nature research

Corresponding author(s): Phillip A. Sharp

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	-FACS data was collected with BD FACS Diva Version 8.0.1 -Comet assay slides were imaged using a Nikon 80i fluorescence microscope and Nikon NIS Elements Advanced Research Miscroscope Imagine Software v.3.22.14 (Build 736) -Western blots were developed on film and electronically scanned using a Canon CanoScan 880F scanner with ArcSoft photoSutdio4 software or a Epson Perfection v.19 scanner with Epson Scan2 v.6.2.8 software.
Data analysis	-Western blots were cropped in Adobe Photoshop CC 2018 and Adobe Illustrator CC 2018. -FACS data was analyzed with FlowJo Version 1.0.1 -Comet Assay slide images were analyzed with ImageJ version 2.0.0-rc-43/1.51w using the OpenComet v1.3.1 plugin (http://www.cometbio.org/); plots of olive tail moment and corresponding statistical analysis were made in GraphPad PRISM 7 software. -Reads were mapped with Tophat v. 2.0.12 or STAR aligner version 2.4.1d (RNA sequencing) or BWA aligner version 0.7.12 (ChIP Sequencing) -Gene-level quantification was performed using Rsem version 1.2.26, EBSeq version 1.10.1, and DESeq version 1.22.1 -Quantification of alternative splicing events was performed using MISO version 0.4.1 -IPA transcript quantification was performed using DEXSeq version 1.16.10 -Plots quantifying IPA site usage in human tumors from TCGA (along with corresponding statistical analysis) and were made using GraphPad PRISM 7 software -IGV browser shots were made using IGV v.2.3.82 or v.2.4.13

-Custom scripts were used throughout the bioinformatics analysis as indicated in the materials and methods and will be made available upon 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

RNA sequencing and ChIP sequencing data is deposited in the Gene Expression Omnibus under accession number GSE116017. Custom scripts used for analysis will be made available upon 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	Sample sizes were not predetermined. Instead, sample sizes were selected based upon the availability of previously published data as outlined below. TCGA cohorts of prostate adenocarcinoma and ovarian serous cystadenocarcinoma were assessed for the presence of missense or truncating point mutations as well as copy-number variations (amplifications or deletions) in CDK12 and BRCAness genes using cBioPortal (www.cbioportal.org). Additionally, normalized CDK12 mRNA expression levels were considered. We included all tumors from these cohorts in our analysis that were predicted to have CDK12 loss-of-function (LOF) mutation(s) as annotated in cBioPortal. We considered tumors to be likely CDK12 LOF if they carried at least one truncating or missense putative driver mutation or if the copy number analysis classified the tumor as carrying a deep deletion (indicating a likely homozygous deletion across the locus) and if the mRNA expression levels of CDK12 were significantly downregulated compared to the mean expression of WT, diploid CDK12 tumors. We additionally included 4-5 shallow deletions from each tumor type that exhibited the lowest mRNA expression levels of CDK12, and one in-frame deletion mutant of unknown functional consequence. As an additional control for specificity, we included a single tumor carrying a K975E mutation that had been previously validated as a missense mutation with no LOF. The only CDK12 mutated tumors we excluded from consideration were three tumors in prostate carrying missense mutations of unknown consequence as we could not accurately classify them as WT or LOF. 24 total CDK12 mutated tumors fit this criteria (n=12 prostate; n=12 ovarian). A set of patients from the prostate and ovarian cohorts with wild-type, diploid CDK12 loci were selected along with one or two samples from each tumor type carrying an amplified locus (prostate n=1; ovarian n=2). Among all such tumors, this control set was selected by ranking the tumors in order of normalized CDK12 mRNA expression and taking the subset with the highest expression (9 from each cohort). For the BRCAness control subset (n=23; ovarian n=12 and prostate n=12), we selected a set of tumors that carried only “likely oncogenic” missense or truncating mutations and selected a set that contained all available BRCAness genes, as well as larger samples for genes that are more frequently mutated (e.g. BRCA1/2 and CHEK2). These tumors were selected without considering CDK12 gene expression levels. Retrospectively, we also determined that two of the CDK12 WT diploid tumors carried putative deep deletions with mRNA loss in FANCA, CHEK1, and CHEK2 in one tumor and BRCA2 in the other; these tumors exhibited identical low IPA usage consistent with the other CDK12 WT tumors. Once the tumor sets were selected based on these genomic characteristics, sequencing data from all of these tumors and only these tumors were downloaded and included in the quantifications performed.
Data exclusions	Data exclusion criteria for the TCGA analysis (Figure 4F,G and Extended Data Fig. 10A,B) was predetermined in that only tumors with RNA sequencing reads mapping to the genes of interest (ATM, FANCD2, or WRN) would be included in the isoform analysis. Two tumors with CDK12 mutations were excluded from the analysis of ATM IPA #1 and #2 because they showed no RNA sequencing reads mapping across the entire ATM locus (perhaps due to a deletion at the ATM locus).
Replication	All attempts at replication were successful. All results were replicated in two independent clonal cell lines (ie. two different Cdk12 knockout clones, two different V5 epitope-tagged clones, etc) in addition to replication within the same clonal cell line. For THZ531 treatments of human cancer cell lines (Extended Data Fig. 10C,D) we used three independent cancer cell lines, 2 cell lines derived from patients with prostate carcinoma (22RV1 and PC-3) and one cell line derived from an ovarian carcinoma patient (OVCAR4), to increase the robustness of our results.

