bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



# Stochastic Protein Labeling Enables Long-term Single Molecule Observation *in vivo*

Hui Liu, Peng Dong, Maria S. Ioannou, Li Li, Jamien Shea, H. Amalia Pasolli, Jonathan Grimm, Pat Rivlin, Luke D. Lavis, Minoru Koyama, Zhe Liu *

Janelia Research Campus

Howard Hughes Medical Institute

19700 Helix Dr. Ashburn, VA 20147

* Corresponding Author: liuz11@janelia.hhmi.org

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



## ABSTRACT

Our ability to unambiguously image and track individual molecules in live cells is limited by packing of multiple copies of labeled molecules within the resolution limit. Here we devise a universal genetic strategy to precisely control protein copy number in a cell. This system has a dynamic titration range of more than 10,000 fold, enabling sparse labeling of proteins expressed at widely different levels. Combined with fluorescence signal amplification tags, this system extends the duration of automated single-molecule tracking by 2 orders of magnitude. We demonstrate long-term imaging of synaptic vesicle dynamics in cultured neurons as well as in live *zebrafish*. We found that axon initial segment utilizes a 'waterfall' mechanism gating synaptic vesicle transport polarity by promoting anterograde transport processivity. Long-time observation also reveals that transcription factor Sox2 samples clustered binding sites in spatially-restricted sub-nuclear regions, suggesting that topological structures in the nucleus shape local gene activities by a sequestering mechanism. This strategy thus greatly expands the spatiotemporal length scales of live-cell single-molecule measurements for a quantitative understanding of complex control of molecular dynamics *in vivo*.

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



## INTRODUCTION

Life is orchestrated by animate molecular processes in living cells. Decoding the dynamic behavior of single molecules enables us to investigate biology at this fundamental level that is often veiled in ensemble experiments. Sparse labeling is critical for live-cell single-molecule observation[1,2], because high copy numbers ($10^3$ ~ $10^7$) of protein molecules co-occupy the small volume inside of a cell (10 ~ 20 µm in diameter). Thus, multiple protein molecules are usually packed within the light diffraction limit (200 ~ 300 nm). For example, in a small area (10 µm × 10 µm) comparable to the size of a mammalian cell nucleus, single-molecule images begin to overlap with each other at low concentrations (> $10^2$ ~ $10^3$ copies), preventing precise localization (**Fig. S1A-B**). Accurate tracking of single molecules requires labeling at much lower densities (< $10^2$ copies) to link positions of a single molecule across multiple frames with high fidelity (**Fig. S1B**). Currently, the most reliable live-cell single-molecule imaging method is by stochastic switching of photoactivatable fluorescence proteins (FPs) or dyes (*e.g.* single particle tracking photo-activated localization microscopy - sptPALM or direct stochastic optical reconstruction microscopy - dSTORM)[2-9]. The stochastic activation separates excitations of individual fluorophores temporally, allowing single-molecule imaging in densely labeled samples. However, under these conditions, only one fluorophore per molecule is utilized for imaging. Thus, high laser powers (> $10^3$ W/cm$^2$) potentially damaging to live specimens are applied to generate sufficient photon counts for localization[10]. As a result, fluorescent molecules are quickly bleached and intracellular single-molecule dynamics can only be probed on relatively short time scales. Although sub-cellular diffusivity maps can be generated from dense short trajectories[11,12], important dynamic aspects (*e.g.* active transport processes and long-lived transcription factor chromatin binding dynamics) of molecular systems in live cells are being overlooked due to lack of tools to observe single molecules for extended periods of time.

To address this general challenge in live imaging, we set to devise a universal genetic tool to precisely titrate protein labeling densities in live cells and thus we can bypass the dependence on photoactivatable XFPs for single-molecule imaging. Several desirable features for such a system are 1) simple to implement genetically, 2) highly-tunable and 3) cell-type independent. 4) Most importantly, this system should be compatible with any fluorescence signal amplification system currently available or developed in the future (*e.g.* 3XGFP, 3XHaloTag, GFP11[13], sfCherry11[13], SunTag[14] and

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



ArrayG[15]) for multicolor single-molecule imaging experiments. It's worth noting that current methods[16] to fine tune protein labeling density include, 1) using weak promoters (*e.g.* CMV*d3*) to decrease mRNA transcript levels[17], 2) inserting an IRES element upstream of a gene to reduce the translation efficiency[18] and 3) utilizing a degron to destabilize the protein[19]. The drawbacks of these methods are apparent. First, the strength of a promoter can be highly cell-type dependent[20] and weak promoters are usually associated with high gene expression heterogeneity in single cells[21,22]. Moreover, we found that even leaky expression from a Tet-on mini CMV promoter was unable to sparsely label dCas9-HaloTag molecules for direct single-molecule imaging[23]. Secondly, the labeling densities generated by weak promoters or the IRES are not tunable and thus these methods might not be generally applicable to sparsely label proteins expressed at different levels. Finally, a degron alters the protein stability and might influence the function of a protein by ubiquitination.

## RESULTS

### A Stochastic Protein Labeling Strategy

One method that we identified as suitable to control labeling density is by translational read-through (RT). Specifically, during protein translation, ribosome sometimes mis-incorporates an amino acid at the stop codon and continues to translate the downstream mRNA sequence. The RT mechanism is conserved from bacteria to mammals[24-26]. Interestingly, the probability that a ribosome translates beyond the stop codon (the RT efficiency) is largely dictated by the sequence of the stop codon and its immediate flanking regions[26]. Because the copy number of mRNA is orders of magnitude higher than that of DNA, simulation experiments suggest that, compared to using a weak promoter, sparse labeling via a RT strategy would benefit from averaging effect at the translational level and give rise to much less labeling heterogeneity in single cells (**Fig. S2A-D**). To test this approach, we performed live-cell PALM[27] experiments with cells expressing Histone H2B and mEOS4b separated by different RT sequences (**Fig. 1A-B**). Although mRNAs were expressed at similar levels (**Fig. S3B**), we observed dramatic reductions of mEOS4b labeling densities upon the RT sequence insertion (**Fig. 1B-C, Fig. S3A** and **Movie S1**). We found that each RT sequence has a distinct labeling probability (RT1; ~5%, RT2; ~ 1.5%, RT3; ~0.4%) (**Fig. 1C**). Importantly, these RT

