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

31   Single-molecule tracking allows the study of transcription factor dynamics in the

32   nucleus, giving important information regarding the search and binding behavior of

33   these proteins with chromatin *in vivo*. However, these experiments suffer from

34   limitations due to photobleaching of the tracked protein and assumptions about the

35   exponential behavior required for data interpretation, potentially leading to serious

36   artifacts. Here, we developed an improved method to account for photobleaching

37   effects, theory-based models to accurately describe transcription factor dynamics, and

38   an unbiased model selection approach to determine the best predicting model. A new

39   biological interpretation of transcriptional regulation emerges from the proposed models

40   wherein transcription factor searching and binding on the DNA results in a broad

41   distribution of binding affinities and accounts for the power-law behavior of transcription

42   factor residence times.

Introduction

44   Transcription factors (TFs) are key regulatory proteins responsible for turning

45   genes "on" and "off" by binding to enhancer or promoter elements across the genome[1].

46   Fluorescence microscopy techniques have revolutionized our understanding of how TFs

47   search and interact with chromatin[2]. Fluorescence recovery after photobleaching

48   (FRAP) in live-cell systems unveiled the dynamic nature of these proteins, in contrast to

49   the long-standing static model of TF-chromatin interactions[3]. The combined innovative

50   technological improvements in fluorophore brightness and stability[4], optical set-ups to

51   increase the signal-to-noise ratio[5], and camera speed and sensitivity now allow the

2

52   study of single-molecules at an unprecedent temporal and spatial resolution. Single-

53   molecule tracking (SMT) is a powerful technique that allows the characterization of

54   protein dynamics in single, live cells. It is based on detecting and following through time

55   the traces produced by the light emitted from a single fluorophore. When applied to the

56   study of TFs, important information regarding the search and binding dynamics of these

57   proteins can be extracted[2].

58   The SMT approach (**Fig. 1**) has now been applied for approximately two-dozen

59   TFs in a variety of cellular systems[6]. The current consensus, based on empirical

60   residence time distributions, describes TFs as able to transition between three different

61   states: 1) unbound from DNA (diffusing in the nucleus), 2) non-specifically bound and 3)

62   specifically bound to chromatin (i.e. interacting with specific response elements)[7]. This

63   three-population model is based on the observation that TFs can intermittently stop, and

64   then resume rapid diffusion[8, 9]. Thus, the empirical residence time distribution of the

65   "stopped molecules" have been phenomenologically fitted to families of exponential

66   distributions[10-19], with no underlying normative model for the origins of these

67   distributions[8]. Nevertheless, the bi-exponential fits to the distribution suggest that the

68   DNA-bound population of molecules includes two distinct subpopulations: a short-lived

69   fraction ('fast stops') and a longer-lived fraction ('slow stops'). The slow fraction has

70   been hypothesized to represent specific binding events associated with enhancers or

71   promoters, while the fast fraction is hypothesized to represent non-specific binding to

72   chromatin[11, 12, 15, 17]. Experiments wherein the DNA-binding domain of TFs has been

73   mutated are consistent with this hypothesis as the longer events were reported to be

74   dramatically reduced[10, 12, 17, 18].

3

75   Despite the technological advances in studying TF dynamics, a major remaining

76   limitation of the SMT approach is the relatively short fluorescence stability (or tendency

77   to photobleach) of any fluorophore dye. This makes it difficult to distinguish whether a

78   loss of signal is due to an unbinding event, drifting or photobleaching. Hence, careful

79   corrections must be performed to accurately estimate TF dynamics. Unfortunately,

80   photobleaching correction methods vary widely among research groups[17, 20, 21] which

81   highlights the lack of a standard approach to overcome the photobleaching bias of SMT

82   strategies[9, 12, 18, 22].

83   In this work, we propose a new and improved photobleaching correction method

84   that uses fluorescence dynamics of histones measured at the focal plane under precise

85   SMT acquisition conditions. We then use the glucocorticoid receptor (GR), a ligand

86   regulated transcription factor[23], as a model TF to validate and test the new

87   methodology. This approach unexpectedly reveals a novel, power-law behavior of GR

88   residence time distributions. We then derive theory-based models for TF dynamics and

89   a principled method to obtain optimal model parameters from empirical residence time

90   distributions, using Bayesian statistics. We show that a model of TF searching and

91   binding or a model of a nucleus with broad distribution of binding affinities accounts for

92   the power law behavior of GR residence times obtained after implementing the modified

93   photobleaching correction method. These models exhibit a broad effective distribution of

94   binding affinities, thus challenging the established model with two discrete, bound

