A New Model for Single-Molecule Tracking Analysis of Transcription Factor Dynamics

David A. Garcia[1,2,7], Gregory Fettweis[1,7], Diego M. Presman[1,3,7], Ville Paakinaho[1,4], Christopher Jarzynski[2,5,6], Arpita Upadhyaya[2,6], and Gordon L. Hager[1,*]

[1]Laboratory of Gene Expression and Receptor Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20893, USA.

[2]Department of Physics, University of Maryland, College Park, MD 20742, USA

[3]IFIBYNE, UBA-CONICET, Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Argentina

[4]Institute of Biomedicine, University of Eastern Finland, Kuopio, PO Box 1627, FI-70211 Kuopio, Finland.

[5]Department of Chemistry, University of Maryland, College Park, MD 20742, USA

[6]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

[7] These authors contributed equally to this work

* Correspondence to: hagerg@exchange.nih.gov

Running title: Power-law behavior in transcription factor dynamics

Keywords: Transcription factor, glucocorticoid receptor, power-law, Bayesian Statistics, chromatin, DNA binding, photobleaching, bi-exponential.

Abstract

Single-molecule tracking allows the study of transcription factor dynamics in the nucleus, giving important information regarding the search and binding behavior of these proteins with chromatin *in vivo*. However, these experiments suffer from limitations due to photobleaching of the tracked protein and assumptions about the exponential behavior required for data interpretation, potentially leading to serious artifacts. Here, we developed an improved method to account for photobleaching effects, theory-based models to accurately describe transcription factor dynamics, and an unbiased model selection approach to determine the best predicting model. A new biological interpretation of transcriptional regulation emerges from the proposed models wherein transcription factor searching and binding on the DNA results in a broad distribution of binding affinities and accounts for the power-law behavior of transcription factor residence times.

Introduction

Transcription factors (TFs) are key regulatory proteins responsible for turning genes “on” and “off” by binding to enhancer or promoter elements across the genome[1]. Fluorescence microscopy techniques have revolutionized our understanding of how TFs search and interact with chromatin[2]. Fluorescence recovery after photobleaching (FRAP) in live-cell systems unveiled the dynamic nature of these proteins, in contrast to the long-standing static model of TF-chromatin interactions[3]. The combined innovative technological improvements in fluorophore brightness and stability[4], optical set-ups to increase the signal-to-noise ratio[5], and camera speed and sensitivity now allow the

study of single-molecules at an unprecedent temporal and spatial resolution. Single-molecule tracking (SMT) is a powerful technique that allows the characterization of protein dynamics in single, live cells. It is based on detecting and following through time the traces produced by the light emitted from a single fluorophore. When applied to the study of TFs, important information regarding the search and binding dynamics of these proteins can be extracted[2].

The SMT approach (**Fig. 1**) has now been applied for approximately two-dozen TFs in a variety of cellular systems[6]. The current consensus, based on empirical residence time distributions, describes TFs as able to transition between three different states: 1) unbound from DNA (diffusing in the nucleus), 2) non-specifically bound and 3) specifically bound to chromatin (i.e. interacting with specific response elements)[7]. This three-population model is based on the observation that TFs can intermittently stop, and then resume rapid diffusion[8, 9]. Thus, the empirical residence time distribution of the "stopped molecules" have been phenomenologically fitted to families of exponential distributions[10-19], with no underlying normative model for the origins of these distributions[8]. Nevertheless, the bi-exponential fits to the distribution suggest that the DNA-bound population of molecules includes two distinct subpopulations: a short-lived fraction ('fast stops') and a longer-lived fraction ('slow stops'). The slow fraction has been hypothesized to represent specific binding events associated with enhancers or promoters, while the fast fraction is hypothesized to represent non-specific binding to chromatin[11, 12, 15, 17]. Experiments wherein the DNA-binding domain of TFs has been mutated are consistent with this hypothesis as the longer events were reported to be dramatically reduced[10, 12, 17, 18].

Despite the technological advances in studying TF dynamics, a major remaining limitation of the SMT approach is the relatively short fluorescence stability (or tendency to photobleach) of any fluorophore dye. This makes it difficult to distinguish whether a loss of signal is due to an unbinding event, drifting or photobleaching. Hence, careful corrections must be performed to accurately estimate TF dynamics. Unfortunately, photobleaching correction methods vary widely among research groups[17, 20, 21] which highlights the lack of a standard approach to overcome the photobleaching bias of SMT strategies[9, 12, 18, 22].

In this work, we propose a new and improved photobleaching correction method that uses fluorescence dynamics of histones measured at the focal plane under precise SMT acquisition conditions. We then use the glucocorticoid receptor (GR), a ligand regulated transcription factor[23], as a model TF to validate and test the new methodology. This approach unexpectedly reveals a novel, power-law behavior of GR residence time distributions. We then derive theory-based models for TF dynamics and a principled method to obtain optimal model parameters from empirical residence time distributions, using Bayesian statistics. We show that a model of TF searching and binding or a model of a nucleus with broad distribution of binding affinities accounts for the power law behavior of GR residence times obtained after implementing the modified photobleaching correction method. These models exhibit a broad effective distribution of binding affinities, thus challenging the established model with two discrete, bound populations.

RESULTS

Photobleaching: a source of error in single-molecule tracking of TFs

When tracking TFs at the single-molecule level, binding events can be observed as stationary spots (**Fig. 1a-c, Supplementary Video 1**).