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



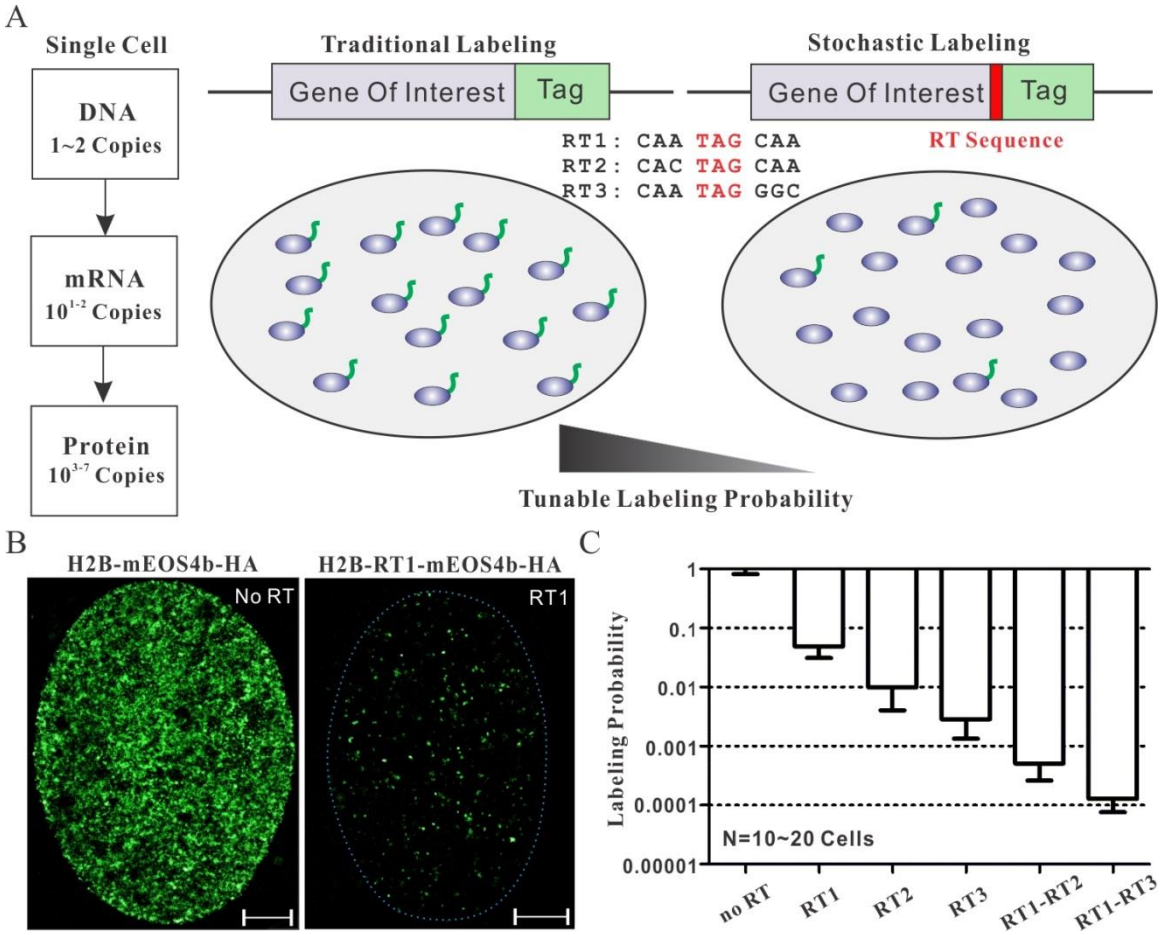



**Figure 1, Sparse Labeling with Translational Read-Through**

(A) Schematics for precise label density control by translational read-through

Left: After transcription and translation, high copy numbers of functional protein molecules ($10^3$ ~ $10^7$) are typically generated in the cell.

Middle: In the traditional labeling strategy, all molecules are labeled and high labeling densities make it challenging to observe single-molecule dynamics (**Fig. S1**)

Right: By placing a short RT sequence between the target protein and the tag, we can accurately control the probability of protein fusion and sparsely label a faction of molecules for single-molecule imaging.

(B) Representative single-cell PALM localization density map for H2B-mEOS4b in U2OS cells without (Left) and with (Right: RT1) read-through (**Movie S1**). In average, 328 localizations per $\mu m^2$ (no RT) and 17.2 localizations per $\mu m^2$ (RT1) were detected within 4500 frames. Bar: 2 µm.

(C) Labeling probabilities for different RT sequences have a large dynamic range (from ~5% down to ~0.001%). Specially, for the probability calculation, H2B-mEOS4b localization density in each RT condition were normalized to the average density in no RT control. Error bars represent standard deviation (SD).

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



probabilities are consistent across 4 different cell types that we tested (**Fig. 1C** and **S3C**). We also discovered that RT sequences work synergistically in tandem, giving rise to even lower labeling probabilities (RT1-RT2; ~0.06%, RT1-RT3; ~0.01%) (**Fig. 1C**). Therefore, the RT system can control labeling density in a highly tunable fashion with a large dynamic titration range (~10,000 fold). If necessary, this range can be further extended by concatenation of RT sequences. It is important to note that, in addition to RT sequences listed here, past studies identified a repertoire of short sequences with RT efficiencies ranging from 0.1% up to ~10%[28]. If necessary, these sequences can be adapted to make fine adjustments to the labeling density. Based on these data, we propose that the highly tunable RT system can be used to stochastically label proteins in live cells for single-molecule imaging (**Fig. 1A**).

## Long-term Observation of Vesicle Trafficking Dynamics in Cultured Neurons and in *zebrafish*

We first sought to apply this stochastic labeling strategy to an important question in neuroscience - how neuronal trafficking polarity is dynamically established. Specifically, synaptic vesicle precursors (SVPs) packaged at the soma must be specifically targeted to presynaptic regions for release and recycle as synaptic vesicles (SVs)[29]. Airyscan super-resolution imaging (**Fig. 2B** and **Fig. S4A**) and electron microscopy (EM) studies [30,31] reveal that high copy numbers of SVPs/SVs are densely packed below the diffraction limit in mammalian neurons, particularly in boutons and presynaptic regions. Consistent with these results, when we labeled a classical synaptic vesicle marker – Synaptophysin (Syp) with traditional strategies, individual SVPs/SVs cannot be distinguished from each other (**Fig. 2A-B**). Next, we implemented the RT strategy to sparsely label Syp with the SunTag (**Fig. S4B**) - a fluorescence signal amplification system capable of directing 24 copies of fluorescence protein to the target[14]. After optimizing the RT sequence, we can directly observe movements of bright, sparsely-spaced SVPs/SVs in different compartments of the cell (**Fig. 2B** and **Movie S2**). The advantages for this strategy are *1)* there are ~ 30 copies of Syp in each SVP/SV[32,33] and on average only one copy of Syp is labeled (**Fig. 2A**). Thus, we expect that the large size (> 1.4 Md) of the SunTag would not significantly affect the SVP/SV function. Indeed, both immunofluorescence staining and high-resolution EM experiments demonstrated that SVPs/SVs labeled with SunTag were correctly targeted to presynaptic regions (**Fig.**

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



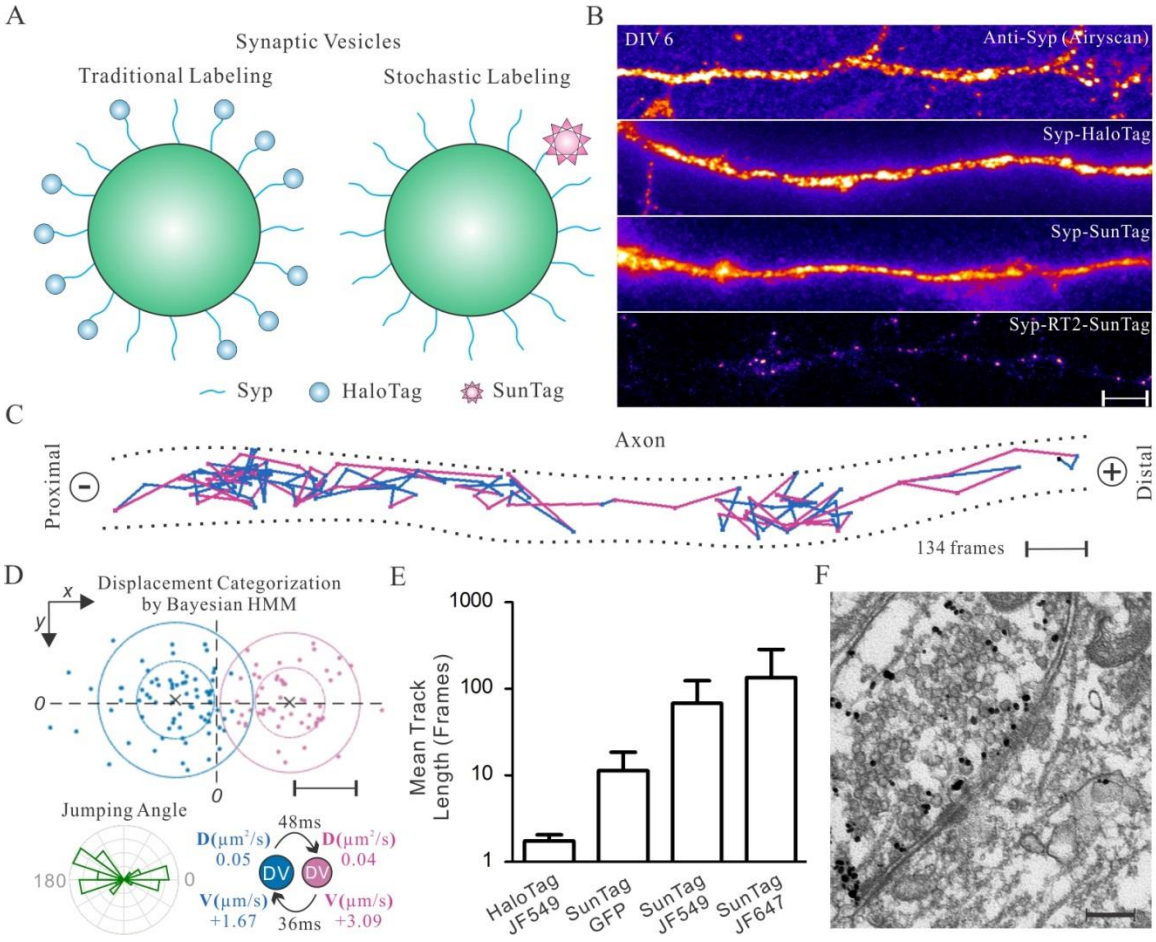



**Figure 2, Automated Long-term Tracking of Individual Synaptic Vesicle Precursors in Live Neurons**

(A) SVP/SV labeling either by tagging Syp with the HaloTag (Left) or sparsely with the SunTag via RT sparse labeling (Right).