95   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

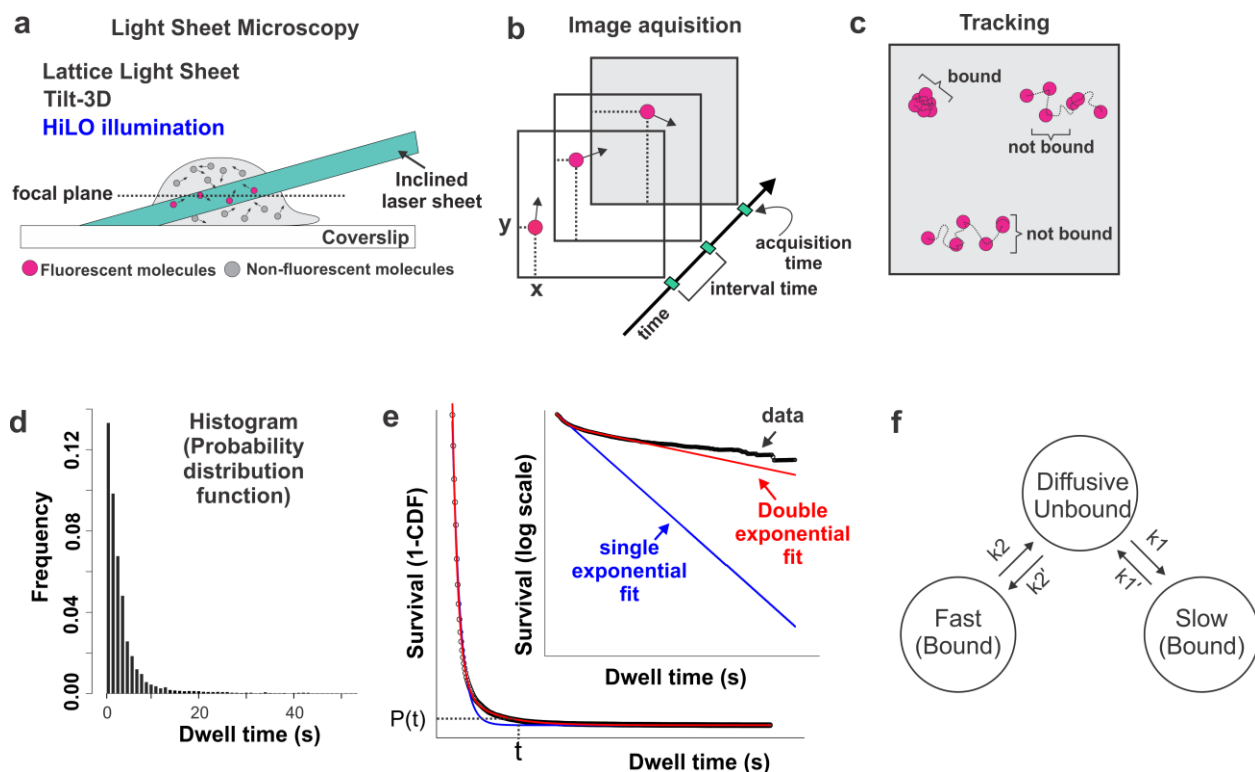as stationary spots (**Fig. 1a-c, Supplementary Video 1**).

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

129   obtain a local dwell time for TFs which is defined as the time interval between a single

130   molecule transitioning from a diffusive state to a bound state and its subsequent

131   unbinding from DNA. The definition of dwell time presented here differs of the one used

132   in the field (residence time) in that we are considering any process occurring in a point

133   spread function as a single observable binding event in the calculation of the dwell time

134   distribution, not just discrete binding states. The dwell time distribution is generated by

135   integrating the ensemble-averaged distribution of bound times (**Fig. 1d,**

136   **Supplementary Note 1.1**). Most often, a "survival" distribution, defined as 1-CDF,

137   where CDF is the empirical cumulative distribution function of dwell times, is used for

138   further analysis (**Fig. 1e**). This survival distribution is usually fit to a bi-exponential

139   distribution, interpreted as the "three population model" (i.e. diffusive, fast bound, slow

140   bound) illustrated in **Fig. 1f**.

141   The upper temporal limit in SMT experiments is ultimately determined by the

142   intrinsic photostability of the chosen fluorophore. When the affinity of bound TFs leads

143   to dwell times longer than the averaged photostability of their fluorescent dyes,

144   residence times cannot be resolved. Importantly, even when bound molecules have

145   relatively lower affinities, they will appear to have shorter experimental dwell times due

146   to photobleaching (PB) bias. To demonstrate this, we conducted single-molecule

147   imaging of the glucocorticoid receptor (GR), a ligand-dependent transcription factor[23],

148   tagged with HaloTag-Janelia Fluor 549 (JF_{549})[7]. If we artificially modulate the PB

149   conditions by changing acquisition parameters (exposure time, interval time, laser

150   power), the resulting kymographs (**Supplementary Fig. 1, Supplementary Videos 1-4)**

151   appear to have originated from different TFs. Therefore, PB must be properly corrected

152   to prevent artifacts in the analysis of SMT data.

153   Currently, there is no consensus on the proper method to estimate PB rates. One

154   common approach is to simply count, frame-by-frame, the number of particles in the

155   focal plane and then fit the time-dependent number to a bi-exponential model[8, 10, 11, 16,

156   17]. However, this method underestimates the real PB that stably-bound proteins

157   undergo at the focal plane. The dwell time of diffusive (unbound) molecules is much

158   shorter than that of bound molecules (**Fig. 2a**) in the focal plane. Therefore, stably-

159   bound molecules will photobleach faster than diffusive particles. Moreover, the number

160   of particles in the focal plane will be heavily dominated by the diffusive component,

161   which will also distort PB rates derived from the fits.

162   Another strategy uses histones as a proxy for obtaining PB rates. Histones are a

163   good representation of stably-bound proteins because, after integration into chromatin,

164   they have a residence time much longer than the photostability of any currently

165   available organic fluorophore[24]. Therefore, by measuring the residence time of histones,

166   we can obtain a direct representation of PB for particles in the focal plane, as the

167   disappearance of a long-lived particle will most likely represent a PB event. This method

168   is widely used in the SMT field[12, 19, 21 13, 14, 25] but, as we will demonstrate below, it has

169   not been properly implemented. In current models, photobleaching kinetics are

170   characterized by an exponential parameter k_{his} and the slow component of a TF is

171   characterized by another exponential parameter k_{TF} (see **Supplementary Note 2** for

172   details). This rate k_{TF} is corrected for PB to give the "real residence time" (k_{TFreal}) of the

173   transcription factor by: $k_{TFreal} = k_{TF} - k_{his}$. However, the correction assumes that TF

174   kinetics come from an exponential family since PB kinetics is exponential. Therefore,

