins to their correct destinations. Significant advances in live cell microscopy have allowed tracking of these pathways by following the dynamics of fluorescently labelled proteins in living cells.

Objectives—This paper explores intelligent data analysis techniques to model the dynamic behavior of proteins in living cells as well as to classify different experimental conditions.

Methods—We use a combination of decision tree classification and hidden Markov models. In particular, we introduce a novel approach to “align” hidden Markov models so that hidden states from different models can be cross-compared.

Results—Our models capture the dynamics of two experimental conditions accurately with a stable hidden state for control data and multiple (less stable) states for the experimental data recapitulating the behaviour of particle trajectories within live cell time-lapse data.

Conclusions—In addition to having successfully developed an automated framework for the classification of protein transport dynamics from live cell time-lapse data our model allows us to understand the dynamics of a complex trafficking pathway in living cells in culture.

MeSH Keywords

Medical informatics; Cellular Dynamics; Stochastic Processes; Decision Trees

1 Introduction

One important aspect of cellular function, which is at the basis of tissue homeostasis, is the delivery of proteins to their correct destinations. Impairment of this process leads to a number of diseases and can be observed, amongst others, in the aging brain where axonal and dendritic transport is impaired [1] and in the altered extracellular matrix associated with osteoarthritis and aged skin [2]. The underlying causes of this impairment are most likely due to changes at the cellular level, specifically to the intracellular organelles that process

allan.tucker@brunel.ac.uk.

the proteins for secretion. In recent years significant advances in live cell microscopy have allowed tracking of single fluorescent molecules and fluorescently labelled proteins inside of living cells [3]. With these methods it is now possible not only to follow protein secretion but to unravel the mechanisms driving the motion of a wide variety of cellular components ranging from organelles to protein molecules. Several tracking algorithms and programs are available [4,5]. After automated detection and quantitative analysis of particle trajectories, simple speed and kinetic analysis can give some insight into the mechanisms that control cell secretion but a more sophisticated, quantitative analysis is needed to understand how signals generated outside the cell are relayed to the transport machinery. In light of the large data sets obtained from a single cell combined with the need to compare multiple data sets derived from cells under different experimental conditions, both qualitative and quantitative statistical analysis is necessary to extract and validate the data and to develop models of these complex biological systems.

The data described above is characterised by its uncertain nature and probabilistic models such as dynamic Bayesian networks [6] and hidden Markov models (HMM) [7] are very well suited for data mining and model development because of their ability to combine dynamic modelling with classification. The use of specialised models such as these allows us to understand explicitly the characteristics of the trajectories that are discovered. For example, some particles may take a more direct route to the cell membrane whilst others maintain an “explorative” behaviour, resulting in a delay on arrival of the cargo at the plasma membrane and an increased possibility of mistargeting [8, 9]. We are analysing datasets that measure tracks generated in living cells that have been transfected with a green fluorescent protein chimera. Cells are imaged before and after exposure to growth factors and exposure to signalling inhibitors. Tracks generated with the particle tracker plug-in for Image J [5] are used. See Figure 1 for an example of particle trajectories within a living cell.

The use of intelligent data analysis is not new to biological and medical data [10, 11]. Previously, we and others have classified time-series based upon their dynamics [12,13,14,15] and in [14] in particular attempts were made to cluster time-series based upon dynamic model parameters. In [16] HMMs were used to classify particle trajectories of a membrane-associated protein interacting with a homogeneously distributed binding partner. More generally, state space models have a very successful history of modelling trajectories in many disciplines [17]. In this paper we apply for the first time supervised classifier learning to the parameters of dynamic models in order to identify key features of particle trajectories under two experimental conditions. The next section describes the methods before our results are outlined in Section 3 and conclusions made in Section 4.