(B) High packing densities prevent imaging single SVP/SV particles with traditional labeling methods in DIV 6 cultured hippocampal neurons (From top to bottom, Anti-Syp staining (Airyscan); HaloTag fusion (total internal reflection fluorescence - TIRF); SunTag without RT sparse labeling (TIRF)). When sparsely labelled with the SunTag (lower panel, TIRF) using RT2, individual and sparsely-spaced SVPs/SVs (scFv-HaloTag-JF646) can be observed and imaged. DIV is short for Days In Vitro. Scale bar: 5 μm

(C) One representative trajectory tracking SVP movements in axon: steps are color-coded according to transport modes classified by the HMM-Bayes algorithm shown in (D). The '+' and '–' denote anterograde and retrograde transport directions, respectively. Scale bar: 200 nm. See **Movie S3** for raw imaging data.

(D) Displacements between frames for the trajectory shown in (C) can be classified into two main categories with one transport component pointing to the '+' direction and the other to the '–' direction. Bottom left, Rose histogram for jumping angles for trajectory shown in (C) (See

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



methods for calculation details). Bottom right, transport velocities, diffusion coefficients and mean lifetimes for the two states defined in the upper panel. Scale bar: 100 nm

(E) Track length comparison for different labeling strategies. Data in lane 1 were acquired using the dSTORM strategy with high laser powers (>1 kW/cm$^2$). Data in Lane 2~4 were acquired with relatively low laser powers (~50 W/cm$^2$) with sparse SunTag labeling. Error bars represent SD.

(F) Immuno-EM images showing SunTag labeled SVPs/SVs (RT1) were targeted to presynaptic regions in DIV 12 hippocampal neurons. Scale bar: 200 nm

**2F** and **Fig. S5A-B**). *2)* Due to sparse labeling, the movements of single SVPs/SVs covering large distances in the cell can be unambiguously imaged and tracked with high accuracy (**Fig. 2C, Movie S3** and **S6**). *3)* Because of the signal amplification from the SunTag, ~20 fold less laser powers (~50 mW/cm$^2$) can be used compared to the sptPALM or dSTORM method for single-molecule imaging at high speed (50 Hz). Thus, imaging induced photo-toxicity is greatly reduced without trading off imaging speed. *4)* By incorporating HaloTag and bright Janelia Fluor (JF) dyes[5] into the SunTag system, average trajectory lengths showed > 100 fold improvement over the traditional single HaloTag strategy (**Fig. 2E, Movie S2**). To further validate the applicability of this labeling method, we sparsely labeled SVPs/SVs in live *zebrafish* using the Syp-SunTag RT system (**Fig. S6A**). We can observe dynamic transport of single SVPs along neurites *in vivo* (**Fig. S6B-C** and **Movie S4-S5**).

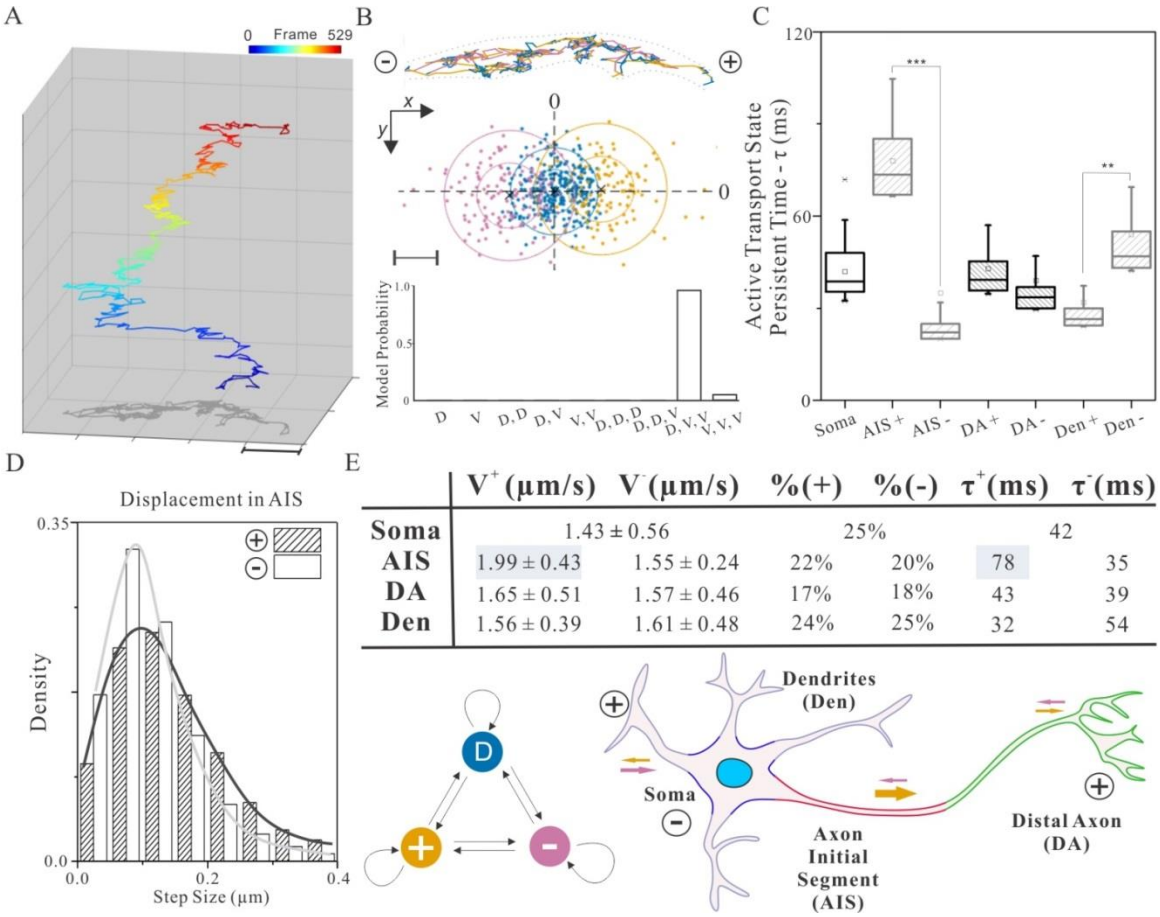
## AIS Promotes SV/SVP Anterograde Transport Dynamics

It is worth noting that SVP/SV diffusion dynamics has been imaged before with traditional GFP labeling[34,35] or extracellular loading of antibody-dye conjugates[36,37]. However, because of dense labeling, these methods do not allow long-term, fast tracking of SVPs/SVs, which is essential for quantifying SVP/SV transport dynamics. Since reliable classification of transitions between active transport and random diffusion requires continuous analysis of displacements in long trajectories[38,39], we applied a recently developed HMM-Bayes algorithm to analyze our SVP/SV tracking data[38]. This program effectively separates SVP/SV diffusion, anterograde and retrograde transport processes (**Fig. 2C-D, Fig. 3B** and **Fig. S6C**) and provides detailed kinetic information, such as diffusion coefficients, transport velocities and lifetimes for each state. Here for

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



the first time, long-term dynamics of individual SVPs/SVs can be imaged and quantitatively analyzed in live mammalian neurons with high spatiotemporal resolution.




| | $V^+$ (μm/s) | $V^-$ (μm/s) | %(+) | %(-) | $\tau^+$(ms) | $\tau^-$(ms) |
|---|---|---|---|---|---|---|
| **Soma** | 1.43 ± 0.56 | | 25% | | 42 | |
| **AIS** | 1.99 ± 0.43 | 1.55 ± 0.24 | 22% | 20% | 78 | 35 |
| **DA** | 1.65 ± 0.51 | 1.57 ± 0.46 | 17% | 18% | 43 | 39 |
| **Den** | 1.56 ± 0.39 | 1.61 ± 0.48 | 24% | 25% | 32 | 54 |

**Figure 3, Axon Initial Segment Gates Synaptic Vesicle Precursor Transport Polarity**

(A) A representative SVP trajectory consists of 529 frames showing dynamic transitions between diffusion and active transport states. See **Movie S6** for the raw imaging data. The frame number is color coded as indicated by the color bar above. Scale bar: 1 μm

(B) Top, individual steps in the trajectory showed in (A) are colored coded by the three hidden states classified in the middle panel. Middle, displacements between frames for the trajectory shown in (A) can be classified by the HMM-Bayes algorithm into one diffusion state (blue) and two transport states with opposing directions - anterograde (+, yellow) and retrograde (-, pink). Bottom, the probabilities of competing models for correctly describing the hidden states in the trajectory. D stands for diffusion and V for active transport. Scale bar: 100 nm

(C) The persistent lifetime for each active transport state in the indicated compartment. '+' and '–' denote anterograde and retrograde transport respectively (Number of cells analyzed > 10, Number of trajectories analyzed > 300). In the box chart, top and bottom error bars represent 95% and 5% percentile, respectively; box represents the range from 25% to 75% percentile; center line represents the median and the small square represents mean value. ***, *p*-value < 0.01; **, *p*-value < 0.05

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(D) The step size distribution for anterograde (+) and retrograde (-) transport processes in the AIS.