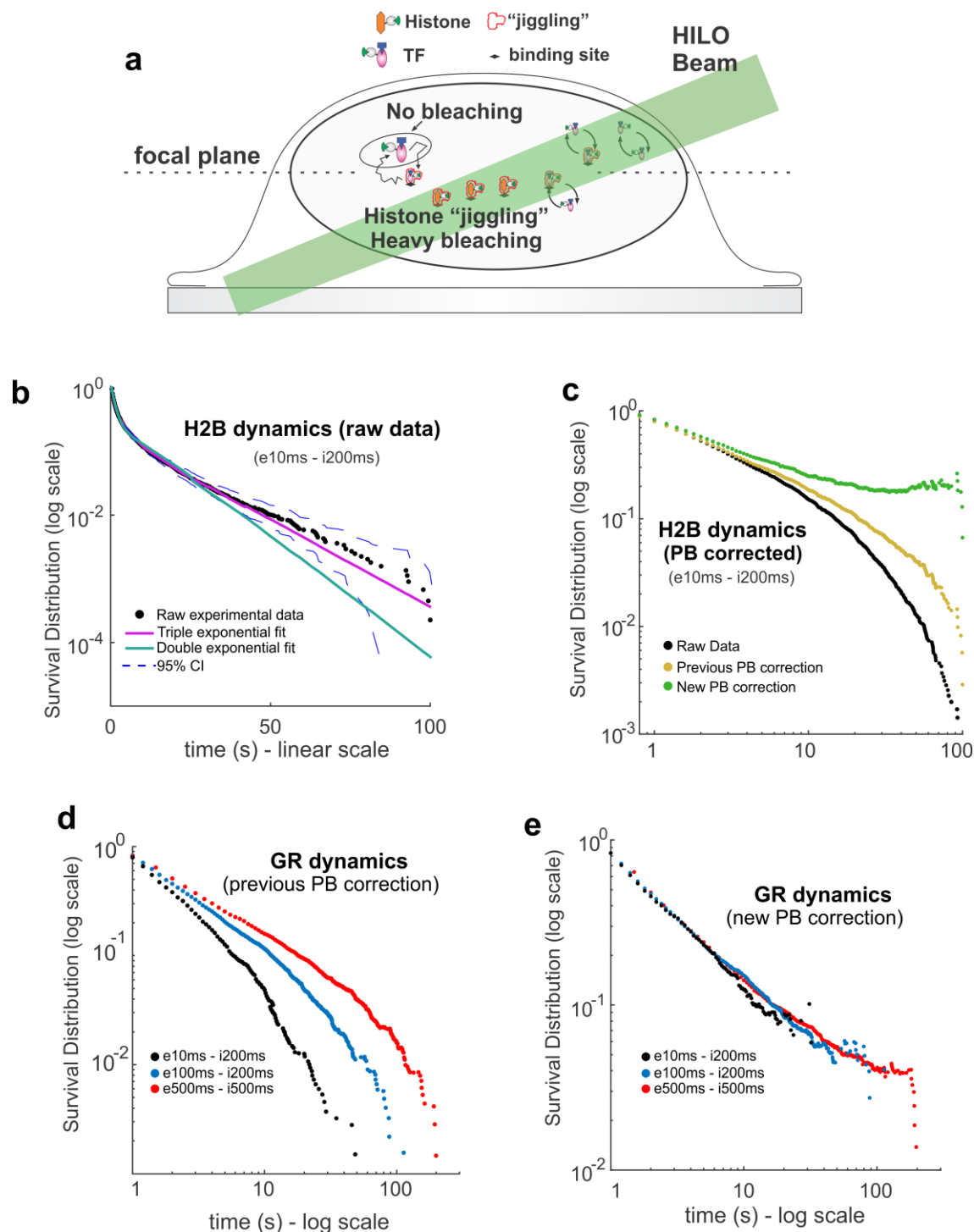175   the parameter k_{TFreal} may emerge as an artifact of this assumption.

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

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

8

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

211

212 **Improving photobleaching correction**

213   We propose a new PB correction method also based on histone data as a proxy

214 of the fluorophore stability. A detailed derivation is described in **Supplementary Note 2**.

215 The main step involves SMT of histones under the same conditions that the TF of

216 interest will be imaged. We tracked individual H2B, H3 or H4 molecules using highly

217 inclined and laminated optical sheet (HiLO) illumination[5] by sub-optimal transient

218 transfection of HaloTag-fused histones, labeled with JF_{549} HaloTag ligand[26] (**see online**

219 **Methods**). The three histone variants we tested presented statistically similar dynamics

220 (**Supplementary Fig. 2a**). We continued with H2B for all further experiments.

221   Histone genes are primarily transcribed upon entry into S-phase of the cell

222 cycle[27]. Due to our transient transfection approach, HaloTag-H2B proteins will be

223 translated during interphase and therefore some histones will not be incorporated into

224 chromatin at the time of acquisition (**Supplementary Video 5**). Hence, survival

225 distribution of H2B will be composed of PB kinetics and a diffusive/transient binding

226 component. To account for this behavior and assuming PB kinetics at the single-

227 molecule level is exponentially distributed, the survival distribution of H2B is fit to an

228 exponential family with three components (**Fig. 2b**, **Supplemental Note 2**). The faster

229   components characterize the dynamics of histones that have not been stably

230   incorporated into chromatin, while the third (slower) component describes the PB

231   kinetics of the fluorophore. To confirm that our method quantifies PB kinetics and not

232   intrinsic dynamics of the histone H2B, we calculated PB lifetimes using histones H3 and

233   H4 with the same statistical results (**Supplementary Fig. 2b**). Consequently, artificially

234   modifying PB kinetics (by changing acquisition conditions) modulates the H2B survival

235   distribution (**Supplementary Fig. 2c**) and, accordingly, the mean PB lifetimes

236   (**Supplementary Fig. 2d, Supplementary Videos 5-8**). This method for estimating PB

237   rates has some drawbacks when a non-uniform illumination in the focal plane is used

238   (as in HiLO microscopy), since the calculated rate corresponds to the average

239   fluorophore decay in the focal plane, which may differ slightly in each location[7].

240   To test this new PB correction method, we first analyzed how histone data itself

241   changes after correction. Comparisons between H2B survival distribution with either our

242   previous correction methods (**Supplementary Fig. 2e**) or with no correction at all,

243   reveals a predictable upward shift of the distribution (**Fig. 2c,** compare yellow and black

244   datapoints). However, H2B data still artifactually resembles the dynamics of a TF with a

245   relatively short residence time. Conversely, if the newly proposed PB correction is used

246   (**Fig. 2c**, green datapoints), a plateau in survival probability appears, indicating that H2B

247   dwell times are now longer than those resulting from the photostability of the

248   fluorophore. The high fluctuations at the tail of the distribution are likely due to noise in

249   the data and the appearance of multiple particles within the point spread function

250   (**Supplementary Fig. 2f**). For reproducibility and reliability purposes, multiple biological

251   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

11

275   plateau in the tail due to a few molecules staying immobilized longer than the