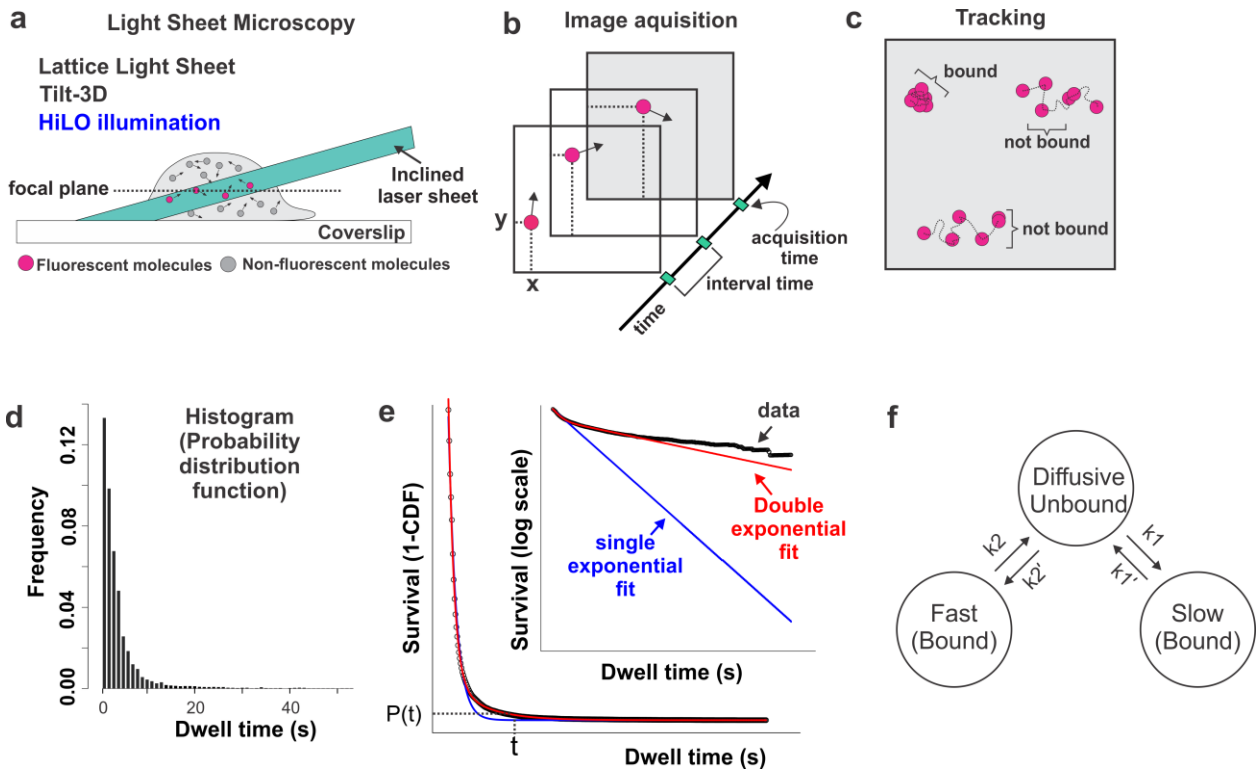

Figure 1. **The current SMT pipeline and interpretation of TF dynamics**. (**a**) A HiLO set-up is most commonly implemented to increase signal-to-noise ratio (shown in panel). A laser beam is tilted and hits the sample creating a thin illumination layer in the focal plane. Any light sheet microscope may be used. (**b**) Several images are taken at specific yet variable acquisition and interval time conditions. (**c**) A tracking algorithm is used to follow each individual molecule and classify them as either bound or unbound. (**d**) From the bound population, a histogram is usually plotted which shows the frequency of TF molecules that are bound for a specific time (dwell time). (**e**) Fitting of the data is performed on the survival plot, which corresponds to 1-CDF, where CDF is the cumulative distribution function. This plot represents the probability P that a molecule will last t number of time points, or longer. This distribution has been phenomenologically fitted to either a single or a bi-exponential distribution. The better fit to a bi-exponential function gave rise to the notion that two-distinct bound populations (fast and slow) exist. (**f**) The current model states that TFs can be found in three different states: unbound from the DNA (diffusing in the nucleus), specifically bound (slow stops), and non-specifically bound (fast stops). The latter can be composed of TFs sliding and hopping on the DNA to facilitate searching of specific sites. Due to the resolution limit, any transition between specific and non-specific bound states cannot be distinguished.

The experimental information recovered is the time the molecule “remains” visible before it bleaches or moves out of the focal plane. From these observations, one can

obtain a local dwell time for TFs which is defined as the time interval between a single molecule transitioning from a diffusive state to a bound state and its subsequent unbinding from DNA. The definition of dwell time presented here differs of the one used in the field (residence time) in that we are considering any process occurring in a point spread function as a single observable binding event in the calculation of the dwell time distribution, not just discrete binding states. The dwell time distribution is generated by integrating the ensemble-averaged distribution of bound times (**Fig. 1d, Supplementary Note 1.1**). Most often, a "survival" distribution, defined as 1-CDF, where CDF is the empirical cumulative distribution function of dwell times, is used for further analysis (**Fig. 1e**). This survival distribution is usually fit to a bi-exponential distribution, interpreted as the "three population model" (i.e. diffusive, fast bound, slow bound) illustrated in **Fig. 1f**.

The upper temporal limit in SMT experiments is ultimately determined by the intrinsic photostability of the chosen fluorophore. When the affinity of bound TFs leads to dwell times longer than the averaged photostability of their fluorescent dyes, residence times cannot be resolved. Importantly, even when bound molecules have relatively lower affinities, they will appear to have shorter experimental dwell times due to photobleaching (PB) bias. To demonstrate this, we conducted single-molecule imaging of the glucocorticoid receptor (GR), a ligand-dependent transcription factor[23], tagged with HaloTag-Janelia Fluor 549 (JF_{549})[7]. If we artificially modulate the PB conditions by changing acquisition parameters (exposure time, interval time, laser power), the resulting kymographs (**Supplementary Fig. 1, Supplementary Videos 1-4)**

appear to have originated from different TFs. Therefore, PB must be properly corrected to prevent artifacts in the analysis of SMT data.

Currently, there is no consensus on the proper method to estimate PB rates. One common approach is to simply count, frame-by-frame, the number of particles in the focal plane and then fit the time-dependent number to a bi-exponential model[8, 10, 11, 16, 17]. However, this method underestimates the real PB that stably-bound proteins undergo at the focal plane. The dwell time of diffusive (unbound) molecules is much shorter than that of bound molecules (**Fig. 2a**) in the focal plane. Therefore, stably-bound molecules will photobleach faster than diffusive particles. Moreover, the number of particles in the focal plane will be heavily dominated by the diffusive component, which will also distort PB rates derived from the fits.

Another strategy uses histones as a proxy for obtaining PB rates. Histones are a good representation of stably-bound proteins because, after integration into chromatin, they have a residence time much longer than the photostability of any currently available organic fluorophore[24]. Therefore, by measuring the residence time of histones, we can obtain a direct representation of PB for particles in the focal plane, as the disappearance of a long-lived particle will most likely represent a PB event. This method is widely used in the SMT field[12, 19, 21 13, 14, 25] but, as we will demonstrate below, it has not been properly implemented. In current models, photobleaching kinetics are characterized by an exponential parameter k_{his} and the slow component of a TF is characterized by another exponential parameter k_{TF} (see **Supplementary Note 2** for details). This rate k_{TF} is corrected for PB to give the "real residence time" (k_{TFreal}) of the transcription factor by: $k_{TFreal} = k_{TF} - k_{his}$. However, the correction assumes that TF

kinetics come from an exponential family since PB kinetics is exponential. Therefore, the parameter k_{TFreal} may emerge as an artifact of this assumption.