(E) Top, statistic summary for anterograde (+) and retrograde (-) transport processes in different compartments of the neuron. V, transport velocity; %, time fraction; τ, state lifetime; the velocities were shown in mean ± standard error of the mean (SEM). Bottom left, hidden markov model diagram describing the dynamic behavior of SVP in the cell. Bottom right, schematics explaining the abbreviations for different functional compartments of the neuron and the directions of active transport. The size of arrows indicates the degree of cargo transport processivity.

All tracks from different compartments were taken from the same datasets to exclude the bias from stage or sample motion.

Next, we divided each neuron into four distinct functional compartments (Dendrites (Den), Soma, Axon initial segment (AIS) and Distal Axon (DA)) based on the localization of Ankyrin G - an AIS specific scaffold protein (**Fig. S7A-B**). The trajectories were rotated to align along the ***x*** axis with anterograde transport pointing to the '+' direction (**Fig. S8** and **Fig. S9)**. Analysis of region-specific SVP/SV dynamics reveal several kinetic features that explain the steady-state SVP/SV distributions and target search mechanisms in neurons: *1)* we found that SVPs/SVs display compartment-specific search kinetics by dynamically partitioning between short-lived (30~80 ms) anterograde and retrograde transport movements interspersed by relatively long-lived slow diffusion states (100~200 ms). (**Fig. 3B, 3C, 3E** and **Fig. S10C**). *2)* In the AIS, the processivity of the anterograde transport is substantially enhanced, with anterograde transport lifetime and average movement velocity significantly increased (**Fig. 3C-E**). At the same time, the slow diffusion state has a substantially shortened lifetime in the AIS (**Fig. S10C**). These observations suggest that specialized functional components in the AIS promote SVP anterograde transport while limiting transient binding and retrograde backtrack movements, in analogy to how a waterfall facilitates unidirectional water flows. *3)* Interestingly, in dendrites and distal axons, the strengths of anterograde and retrograde transport are balanced (**Fig. 3C, 3E** and **Fig. S10A-B**), suggesting that SVPs unbiasedly sample these regions, dynamically searching for binding sites. The slow diffusion fraction is significantly higher in distal axons than in dendrites (**Fig. S10C**), consistent with selective recruitment of SVPs/SVs to boutons and pre-synaptic regions. Previous studies show that the AIS consists of specifically arranged nanoscale cytoskeleton architectures[40,41] and the AIS serves as a passive filter controlling the selective entry of axonal cargos[42]. Data from our high-resolution kinetics analysis are complementary to these observations, highlighting a crucial role of the AIS in actively

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



enhancing the axonal transport polarity by promoting the anterograde transport processivity (**Fig. 3E**).

**Imaging Long-lived TF Chromatin Binding Events**

We next sought to apply this method to imaging an entirely different molecular phenomenon - the binding of transcription factors (TFs) to chromatin. One highly debated area in this field is how long individual TF dwells on chromatin. Two popular single-molecule residence time measurement methods were developed recently: one is based on time lapse imaging[4,43] and the other relies on motion blur by using a long acquisition time of 500ms[1,3]. However, technical problems associated with dense labeling and photo-bleaching prevent these methods from reliably measuring the long-lived chromatin binding events at the single-molecule level. To address this problem, we devised an N-terminal RT sparse labeling strategy, in which the RT sequence is placed upstream of the 3XHaloTag-TF sequence (**Fig. 4A**) and therefore protein over-expression is minimized. It's worth noting that the reason for us to switch from the SunTag to the 3XHaloTag for TF imaging is because the SunTag is about 20~60 fold larger in size (> 1.4 Md) compared to a typical TF molecule. Not surprisingly, it was shown that the SunTag could affect protein dynamics in live cells[14]. In addition, free scFv-GFP/HaloTag molecules gave rise to high fluorescence background in the nucleus, obscuring reliable single-molecule observation in most cells.

After optimizing RT sequence for each TFs, we can visualize single molecules from the first imaging frame without high laser power pre-bleaching step (dSTORM) (**Fig. 4B** and **Fig. S11B**). This is important for two reasons: *1)* because of the sparse labeling, in average only one labeled molecule exists in one diffraction limited area. Therefore, we can safely introduce long dark times (0 ~ 4.5 *s*) between imaging frames without the risk of tracking errors where different nearby molecules across frames are linked as one (**Fig. S11A**) *2)* our measurements do not rely on a single fluorophore switched back from the triplet state (dSTORM). Three copies of bright JF549 dye molecules that are not pre-damaged ensure blinking/switching-proof, long-term observation of single-molecule dynamics (**Fig. 4B, Fig. S11B** and **Movie S7**). When we combined time lapse imaging ($t_{lp}$, 0.5 ~ 5 *s*) and motion blur by using a 500ms exposure time (**Fig. 4A**), we were able to extend the temporal length scale of single-molecule dwell time measurement by ~2

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



orders of magnitude into the 10 ~ 20 mins range (**Fig. 4B-D** and **Movie S7**). For example, average residence times for different classes of chromatin binding proteins (Sox2; 20 ± 7.5 *sec*, Sp1; 76 ± 24.3 *sec*, CTCF; 161 ± 36.2 *sec* and H2B; 1135 ± 267.7 *sec*) measured by this approach are well separated in the temporal domain (**Fig. 4D**), consistent with their distinct roles in gene regulation. These results are in contrast to 12.1 ± 4.5 *sec* (Sox2), 16.5 ± 6.4 *sec* (Sp1), 18.1 ± 3.7 *sec* (CTCF) and 44.6 ± 12.6 *sec* (H2B) measured by the traditional strategy used in previous reports[3,44].