276   characteristic photobleaching time of the fluorophore (illustrated in **Supplementary Fig.**

277   **2f**). An acquisition rate of 1000 ms allows the observation of the long-lived events in the

278   tail of the power law distribution (**Supplementary Fig. 3d,** independent replicates are
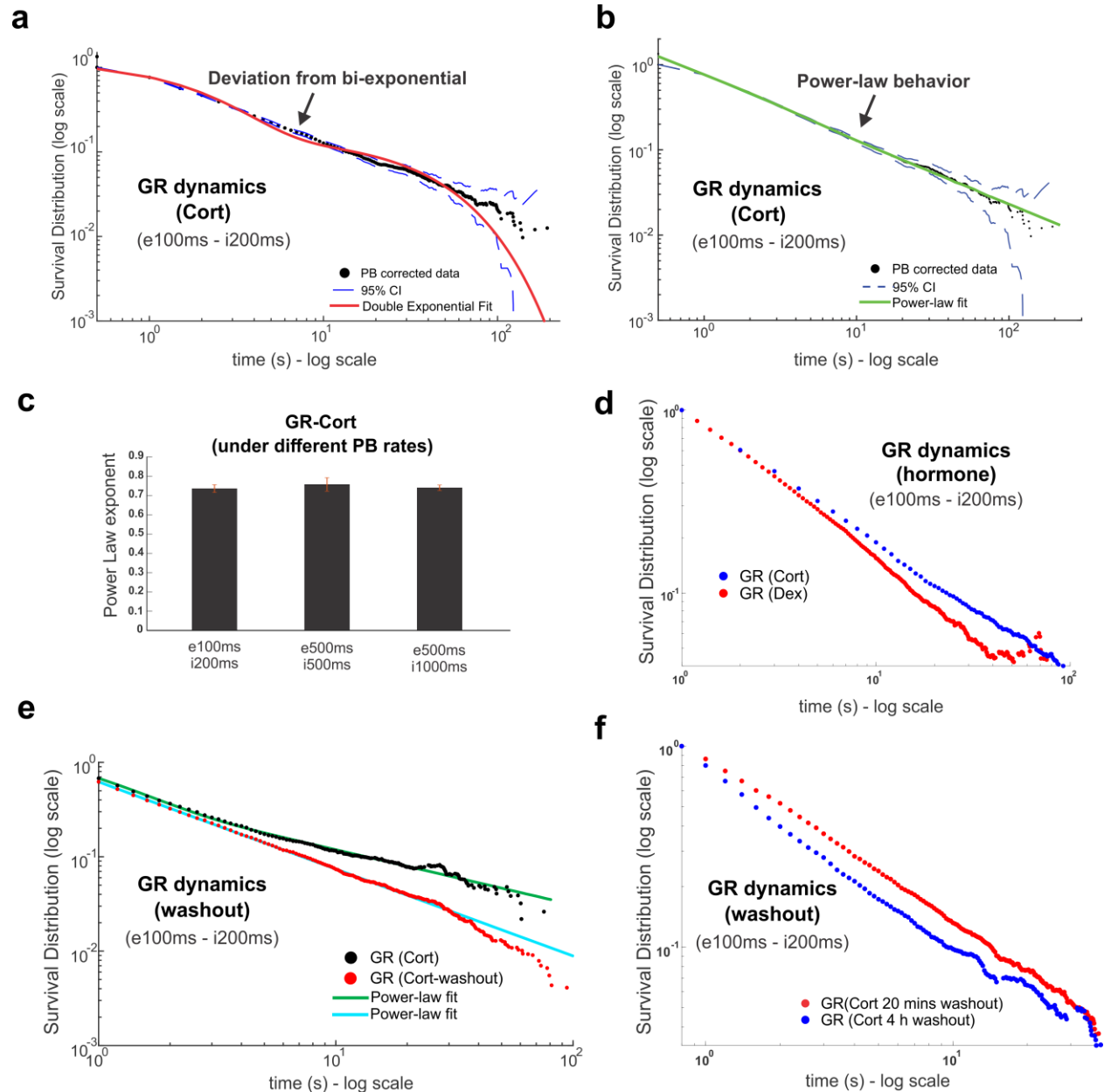
279   plotted as well to demonstrate reproducibility).

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

12

300   (**c**) Power law exponent of GR-Cort SMT data under different photobleaching rates, artificially generated
301   by modulation of acquisition conditions. (**d**) GR dynamics under Cort or dexamethasone (Dex) activation.
302   Power-law exponents of 0.71 +/- 0.05 and 0.81 +/- 0.01 respectively. The natural hormone exhibits longer
303   dwell times. (**e**) Comparison of GR dynamics under Cort or washout of the hormone, which inactivates the
304   TF. (**f**) GR dynamics under a standard 20 min or a more stringent 4h washout protocol.
305

306   A random variable t follows a power law[28] for $t > t_{min}$ if $f(t) = At^{-\beta}$, where A is

307   a constant and $\beta \in \mathbb{R}^+$ corresponds to the critical exponent, also known as the scaling

308   parameter. Power laws are heavy tailed (right-skewed) and β is a measure of this

309   skewness. Quantification of the critical exponent of the power-law shows that GR under

310   different acquisition conditions (i.e. under different PB rates) exhibits the same

311   dynamics (**Fig. 3c**). Surprisingly, GR activated with Cort shows an upward-shifted

312   (lower β) distribution compared to the more potent hormone dexamethasone (**Fig. 3d**),

313   suggesting longer residence times for the less potent ligand. These differences were not

314   apparent in previous studies[10, 17] likely due to inaccurate PB correction and/or a partial

315   exploration of the dynamic range of the protein of interest in the SMT experiments.

316   Previous research has assumed that the dynamics of non-specific binding is well

317   described by a single exponential component with a much shorter dwell time than

318   specific binding[8, 12, 17]. To understand the dynamics of non-specific binding, we

319   inactivated GR by washing out the hormone[29] for 20 minutes, which greatly reduces

320   specific binding measured by chromatin immunoprecipitation[30]. Interestingly, GR still

321   exhibits power law behavior, although with shorter dwell times as indicated by an

322   increase in the power law exponent (**Fig. 3e**). While longer washouts (four hours) shows

323   further reduction in dwell times, GR still exhibits power-law dynamics (**Fig. 3f,**

324   **Supplementary Fig. 3e, Supplementary Video 9**). In conclusion, the new PB

325   correction method has a major impact on the *distribution* of TF residence times (power-

326   law instead of bi-exponential). This, in turn, questions the interpretation of specific and

327   non-specific binding as two distinct populations with discrete (and measurable)

328   residence times.