2 Methods

Osteoblastic ROS17/2.8 cells were grown in Dulbecco's minimal essential medium supplemented with Glutamax, 100 units/ml penicillin/streptomycin and 10% fetal calf serum at 37°C and 5% CO₂. For live imaging, 10⁵ cells were seeded in a 14 mm glass-bottomed dish (MatTek) and transfected with pEGFP-SPARC [18] using Lipofectamine 2000 (Invitrogen) 48 h before imaging. For imaging, cells were transferred to pre-warmed Opti-MEM medium supplemented with 0.5% bovine serum albumin, 30 mM HEPES/NaOH pH

7.4 and 1 mM CaCl₂. Cells were then placed in a heated chamber at 37 °C on a Nikon TE2000-S microscope fitted with an objective heater, and allowed to equilibrate for 30 min. Images were collected under total internal fluorescence mode, which utilises the unique properties of an induced evanescent wave that only illuminates the area of the cell that is in close contact to the glass surface of the dish [19], using the 60 x TIRFM objective with a 488 nm laser. 101 images of each field perpendicular to the cell-glass interface were collected at 300 ms intervals. Individual fields were imaged immediately before and after treatment with 25ng/ml TGFβ or vehicle for 10 min. No field was imaged more than twice to avoid phototoxicity.

Trajectories of vesicles containing fluorescent protein were detected in the recorded image sequences with the particle tracker plug-in for Image J, which after detection records the x and y coordinates of the tracked vesicles [5]. For our experiments we use two continuous states to represent the x and y coordinates of each vesicle. We have analysed two different experimental conditions, control data (or basal) where the cells have been mock treated (addition of solvent without TGFβ), and data where the cells have been exposed to TGFβ, an ubiquitous growth factor that plays an important role in bone development and maintenance [20]. It is known to control matrix protein secretion and the cytoskeleton both through transcriptional and immediate early signalling events [21, 22, 23]. It is therefore expected to impact on vesicle trajectories. We have data from 28 independent experiments, 10 of which were TGFβ-treated experiments. This involves 458 detected trajectories with duration of greater than 1.5 s per field of view.

The *x and y coordinates* of each particle were recorded, having first been zero-centred so that each coordinate is relative to its starting position. As well as recording particle *x and y coordinates*, a *distance* metric was calculated which represents how far a particle has moved overall. In addition *displacement* was measured that represents the overall distance from a particle's starting point. The calculations for these metrics for a point in one trajectory are shown in Equations 1 and 2.

Equation 1:

$$distance = \sum_{i=2}^n \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$$

where x_i represents the x co-ordinate in the i th position in the trajectory and the previous co-ordinate is represented as x_{i-1} . The same definition applies to the variables and len is the length of the trajectory.

Equation 2:

$$displacement = \sum_{i=2}^n \sqrt{(x_i - x_1)^2 + (y_i - y_1)^2}$$

where x_{start} and y_{start} represent the first coordinates in a trajectory.

Using these values as data, we construct HMMs in order to model the dynamics of the trajectories. Our approach takes the parameters of HMMs that have been learnt from this data, and uses them to classify the experiments to see if we can separate out the key features that discriminate between the experimental conditions. For this we use a Decision Tree (DT) [24] as this transparently classifies the data allowing us to explore what the key features of the trajectory dynamics are. The HMM is a tool for representing probability distributions over sequences of observations. It encompasses a discrete hidden variable (node) and a discrete or continuous observed node per slice (as shown in Figure 2) where there are conditional distributions to model the transition between hidden states as well as the probability of seeing observations given a hidden state. Figure 2 shows the architecture of a HMM where square nodes denote the discrete hidden process and circles denote the observed variables. We use 5 hidden states in our experiments as we want to capture a diverse number of individual characteristics of the trajectories in both basal and TGF β -induced states, which also recapitulate the complexity of the cellular response.