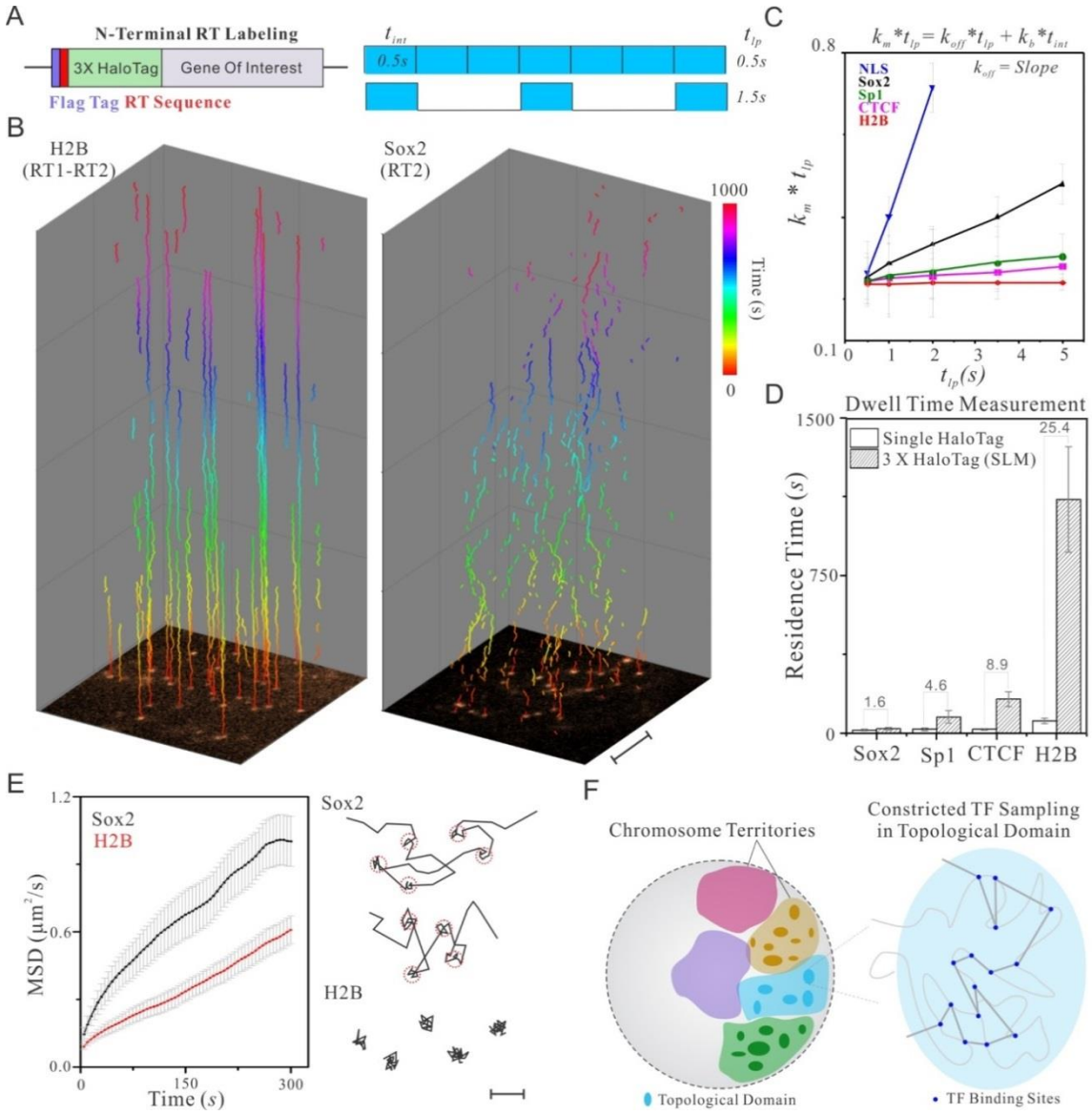



**Figure 4, Long-time Observation of TF-chromatin binding Dynamics in the Nucleus**

(A) Left, the RT sequence is placed between a FLAG tag and a 3X HaloTag to achieve sparse labeling of transcription factors in live cells. Right, Due to the low labeling density, relatively long

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



dark times between imaging frames can be introduced to probe long-lived binding events without tracking contaminations (See **Fig. S11A** for details).

(B) With sparse labeling, single molecules are visible from the first frame shown at the bottom and can be faithfully tracked up to 1000 *sec* for H2B molecules (Left) with a lapse time of 5 *sec*. On the contrary, Sox2 binding is more dynamic (Right). In each trajectory, the time component displayed in the *z* direction is color coded as indicated by the color bar. Scale bar: 5 µm

(C) By using a varied lapse-time imaging method as previously described[4], the true chromatin dissociation rate ($k_{off}$) of a TF can be calculated by using the slope of the linear regression of $k_m \times t_{lp}$ as a function of $t_{lp}$. This calculation renders dissociation rate ($k_{off}$) independent of photo-bleaching rate ($k_b$). See **Eqn. S1-2** for calculation details. Tracking data from 8~10 cells were used in this analysis. Average number of trajectories for each data point: ~1850

(D) Dwell time measurement comparison by using the traditional dense single HaloTag labeling strategy and the 3XHaloTag RT labeling strategy. The number above each bar reflects the fold of dwell time differences.

(E) Left, the mean square displacement (MSD) plots show that, even when we selectively image less mobile fractions using long lapse times (5 *s*) and a long acquisition time of 500ms, Sox2 molecules are more mobile than H2B (Number of Trajectories: 1692 (H2B) ; 1049 (Sox2)). Error bars in (C), (D), and (E) represent SD. Right, a close examination of Sox2 and H2B trajectories shows distinct binding behaviors. Specially, Sox2 hops between adjacent stable binding sites (highlighted by 50 nm red circles). On contrast, H2B molecules remain relatively static at fixed locations. Scale bar: 100 nm

(F) Long-term tracking suggests that Sox2 molecules translocate among binding sites in spatially-restricted regions inside the nucleus, supporting a sequestering mechanism of topological chromosome domains in regulating local gene activities.

## Sox2 Hops Locally in the Nucleus

One long-standing hypothesis in gene regulation is that topological chromosome domains might compartmentalize local gene activities by physically sequestering gene regulatory machinery. This notion was derived from extensive HiC (chromosome conformation capture coupled with high throughput sequencing) experiments[45-47] and has yet to be substantiated by live cell data. Previously, we showed that the stable binding sites of a pluripotent factor - Sox2 are spatially clustered in living embryonic stem (ES) cells[48], suggesting the existence of high-order organization of cis-regulatory elements in the nucleus. However, due to technical problems associated with dense labeling and fast photo-bleaching, we were unable to directly perform long-term tracking experiments to investigate whether Sox2 dynamics is altered in these binding site clusters. Here, long-time observation of slow diffusion events reveals distinctive dynamic differences between Sox2 and a chromatin marker (H2B) at the single-molecule level

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(**Fig. 4B**). Specifically, Mean Square Displacement (MSD) plots suggest that Sox2 molecules are more mobile than chromatin (H2B) even when we selectively imaged slow diffusion events by long lapse times (2 ~ 5 *s*) (**Fig. 4E and Fig. S12**). A detail analysis of individual trajectories revealed that Sox2 dynamically hops locally in the nucleus (**Fig. 4B, 4E**, **Fig. S12** and **Movie S7-S8**). In contrast, H2B molecules are generally more static, bound to fixed positions (**Fig. 4B, 4E and Fig. S12A)**. We note that Sox2 molecules in the same field of view displayed both stable binding and hopping behaviors (**Fig. S12B** and **Movie S7**), excluding the possibility that the hopping behavior is an artifact of cell movement or imaging platform drift. These results suggest that topological structures in the nucleus kinetically facilitate Sox2 molecules to explore locally, confining target search and gene regulatory activities. Our studies thus provide the first evidence supporting a dynamic sequestering mechanism of topological structures in the nucleus of a live cell (**Fig. 4F**).

## CONCLUSION

In summary, here we demonstrate a universal titration handle to control protein labeling density in live cells. Pairing with genetically-encoded fluorescence signal amplification systems, this method enables long-term single-particle tracking of densely-packed cellular organelles and proteins. In addition to imaging applications, the N-terminal RT strategy can also be broadly used to fine tune protein copy number in a cell. Ideally, the C-terminal RT strategy should be combined with genome editing to stochastically label a protein from the endogenous gene locus, while the N-terminal RT strategy should be used by exogenous expression. We note that the strategy outlined here aims to unambiguously track one molecule as long-term as possible to delineate its exquisite dynamic modes of action in live cells. However, if desirable, one could generate local diffusivity maps by sptPALM while monitoring long-term dynamics in a simultaneous two-color imaging experiment. The beauty of this method lies in its simplicity. Only basic microscopy set-ups are required for imaging. However, this strategy can be readily combined with more advanced optical imaging modalities[10], such as lattice light-sheet microscopy[49], adaptive optics microscopy[50] and the MINFLUX method[51] to study the kinetics and control logics of molecular systems with higher spatial and temporal resolution.

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



## METHODS

### Cell Culture

Mouse D3 (ATCC) ES cells were maintained on 0.1% gelatin coated plates in the absence of feeder cells. The ES cell medium was prepared by supplementing knockout DMEM (Invitrogen) with 15% FBS, 1 mM glutamax, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol and 1000 units of LIF (Millipore). NIH/3T3 cells and U2OS cells were maintained in DMEM medium (Corning, without phenol-red) supplemented with 10% FBS, 1 mM glutamax. SH-SY5Y cells were cultured in DMEM/F12 medium (Gibco, without phenol-red) supplemented with 10% FBS.

Electroporation was used to transfect ES cells. Specifically, the Amaxa 4D-Nucleofector System and the kit for Mouse Embryonic Stem Cells (Lonza) were used according to manufacturer's suggestions. We used Lipofectamine 3000 reagent (Invitrogen) for transfecting NIH/3T3 cells, U2OS cells and SH-SY5Y cells. For generating stable cell lines, cells were co-transfected with Piggybac transposon vector (neomycin-resistant) and a helper plasmid that encodes Piggybac transposase (Supper Piggybac Transposase, System Biosciences). Two days post-transfection, the cells were subjected to G418 (Thermo Fisher Scientific) selection.