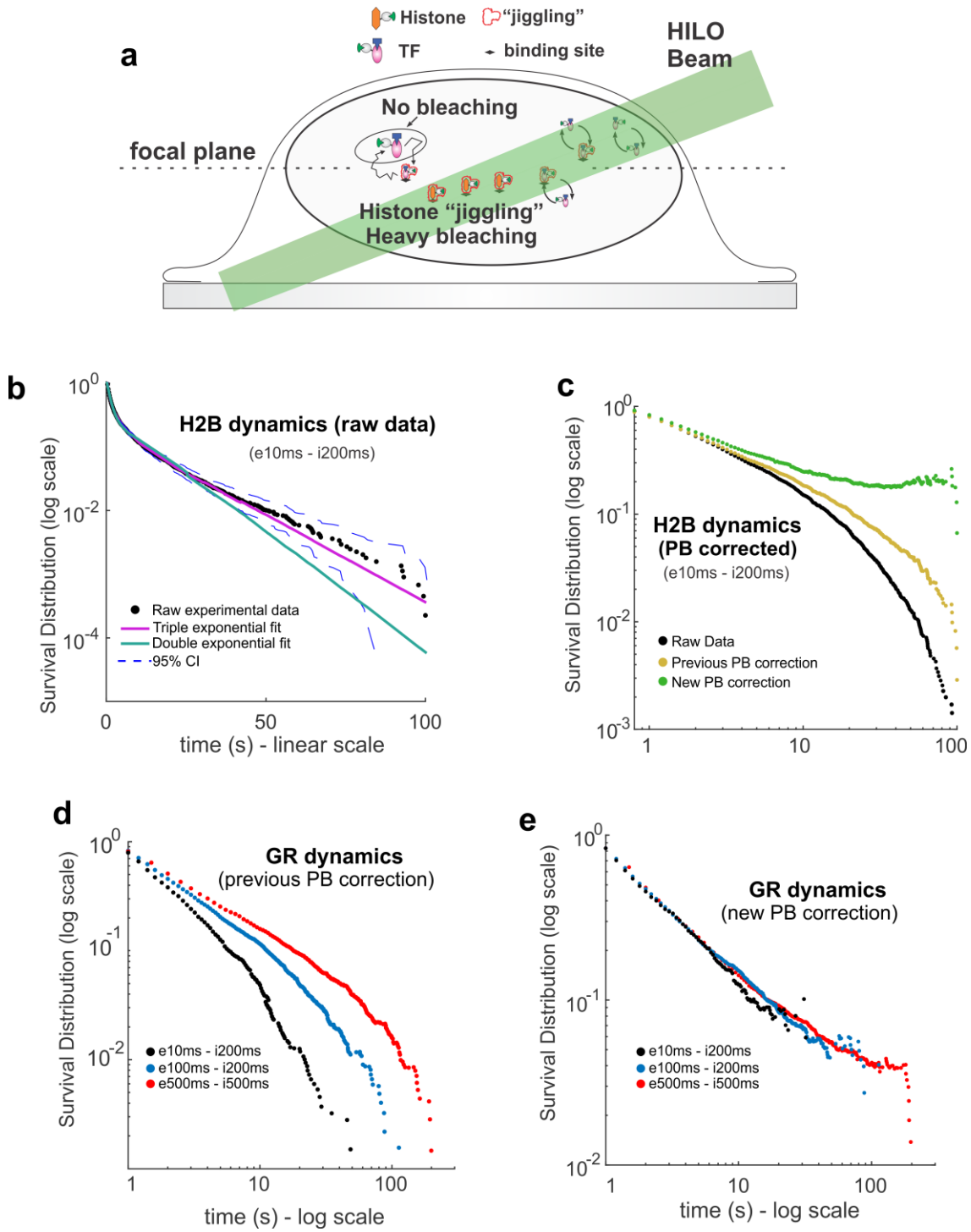

Figure 2. **The new proposed photobleaching correction method**. (**a**) TFs will not photobleach uniformly in the nucleus. A group of TFs stay bound in the focal plane until they bleach, another group stay bound in the focal plane for a short time and diffuse away from it, another group never enters the

focal plane and a final group diffuses through the focal plane. (**b**) Fit of H2B survival distribution to a double exponential and triple exponential. A triple exponential better represents the experimental data where the slower component corresponds to the photobleaching rate in the focal plane. (**c**) Raw data for H2B survival distribution (black), previous correction using number of molecules decayed (yellow) which shows that H2B after correction still have a finite dwell time. Upgraded photobleaching correction for histone H2B representing two different regimes: stably incorporated histones that have a longer residence time that photobleaching and a dynamic regime representing unincorporated histones nonspecifically interacting in the nucleus. (**d**) Single-molecule tracking data of the glucocorticoid receptor (GR) activated with corticosterone (Cort). GR acquired with different photobleaching kinetics and corrected using the number of particles decayed in the focal plane (previous PB correction). (**e**) Same experimental data as (d) but corrected using the upgraded photobleaching correction, showing similar dynamics independent of photobleaching.

Improving photobleaching correction

We propose a new PB correction method also based on histone data as a proxy of the fluorophore stability. A detailed derivation is described in **Supplementary Note 2**. The main step involves SMT of histones under the same conditions that the TF of interest will be imaged. We tracked individual H2B, H3 or H4 molecules using highly inclined and laminated optical sheet (HiLO) illumination[5] by sub-optimal transient transfection of HaloTag-fused histones, labeled with JF_{549} HaloTag ligand[26] (**see online Methods**). The three histone variants we tested presented statistically similar dynamics (**Supplementary Fig. 2a**). We continued with H2B for all further experiments.

Histone genes are primarily transcribed upon entry into S-phase of the cell cycle[27]. Due to our transient transfection approach, HaloTag-H2B proteins will be translated during interphase and therefore some histones will not be incorporated into chromatin at the time of acquisition (**Supplementary Video 5**). Hence, survival distribution of H2B will be composed of PB kinetics and a diffusive/transient binding component. To account for this behavior and assuming PB kinetics at the single-molecule level is exponentially distributed, the survival distribution of H2B is fit to an exponential family with three components (**Fig. 2b**, **Supplemental Note 2**). The faster

components characterize the dynamics of histones that have not been stably incorporated into chromatin, while the third (slower) component describes the PB kinetics of the fluorophore. To confirm that our method quantifies PB kinetics and not intrinsic dynamics of the histone H2B, we calculated PB lifetimes using histones H3 and H4 with the same statistical results (**Supplementary Fig. 2b**). Consequently, artificially modifying PB kinetics (by changing acquisition conditions) modulates the H2B survival distribution (**Supplementary Fig. 2c**) and, accordingly, the mean PB lifetimes (**Supplementary Fig. 2d, Supplementary Videos 5-8**). This method for estimating PB rates has some drawbacks when a non-uniform illumination in the focal plane is used (as in HiLO microscopy), since the calculated rate corresponds to the average fluorophore decay in the focal plane, which may differ slightly in each location[7].