We adopt three approaches to classifying the trajectories based upon the discovered HMMs. In all approaches we use the EM (Baum Welch) algorithm to train the HMM [7]:

Set Up 1 - Pre-processing the variables

Firstly, we process the parameters. Processing ensures that states learnt using the EM algorithm (during HMM learning) are comparable across experiments. For example, state A in a HMM trained on one experiment may represent state B in another HMM trained on another experiment. To process we firstly calculate the probability of remaining in the same state by summing over all probabilities in the transition matrix that result in no switch to another state, in other words every $P(H_{t+1}|H_t)$ where $H_{t+1} = H_t$ (H_t represents hidden state at time t). We also calculate the probability of switching to another state by summing over all probabilities that involve a switch to a different state, that is all $P(H_{t+1}|H_t)$ where $H_{t+1} \neq H_t$. We then calculate the mean, maximum and minimum values of the mean and covariance for distance and displacement. This results in the following parameter features being used to train the classifier:

- Probability of remaining in the same state, $P(H_{t+1} = i | H_t = i)$.
- Probability of switching to one of the other remaining states, $P(H_{t+1} \neq i | H_t = i)$
- Covariance of X coordinate averaged over all states
- Covariance of Y coordinate averaged over all states
- Mean Displacement
- Mean Distance
- Covariance of Displacement
- Covariance of Distance

Unlike the individual HMMs, these processed parameters are now comparable across the HMMs trained from the different experiments.

Set Up 2 - Realigning the HMMs

Secondly, we developed a simple heuristic search to realign the HMM parameters so that the discovered states fitted one another as best they could by minimizing the difference between their probability matrices. The number of possible combinations of alignments grows exponentially with the number of experiments (the number of possible alignments is K^M where K is the number of hidden states and M is the number of experiments). As a result, we exploited a simulated annealing approach to optimize the realigned HMMs where the score was calculated using the Euclidean distance between the parameter matrices. This means that the parameters that are used by the classifier are simply the transition probabilities, the means and covariances of the x and y coordinates, the distance and the displacement for each state (unlike in set up 1 where the parameters are processed to deal with issues of HMM alignment).

For both Set Up 1 and Set Up 2, once the parameters are processed or aligned depending on the set up adopted, a simple decision tree is used to classify the HMMs based upon the experiment they were learnt from using their parameters as features. Here we use the C4.5 Decision Tree induction algorithm [24].

Set up 3 - A HMM Classifier

We also compared the two approaches above with a HMM classifier [6]. This works by simply training a HMM for each class of experiment, TGF β and Basal. Therefore a HMM was built that modelled trajectories from the TGF β experiments and another from the Basal experiments. A measure of fit can then be applied to new trajectories to determine the most likely classifier that generated the data and therefore, classified the data. We calculated the log likelihood using the junction tree algorithm [25] to determine class in our experiments.

3 Results

Firstly, we experimented with the use of the standard HMM architecture to model the data but found through simulations and inspection of the parameters of the learnt models, that the smooth nature of the vesicle movements was not captured. Standard HMMs resulted in trajectories that jumped from one region to another unrealistically. Therefore, we applied a variation known as the Auto Regressive HMM (ARHMM). The ARHMM is illustrated in Figure 3, where the structure is similar to the standard HMM but the distribution for the observed nodes are now conditioned upon the previous value of the observed variables as well as the hidden variable. We experimented with different numbers of hidden states and found that 3 appeared to be the minimum for capturing typical trajectory behaviour. However, to capture the behaviour of vesicles in our experimental data we found 5 hidden states were necessary to model more specific behaviours without averaging over a number of different behaviours.

By simulating the models learnt from different experimental conditions (basal and TGF β exposure) we have found that there are clear differences in their dynamics. For control data, one hidden state is dominant throughout and there is no considerable change between the classes (see Figure 4 for sample data). However, the TGF β data generates models that are far

less stable with a high probability of state changes (see Figure 5). It seems that the steady state basal data involves relatively small movements whereas the TGF β data involves a mixture of states. This results in differing movements of the vesicles- small jumps and larger jumps. One state appears to involve massive jumps, which could be also due to errors in the tracking process where a particle disappears from the image and reappears later in a distant position. However, this is not very likely as both control and TGF β samples were run under identical conditions.

Despite these circumstantial differences, multidimensional plotting methods indicate that the basal data and TGF β data are not as easily separable as hoped with a large degree of overlap between data points from the two classes of experiment (compare also figure 4 and 5 where most states cluster around the 0 to -2 x values). Therefore, we use supervised learning of a Decision Tree (DT) classifier based upon the C4.5 algorithm [24] to classify the models based upon their parameters.