### Primary Culture of Hippocampal Neurons

We prepared dissociated hippocampal neurons from P0 to 1 Sprague-Dawley rat pups. Briefly, the hippocampi were dissected out and digested with papain (Worthington Biochemical). After digestion, the tissues were gently triturated and filtered with the cell strainer. The cell density was counted. ~2.5 X $10^5$ cells was transfected with indicated constructs by using P3 Primary Cell 4D-Nucleofector X kit (Lonza). After transfection, neurons were plated onto poly-D-lysine (PDL, Sigma)-coated coverslips and maintained in NbActiv4 medium (BrainBits) at 37 °C for indicated days before imaging.

### Plasmids and Molecular Cloning

H2B, Sox2, and Sp1 cDNA were amplified from constructs used in our previous studies[3,52]. CTCF cDNA was amplified from ES cell cDNA libraries. Synaptophysin cDNA

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



and 24xGCN4 (SunTag) cDNA were obtained from Addgene (Plasmid #51509 and Plasmid #72229). The scFv-GCN4-sfGFP-NLS, ankG-mCherry, and SEP-NR1 constructs were obtained from Addgene (Plasmid #60906, Plasmid #42566, and Plasmid #23999). The DNA fragment of 3xHaloTag with M to L substitutions was synthesized by Integrated DNA Technologies. The cDNA fragments were cloned into Piggybac transposon vector (PB533A-2, System Biosciences) or modified Piggybac transposon vector with human synapsin 1 promoter for neuronal specific expression. HaloTag (Promega) was used to replace the sfGFP region to get scFv-GCN4-halo-NLS construct. The plasmids used for fish microinjection were made using the Tol2-Gateway system and the pDEST vector was used as the destination construct.

**Real time PCR**

Total RNA was extracted with Trizol Reagent and reverse-transcribed by SuperScript III Reverse Transcriptase with oligo-dT primer (Thermo Fisher Scientific). cDNA corresponding to 5 ng of total RNA was used in each SYBR Select Master Mix (Thermo Fisher Scientific) reaction. Reactions were performed in duplicates and the results were collected on a CFX96 Touch Real Time PCR Detection System (Bio-Rad).

The PCR primers for H2B-mEOS4b-HA construct are:

Forward: 5-AATGTATGTGCGTGATGGAGTG-3

Reverse: 5-TATGGGTAACCTGAACCTGATC-3

The primers for human GAPDH are:

Forward: 5- AACGGATTTGGTCGTATTGGGC-3

Reverse: 5- CCTGGAAGATGGTGATGGGAT-3

**Western Blot**

Whole cell extracts from U2OS cells were isolated using RIPA buffer that contained Complete Protease Inhibitor Cocktail (Roche). Protein concentrations were measured using Bio-Rad Protein Assay against BSA standards. Protein from each sample was resolved by SDS-PAGE. Primary antibodies used: HA tag (ab9134, Abcam, 1:500) and

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



beta-tubulin (#2128, Cell Signaling Technology, 1:1000). HRP conjugated secondary antibodies (Pierce) were used at a dilution of 1:2000. LumiGLO chemiluminescent substrate (Cell Signaling Technology) was used for HRP detection and light emission was captured by films.

**Immunofluorescence Staining**

Cells were first fixed with 4% paraformaldehyde, permeabilized and blocked with 10% goat serum, 1% BSA, and 0.25% Triton in PBS. Samples were stained with synaptophysin (ab8049, Abcam, 1:500) or synapsin I (AB1543P, Millipore, 1:5000) primary antibodies in PBS containing 10% goat serum, 1% BSA, and 0.1% Triton. Secondary antibodies: DyLight 488 conjugated secondary antibodies (Jackson ImmunoResearch, 1:400).

**Immuno-EM Experiment**

Primary hippocampal neurons were grown on PDL-coated 8-well chamber slide (Thermo Fisher Scientific) for 10~14 days. Pre-embedding immunogold labeling was performed on the wells as previously reported[53]. Briefly, Neuronal cultures were fixed with 4% paraformaldehyde/0.05% glutaraldehyde in sodium cacodylate buffer (0.1 M, pH 7.2). After washing with cacodylate buffer and PBS, samples were blocked with 50 mM glycine in PBS and permeabilized in PBS containing 5% normal goat serum and 0.1% saponin. Then they were incubated with anti-GFP antibody (A-11122, Thermo Fisher Scientific, 1:200) first and Nanogold-Fab' fragment of goat anti-rabbit IgG (Nanoprobes; 1:150) later. After silver enhancement (HQ kit, Nanoprobes) for 6 minutes, the samples were treated with 0.5% osmium tetroxide in phosphate buffer, dehydrated in ethanol and finally embedded in Eponate 12 resin (Ted Pella, Inc) for ultrathin sectioning. Ultrathin sections were contrasted with aqueous uranyl acetate and lead citrate and imaged on a Tecnai Spirit electron microscope (FEI, Hillsboro, OR) using an Ultrascan 4000 (Gatan, Inc) digital camera.

**Microinjection and imaging of *zebrafish***

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Using a *zebrafish* gal4-vp16 driver line under the control of the pan-neuronal Huc promoter, embryos were co-injected at the 1 or 2 cell stage with plasmid DNA of UAS (5X)-syp-RT1-24xGCN4 at 25 ng/µl and 8 ng/µl for UAS (10X)-scFv-GCN4-sfGFP-NLS along with 25 ng/µl Tol2 mRNA. Fish were examined 48~60 hours later for GFP fluorescence and positive ones were picked to do the HILO imaging on a Nikon Eclipse TiE Motorized Inverted microscope equipped with a 100X Oil-immersion Objective lens (Nikon, N.A. = 1.49).

**Cell Labeling Strategy and Preparation for Imaging**

Transfected hippocampal neurons were plated onto an ultra-clean cover glass pre-coated with PDL and cultured for indicated days (DIV 6 or DIV 12). For live imaging, culture medium was replaced with Tyrode's solution (140 mM NaCl, 5 mM KCl, 3 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES, 10 mM glucose, pH 7.35).

Stable ES cell lines were plated onto an ultra-clean cover glass pre-coated with IMatrix-511 (Clontech). ES cell imaging experiments were performed in the ES cell imaging medium, which was prepared by supplementing FluoroBrite medium (Invitrogen) with 10% FBS, 1 mM glutamax, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate, 10 mM HEPES (pH 7.2~7.5), 0.1 mM 2-mercaptoethanol and 1000 units of LIF (Millipore).

Transfected NIH3T3 and U2OS cells were plated onto an ultra-clean cover glass without coating. The NIH3T3 and U2OS imaging medium was prepared by supplementing FluoroBrite medium (Invitrogen) with 10% FBS, 1 mM glutamax, 0.1 mM nonessential amino acids and 1 mM sodium pyruvate.

For the saturated labeling of SunTag (24XHaloTag) or 3X HaloTag, cells were incubated with JF646-HTL or JF549-HTL with final concentration of ~10 nM for 30 mins. HTL stands for HaloTag Ligand. Chemical structures and synthesis procedures of JF549-HTL and JF646-HTL were described previously[5].

To image single molecules without using the sparse labeling strategy, we first tested the optimal HaloTag-JF549 and HaloTag-JF647 labeling concentrations. Several concentrations of JF549-HTL and JF646-HTL (0.5 nM, 1 nM, 2 nM and 5 nM) were used to treat cells for 15 mins and then cells were washed with imaging medium for 3 times. The cover glasses were then transferred to live-cell culturing metal holders and mounted

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



onto the microscope one by one. Proper HaloTag-JF549 or HaloTag-JF646 labeling concentrations were determined by the criterion that single-molecules can be easily detected after no or a minimal 2 ~ 5 *sec* pre-bleaching. After fixing the labeling concentration for each cell line, we then proceeded to perform the 2D single-molecule imaging experiments.

**PALM Imaging to Estimate RT Efficiency**

Live-cell PALM imaging experiments were used to estimate the H2B-mEOS labeling densities (**Fig. 1B-C, Fig. S3C** and **Movie S1**). Specifically, imaging was performed on a Nikon Eclipse TiE Motorized Inverted microscope equipped with a 100X Oil-immersion Objective lens (Nikon, N.A. = 1.49), four laser lines (405/488/561/642), an automatic TIRF illuminator, a perfect focusing system, a tri-cam splitter, three EMCCDs (iXon Ultra 897, Andor) and Tokai Hit environmental control (humidity, 37 °C, 5% $CO_2$). Proper emission filters (Semrock) were switched in front of the cameras for GFP, JF549 or JF646 emission and a band mirror (405/488/561/633; BrightLine quad-band bandpass filter, Semrock) was used to reflect the laser into the objective. For sparse single-molecule imaging, mEOS4b moiety was stochastically converted from 'Green' to 'Red' state with low-dose 405 nm illumination (5~10 W/cm$^2$). Then, single molecules were imaged at 50 Hz using a 561 nm laser with the excitation intensity of ~ 1000 W/cm$^2$. 5000 frames were collected for each cell. For each RT condition, data from 10~20 cells were collected.