To test this new PB correction method, we first analyzed how histone data itself changes after correction. Comparisons between H2B survival distribution with either our previous correction methods (**Supplementary Fig. 2e**) or with no correction at all, reveals a predictable upward shift of the distribution (**Fig. 2c,** compare yellow and black datapoints). However, H2B data still artifactually resembles the dynamics of a TF with a relatively short residence time. Conversely, if the newly proposed PB correction is used (**Fig. 2c**, green datapoints), a plateau in survival probability appears, indicating that H2B dwell times are now longer than those resulting from the photostability of the fluorophore. The high fluctuations at the tail of the distribution are likely due to noise in the data and the appearance of multiple particles within the point spread function (**Supplementary Fig. 2f**). For reproducibility and reliability purposes, multiple biological replicates are always taken, hence the ensemble average is used as the survival

252 distribution for analysis (**Supplementary Fig. 2g,** all data). Even though the decay in
253 number of particles at the focal plane (previous method) could potentially be used to
254 correct for differences in laser illumination between replicates, we found that it makes
255 the data more heterogenous (*c.f.* **Supplementary Fig. 2h** and **2i**), suggesting this is not
256 a suitable measure for PB estimates.

257 Next, we apply our new PB correction to GR as a model TF. As with histones, we
258 transiently transfected 3617 cells with GR fused to HaloTag (HaloTag-GR), incubated
259 the cells with JF_{549}, and activated the receptor with its natural ligand, corticosterone
260 (Cort, 600 nM). We artificially varied the PB kinetics of the experiment by changing the
261 exposure time and laser power (**see Methods**). When the previous PB correction was
262 applied, GR survival curves show clear dependence on acquisition conditions (**Fig. 2d**),
263 illustrating the artifact produced by PB. Conversely, with the new PB correction, we can
264 now retrieve the same underlying distribution (**Fig. 2e**), independently of the acquisition
265 conditions.

266 **New photobleaching correction reveals power-law behavior of TF dynamics**

267 Surprisingly, after correcting for PB with the new method, GR's SMT data now
268 deviates clearly from a bi-exponential distribution (**Fig. 3a**). However, the data looks
269 strikingly linear in a log-log plot (**Fig. 3b)**, which suggests a power-law behavior. A
270 similar distribution is observed upon GR activation with dexamethasone (Dex, 100 nM),
271 a more potent, synthetic hormone (**Supplementary Fig. 3a**). To rule out any artifacts
272 from acquisition conditions, SMT experiments for GR-Cort complexes were acquired at
273 different time intervals and exposure times (**Supplementary Fig. 3b-d**). Independent of
274 the sampling times, GR's survival distribution appears power-law distributed, with a

plateau in the tail due to a few molecules staying immobilized longer than the characteristic photobleaching time of the fluorophore (illustrated in **Supplementary Fig. 2f**). An acquisition rate of 1000 ms allows the observation of the long-lived events in the tail of the power law distribution (**Supplementary Fig. 3d,** independent replicates are plotted as well to demonstrate reproducibility).

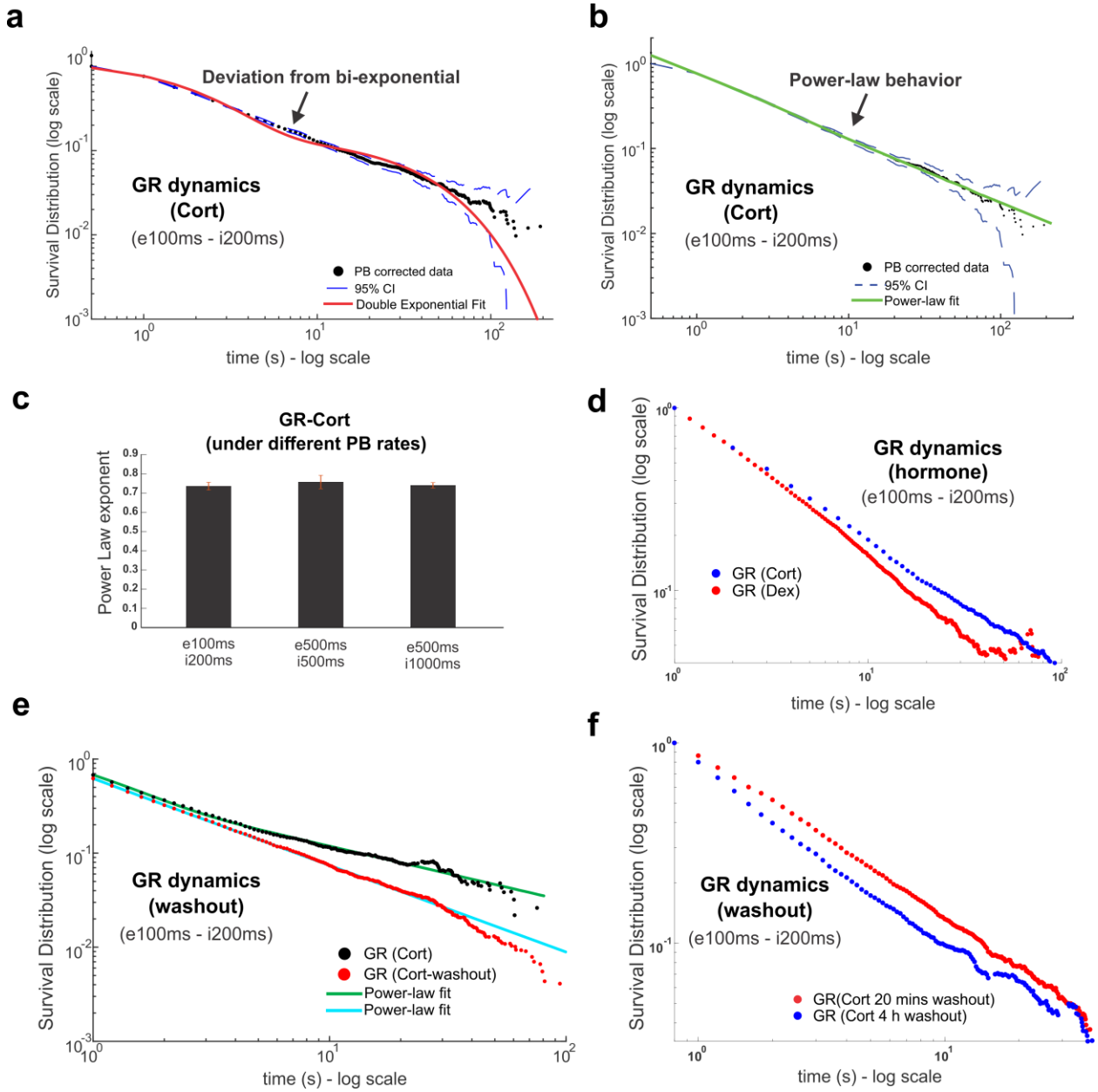

Figure 3. **Impact of photobleaching correction on GR dynamics**. (**a-b**) Single-molecule tracking data of GR activated with corticosterone (Cort). Data was acquired at 100 ms exposure time (e100ms) in 5Hz interval (i200ms). The survival distribution is shown, fitted to either a bi-exponential (a) or a power-law (b).