Randomization	No randomization was used throughout this study. Randomization was not relevant to our study because we are not assigning samples, organisms, or participants to treatment groups.
Blinding	The researcher was blinded during imagining of the data reported in Extended Data Fig. 1G.

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All materials are commercially available or are available from the Sharp Lab upon request.

Antibodies

Antibodies used

Primary Antibodies Used:
(1) Anti-HA High Affinity Antibody (Roche 11867423001):
-Rat IgG1 Monoclonal Antibody, Clone 3F10
-Multiple Lots Used
-200ng/mL dilution used for Western Blots

(2) Enolase I (CST 3810S)
-Rabbit polyclonal antibody
-Multiple Lots Used
-1:1000-1:5000 dilution used for Western Blots

(3) Vinculin (Sigma V9131)
-Mouse IgG1 Monoclonal Antibody
-Lot #129K4849
-1:5,000-1:10,000 dilution used for Western Blots

(4) p53 (1C12) (CST 2524)
-Mouse IgG1 Monoclonal Antibody, Clone 1C12
-Lot #10
-1:1000-1:2000 dilution used for Western Blots

(5) P-p53 Serine15 (CST 9284S)
-Rabbit Polyclonal Antibody
-Lot# 18
-1:1000-1:2000 dilution used for Western Blots

(6) ATR (CST 13934S)
-Rabbit IgG Monoclonal Antibody, Clone E1S3S
-Lot# 1
-1:1000 dilution used for Western Blots

(7) FANCD2 (Abcam ab108928)
-Rabbit IgG monoclonal, EPR2302
-Lot# GR130039-19
-1:1000 dilution used for Western Blots

(8) FITC Rabbit Anti-Active Caspase 3 (BD Biosciences 550480)
-Rabbit IgG monoclonal, clone C92-605
-Multiple lots used
-Manufacturer's protocol was followed for dilution used in FACS (20uL of antibody in 100uL of 1X Perm/Wash Buffer per 1*10^6

cells)

(9) V5 (Life Technologies R96025)
-Mouse IgG2a monoclonal
-Multiple lots used
-1:1000 dilution used for Western Blot

(10) HSP90 (BD 610418)
-Mouse IgG1 monoclonal
-Lot# 5183592 used
-1:5,000-10,000 dilution used for Western Blots

(11) 8WG16 (Abcam ab817)
-Mouse IgG2a monoclonal, clone 8WG16
-Multiple lots used
-1:2000 used for Western Blots

(12) 3E10 (Millipore 04-1571)
-Rat IgG1k monoclonal, clone 3E10
-Multiple lots used.
-1:3000 dilution used for Western Blots