329

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

331   To better understand the link between TF binding and the observed residence

332   time distributions, we explored different theoretical models that may explain the

333   emergence of different behaviors in the survival distribution. Calculation of dwell time

334   distributions is a first-passage time problem in stochastic analysis and has been widely

335   used to characterize the kinetic properties of molecular motors and ion channels[31].

336   When simple kinetic schemes are involved, dwell time distributions can be calculated

337   analytically. However, for more complex systems, other methods must be used. One

338   particularly powerful approach is to assign one or more states to "act" as an absorbing

339   boundary, and then solve the associated first-order kinetic equations to obtain dwell

340   time distributions[32] (**Supplementary Note 1.1)**. We assume that the diffusive state

341   (unbound) corresponds to an absorbing boundary state since tracked particles end with

342   such transitions. The single-molecule either photobleaches, disappears from the focal

343   plane or begins diffusing. Any rebinding of the TF is considered an independent event.

344   We first examined the widely used bi-exponential model under this framework

345   (**Fig. 4a**). According to this model, TFs can occupy three different states: diffusive, slow

346   and fast. The diffusive state plays the role of an absorbing boundary state (i.e. exit from

347   a bound state). In the current literature, it has been interpreted that the slow and fast

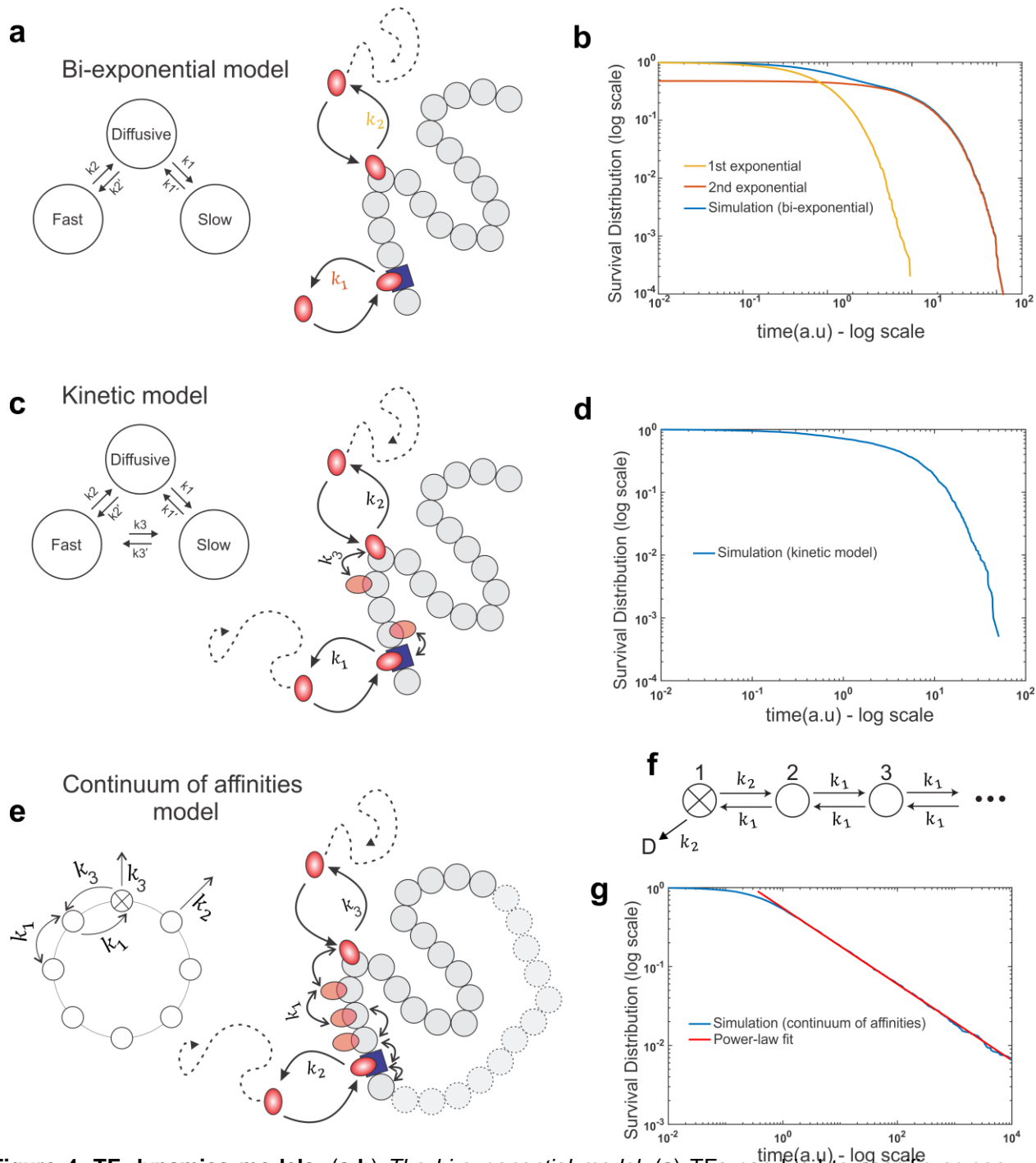348   states correspond to specific and nonspecific binding, respectively[33].

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**)

15

377 Numerical simulation of (e) showing the emergence of power-law behavior (red line). See
378 **Supplementary Note 1** for details.
379

380 With this assumption of a well separated and narrow distribution of affinities, the

381 expected behavior of the survival distribution corresponds to a double exponential with

382 exponential parameters determining the average residence time of each state, as

383 confirmed with stochastic simulations (**Fig. 4b**) using the Gillespie algorithm[34].

384 Importantly, this model does not allow for transitions between fast and slow states. For

385 a complete derivation, see **Supplementary Note 1.2**.

