

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](#).

Statistics

For all statistical analyses, confirm that the following items are present in the 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
- ☐ ☒ A statement on 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 statistical parameters 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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used.

Data analysis

We used MOIRAI, Trimmomatic v 0.36 for trimming reads and removing ribosomal reads. STAR v 2.5.0a was used to align reads. RseQC package v 2.6.4 was used to calculate read distribution for RNA-seq data. To merge and subsample bam files, SAMtools v 0.1.19 was used. Coverage at single-base-pair resolution was calculated in BEDTools v 2.2.25. BedGraph files were converted into bigWig files using bedGraphToBigWig. To merge bigWig files, bigWigMerge was used. Reads mapping to promoters and enhancers were counted using bigWigAverageOverBed. Normalization to counts per million was performed using edgeR v 3.16.5. Transcription start sites were identified as previously described: <http://fantom.gsc.riken.jp/5/ssstar/> Protocols: HeliScopeCAGE_read_alignment. Decomposition peak identification was used for tag clustering: <https://github.com/hkawaji/dpi1>. Bi-directional enhancer identification was performed as previously described: <https://github.com/anderssonrobin/enhancers>. Metagene analysis was performed using deepTools v 2.5.2. UCSC LiftOver tool was used to convert genome coordinates between assemblies. Plots were made in R v 3.3.2 using ggplot2 v 2.2.1 and gplots v 3.0.1.1 packages. Trend analysis was performed using trend v 1.1.0 R package.

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. 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

Raw and processed data are available at the Gene Expression Omnibus (GEO) under accession GSE118075.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No sample-size calculation was performed. We performed our time-course experiment using three biologically independent samples on the basis of the previous study (Arner et al. Science 2015). Other experiments were performed with at least two biologically independent samples to confirm the reproducibility of our data.
Data exclusions	The replicate 1 at 120 minutes was not experimentally successful, and hence was not added to the time course analysis (although we prepared replicate 4 in an independent time course experiment). The analysis for 120 minutes was performed with the remaining two biological replicates.
Replication	All experiments were reliably reproduced in biologically independent samples. Biological independent samples were well correlated.
Randomization	Randomization was not relevant to our study because our aim was to study 5'-ends of nascent RNAs genome-wide.
Blinding	Blinding was not relevant to our study because our aim was to study 5'-ends of nascent RNAs genome-wide.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	GAPDH Monoclonal Antibody (6C5) (Thermo Fisher Scientific, AM4300, lot 00439919, 1:2000), Anti-SNRP70 antibody (abcam, ab83306, lot GR3178496-1, 1:1000), Pol II Antibody (F-12) (Santa Cruz Biotechnology, sc-55492, lot F2416, 1:1000), Histone H3 Antibody (24HC2 LC12)(Thermo Fisher Scientific, 701517, lot 00439919, 1:2000), Anti-Rabbit IgG, HRP-Linked Whole Ab Donkey (GE Healthcare, NA934-100UL, lot 9804676, 1:5000) and Anti-Mouse IgG, HRP-Linked Whole Ab Sheep (GE Healthcare, NA931-100UL, lot 9833235, 1:5000).
Validation	GAPDH Monoclonal Antibody (6C5) was validated in western blot using human cell lysate on (https://www.thermofisher.com/antibody/product/GAPDH-Antibody-clone-6C5-Monoclonal/AM4300). Anti-SNRP70 antibody was validated in western blot using human cell lysate on (https://www.abcam.co.jp/snrp70-antibody-ab83306.html). Histone H3 Antibody (24HC2 LC12) was validated in western blot using human cell lysate on (https://www.thermofisher.com/antibody/product/Histone-H3-Antibodyclone-24HC2-LC12-Monoclonal/701517). Pol II Antibody (F-12) was validated in western blot using human cell lysate on

(<https://www.scbt.com/scbt/ja/product/pol-ii-antibody-f-12>). Relevant references were cited on the manufacturer's website.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	MCF-7, GM12878, K562, HeLa-S3 and HepG2 were obtained from the FANTOM5 consortium at RIKEN, Yokohama, Japan.
Authentication	Comparative transcriptomic analysis using published data confirmed the authenticity of the cell lines.
Mycoplasma contamination	RNA-seq data in MCF-7, GM12878, K562, HeLa-S3 and HepG2 cells indicated no sign of mycoplasma contamination. No obvious sign of contamination was seen during cell culture.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	In this study, 8 weeks female mice were used and the strain of mice was C57BL/6NJcl.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	All mouse tissues used in this study were obtained from experiments approved by the RIKEN Yokohama Animal Experiment Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.