Secondary Antibodies Used (all blots except the V5 epitope tag):
(1) ECL Anti-Rat IgG (GE Healthcare NA935V)
-1:5000-1:50,000 dilution for Western Blots

(2) ECL Anti-Mouse IgG (GE Healthcare NA931V)
-1:5000-1:50,000 dilution for Western Blots

(3) ECL Anti-Rabbit IgG (GE Healthcare NA934V)
-1:5000-1:50,000 dilution for Western Blots

Secondary Antibody for blots with the V5 epitope tag:
Anti-Mouse IgG, HRP-linked antibody (CST 7076S):
-1:1000-1:2000 dilution for Western Blots

Note: Please see the ChIP Sequencing section of the form for information on antibodies used in the ChIP sequencing experiments.

Validation

All antibodies were validated by the manufacturer and used extensively throughout the literature. For all Western Blots where V5-epitope antibody was used, lysates from parental cells without introduction of the V5 epitope tag were run in parallel as a control for antibody specificity. All ChIP sequencing samples were performed alongside ChIP Western Blots (data not shown) to make sure that the antibody was appropriately recognizing the phosphorylation status of the CTD in our immunoprecipitations as predicted by mobility shifts.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Cdk12 knockout clones used in this study were derived from V6.5 (C57Bl/6-129) mouse embryonic stem cells, which were a gift from Rudolf Jaenisch's lab. 22RV1 and PC-3 cells were purchased directly from ATCC. OVCAR4 cells were a gift from Paula Hammond's lab and obtained from the Koch Institute's High Throughput Sciences Facility's Cell Line Repository.

Authentication

All human cell lines were authenticated upon arrival by STR profiling performed by ATCC using 17 short tandem repeat (STR) loci plus the gender determining locus, Amelogenin. Cells were considered authenticated if they were a 80% or greater allelic match to published STR profiles. The results are summarized below for each individual cell line.
(1) OVCAR-4 cells authenticated with a 100% match to the STR profile for OVCAR-4 cells reported in Yu, M. et al. A resource for cell line authentication, annotation, and quality control. Nature 520, 307-311 (2015).
(2) 22RV1 cells authenticated with a 94% allelic match to ATCC's internal STR profile for the 22RV1 (ATCC CRL-2505) cell line and with a 94% allelic match to the STR profile for 22RV1s from the ATCC "source" reported in Yu, M. et al. A resource for cell line authentication, annotation, and quality control. Nature 520, 307-311 (2015).
(3) PC-3 cells authenticated with a 100% allelic match to ATCC's internal STR profile for their PC-3 (ATCC CRL-1435) cell line and with a 100% allelic match to the STR profile for PC-3 cells reported in Lorenzi, P.L. et al. DNA fingerprinting of the NCI-60 cell line panel. Molecular Cancer Therapeutics 8, 713-724 (2009)."

Mycoplasma contamination

All cell lines were tested for mycoplasma contamination periodically, including immediately upon receipt via the MycoAlert Mycoplasma Testing Kit (Lonza). Results were always negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cells lines were used in this study.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

ChIP sequencing data is deposited in the Gene Expression Omnibus under accession number GSE116017.