386   We next extended the bi-exponential model to allow for transitions between the

387 slow and fast components (**Fig. 4c**). Due to the resolution limit (~30nm), any transitions

388 between specific and non-specific bound states cannot be distinguished. The resulting

389 survival distribution corresponds to a family of exponentials; we call this a *kinetic model*

390 (**Supplementary Note 1.3**). Simulations were performed as before, and the expected

391 distribution is displayed in **Fig. 4d**.

392   Finally, several theoretical studies have posited that TF search and "final" binding

393 to its cognate site on the DNA involves a combination of bulk diffusion in the nucleus,

394 1D sliding along the DNA, hopping and translocation, and theoretical search times for

395 the TF to find specific sites in this framework have been estimated[35-37]. In this model,

396 TFs will have a multiplicity of fast bound states that must be accounted for in the

397 analysis of dwell time data. To do so, we modeled TF movement on the DNA as

398 hopping on a circular chain composed of specific sites and non-specific sites (**Fig. 4e**).

399 The main assumption in our "*continuum of affinities*" model derivation (**Supplementary**

400 **Note 1.4**) is that the number of non-specific sites on the DNA is much larger than the

401 number of specific sites. This is a biologically reasonable assumption as only a few to

16

402   tens of thousands of specific sites are bound by any TF according to genome wide

403   studies[38], while the entire genome contains millions of "other" potential chromatin sites.

404   An analytical solution can be found for the simplest case where there is a single specific

405   binding site and the TF unbinds from the specific site (**Fig. 4f, Supplementary Note**

406   **1.4.3**). A simulation based on the model gives rise to asymptotic power law behavior at

407   time scales compatible with specific binding (**Fig. 4g**).

408   To test which model better represents the observed behavior of GR dynamics,

409   we used the Bayesian information criterion (BIC)[39] and Kolmogorov–Smirnov (KS) test

410   to choose the best-predicted model (**see Methods).** Clearly, power-law behavior

411   emerges as the best fit, as illustrated for both GR-Cort (**Supplementary Fig. 4a-c)** and

412   GR-Dex (**Supplementary Fig. 4d-f)** data. Interestingly, power law distributions have

413   infinite variance, which implies a finite probability of long-lived binding events. This

414   could imply that productive binding events may be rare with dwell times much longer

415   than previously appreciated, as indicated by the right-skewness of the distribution.

416   Another possibility corresponds to a wide distribution of binding affinities in the nucleus

417   due to heterogeneity in 1) binding affinities of individual response elements, and/or 2)

418   local nuclear microenvironment.

419   In summary, by incorporating an improved PB correction method, we discovered

420   that the survival distribution of GR dwell times does not follow a bi-exponential model. In

421   fact, the data follows a power-law distribution, which we can derive using two theoretical

422   models (**Supplementary Note 1.4-1.5**). Ultimately, if there is a way to define or

423   distinguish non-specific from specific binding, our results indicate that it cannot be

424   based on their global residence times. However, the slope of the residence time

17

425   distribution does provide an estimate of the overall affinity and can be used for

426   comparing TFs and their function under different conditions.

427

428   **Discussion**

429   In the present study, we describe a new photobleaching correction method to

430   prevent photobleaching-related bias of the dwell time distribution of TFs. This method is

431   based on histone SMT dynamics, used here as a proxy for photobleaching kinetics

432   within the focal plane (**Fig. 2**). When correcting H2B SMT data by itself with this new

433   approach, we showed that the survival distribution of H2B exhibits a "plateau" (**Fig. 2c**).

434   This indicates very slow dynamics for the core histone, much longer than our

435   experimental time scale, and is fully compatible with previously published FRAP data[40].

436   Additionally, we are now able to reconcile data acquired under different experimental

437   conditions (**Fig. 2d-e**) whereas previous attempts were not successful[10, 20, 41, 42].

438   Given the current phenomenological interpretation of SMT data and the general

439   lack of a model-based approach in the field, we derived theory-based models in an

440   attempt to explain TF dynamics more accurately. We explored three different models:

441   the classic bi-exponential model, a kinetic three-state model and a power-law model

442   (**Fig. 4**). After correcting for photobleaching and implementation of Bayesian Information

443   Criterion, we could identify the best predictive model that explained the residence time

444   distribution of a paradigmatic TF, the GR. Surprisingly, the dwell time distribution of

445   activated-GR is best described by a power-law. Although a recent study corrected PB of

446   NF-κB dynamics by normalization with the H2B dwell time distribution, they still reported

447   exponential behavior of the TF[43]. However, they did not consider alternative models and

448   the range of detected experimental dwell times is relatively short (maximum ~10 s),

449   which might have prevented observation (and verification) of asymptotic power law

450   behavior.

451   Our observation of power law behavior of GR residence times suggests a model

452   with a continuum of DNA-bound states rather than discrete non-specific/specific binding

453   of TFs. Consistent with this model, wash-out of the hormone (**Fig. 3e**) revealed that the

454   dwell time distribution also follows a power law, indicating no apparent dynamical

455   differences between the so-called specific and non-specific binding. Nevertheless, the

456   overall residence time decreases when the receptor is less active, suggesting that a

457   majority of the longer events observed with the fully activated receptor are associated

458   with productive transcription as previously reported[8, 10, 12, 17, 43]. However, non-specific

459   binding can also result in TF binding events with long residence times, the implications

460   of which are still not known. Critical efforts are required to investigate whether the

461   slow(er) stops seen in SMT are matched exclusively to specific interactions with

462   chromatin. Alternatively, a sub-population of these "stops" could correspond to

463   microscopic regions in the nucleus where diffusion is severely impaired, or transient

464   interaction with "clustered" structures such as foci observed for GR[44], or another hitherto

465   unknown mechanism.