(**c**) Power law exponent of GR-Cort SMT data under different photobleaching rates, artificially generated by modulation of acquisition conditions. (**d**) GR dynamics under Cort or dexamethasone (Dex) activation. Power-law exponents of 0.71 +/- 0.05 and 0.81 +/- 0.01 respectively. The natural hormone exhibits longer dwell times. (**e**) Comparison of GR dynamics under Cort or washout of the hormone, which inactivates the TF. (**f**) GR dynamics under a standard 20 min or a more stringent 4h washout protocol.

A random variable t follows a power law[28] for $t > t_{min}$ if $f(t) = At^{-\beta}$, where A is a constant and $\beta \in \mathbb{R}^{+}$ corresponds to the critical exponent, also known as the scaling parameter. Power laws are heavy tailed (right-skewed) and β is a measure of this skewness. Quantification of the critical exponent of the power-law shows that GR under different acquisition conditions (i.e. under different PB rates) exhibits the same dynamics (**Fig. 3c**). Surprisingly, GR activated with Cort shows an upward-shifted (lower β) distribution compared to the more potent hormone dexamethasone (**Fig. 3d**), suggesting longer residence times for the less potent ligand. These differences were not apparent in previous studies[10, 17] likely due to inaccurate PB correction and/or a partial exploration of the dynamic range of the protein of interest in the SMT experiments.

Previous research has assumed that the dynamics of non-specific binding is well described by a single exponential component with a much shorter dwell time than specific binding[8, 12, 17]. To understand the dynamics of non-specific binding, we inactivated GR by washing out the hormone[29] for 20 minutes, which greatly reduces specific binding measured by chromatin immunoprecipitation[30]. Interestingly, GR still exhibits power law behavior, although with shorter dwell times as indicated by an increase in the power law exponent (**Fig. 3e**). While longer washouts (four hours) shows further reduction in dwell times, GR still exhibits power-law dynamics (**Fig. 3f, Supplementary Fig. 3e, Supplementary Video 9**). In conclusion, the new PB correction method has a major impact on the *distribution* of TF residence times (power-

law instead of bi-exponential). This, in turn, questions the interpretation of specific and non-specific binding as two distinct populations with discrete (and measurable) residence times.

Re-interpretation of SMT data: new models for TFs kinetics

To better understand the link between TF binding and the observed residence time distributions, we explored different theoretical models that may explain the emergence of different behaviors in the survival distribution. Calculation of dwell time distributions is a first-passage time problem in stochastic analysis and has been widely used to characterize the kinetic properties of molecular motors and ion channels[31]. When simple kinetic schemes are involved, dwell time distributions can be calculated analytically. However, for more complex systems, other methods must be used. One particularly powerful approach is to assign one or more states to “act” as an absorbing boundary, and then solve the associated first-order kinetic equations to obtain dwell time distributions[32] (**Supplementary Note 1.1)**. We assume that the diffusive state (unbound) corresponds to an absorbing boundary state since tracked particles end with such transitions. The single-molecule either photobleaches, disappears from the focal plane or begins diffusing. Any rebinding of the TF is considered an independent event.

We first examined the widely used bi-exponential model under this framework (**Fig. 4a**). According to this model, TFs can occupy three different states: diffusive, slow and fast. The diffusive state plays the role of an absorbing boundary state (i.e. exit from a bound state). In the current literature, it has been interpreted that the slow and fast states correspond to specific and nonspecific binding, respectively[33].