Set up 1 (Parameter Processing)

In order to see to what degree the DT models can be used for predicting whether TGF β was introduced, we applied leave-one-out cross-validation (LOO-CV) to the processed parameters. The resulting accuracy was 75% and the resulting DTs (see Figure 6) showed that the displacement associated with the most probable hidden state appears the most predictive. In other words, states with high dispersion were more likely to be from HMMs learnt from trajectories derived from cells after exposure to TGF β . Other predictive features were the probabilities associated with hidden state change / stability and the covariance of coordinates; the unstable models seem to be associated with less obvious trajectory behaviour, often switching between directed and undirected movements, and between large and small jumps. The full statistical results of the sensitivity analysis are documented in Table 1.

Set Up 2 (HMM realignment)

We now turn to the HMM state alignment results where we relabel states from each HMM to "best fit" one another. This allows us to say, for example, that state A represents a highly stable small displacement state across all experiments. It also means that we do not need to apply any processing to the parameters as we did in the previous DT. The results of applying the alignment are illustrated in the form of a transition diagram for basal experiments and TGF β -treated experiments that are applicable to all HMMs. Figure 7 shows these transition diagrams where links represent transition probabilities greater than 0.5 that appear in at least half of each set of experiments. The figure also shows the general characteristics of the state in terms of the expected distance (Dist in the figure) and mean displacement (Disp in the figure) parameters. Interestingly the only difference between the transitions for basal and TGF β -treated experiments is the link between state D and state E. This implies that the E, A, and B states capture TGF β -induced behaviour more. State B is the most interesting as it seems to capture a very distinct behaviour between the two sets of experiments with high displacement but low distance for TGF β data and the opposite for basal. The full breakdown of the expected parameters for each state is shown in Figure 9.

The decision tree learnt from the aligned HMMs (Figure 8) highlights the importance of State C. It features in the first three decision nodes because they either involve transitions to State C or the overall probability of being in state C (p_C in the Figure). However, it is also evident that some of the less probable transitions are important in differentiating between TGF β -treated and basal trajectories (as these probabilities do not appear in Figure 7). For example, the AC transition is the most significant feature in the tree which captures a change from medium distance and displacement to low in the basal data but not in the TGF β . This implies that medium distance and displacement is evident in the basal experiments but “not for long” as the transition probability table implies that there will be a return to smaller distances and displacement (as characterised by state C). To further support this hypothesis, the probability of being in state C is lower for TGF β data (as demonstrated in the second decision node of Figure 8). The decision tree achieved an accuracy of 65% when applying LOO-CV which, albeit lower than the processed data, is considerably better than chance and what is more, the resulting models offer the power to explore individual states more fully. The full statistical results of the sensitivity analysis are documented in Table 2.

Trajectories from three independent experiments before and after TGF β treatment were also classified according to the shape of their Mean Square Displacement (MSD) curves [26, 8]. 18 trajectories from each of the six image sequences were assigned one of four categories - simple diffusion (SD), diffusion plus directed motion (DD), restricted diffusion (RD), or stationary (ST) (see Figure 10 for examples). Trajectories with linear displacement curves were classified as SD. Trajectories whose MSD plot had an upward curve were classified as DD, and those with a downward curve were classified as RD. It seems clear that the discovered states in the realigned HMMs map directly onto these trajectory characterisations. For example, SD appears to be represented in state A (both basal and TGF β) and E (TGF β only). DD is represented in state D (both) and state E (basal only). RD is represented in state B (TGF β only). In our models, these states have been found automatically using a realignment strategy and would not have been possible otherwise as the states would represent different behaviours in the different HMM models as they are learnt using a different execution of the EM algorithm (on different experimental data).

Set Up 3 (HMM Classifier)

For the HMM classifier we trained a HMM for each class of experiment, TGF β and Basal, and applied a measure of fit to new trajectories to determine which is the most likely classifier that generated the data. We calculated the log likelihood using the junction tree algorithm [25].

The results of the HMM Classifier are shown in Table 3. These statistics are similar to those generated from Set Up 2 except that the classifier is more specific than sensitive. Overall accuracy is almost identical, however.