466   The broad distribution of affinities is puzzling but may be explained by

467   heterogeneity in the nuclear structure and chromatin environment. Targets for a

468   searching TF certainly exist in a wide variety of chromatin states (compacted fibers,

469   different nucleosome modification conditions, etc.). Also, affinities for the thousands of

470   alternative binding sites in response elements must vary significantly. Furthermore,

19

471   recent work points to the presence of transcriptional hubs and liquid-liquid phase

472   separation domains[45] that contribute to the complexity of nuclear organization. If TFs

473   exhibit different dynamical properties in these structures, it is not surprising to find a

474   broad variation in binding affinities. The resulting broad distribution of binding affinities

475   in these scenarios goes against the widely-held assumption that TF dynamics on

476   chromatin results from well-separated and narrow distributions of specific and non-
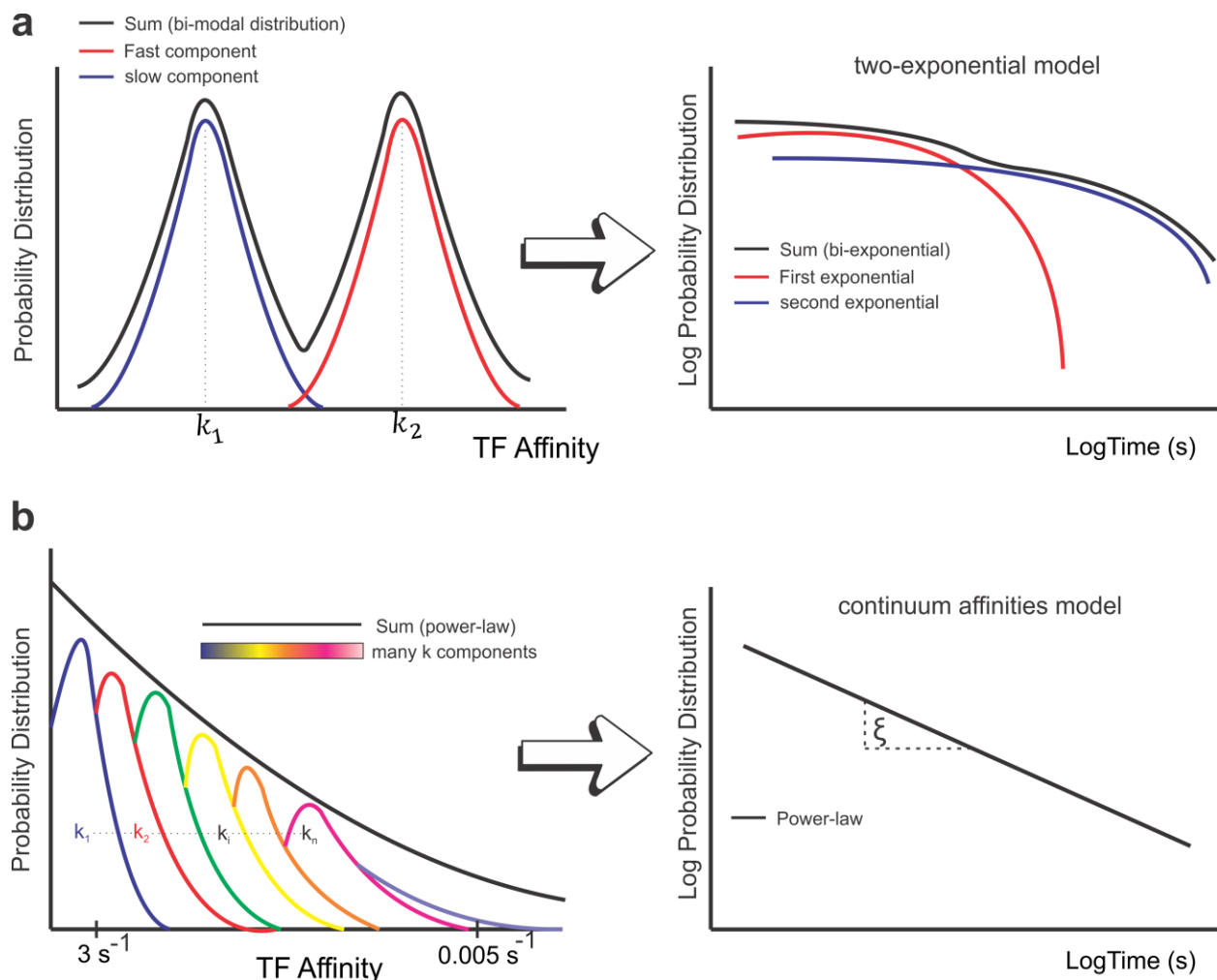
477   specific binding (**Fig. 5**).

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

20

497 affinities may explain the emergence of power law behavior (characterized by the exponent, ξ) in the
498 residence time of TFs (right graph).
499
500
501 Given the heterogeneities in local organization and nuclear structure, TF binding sites

502 on chromatin can be viewed as a collection of traps with a distribution of trap depths

503 (analogous to binding affinities). In such a finite disordered system, the distribution of

504 trapping times asymptotically approaches a power law[46, 47] (**Fig. 5b, Supplementary**

505 **Note 1.5**)

506   Alternatively, but not mutually exclusive, heterogeneity in the searching

507 mechanism of TFs may affect the effective affinity constant observed in SMT

508 experiments. In support of the latter, the tetracycline repressor (TetR), a chimera

509 between a bacterial and a viral protein with no known endogenous targets in

510 mammalian cells, when used as a proxy to emulate TF dynamics, also showed power-

511 law behavior for "non-specific binding"[48, 49]. However, it could still be described as an

512 exponential on an artificially (and single) specific DNA binding array[49]. Thus, the intrinsic

513 nature of the searching mechanism of any DNA-binding protein may be governed by

514 power-law dynamics. In addition, the heterogeneity of dwell times in the thousands of

515 response elements for an endogenous TF could explain why GR can present power-law

516 tails as opposed to TetR, which can only bind to one artificial array site. Interestingly, a

517 recent study in yeast[50] reports that both the TF Ace1p and the chromatin remodeler

518 RSC binding follow a bi-exponential binding distribution in cells containing a natural

519 tandem of ten *CUP1* (Ace1p responsive) *genes.* This dynamic and discrete behavior, in

520 contrast with our GR data, can be explained by the particular and homogeneous

521 chromatin environment of single array of specific sites. Consequently, we speculate that

21

522   a broad distribution of binding affinities coming from a whole population of different

523   binding sites (thousands in the case of GR) may result in a power-law behavior (**Fig. 5**).