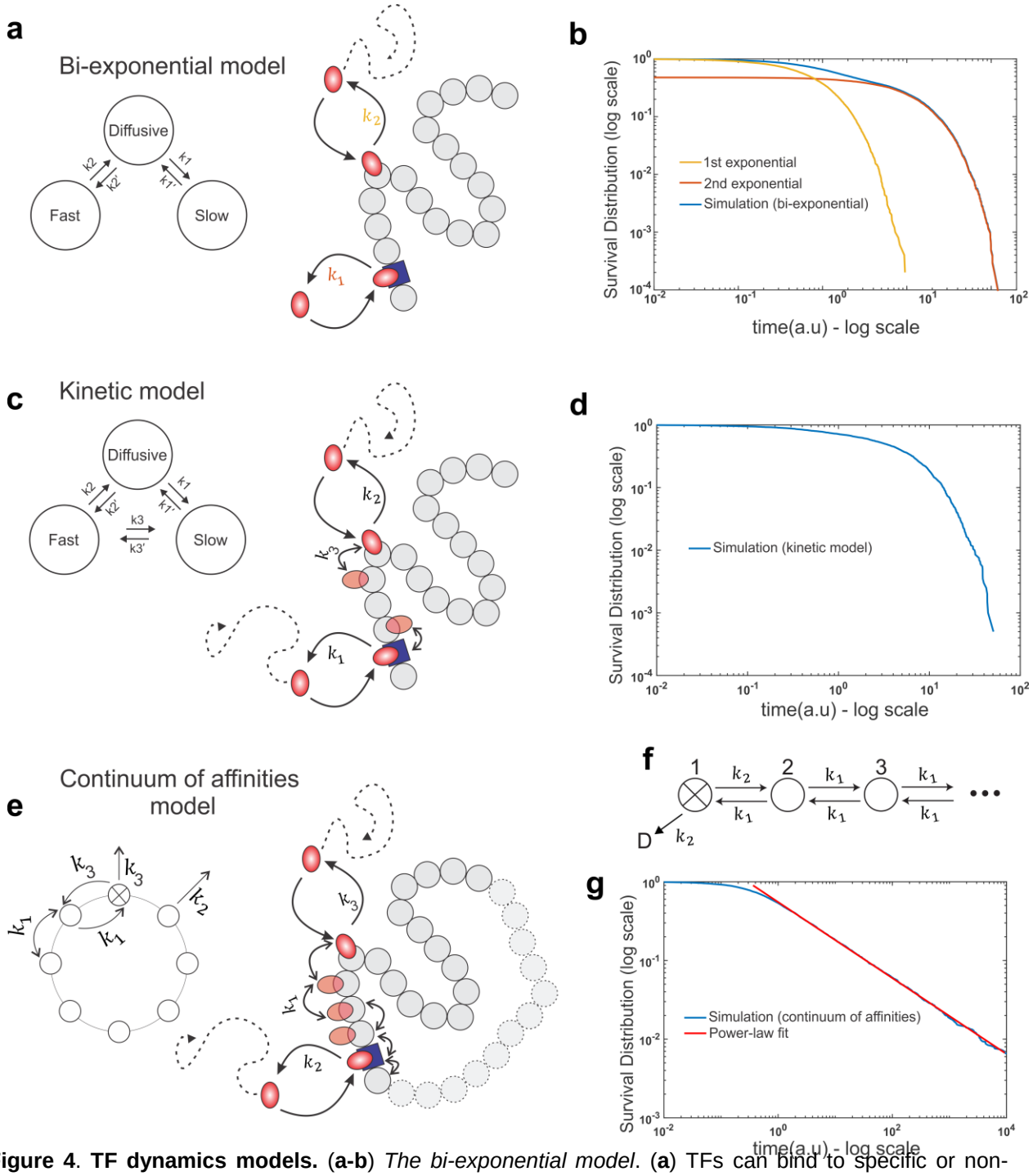

Figure 4. **TF dynamics models.** (**a-b**) *The bi-exponential model*. (**a**) TFs can bind to specific or non-specific sites with different affinities. After binding to DNA, they will unbind from DNA; transitions between specific and non-specific sites are forbidden. (**b**) Numerical simulation showing the emergence of bi-exponential behavior. (**c-d**) *The Kinetic model*. (**c**) TFs can transition from non-specific sites to specific sites and vice versa. Transitions in the DNA are considered indistinguishable. (**d**) Simulation results of the Kinetic model. (**e-g**) *The continuum of affinities model*. (**e**) TFs can diffuse on the DNA, and transition between any state (Diffusive, specifically bound, nonspecifically bound). Dwell time is defined as the time spent on the DNA, either bound or sliding. (**f**) Cartoon-simplification of (e), a TF arrives to a random site, scans the DNA until it finds a specific site and it unbinds. This model can be solved analytically. (**g**)

Numerical simulation of (e) showing the emergence of power-law behavior (red line). See **Supplementary Note 1** for details.

With this assumption of a well separated and narrow distribution of affinities, the expected behavior of the survival distribution corresponds to a double exponential with exponential parameters determining the average residence time of each state, as confirmed with stochastic simulations (**Fig. 4b**) using the Gillespie algorithm[34]. Importantly, this model does not allow for transitions between fast and slow states. For a complete derivation, see **Supplementary Note 1.2**.

We next extended the bi-exponential model to allow for transitions between the slow and fast components (**Fig. 4c**). Due to the resolution limit (~30nm), any transitions between specific and non-specific bound states cannot be distinguished. The resulting survival distribution corresponds to a family of exponentials; we call this a *kinetic model* (**Supplementary Note 1.3**). Simulations were performed as before, and the expected distribution is displayed in **Fig. 4d**.

Finally, several theoretical studies have posited that TF search and “final” binding to its cognate site on the DNA involves a combination of bulk diffusion in the nucleus, 1D sliding along the DNA, hopping and translocation, and theoretical search times for the TF to find specific sites in this framework have been estimated[35-37]. In this model, TFs will have a multiplicity of fast bound states that must be accounted for in the analysis of dwell time data. To do so, we modeled TF movement on the DNA as hopping on a circular chain composed of specific sites and non-specific sites (**Fig. 4e**). The main assumption in our “*continuum of affinities*” model derivation (**Supplementary Note 1.4**) is that the number of non-specific sites on the DNA is much larger than the number of specific sites. This is a biologically reasonable assumption as only a few to

tens of thousands of specific sites are bound by any TF according to genome wide studies[38], while the entire genome contains millions of "other" potential chromatin sites. An analytical solution can be found for the simplest case where there is a single specific binding site and the TF unbinds from the specific site (**Fig. 4f, Supplementary Note 1.4.3**). A simulation based on the model gives rise to asymptotic power law behavior at time scales compatible with specific binding (**Fig. 4g**).

To test which model better represents the observed behavior of GR dynamics, we used the Bayesian information criterion (BIC)[39] and Kolmogorov–Smirnov (KS) test to choose the best-predicted model (**see Methods).** Clearly, power-law behavior emerges as the best fit, as illustrated for both GR-Cort (**Supplementary Fig. 4a-c)** and GR-Dex (**Supplementary Fig. 4d-f)** data. Interestingly, power law distributions have infinite variance, which implies a finite probability of long-lived binding events. This could imply that productive binding events may be rare with dwell times much longer than previously appreciated, as indicated by the right-skewness of the distribution. Another possibility corresponds to a wide distribution of binding affinities in the nucleus due to heterogeneity in 1) binding affinities of individual response elements, and/or 2) local nuclear microenvironment.

In summary, by incorporating an improved PB correction method, we discovered that the survival distribution of GR dwell times does not follow a bi-exponential model. In fact, the data follows a power-law distribution, which we can derive using two theoretical models (**Supplementary Note 1.4-1.5**). Ultimately, if there is a way to define or distinguish non-specific from specific binding, our results indicate that it cannot be based on their global residence times. However, the slope of the residence time

distribution does provide an estimate of the overall affinity and can be used for comparing TFs and their function under different conditions.

Discussion

In the present study, we describe a new photobleaching correction method to prevent photobleaching-related bias of the dwell time distribution of TFs. This method is based on histone SMT dynamics, used here as a proxy for photobleaching kinetics within the focal plane (**Fig. 2**). When correcting H2B SMT data by itself with this new approach, we showed that the survival distribution of H2B exhibits a "plateau" (**Fig. 2c**). This indicates very slow dynamics for the core histone, much longer than our experimental time scale, and is fully compatible with previously published FRAP data[40]. Additionally, we are now able to reconcile data acquired under different experimental conditions (**Fig. 2d-e**) whereas previous attempts were not successful[10, 20, 41, 42].

Given the current phenomenological interpretation of SMT data and the general lack of a model-based approach in the field, we derived theory-based models in an attempt to explain TF dynamics more accurately. We explored three different models: the classic bi-exponential model, a kinetic three-state model and a power-law model (**Fig. 4**). After correcting for photobleaching and implementation of Bayesian Information Criterion, we could identify the best predictive model that explained the residence time distribution of a paradigmatic TF, the GR. Surprisingly, the dwell time distribution of activated-GR is best described by a power-law. Although a recent study corrected PB of NF-κB dynamics by normalization with the H2B dwell time distribution, they still reported exponential behavior of the TF[43]. However, they did not consider alternative models and

the range of detected experimental dwell times is relatively short (maximum ~10 s), which might have prevented observation (and verification) of asymptotic power law behavior.

Our observation of power law behavior of GR residence times suggests a model with a continuum of DNA-bound states rather than discrete non-specific/specific binding of TFs. Consistent with this model, wash-out of the hormone (**Fig. 3e**) revealed that the dwell time distribution also follows a power law, indicating no apparent dynamical differences between the so-called specific and non-specific binding. Nevertheless, the overall residence time decreases when the receptor is less active, suggesting that a majority of the longer events observed with the fully activated receptor are associated with productive transcription as previously reported[8, 10, 12, 17, 43]. However, non-specific binding can also result in TF binding events with long residence times, the implications of which are still not known. Critical efforts are required to investigate whether the slow(er) stops seen in SMT are matched exclusively to specific interactions with chromatin. Alternatively, a sub-population of these “stops” could correspond to microscopic regions in the nucleus where diffusion is severely impaired, or transient interaction with “clustered” structures such as foci observed for GR[44], or another hitherto unknown mechanism.

