Corresponding author(s): Dr. Patrick Cramer
Dr. Xavier Darzacq
Dr. Markus Zweckstetter

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The in vitro datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The PALM and SPT data are available at http://doi.org/10.5281/zenodo.1188488. The raw spaSPT data is available in Spot-On readable CSV format in the form of single-molecule trajectories.
Data analysis	The Spot-On Matlab code is available together with a step-by-step guide at Gitlab: https://gitlab.com/tjian-darzacq-lab/spot-on-matlab. For additional documentation, please see also the Spot-On website https://SpotOn.berkeley.edu and previous publications 28,30. All other codes are available from corresponding author on reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The in vitro datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The PALM and SPT data are available at http://doi.org/10.5281/zenodo.1188488. The raw spaSPT data is available in Spot-On readable CSV format in the form of single-molecule trajectories. The Spot-On Matlab code is available together with a step-by-step guide at Gitlab: https://gitlab.com/tjian-darzacq-lab/spot-on-matlab. For additional documentation, please see also the Spot-On website https://SpotOn.berkeley.edu and previous publications (Ref. 34, Ref. 36). All other codes used in this study are available from corresponding author on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For single molecule analysis, sample size was analysed and described in Hansen et al, 2018. For FRAP and Clustering analysis statistical reproducibility was determined by datas resampling.
Data exclusions	No data were excluded.
Replication	All attempts at replication were successful.
Randomization	Randomization was not required in this study.
Blinding	Blinding was not required in this study and investigators were not blinded to group allocation.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-Ser5P CTD antibody (kind gift of Dirk Eick (Helmholtz Center Munich), monoclonal antibody, clone 3E8, supplied as hybridoma supernatant) anti-MBP conjugate (Abcam, ab49923, HRP-coupled, monoclonal antibody, MBP-17, GR251034-1) anti-GST antibody (GE healthcare, RPN1236, HRP-coupled, monoclonal antibody) anti-rat antibody (Sigma Aldrich, A9037, HRP-coupled, polyclonal antibody, SLBL0586V) anti-Pol II antibody (N20, Santa Cruz , sc-899) anti-Lamin A antibody (abcam, ab26300)
Validation	Anti-Ser5P CTD antibody, anti-GST-HRP antibody, and anti-MBP-HRP antibody were used to probe TFIIH kinase activity. In the

study antibodies were used to detect highly purified proteins. Validity of detection was cross-confirmed by simultaneous polyacrylamide gel electrophoresis and subsequent coomassie staining, based on the known migration behavior of the unphosphorylated protein species. Identity of size-shifted (i.e. phosphorylated) CTD-proteins was confirmed through the use of two antibodies, which bind the target in an orthogonal manner (e.g. anti-MBP and anti-Ser5P). Anti-Ser5P CTD antibody is of monoclonal origin (clone 3E8) and was thoroughly validated with synthetic peptides to require phosphorylated Ser5 for binding (Chapman et al. (2007), Science 318, 1780-1782). The antibody was further characterized in previous, peer-reviewed publications (e.g. Clemente-Blanco et al. (2011), Nat Cell Biol. 13, 1450-1456; Xu et al. (2012), Dev. Cell 23, 1059-71; Manafra et al. (2014), Plos One 9, e99603).
N20 anti-Pol II antibody was used to assess the RPB1 level expressed in the different human cell lines. Another anti-RPB1 antibody was used to compare the size of the detected bands with the ones detected by the N20 antibody. The result was consistent. Due to changes in the CTD length between the different cell lines, antibodies against the RPB1 C-terminal domain cannot be used. The N20 antibody was additionally successfully used in numerous peer-reviewed studies (e.g. Emmett et al., Nature 546, 544-548, 2017; Dieuleveult et al., Nature 530, 113-6, 2016; Forget et al., NAR 41, 6881-6891, 2015).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human U2OS osteosarcoma cells (Research Resource Identifier: RRID:CVCL_0042)
Authentication	The parental U2OS cell line was authentificated by the UC Berkeley Cell culture facility on 5/5/2017 by STR analysis . The result was a 100% match with U2OS cell line.
Mycoplasma contamination	The parental U2OS cell lines were tested for mycoplasma contamination before establishing the RPB1 cell lines. Then the cells are tested every 6 months or so. We never found mycoplasma contamination for these lines.
Commonly misidentified lines (See ICLAC register)	NA

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For the Halo-tagged line, Halo-TMR labelling (500 nM) was performed for 30mn at 37°C before harvesting the cells.The cells were harvested by Trypsinization after wash with PBS. Then Cells were resuspended in media, span down and resuspended in PBS for analysis.
Instrument	BD LSRFortessa
Software	Diva software
Cell population abundance	10,000 cells were analysed per sample
Gating strategy	Gate between positive and negatively labelled cells were fixed running WT non labelled cells treated the same way the day of each experiment. The gating strategy was used only to compare the level of expression in between cells. Same gate used for all the conditions. The gate is represented on each graph. This is more a qualitative than quantitative analysis.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

nature research

Corresponding author(s): Prof. Markus Zweckstetter

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	For the Halo-tagged line, Halo-TMR labelling (500 nM) was performed for 30mn at 37°C before harvesting the cells.The cells were harvested by Tripsynisation after wash with PBS. Then Cells were resuspended in media, span down and resuspended in PBS for analysis
6. Identify the instrument used for data collection.	BD LSRFortessa
7. Describe the software used to collect and analyze the flow cytometry data.	Diva software
8. Describe the abundance of the relevant cell populations within post-sort fractions.	10,000 cells were analysed per sample
9. Describe the gating strategy used.	gate between positive and negatively labelled cells were fixed running WT non labelled cells treated the same way the day of each experiment. The gating strategy was used only to compare the level of expression in between cells...same gate used for all the conditions. The gate is represented on each graph. This is more a qualitative than quantitative analysis.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