524   In summary, by the implementation of proper photobleaching kinetics, we reveal a new

525   model of TF dynamics. Our findings suggest that, contrary to the established paradigm,

526   TF dwell times follow a broad distribution with no evidence of binary, discrete

527   populations.

528

Acknowledgments

Author Contributions

540   D.A.G, G.F., D.M.P., and G.L.H. conceived the experiments and wrote the

541   original draft with subsequent input and editing from all authors. D.A.G, G.F., and

542   D.M.P. performed imaging experiments, analyzed data and prepared figures. V.P.

543   participated in the early stages of the project. D.A.G conceived and wrote in MATLAB®

544   the photobleaching correction method (with constant feedback from G.F and D.M.P.).

545   D.A.G and C.J. derived the theoretical models, performed the simulations, and

546   implemented the Bayesian statistical analysis. A.U. supervised data analysis, theoretical

547   models and participated in the manuscript preparation; G.L.H. supervised the project.

548   **Competing Interests statement**

549   The authors declare no competing interests

550   **References**

551   1.   Lambert, S.A. et al. The Human Transcription Factors. *Cell* **172**, 650-665 (2018).
552   2.   Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J Cell*
553   *Biol* **217**, 1181-1191 (2018).
554   3.   Coulon, A., Chow, C.C., Singer, R.H. & Larson, D.R. Eukaryotic transcriptional
555   dynamics: from single molecules to cell populations. *Nat Rev.Genet.* **14**, 572-584
556   (2013).
557   4.   Grimm, J.B. et al. Bright photoactivatable fluorophores for single-molecule
558   imaging. *Nat Methods* **13**, 985-988 (2016).
559   5.   Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination
560   enables clear single-molecule imaging in cells. *Nat Methods* **5**, 159-161 (2008).
561   6.   Goldstein, I. et al. Transcription factor assisted loading and enhancer dynamics
562   dictate the hepatic fasting response. *Genome Res* **27**, 427-439 (2017).
563   7.   Presman, D.M. et al. Quantifying transcription factor dynamics at the single-
564   molecule level in live cells. *Methods* **123**, 76-88 (2017).
565   8.   Mazza, D., Abernathy, A., Golob, N., Morisaki, T. & McNally, J.G. A benchmark
566   for chromatin binding measurements in live cells. *Nucleic Acids Res* **40**, e119
567   (2012).
568   9.   Izeddin, I. et al. Single-molecule tracking in live cells reveals distinct target-
569   search strategies of transcription factors in the nucleus. *Elife* **3**, e02230 (2014).
570   10.   Paakinaho, V. et al. Single-molecule analysis of steroid receptor and cofactor
571   action in living cells. *Nat Commun* **8**, 15896 (2017).
572   11.   Ball, D.A. et al. Single molecule tracking of Ace1p in Saccharomyces cerevisiae
573   defines a characteristic residence time for non-specific interactions of
574   transcription factors with chromatin. *Nucleic Acids Res* **44**, e160 (2016).
575   12.   Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in
576   embryonic stem cells. *Cell* **156**, 1274-1285 (2014).
577   13.   Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and
578   cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6** (2017).
579   14.   Kieffer-Kwon, K.-R. et al. Myc regulates chromatin decompaction and nuclear
580   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).

626 34. Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J Phys*
627   *Chem* **81**, 2340-2361 (1977).
628 35. Bauer, M. & Metzler, R. Generalized facilitated diffusion model for DNA-binding
629   proteins with search and recognition states. *Biophys J* **102**, 2321-2330 (2012).
630 36. Berg, O.G. & Blomberg, C. Association kinetics with coupled diffusional flows.
631   Special application to the lac repressor--operator system. *Biophysical chemistry*
632   **4**, 367-381 (1976).
633 37. Marklund, E.G. et al. Transcription-factor binding and sliding on DNA studied
634   using micro- and macroscopic models. *Proc Natl Acad Sci U S A* **110**, 19796-
635   19801 (2013).
636 38. Wang, J. et al. Sequence features and chromatin structure around the genomic
637   regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798-1812
638   (2012).
639 39. Schwarz, G. Estimating the dimension of a model. *Ann Stat.* **6**, 461-464 (1978).
640 40. Kimura, H. Histone dynamics in living cells revealed by photobleaching. *DNA*
641   *repair* **4**, 939-950 (2005).
642 41. Agarwal, H., Reisser, M., Wortmann, C. & Gebhardt, J.C.M. Direct Observation
643   of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. *Biophys J*
644   **112**, 2051-2055 (2017).
645 42. Clauss, K. et al. DNA residence time is a regulatory factor of transcription
646   repression. *Nucleic Acids Res* **45**, 11121-11130 (2017).
647 43. Callegari, A. et al. Single-molecule dynamics and genome-wide transcriptomics
648   reveal that NF-kB (p65)-DNA binding times can be decoupled from transcriptional
649   activation. *PLoS Genet* **15**, e1007891 (2019).
650 44. Stortz, M. et al. Mapping the Dynamics of the Glucocorticoid Receptor within the
651   Nuclear Landscape. *Sci Rep* **7**, 6219 (2017).
652 45. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase
653   Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
654 46. Bouchaud, J.-P. & Georges, A. Anomalous diffusion in disordered media:
655   Statistical mechanisms, models and physical applications. *Physics Reports* **195**,
656   127-293 (1990).
657 47. J., B. Weak ergodicity breaking and aging in disordered systems. *Journal de*
658   *Physique I* **2**, 1705-1713 (1992).
659 48. Caccianini, L., Normanno, D., Izeddin, I. & Dahan, M. Single molecule study of
660   non-specific binding kinetics of LacI in mammalian cells. *Faraday discussions*
661   **184**, 393-400 (2015).
662 49. Normanno, D. et al. Probing the target search of DNA-binding proteins in
663   mammalian cells using TetR as model searcher. *Nat Commun* **6**, 7357 (2015).
664 50. Mehta, G.D. et al. Single-Molecule Analysis Reveals Linked Cycles of RSC
665   Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. *Mol*
666   *Cell* **72**, 875-887.e879 (2018).

667