The broad distribution of affinities is puzzling but may be explained by heterogeneity in the nuclear structure and chromatin environment. Targets for a searching TF certainly exist in a wide variety of chromatin states (compacted fibers, different nucleosome modification conditions, etc.). Also, affinities for the thousands of alternative binding sites in response elements must vary significantly. Furthermore,

recent work points to the presence of transcriptional hubs and liquid-liquid phase separation domains[45] that contribute to the complexity of nuclear organization. If TFs exhibit different dynamical properties in these structures, it is not surprising to find a broad variation in binding affinities. The resulting broad distribution of binding affinities in these scenarios goes against the widely-held assumption that TF dynamics on chromatin results from well-separated and narrow distributions of specific and non-specific binding (**Fig. 5**).

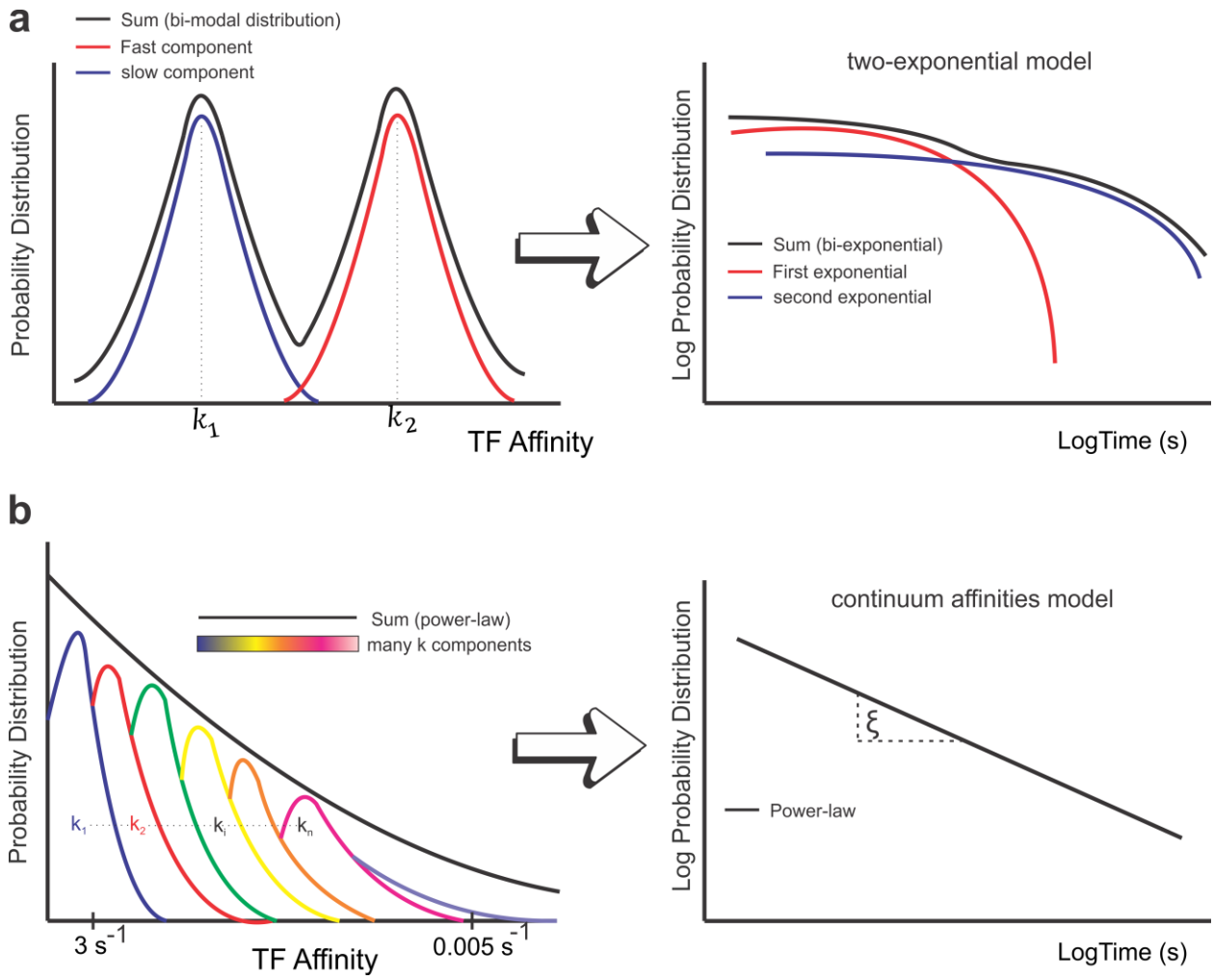

Figure 5. **Heterogeneity in binding affinities**. (**a**) In the bi-exponential model, specific sites and nonspecific sites have a well separated and narrow distribution of affinities (left graph), which results in a double exponential model for the survival distribution (right graph). (**b**) The cell nucleus may have a broad distribution of affinities due to its heterogeneity (black line). This distribution is composed of local dwell times with a well-defined affinity distribution (depicted in different colors, left graph). This distribution of

affinities may explain the emergence of power law behavior (characterized by the exponent, ξ) in the residence time of TFs (right graph).

Given the heterogeneities in local organization and nuclear structure, TF binding sites on chromatin can be viewed as a collection of traps with a distribution of trap depths (analogous to binding affinities). In such a finite disordered system, the distribution of trapping times asymptotically approaches a power law[46, 47] (**Fig. 5b, Supplementary Note 1.5**)

Alternatively, but not mutually exclusive, heterogeneity in the searching mechanism of TFs may affect the effective affinity constant observed in SMT experiments. In support of the latter, the tetracycline repressor (TetR), a chimera between a bacterial and a viral protein with no known endogenous targets in mammalian cells, when used as a proxy to emulate TF dynamics, also showed power-law behavior for "non-specific binding"[48, 49]. However, it could still be described as an exponential on an artificially (and single) specific DNA binding array[49]. Thus, the intrinsic nature of the searching mechanism of any DNA-binding protein may be governed by power-law dynamics. In addition, the heterogeneity of dwell times in the thousands of response elements for an endogenous TF could explain why GR can present power-law tails as opposed to TetR, which can only bind to one artificial array site. Interestingly, a recent study in yeast[50] reports that both the TF Ace1p and the chromatin remodeler RSC binding follow a bi-exponential binding distribution in cells containing a natural tandem of ten *CUP1* (Ace1p responsive) *genes*. This dynamic and discrete behavior, in contrast with our GR data, can be explained by the particular and homogeneous chromatin environment of single array of specific sites. Consequently, we speculate that

a broad distribution of binding affinities coming from a whole population of different binding sites (thousands in the case of GR) may result in a power-law behavior (**Fig. 5**). In summary, by the implementation of proper photobleaching kinetics, we reveal a new model of TF dynamics. Our findings suggest that, contrary to the established paradigm, TF dwell times follow a broad distribution with no evidence of binary, discrete populations.