Files in database submission

GSM3207205 Cdk12△_Clone36_plusDox_8WG16_rep1_ChIPseq
GSM3207206 Cdk12△_Clone36_plusDox_8WG16_rep2_ChIPseq
GSM3207207 Cdk12△_Clone36_plusDox_Rpb3_rep1_ChIPseq
GSM3207208 Cdk12△_Clone36_plusDox_Rpb3_rep2_ChIPseq
GSM3207209 Cdk12△_Clone36_plusDox_H5_rep1_ChIPseq
GSM3207210 Cdk12△_Clone36_plusDox_H5_rep2_ChIPseq
GSM3207211 Cdk12△_Clone36_plusDox_3E10_rep1_ChIPseq
GSM3207212 Cdk12△_Clone36_plusDox_H14_rep1_ChIPseq
GSM3207213 Cdk12△_Clone36_plusDox_H14_rep2_ChIPseq
GSM3207214 Cdk12△_Clone36_plusDox_3E8_rep1_ChIPseq
GSM3207215 Cdk12△_Clone36_plusDox_negContIP_antiMsIgM_ChIPseq
GSM3207216 Cdk12△_Clone36_plusDox_negContIP_msIgG2a_IsotypeControl_ChIPseq
GSM3207217 Cdk12△_Clone36_plusDox_negContIP_antiRatIgG_ChIPseq
GSM3207218 Cdk12△_Clone36_plusDox_InputControl
GSM3207219 Cdk12△_Clone36_minusDox_8WG16_rep1_ChIPseq
GSM3207220 Cdk12△_Clone36_minusDox_8WG16_rep2_ChIPseq
GSM3207221 Cdk12△_Clone36_minusDox_Rpb3_rep1_ChIPseq
GSM3207222 Cdk12△_Clone36_minusDox_Rpb3_rep2_ChIPseq
GSM3207223 Cdk12△_Clone36_minusDox_H5_rep1_ChIPseq
GSM3207224 Cdk12△_Clone36_minusDox_H5_rep2_ChIPseq
GSM3207225 Cdk12△_Clone36_minusDox_3E10_rep1_ChIPseq
GSM3207226 Cdk12△_Clone36_minusDox_3E10_rep2_ChIPseq
GSM3207227 Cdk12△_Clone36_minusDox_H14_rep1_ChIPseq
GSM3207228 Cdk12△_Clone36_minusDox_3E8_rep1_ChIPseq
GSM3207229 Cdk12△_Clone36_minusDox_3E8_rep2_ChIPseq
GSM3207230 Cdk12△_Clone36_minusDox_negContIP_antiMsIgM_ChIPseq
GSM3207231 Cdk12△_Clone36_minusDox_negContIP_msIgG2a_IsotypeControl_ChIPseq
GSM3207232 Cdk12△_Clone36_minusDox_negContIP_antiRatIgG_ChIPseq
GSM3207233 Cdk12△_Clone36_minusDox_InputControl
GSM3207234 Cdk12△_Clone36_minusDox_H14_rep2_ChIPseq
GSM3207235 Cdk12△_Clone36_plusDox_3E8_rep2_ChIPseq
GSM3207236 Cdk12△_Clone36_plusDox_3E10_rep2_ChIPseq
GSE116017_NFH36minus_3E10_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36minus_3E8_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36minus_8WG16_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36minus_H14_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36minus_H5_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36minus_Rpb3_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36minus_negContIPs_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_3E10_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_3E8_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_8WG16_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_H14_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_H5_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_Rpb3_merge_sorted_bigwig_normalized.bw
GSE116017_NFH36plus_negContIPs_merge_sorted_bigwig_normalized.bw

Genome browser session
(e.g. UCSC)

Not applicable to our study.

Methodology

Replicates

We performed ChIP on Cdk12Δ cells in +Dox or –Dox 48 hour conditions for total RNAPII density and Ser2p RNAPII density. Two independent antibodies were used for each protein target as follows: 8WG16 (Abcam ab817) and Rpb3 (Bethyl A303-771A) for total RNAPII density; and H5 Clone (Abcam ab24758) and 3E10 Clone (Millipore 04-1571) for Ser2p RNAPII density. Two biological replicates were processed for each antibody and dox condition (for an example of the experimental setup for one protein target, e.g. RNAPII, see Extended Data Fig. 4). For each dox condition, we also processed 4 negative control libraries: one whole-cell extract (WCE) sample and 3 mock IP samples with the following antibodies: goat anti-mouse IgM (Thermo Fisher 31172), goat anti-rat IgG (Thermo Fisher 31226), mouse IgG2a [MOPC-173] Isotype Control (Abcam ab18413).

Sequencing depth

All reads are 40bp and paired-end. The number of mapped reads per library is shown below. Samtools v. 0.1.19 Flagstat was used to calculate number of mapped reads treating each end of paired reads as a separate read.