To summarise the results, firstly the HMMs learnt from the individual experiments (as in Set Up 1 and 2) seem to capture the general behaviour of the particles very well with stable small jumps in the basal experiments and larger erratic behaviour for the TGF β treated ones. What is more, the pre-processing of these parameters (in Set Up 1) allows us to successfully predict the experiment type from the dynamic models (with an accuracy of 75%). The novel

approach for aligning HMM parameters so that the states discovered from different experiments can be directly compared (Set Up 2) leads to a lower overall accuracy (decision trees learnt from these parameters have 65% accuracy). However, this was no worse than the accuracy of the standard HMM classifier from Set Up 3 (see Table 3 - 64%). Furthermore, the models from Set Up 2 allowed us to build common transition matrices for all HMMs and analyse the resulting key states and transitions, whilst keeping the states and parameters in the individual HMMs from each experiment explicit within our decision tree classifier. As a result, the discovered states appear to fit well into previously documented typical trajectory types known as directed, simple and restricted. In contrast, interpretation of Set Up 3 is difficult without an explicit transparent classifier (such as a DT) to explore.

4 Conclusions and Future Work

This paper explores the use of a combination of auto-regressive hidden Markov models with supervised learning of the model parameters to predict the experimental conditions of particle trajectory data. We explore three approaches: Firstly we explore a method for summarising HMM parameters that have been learnt on a different experiment with an associated class label. These processed parameters are then fed into a decision tree to generate trees that help to explain the classifications. This approach has proved very good at modelling this data and has captured the dynamics of the two experimental conditions nicely, with a stable hidden state for control data and multiple (less stable) states for the experiments that involve cells that have been exposed to a growth factor, known to elicit a wide range of cellular responses. Even less stable states could identify possible noise in the experimental set up where vesicles shift coordinates dramatically. What is more, the features identified by the decision tree classifier distinguish the conditions, focussing on the probability of changing the internal hidden state, the dispersion and the distance metrics. Secondly, we explore a method that automatically aligns the individual HMMs without a need for summarising the parameters, prior to decision tree learning. This means that we can analyse the behaviour of each state across all experiments to understand general characteristics of the two classes of experiment. The approach resulted in interesting decision trees involving state transitions that help to understand the different trajectories. However, classification accuracy was reduced. Finally we compare the two novel methods to a standard HMM classifier, which generates similar accuracies to the aligned HMM but without the explicit explanation of the underlying states and classifications in the form of a decision tree.

Future work will explore other dynamic models such as dynamic Bayesian networks [6, 27], where the relationship between variables will be made more explicit in the directed links of the model structure, and hierarchical Bayesian networks [28], where different levels of latent variable can be incorporated. The use of variational methods [29] in state space models may also be worth exploring to see if more complex dynamics occur in the trajectories over time. Bayesian networks are themselves well-suited to explanation and a unification of the dynamic model with the explanatory classifier would be a sensible way forward thus removing the need for the decision tree. Bayesian classifiers [30, 31] are already very popular and some way to integrate these with dynamic Bayesian networks will be explored. Furthermore, we will be testing our approach on other live cell time-lapse data to determine

its robustness as a predictor of cellular responses. As outlined in the introduction some vesicles may take a more direct route to the cell membrane whilst others are maintaining an “explorative” behaviour, which may result in a delay on arrival of the cargo at the plasma membrane and an increased possibility of mistargeting. Differences in vesicular behaviour maybe a direct consequence of growth factor signalling or treatment with inhibitors. In this paper we have shown that our approach highlights these differences very well and will enable us to create a “fingerprint” of the specific treatment/condition to facilitate the study of alterations to vesicular trafficking in health and disease.

Acknowledgements

The work of DR and GS is supported by an Arthritis Research UK Grant (18197).