Acknowledgments

We thank Tatiana Karpova and David Ball from the Optical Microscopy Core at the NCI, NIH for the assistance in imaging and data processing. We thank Luke Lavis (Janelia Research Campus) for providing HALO dyes. We thank Kaustubh Wagh for helpful discussions of the analytical results. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. V.P. was supported, in part, by the Academy of Finland, the University of Eastern Finland strategic funding and the Sigrid Jusélius Foundation. D.M.P was supported, in part, by CONICET. A.U. acknowledges support from the NCI-UMD Cancer Technology Partnership, and the awards NSF PHY 1607645, NSF PHY 1806903.

Author Contributions

D.A.G, G.F., D.M.P., and G.L.H. conceived the experiments and wrote the original draft with subsequent input and editing from all authors. D.A.G, G.F., and D.M.P. performed imaging experiments, analyzed data and prepared figures. V.P. participated in the early stages of the project. D.A.G conceived and wrote in MATLAB® the photobleaching correction method (with constant feedback from G.F and D.M.P.).

D.A.G and C.J. derived the theoretical models, performed the simulations, and implemented the Bayesian statistical analysis. A.U. supervised data analysis, theoretical models and participated in the manuscript preparation; G.L.H. supervised the project.

Competing Interests statement

The authors declare no competing interests

References

1. Lambert, S.A. et al. The Human Transcription Factors. *Cell* **172**, 650-665 (2018).
2. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J Cell Biol* **217**, 1181-1191 (2018).
3. Coulon, A., Chow, C.C., Singer, R.H. & Larson, D.R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev.Genet.* **14**, 572-584 (2013).
4. Grimm, J.B. et al. Bright photoactivatable fluorophores for single-molecule imaging. *Nat Methods* **13**, 985-988 (2016).
5. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods* **5**, 159-161 (2008).
6. Goldstein, I. et al. Transcription factor assisted loading and enhancer dynamics dictate the hepatic fasting response. *Genome Res* **27**, 427-439 (2017).
7. Presman, D.M. et al. Quantifying transcription factor dynamics at the single-molecule level in live cells. *Methods* **123**, 76-88 (2017).
8. Mazza, D., Abernathy, A., Golob, N., Morisaki, T. & McNally, J.G. A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res* **40**, e119 (2012).
9. Izeddin, I. et al. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife* **3**, e02230 (2014).
10. Paakinaho, V. et al. Single-molecule analysis of steroid receptor and cofactor action in living cells. *Nat Commun* **8**, 15896 (2017).
11. Ball, D.A. et al. Single molecule tracking of Ace1p in Saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Res* **44**, e160 (2016).
12. Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285 (2014).
13. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6** (2017).
14. Kieffer-Kwon, K.-R. et al. Myc regulates chromatin decompaction and nuclear architecture during B cell activation. *Molecular Cell* **67**, 566-578.e510 (2017).

15. Kilic, S., Bachmann, A.L., Bryan, L.C. & Fierz, B. Multivalency governs HP1alpha association dynamics with the silent chromatin state. *Nat Commun* **6**, 7313 (2015).
16. Loffreda, A. et al. Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. *Nat Commun* **8**, 313 (2017).
17. Morisaki, T., Muller, W.G., Golob, N., Mazza, D. & McNally, J.G. Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nat Commun.* **5**, 4456 (2014).
18. Sugo, N. et al. Single-Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence. *Sci Rep* **5**, 10662 (2015).
19. Zhen, C.Y. et al. Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. *Elife* **5** (2016).
20. Gebhardt, J.C. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat Methods* **10**, 421-426 (2013).
21. Teves, S.S. et al. A dynamic mode of mitotic bookmarking by transcription factors. *Elife* **5** (2016).
22. Liu, Z. et al. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *Elife.* **3**, e04236 (2014).
23. Presman, D.M. & Hager, G.L. More than meets the dimer: What is the quaternary structure of the glucocorticoid receptor? *Transcription* **8**, 32-39 (2017).
24. Kimura, H. & Cook, P.R. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J.Cell Biol.* **153**, 1341-1353 (2001).
25. Hansen, A.S. et al. Robust model-based analysis of single-particle tracking experiments with Spot-On. *Elife* **7** (2018).
26. Grimm, J.B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* **12**, 244-250 (2015).
27. Ewen, M.E. Where the cell cycle and histones meet. *Genes Dev* **14**, 2265-2270 (2000).
28. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46**, 323-351 (2005).
29. Stavreva, D.A. et al. Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat.Cell Biol.* **11**, 1093-1102 (2009).
30. Stavreva, D.A. et al. Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. *Genome Res.* **25**, 845-857 (2015).
31. Liao, J.C., Spudich, J.A., Parker, D. & Delp, S.L. Extending the absorbing boundary method to fit dwell-time distributions of molecular motors with complex kinetic pathways. *Proc Natl Acad Sci U S A* **104**, 3171-3176 (2007).
32. Van Kampen, N.G. Stochastic processes in physics and chemistry, Vol. 1. (Elsevier, 1992).
33. Goldstein, I. & Hager, G.L. Dynamic enhancer function in the chromatin context. *Wiley interdisciplinary reviews. Systems biology and medicine* (2017).

34. Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 2340-2361 (1977).
35. Bauer, M. & Metzler, R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys J* **102**, 2321-2330 (2012).
36. Berg, O.G. & Blomberg, C. Association kinetics with coupled diffusional flows. Special application to the lac repressor--operator system. *Biophysical chemistry* **4**, 367-381 (1976).
37. Marklund, E.G. et al. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc Natl Acad Sci U S A* **110**, 19796-19801 (2013).
38. Wang, J. et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798-1812 (2012).
39. Schwarz, G. Estimating the dimension of a model. *Ann Stat.* **6**, 461-464 (1978).
40. Kimura, H. Histone dynamics in living cells revealed by photobleaching. *DNA repair* **4**, 939-950 (2005).
41. Agarwal, H., Reisser, M., Wortmann, C. & Gebhardt, J.C.M. Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. *Biophys J* **112**, 2051-2055 (2017).
42. Clauss, K. et al. DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res* **45**, 11121-11130 (2017).
43. Callegari, A. et al. Single-molecule dynamics and genome-wide transcriptomics reveal that NF-kB (p65)-DNA binding times can be decoupled from transcriptional activation. *PLoS Genet* **15**, e1007891 (2019).
44. Stortz, M. et al. Mapping the Dynamics of the Glucocorticoid Receptor within the Nuclear Landscape. *Sci Rep* **7**, 6219 (2017).
45. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
46. Bouchaud, J.-P. & Georges, A. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Physics Reports* **195**, 127-293 (1990).
47. J., B. Weak ergodicity breaking and aging in disordered systems. *Journal de Physique I* **2**, 1705-1713 (1992).
48. Caccianini, L., Normanno, D., Izeddin, I. & Dahan, M. Single molecule study of non-specific binding kinetics of LacI in mammalian cells. *Faraday discussions* **184**, 393-400 (2015).
49. Normanno, D. et al. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun* **6**, 7357 (2015).
50. Mehta, G.D. et al. Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. *Mol Cell* **72**, 875-887.e879 (2018).