GSM3207205 Cdk12△_Clone36_plusDox_8WG16_rep1_ChIPseq (28532770 reads)
GSM3207206 Cdk12△_Clone36_plusDox_8WG16_rep2_ChIPseq (42068228 reads)
GSM3207207 Cdk12△_Clone36_plusDox_Rpb3_rep1_ChIPseq (30061159 reads)
GSM3207208 Cdk12△_Clone36_plusDox_Rpb3_rep2_ChIPseq (40346622 reads)
GSM3207209 Cdk12△_Clone36_plusDox_H5_rep1_ChIPseq (33903383 reads)
GSM3207210 Cdk12△_Clone36_plusDox_H5_rep2_ChIPseq (29391678 reads)
GSM3207211 Cdk12△_Clone36_plusDox_3E10_rep1_ChIPseq (29769307 reads)
GSM3207212 Cdk12△_Clone36_plusDox_H14_rep1_ChIPseq (31165035 reads)
GSM3207213 Cdk12△_Clone36_plusDox_H14_rep2_ChIPseq (37078600 reads)
GSM3207214 Cdk12△_Clone36_plusDox_3E8_rep1_ChIPseq (37621855 reads)
GSM3207215 Cdk12△_Clone36_plusDox_negContIP_antiMsIgM_ChIPseq (18932489 reads)
GSM3207216 Cdk12△_Clone36_plusDox_negContIP_msIgG2a_IsotypeControl_ChIPseq (12657991 reads)
GSM3207217 Cdk12△_Clone36_plusDox_negContIP_antiRatIgG_ChIPseq (15805164 reads)
GSM3207218 Cdk12△_Clone36_plusDox_InputControl (37549659 reads)
GSM3207219 Cdk12△_Clone36_minusDox_8WG16_rep1_ChIPseq (40179760 reads)
GSM3207220 Cdk12△_Clone36_minusDox_8WG16_rep2_ChIPseq (27653176 reads)
GSM3207221 Cdk12△_Clone36_minusDox_Rpb3_rep1_ChIPseq (32304434 reads)
GSM3207222 Cdk12△_Clone36_minusDox_Rpb3_rep2_ChIPseq (66109499 reads)
GSM3207223 Cdk12△_Clone36_minusDox_H5_rep1_ChIPseq (28896925 reads)
GSM3207224 Cdk12△_Clone36_minusDox_H5_rep2_ChIPseq (21841736 reads)
GSM3207225 Cdk12△_Clone36_minusDox_3E10_rep1_ChIPseq (30724266 reads)
GSM3207226 Cdk12△_Clone36_minusDox_3E10_rep2_ChIPseq (34932278 reads)
GSM3207227 Cdk12△_Clone36_minusDox_H14_rep1_ChIPseq (21609399 reads)
GSM3207228 Cdk12△_Clone36_minusDox_3E8_rep1_ChIPseq (25415620 reads)
GSM3207229 Cdk12△_Clone36_minusDox_3E8_rep2_ChIPseq (26933000 reads)
GSM3207230 Cdk12△_Clone36_minusDox_negContIP_antiMsIgM_ChIPseq (17525772 reads)
GSM3207231 Cdk12△_Clone36_minusDox_negContIP_msIgG2a_IsotypeControl_ChIPseq (14321400 reads)
GSM3207232 Cdk12△_Clone36_minusDox_negContIP_antiRatIgG_ChIPseq (15853066 reads)
GSM3207233 Cdk12△_Clone36_minusDox_InputControl (18373247 reads)
GSM3207234 Cdk12△_Clone36_minusDox_H14_rep2_ChIPseq (22614740 reads)
GSM3207235 Cdk12△_Clone36_plusDox_3E8_rep2_ChIPseq (30003179 reads)
GSM3207236 Cdk12△_Clone36_plusDox_3E10_rep2_ChIPseq (30517062 reads)

Antibodies

(1) 8WG16 (Abcam ab817)
-Mouse IgG2a monoclonal, clone 8WG16
-Multiple lots used
-Antibody specificity was validated by several previous publications, one good example:
Bataille, A.R., et al. A Universal RNA Polymerase II CTD Cycle is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. Molecular Cell 45, 158-170 (2012).