References

- [1]. Minoshima S, Cross D. In vivo imaging of axonal transport using MRI: aging and Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2008; 35(Suppl 1):S89–92. [PubMed: 18204931]
- [2]. Meulenbelt I, et al. Clusters of biochemical markers are associated with radiographic subtypes of osteoarthritis (OA) in subject with familial OA at multiple sites. The GARP study. *Osteoarthritis Cartilage*. 2007; 15:379–85. [PubMed: 17052923]
- [3]. Lippincott-Schwartz J. Dynamics of secretory membrane trafficking. *Ann N Y Acad Sci*. 2004; 1038:115–124. [PubMed: 15838105]
- [4]. Jaqaman, et al. Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*. 2008; 5:695–702. [PubMed: 18641657]
- [5]. Sbalzarini IF, Koumoutsakos P. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology*. 2005; 151:182–95. [PubMed: 16043363]
- [6]. Friedman, N; Murphy, KP; Russell, SJ. Proc of the Annual Conference on Uncertainty in AI; 1998. 139–147.
- [7]. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*. 1989; 77(2):257–286.
- [8]. Steyer JA, Almers W. Tracking single secretory granules in live chromaffin cells by evanescent-field fluorescence microscopy. *Biophysical Journal*. 1999; 76:2262–71. [PubMed: 10096921]
- [9]. Manneville JB, et al. Interaction of the actin cytoskeleton with microtubules regulates secretory organelle movement near the plasma membrane in human endothelial cells. *Journal of Cell Science*. 2003; 116:3927–3938. [PubMed: 12928328]
- [10]. Bellazzi R, Diomidous M, Sarkar IN, Takabayashi K, Ziegler A, McCray AT. Data Analysis and Data Mining: Current Issues in Biomedical Informatics. *Methods of Information in Medicine*. 2011; 50(6):536–544. [PubMed: 22146916]
- [11]. Peek N, Combi C, Tucker A. Biomedical Data Mining. *Methods of Information in Medicine*. 2009; 48(3):225–228.
- [12]. Tucker A, Vinciotti V, Garway-Heath D, Liu X. A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artificial Intelligence in Medicine*. 2005; 34:163–177. [PubMed: 15894180]
- [13]. Ramoni M, Sebastiani P, Cohen P. Bayesian Clustering by Dynamics. *Machine Learning*. 2002; 47(1):91–121.
- [14]. North B, Blake A, Isard M, Rittscher J. Learning and Classification of Complex Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(9):1016–1034.
- [15]. Kadous, MW. A general architecture for supervised classification of multivariate time series. Technical Report, School of Computer Science & Engineering, University of New South Wales; 1998. Report No.: UNSW-CSE-TR-9806
- [16]. Das R, et al. A Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between LFA-1 and the Actin Cytoskeleton. *PLoS Comput Biol*. 2009; 5(11):e1000556.

- [17]. Hamilton, JD. Time Series Analysis. Princeton University Press; Princeton: 1994.
- [18]. Ratnayaka A, Paraoan L, Nelson G, Spiller DG, White MR, Hiscott P. Trafficking of osteonectin by retinal pigment epithelial cells: evidence for basolateral secretion. *Int J Biochem Cell Biol.* 2007; 39:85–92. [PubMed: 17035068]
- [19]. Steyer JA, Almers W. A real-time view of life within 100 nm of the plasma membrane. *Nat Rev Mol Cell Biol.* 2001; 2:268–275. [PubMed: 11283724]
- [20]. Kanaan RA, Kanaan LA. Transforming growth factor beta1, bone connection. *Med Sci Monit.* 2006; 12(8):RA164–9. [PubMed: 16865078]
- [21]. Edlund S, et al. Transforming growth factor-beta-induced mobilization of actin cytoskeleton requires signaling by small GTPases Cdc42 and RhoA. *Mol Biol Cell.* 2002; 13:902–914. [PubMed: 11907271]
- [22]. Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell.* 2003; 113:685–700. [PubMed: 12809600]
- [23]. Zhang YE. Non-Smad pathways in TGF-beta signaling. *Cell Res.* 2009; 19:128–139. [PubMed: 19114990]
- [24]. Quinlan, JR. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers; 1993.
- [25]. Jensen, FV. Bayesian Networks and Decision Graphs. Springer-Verlag; New York: 2001.
- [26]. Qian H, et al. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal.* 1991; 60:910–921. [PubMed: 1742458]
- [27]. Ghahramani, Z. Learning Dynamic Bayesian Networks Adaptive Processing of Sequences and Data Structures. Lecture Notes in Artificial Intelligence. Giles, CL, Gori, M, editors. Berlin: Springer-Verlag; 1998. 168–197.
- [28]. Gyftodimos, E; Flach, P. Hierarchical Bayesian Networks: A Probabilistic Reasoning Model for Structured Domains. In: de Jong, Edwin; Oates, Tim, editors. Proceedings of the ICML-2002 Workshop on Development of Representations; 2002 Jul. 23–30.
- [29]. Ghahramani Z, Hinton GE. Variational Learning for Switching State-Space Models. *Neural Computation.* 2000; 12(4):831–864. [PubMed: 10770834]
- [30]. John, GH; Langley, P. Estimating continuous distributions in Bayesian classifiers. UAI'95 Proceedings of the Eleventh conference on Uncertainty in artificial intelligence; San Francisco: Morgan Kaufmann; 1995.
- [31]. Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. *Machine Learning.* 1997; 29:131–163.