**Imaging SVP Dynamics**

SVP particles were imaged using TIRF with our Nikon Eclipse TiE set-up described above. For tracking SunTag-GFP labeled SVPs, we used a 488 nm laser with the excitation intensity of ~ 100 W/cm$^2$. For tracking SunTag-HaloTag-JF549/646 labeled SVPs, we used a laser with the excitation intensity of ~ 50 W/cm$^2$. For imaging HaloTag-JF549 labeled SVPs, we used a 561 nm laser with the excitation intensity of ~1 kW/cm$^2$. An acquisition time of 20 ms was applied for all conditions.

**Single-molecule Imaging of Long-lived Chromatin Binding Events**

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



For imaging immobile TF or H2B binding events in live nuclei, we used the same Nikon Eclipse TiE set-up described above. The TIRF illuminator was adjusted to deliver a highly inclined laser beam to the cover glass with the incident angle smaller than the critical angle. Thus, the laser beam is thus laminated to form a light-sheet (~3 µm) above the cover glass. The oblique illumination (HILO) has much less out-of-focus excitation, compared with the regular Epi-illumination. TFs labeled with 3XHaloTag-JF549 were imaged using a 561 nm laser with the excitation intensity of ~ 50 W/cm$^2$. To specifically probe long-lived binding events, we used 500 ms imaging acquisition time with varied dark times of 0 *sec*, 0.5 *sec*, 1.5 *sec*, 3 *sec* and 4.5 *sec*.

**Single-Molecule Localization, Tracking and Diffusion Analysis**

For 2D single-molecule localization and tracking, the spot localization (x,y) was obtained through 2D Gaussian fitting based on MTT algorithms [54]. The localization and tracking parameters in SPT experiments are listed in the **Table S1**. The resulting tracks were inspected manually. The RT efficiencies (in %, **Fig. 1C** and **Fig. S3C**) were calculated by normalizing single-cell localization density from each RT condition to the average density in the no RT control. Diffusion coefficients (**Fig. 4E** and **Fig. S11C**) were calculated from tracks with at least 5 consecutive frames by the MSDanalyzer [55] with a minimal fitting $R^2$ of 0.8.

**HMM-Bayes Analysis of SVP Dynamics**

The diffusion and transport states of individual SVP trajectories were analyzed by HMM-Bayes program[38] with default parameters. Specifically, maximal 3 states can be inferred from one trajectory. There are possibilities for one or two states as well (*e.g.* **Fig. 2C-D**). The program categorizes each step in the trajectory to either diffusion or transport processes and calculates diffusion coefficient as well as transport velocity. To separate anterograde and retrograde transport processes, we rotate the trajectory and align the neurite along the x axis with the anterograde transport point to '+' direction. Specifically, for each trajectory, we manually defined one proximal and one distal point along the neurite and then the angle ɵ between the vector (from the proximal to the distal point) and the '+' direction of x axis was calculate by *atan2 ()* function in Matlab 2015a.

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Subsequently, the trajectory was rotated by - ɵ. After the rotation correction, the transport processes bifurcated to anterograde and retrograde direction (**Fig. S8**). For each functional compartment (soma, dendrites, distal axon and axon initial segment) in the neuron, the average diffusion coefficients, transport velocities and life-time of different states (**Fig. 2D, Fig. 3C-E** and **Fig. S10**) were calculated. In total, data from 20~30 cells were analyzed for each compartment. The standard deviation reflects cell-to-cell variations.

**TF Residence Time Analysis**

To map stable bound sites in the slow acquisition (500 ms) and time lapse condition, 0.1 $\mu m^2/s$ was set as maximum diffusion coefficient ($D_{max}$) for the tracking. The $D_{max}$ works as a limit constraining the maximum distance ($R_{max}$) between two frames for a particle random diffusing during reconnection. Only molecules localized within $R_{max}$ for at least two consecutive frames will be considered as bound molecules. The duration of individual tracks (dwell time) was directly calculated based on the track length.

All imaging conditions can be fitted well with two-exponential decay model. Only the mean lifetime of the first component was used in this calculation. The measured off rate can be described with the following equation. As described by the previous publication [4].

$$k_m = k_b \times \tau_{\text{int}} / \tau_{lp} + k_{off} \qquad \textbf{Eqn. S1}$$

Where $k_m$ is the measured (observed) off-rate, $\tau_{lp}$ is the lapse time, $k_{off}$ is the TF dissociation constant, $\tau_{\text{int}}$ is camera acquisition time, $k_b$ is photobleaching rate.

Eq.1 can be transformed into Eq.2

$$k_m \times \tau_{lp} = k_b \times \tau_{\text{int}} + k_{off} \times \tau_{lp} \qquad \textbf{Eqn. S2}$$

Therefore, the dissociation rate ($k_{off}$) can be derived through fitting $k_m \times \tau_{lp}$ as a function of $\tau_{lp}$.

**Jumping Angle Analysis**

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



A sliding window of 3 points was applied to each track. The angle between the vectors of the first two and the last two points was calculated by the *acos()* function in the Matlab 2015a. Individual angles were pooled and binned accordingly for the angular Rose histogram (**Fig. 2D** and **Fig. S9B**). The minimal jumping distance between two points is set as 40 nm to ensure that the angle measurement is not significantly affected by the localization uncertainty.

**Labeling Density Simulation**

For a square area (10 µm X 10 µm) of 100 X 100 pixels (**Fig. S1**), the (x, y) positions for *N* single molecules were first randomly generated based on a uniform distribution. The photon counts contributed from each molecule to individual pixels were calculated based on the PSF estimator below.

$$I_i(x,y) = (A_0 e^{\frac{-(x-x_0)^2}{2\sigma_{xy}^2}} e^{\frac{-(y-y_0)^2}{2\sigma_{xy}^2}})_{PSF} \qquad \textbf{Eqn. S3}$$

Specifically, $A_0$ is the signal amplitude; σ is the Standard Deviation (S.D.: ~250 nm) of the Gaussian fit in the indicated direction, in our case S.D. of the x, y direction is the same.

To form the image, we first introduced random white noises ($B_{xy}$) in the area with pixel intensities ranging from 0 ~ $A_0/5$ (Signal-to-Background Ratio > 5). Then, the pixel intensity $I(x,y)$ was calculated by iteratively summing the photon counts contributed from each molecules above the background value, according to equation below:

$$I(x,y) = \sum_1^N I_i\,(x,y) + B_{xy} \qquad \textbf{Eqn. S4}$$

In **Fig. S1**, we used the Einstein's Brownian diffusion equation to calculate the expected jumping distance between frames.

$$< r >= 2\sqrt{D\Delta t} \qquad \textbf{Eqn. S5}$$

$$I_i(x,y) = A_0 e^{\frac{-(x-x_0)^2}{2\sigma_{xy}^2}} e^{\frac{-(y-y_0)^2}{2\sigma_{xy}^2}} \qquad \textbf{Eqn. S6}$$

**Numerical Simulation of Protein Copy Number in the Cell**

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Because the number of reacting DNA molecules is small, random molecular fluctuations become significant and discrete, Gillespie stochastic simulation algorithm (SSA) approach[56] is applied in this case. Such a system can be modeled as a continuous-time Markov process, whose probability distribution obeys what is called a chemical "master equation". We used the Gillespie's Direct Method for the simulation. We first provided a model consisting of a matrix of reaction rates, a function that computes reaction propensities (probability/unit time), an initial state vector, and initial/final times. The SSA functions return the sequence of event times and protein/RNA species amounts.