(2) Rpb3 (Bethyl A303-771A)
-Rabbit polyclonal
-Lot# A303-771A-1
-Validated by manufacturer.
-Previous publication:
Hintermair, C. et al. Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. Scientific Reports 6, 1-13 (2016).

(3) H5 (Abcam ab24758)
- Mouse IgM monoclonal, clone H5
- Multiple lots used.
-Specificity validated in several publications for this particular use, one good example:
Chapman, R.D. et al. Transcribing RNA Polymerase II is Phosphorylated at CTD Residue Serine-7. Science 318,1780-1782 (2007).

(4) 3E10 (Millipore 04-1571)
-Rat IgG1k monoclonal, clone 3E10
-Multiple lots used.
-Specificity validated in several publications for this particular use, one good example:
Chapman, R.D. et al. Transcribing RNA Polymerase II is Phosphorylated at CTD Residue Serine-7. Science 318,1780-1782 (2007).

(5) mouse IgG2a [MOPC-173] Isotype Control (Abcam ab18413)
-Mouse IgG2a monoclonal, clone MOPC-173
-Non-targeting, Isotype Matched Control
-Validated by manufacturer as described here: http://www.abcam.com/mouse-igg2a-mopc-173-isotype-control-chip-grade-ab18413.html

Peak calling parameters

No peak calling was performed. Instead, we analyzed RNAPII density genome-wide by metagene analysis. Please refer to the materials and methods as well as the schematic in Extended Data Figure 4b-d for extensive explanation on our anlaysis pipeline and statistical analysis.

Data quality

To increase robustness in our ChIP sequencing data, we isolated ChIP samples for both RNAPII and Ser2p RNAPII using two

Data quality	independent antibodies for each target and repeated the ChIP in biological duplicate for each one of these antibodies and conditions (16 libraries in total). Comparing metagene profiles of the average read densities of all ChIPs across biological replicates indicates that ChIP profiles from the two independent antibodies are highly reproducible and similar (Extended Data Fig. 3). We therefore aggregated data from both antibodies and both biological replicates throughout our metagene analyses in order to effectively incorporate four independent biological replicates per condition (Extended Data Fig. 4A). Furthermore, we developed a robust statistical framework (Extended Data Fig. 4B-D) to measure the significance of the differences observed in ChIP signal in the presence and absence of Cdk12 in order to achieve a high level of confidence that we can reliably and reproducibly quantify even subtle differences in ChIP read density.
Software	BWA aligner version 0.7.12 was used to map ChIP sequencing reads. Metagene analysis was performed using custom python and R scripts. These scripts will be made available upon request.