Fig. 1.

Image of living cell transfected with a green fluorescent protein chimera that is transported in vesicles towards the plasma membrane (A). Movement of vesicles was observed for 30 seconds with total internal reflection microscopy and individual vesicle tracks were detected by computer software (B).

Fig. 2.

The typical Hidden Markov Model architecture where H_i represents the hidden state in position i and x_i represents the observed variable x in position i .

Fig. 3.

The Auto Regressive Hidden Markov Model architecture where observed variables are conditioned upon previous observations.

Fig. 4.

Top: 10 sampled trajectories of the particles within the cell generated from AR-HMMs learnt from basal data. Bottom: a zoomed-in plot showing some differing characteristics of the trajectories. The different symbols used represent different discovered hidden states.

Fig. 5.
10 sampled trajectories of the particles within the cell generated from AR-HMMs learnt from TGF β data. The different symbols used represent different discovered hidden states. The scale is the same as the top plot in Fig 4 for comparison.

Fig. 6.

Decision Tree learnt from the 5-state HMM parameters using experimental set up 1 with processed parameters. *N* represents the Basal experiments and *Y* represents TGFβ-induced experiments. *probchange* represents the probability of changing state, *probstable* represents the probability of staying in the same state, *mostprobdisp* represents the mean value for displacement when in the state with the highest probability, and *mostprobdist* represents the mean value for distance for the most probable state. Finally, *mostprobcovx* represents the covariance of the *x* coordinate for the most probable state.

Fig. 7.

The “realigned” state transition diagrams (links represent transitions that appear with $p > 0.5$ for the majority of experiments for basal and TGF β). The dashed line only appears in the TGF β transitions.

Fig. 8.

Decision Tree learnt from the 5-state HMMs having been realigned (allowing more exploration of the states). pC represents probability of being in state C , AC represents the probability of switching from state A to state C . Similarly, DC and EB represents the probability of switching from state A and E to state C and B respectively. $covdist_A$ represents the covariance of the distance for state A . N represents the Basal experiments and Y represents TGF β -induced experiments.

Fig. 9.

Summaries of the parameter values for the different discovered states for TGF β -induced and Basal Hidden Markov Models.

Fig 10.
 (Top) Samples of different manually labelled trajectories: Directed, Simple and Restricted.
 (Bottom) Mean Square Displacement (MSD) calculations for these trajectories.

Table 1

Result of leave-one-out cross-validation on Set-Up 1 using the processed HMM parameters to build a Decision Tree where AUC represents the area under the ROC curve.

Accuracy	Sensitivity	Specificity	AUC
0.75	0.9	0.6	0.745

Table 2

Result of leave-one-out cross-validation on Set-Up 2 using the realigned HMM parameters to build a Decision Tree

Accuracy	Sensitivity	Specificity	AUC
0.65	0.7	0.6	0.65

Table 3

Result of leave-one-out cross-validation on Set-Up 3 using the standard HMM Classifier

Accuracy	Sensitivity	Specificity	AUC
0.64	0.59	0.7	0.65