The master gene expression model is defined by the following reaction and parameters:

$$\begin{tikzcd} {[\mathrm{Gene(I)}]} \arrow[r, shift left, "\alpha"] & {[\mathrm{Gene(A)}]} \arrow[l, shift left, "\beta"] \arrow[r, "\theta"] & {[\mathrm{mRNA}]} \arrow[r, "f\cdot\delta"] \arrow[d, "\tau"] & {[\mathrm{Protein}^*]} \arrow[d, "\gamma"] \\ & & \Phi & \Phi \end{tikzcd}$$

$\alpha$, Gene Activation Rate

$\beta$, Gene Inactivation Rate

$\theta$, mRNA Production Rate

$\delta$, Translation Rate

$f$, Read-Through Rate

$\delta$, Translation Rate

$\alpha$, mRNA degradation Rate

$\gamma$, Protein degradation Rates

$*$, Labeled Protein

According to previous modeling calculation[57], the mean copy number of labeled proteins is given by:

$$< Protein^* >= f\frac{\delta}{\gamma}\frac{\theta}{\tau}\frac{\alpha}{\alpha+\beta} \qquad \textbf{Eqn. S7}$$

In the simulation, we kept mRNA Production Rate ($\theta$), mRNA degradation Rate ($\alpha$), Translation Rate ($\delta$) and Protein degradation Rate ($\gamma$) constant. We used a RT Rate ($f$) of 0.5%, similar to the efficiency of RT3. We adjusted the promoter strength by tuning the fraction of the promoter on time (%), which is equal to $\alpha/(\alpha+\beta)$. For reliable single-

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



molecule tracking, we maintained the average protein copy number in the simulation as ~40 copies per cell. Here we defined the protein noise ($\eta$) as the standard deviation divided by the mean.

### Statistics

Comparisons between two groups in **Fig. 3C** were performed with Student's *t*-test. Error bars in all figures represent SD, except the box charts. Differences were considered to reach statistical significance when $p < 0.05$.

## ACKNOWLEDGEMENT

We thank M. Radcliff, S. Moorehead and C. Morkunas for assistance. We thank M.G. Paez-Segala in Looger lab at Janelia research campus for mEOS4b construct. We thank Y. Liang, R.Turcotte in Ji lab and R. Chhetri in Keller lab at Janelia research campus for helping us test imaging conditions. This work is solely funded by Howard Hughes Medical Institute (HHMI).

## AUTHOR CONTRIBUTIONS

Z.L. and H.L. conceived and designed the experiments. H.L. performed and participated in all experiments. P.D., J.S. and M.K. performed the *zebrafish* experiment. P.D., M.S.I. and L.L. participated in the material preparation and initial test. H.A.P. and P.R. performed the EM experiments. J.G. and L.D.L. performed the JF dye synthesis. Z.L. and H.L. analyzed the data and wrote the paper. Z.L. supervised the research.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



## REFERENCE


1 Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194 (2007).

2 Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645 (2006).

3 Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285 (2014).

4 Gebhardt, J. C. *et al.* Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nature methods* **10**, 421-426 (2013).

5 Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature methods* **12**, 244-250 (2015).

6 Grimm, J. B. *et al.* Bright photoactivatable fluorophores for single-molecule imaging. *Nature methods* **13**, 985-988 (2016).

7 Klein, T. *et al.* Live-cell dSTORM with SNAP-tag fusion proteins. *Nature methods* **8**, 7-9 (2011).

8 Manley, S. *et al.* High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature methods* **5**, 155-157 (2008).

9 Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods* **3**, 793-795 (2006).

10 Liu, Z., Lavis, L. D. & Betzig, E. Imaging live-cell dynamics and structure at the single-molecule level. *Mol Cell* **58**, 644-659 (2015).

11 El Beheiry, M., Dahan, M. & Masson, J. B. InferenceMAP: mapping of single-molecule dynamics with Bayesian inference. *Nature methods* **12**, 594-595 (2015).

12 Persson, F., Linden, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature methods* **10**, 265-269 (2013).

13 Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nat Commun* **7**, 11046 (2016).

14 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646 (2014).

15 Ghosh, R., Draper, W., Franklin, J. M., Shi, Q. & Liphardt, J. A fluorogenic nanobody array tag for prolonged single molecule imaging in live cells. *bioRxiv* (2017).

16 Douglass, A. D. & Vale, R. D. Single-molecule imaging of fluorescent proteins. *Methods Cell Biol* **85**, 113-125 (2008).

17 Watanabe, N. & Mitchison, T. J. Single-molecule speckle analysis of actin filament turnover in lamellipodia. *Science* **295**, 1083-1086 (2002).

18 Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. & Hayakawa, T. IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther* **1**, 376-382 (2000).

19 Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods* **6**, 917-922 (2009).

20 Consortium, F. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-470 (2014).

21 Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633-637 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




22 Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).

23 Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823-826 (2015).

24 Bertram, G., Innes, S., Minella, O., Richardson, J. P. & Stansfield, I. Endless possibilities: translation termination and stop codon recognition. *Microbiol* **147**, 255-269 (2001).

25 Dabrowski, M., Bukowy-Bieryllo, Z. & Zietkiewicz, E. Translational readthrough potential of natural termination codons in eucaryotes - The impact of RNA sequence. *Rna Biol* **12**, 950-958 (2015).

26 Cassan, M. & Rousset, J. P. UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. *BMC Mol Biol* **2**, 3 (2001).

27 Shroff, H., Galbraith, C. G., Galbraith, J. A. & Betzig, E. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nature methods* **5**, 417-423 (2008).

28 Schueren, F. & Thoms, S. Functional Translational Readthrough: A Systems Biology Perspective. *PLoS Genet* **12**, e1006196 (2016).

29 Hirokawa, N. & Takemura, R. Molecular motors and mechanisms of directional transport in neurons. *Nat Rev Neurosci* **6**, 201-214 (2005).

30 Gray, E. G. Axo-somatic and axo-dendritic synapses of the cerebral cortex: an electron microscope study. *J Anat* **93**, 420-433 (1959).

31 Sudhof, T. C. The synaptic vesicle cycle. *Annu Rev Neurosci* **27**, 509-547 (2004).

32 Takamori, S. *et al.* Molecular anatomy of a trafficking organelle. *Cell* **127**, 831-846 (2006).

33 Wilhelm, B. G. *et al.* Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* **344**, 1023-1028 (2014).

34 Maeder, C. I., San-Miguel, A., Wu, E. Y., Lu, H. & Shen, K. In vivo neuron-wide analysis of synaptic vesicle precursor trafficking. *Traffic* **15**, 273-291 (2014).

35 Nakata, T., Terada, S. & Hirokawa, N. Visualization of the dynamics of synaptic vesicle and plasma membrane proteins in living axons. *The Journal of cell biology* **140**, 659-674 (1998).

36 Westphal, V. *et al.* Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science* **320**, 246-249 (2008).

37 Kraszewski, K. *et al.* Synaptic vesicle dynamics in living cultured hippocampal neurons visualized with CY3-conjugated antibodies directed against the lumenal domain of synaptotagmin. *J Neurosci* **15**, 4328-4342 (1995).

38 Monnier, N. *et al.* Inferring transient particle transport dynamics in live cells. *Nature methods* **12**, 838-840 (2015).

39 Bouzigues, C. & Dahan, M. Transient directed motions of GABA(A) receptors in growth cones detected by a speed correlation index. *Biophys J* **92**, 654-660 (2007).

40 Leterrier, C. *et al.* Nanoscale Architecture of the Axon Initial Segment Reveals an Organized and Robust Scaffold. *Cell Rep* **13**, 2781-2793 (2015).

41 Xu, K., Zhong, G. S. & Zhuang, X. W. Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal Structure in Axons. *Science* **339**, 452-456 (2013).

42 Song, A. H. *et al.* A selective filter for cytoplasmic transport at the axon initial segment. *Cell* **136**, 1148-1160 (2009).

43 Normanno, D. *et al.* Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun* **6**, 7537 (2015).

44 Swinstead, E. E. *et al.* Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* **165**, 593-605 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/116186; this version posted March 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



45 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).

46 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

47 Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* **16**, 245-257 (2015).

48 Liu, Z. *et al.* 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *Elife* **3**, e04236 (2014).

49 Chen, B. C. *et al.* Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 439-+ (2014).

50 Ji, N., Milkie, D. E. & Betzig, E. Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues. *Nature methods* **7**, 141-147 (2010).

51 Balzarotti, F. *et al.* Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355**, 606-612 (2017).

52 Li, L. *et al.* Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. *Elife* **5**, e17056 (2016).

53 Tao-Cheng, J. H. Activity-related redistribution of presynaptic proteins at the active zone. *Neuroscience* **141**, 1217-1224 (2006).

54 Serge, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nature methods* **5**, 687-694 (2008).

55 Tarantino, N. *et al.* TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures. *The Journal of cell biology* **204**, 231-245 (2014).

56 Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical-Reactions. *Abstr Pap Am Chem S* **173**, 128-128 (1977).

57 Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (2006).