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	All Cdk12 knockout cells were derivatives of V6.5 (C57Bl/6-129) mouse embryonic stem cells as described in Extended Data Figure 1A and the Materials and Methods and were cultured according to the Materials and Methods. Sample preparation was preformed as follows: Growth Curve Analysis Each biological replicate was resuspended in 450uL of ES media followed by addition of 1uL of 50uM calcein-AM in DMSO and 2uL of 2mM ethidium homodimer-1 (Thermo Fisher). Samples were incubated for 15-20 minutes at room temperature protected from light. After staining, 50uL of CountBright Absolute Counting Beads (Thermo Fisher) was added to each sample. Samples were analyzed by flow cytometry such that at least ~5000 CountBright Absolute Counting Beads (~100uL of sample) were recorded per sample, during which samples were vortexed every minute to prevent counting beads from settling out of solution. The number of live cells per replicate was quantified in each sample by counting the number of live (Calcein-AM positive and Ethidium Homodimer-1 negative) cells and comparing it to the number of CountBright Absolute Counting Beads (with known concentration). An example of the flow cytometry gating strategy used is shown in Supplementary Data Figure 1. To calculate the fold change in live cells over the previous 24 hours, the number of live cells in each biological replicate at each time point was compared to the average of the live cells in the 3 biological replicates at the previous time point to give the ratio of live cells every 24 hours. Cell Cycle Profiling Cells were plated at approximately equal cell density 24 hours prior to profiling, such that cells were 50-80% confluent at the time of harvest. Cells were pulsed with 10uM 5-ethynl-2'-deoxyuridine (EdU) for 1 hour under standard growth conditions, then harvested by trypsinization. Collected cell pellets were fixed, permeabilized, and stained for EdU incorporation with Alexa Fluor 647 using Click-iT EdU Flow Cytometry Assay Kit (Thermo Fisher) according to manufacturer's instructions. After EdU staining, cells were resuspended in 1x Click-iT saponin-based permeabilization and wash reagent (Thermo Fisher) with 50ug/mL propidium iodide to label total DNA content and 100ug/mL RNase A. Cells were incubated at room temperature in the dark for 30 minutes, and at least 50,000 cells (Fig. 1C) and 20,000 cells (Extended Data Fig. 1E) gated on P3 (see gating strategy in Supplementary Data Figure 1) were analyzed by flow cytometry for EdU content (AlexaFluor 647) and total DNA content (propidium iodide). An example of the flow cytometry gating strategy used is shown in Supplementary Data Figure 1. Apoptosis Cells were plated at approximately equal cell density at least 24 hours prior to harvest. Cells were harvested by trypsinization. The growth media and HBS wash prior to trypsinization were collected and centrifuged with the trypsinized cell population to collect any apoptosing cells with decreased adherence to the plate. Cell pellets were washed twice with cold PBS. Cells were fixed, permeabilized, and stained for cleaved Caspase-3 (apoptosis) using the FITC Active Caspase-3 Apoptosis Kit (BD Pharmingen) and the recommended protocol. At least 50,000 stained cells (gated on P3, see gating strategy in Supplementary Data Figure 1) were analyzed by FACS. An example of the flow cytometry gating strategy used is shown in Supplementary Data Figure 1.
Instrument	FACS analyzers used throughout this study include: BD FACS Celesta, BD LSRII, BD FACS Canto II, BD FACS LSR Fortessa FACS Sorter Used in Study to isolate single cell clones: BD FACS Aria IIIu
Software	BD FACS Diva Software Version 8.0.1 was used to collect the data. FlowJo Version 1.0.1 was used to analyze the data.
Cell population abundance	Does not apply to any experiments in this study.

Gating strategy

Representative gating strategy for all three FACS experiments is included in Supplementary Data 1.

For Growth Curve Analysis:
To quantify CountBright Absolute Counting Beads (Thermo), we gated on BV421 positive (fluorophore that is not detected in cell population) and high side scatter. For stringency, we re-gated that population on BB515 positive and high side scatter, which produced almost identical (>99.9%) counts. To quantify live cells in the population, we initially gated single cells on SSC-Area versus FSC-Area followed by gating FSC-Height versus FSC-width to remove aggregates. We then gated live cells by plotting Ethidium Homodimer-1 versus Calcein-AM and selecting Calcein-AM positive and Ethidium Homodimer-1 negative cell populations. Single stained populations and unstained populations were run as controls. Cell only and bead only samples were also analyzed as controls.

For Cell Cycle Analysis:
We first gated on single cells by gating on SSC-area versus FSC-area, followed by FSC-height versus FSC-width to remove aggregates, followed by SSC-height versus SSC-width to further remove aggregates. We then selected on the propidium iodide positive population to select stained, single cells. Following that, we plotted EdU incorporation as measured by AlexaFluor 647 (log scale) versus Propidium Iodide (linear scale) as measured on the PE-Texas Red detector and identified the three cycle phases as follows:

Cells that were EdU positive (S-phase)
Cells that were EdU negative and 2n propidium iodide signal (G1 phase)
Cells that were EdU negative and 4n propidium iodide signal (G2 phase).
Single color and unstained controls were analyzed as controls to set the gates.

For Apoptosis Analysis:
We first gated on single cells by gating on SSC-area versus FSC-area, followed by FSC-height versus FSC-width to remove aggregates, followed by SSC-height versus SSC-width to further remove aggregates. We then plotted cells on Pacific Blue (negative control channel) versus FITC (cleaved caspase 3, test channel) to gate for cells that came off the diagonal indicating FITC signal above background, autofluorescence. Unstained cells were compared as a negative control.

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